

جبر خطی عددی و کاربردها

تألیف: بیسوانات داتا
ترجمه: دکتر فائزه توتونیان



انتشارات، شماره ۵۱۸

جبر خطی عددی و کاربردها

تألیف:

یسوا نات داتا

ترجمه:

دکتر فائزه توتونیان

فهرست مطالب

۱۱	در مورد نویسنده.....
۱۳	پیشگفتار مؤلف.....
۲۱	پیشگفتار مترجم.....
۲۳	فصل ۰: مسایل جبر خطی عددی: اهمیت و مشکلات محاسباتی آنها.....
۲۳	۱-۰ مقدمه.....
۲۳	۲-۰ مسائل جبرخطی عددی و اهمیت آنها.....
۲۷	۳-۰ حل مسائل جبرخطی عددی با استفاده از شیوه‌های متداول.....
۲۹	نتیجه‌گیری.....
۳۱	فصل ۱: مروری بر مفاهیم جبر خطی مورد نیاز.....
۳۱	۱-۱ مقدمه.....
۳۱	۲-۱ بردارها.....
۲۳	۳-۱ ماتریسها.....
۴۶	۴-۱ برخی ماتریسهای خاص.....
۵۴	۵-۱ قضیه کیلی - هامیلتون.....
۵۴	۶-۱ مقادیر تکین.....
۵۵	۷-۱ نرمهای برداری و ماتریسی.....
۷۰	۹-۱ مرور و خلاصه.....
۷۰	تمرینها.....
۷۹	پیشنهادها برای مطالعه بیشتر.....
۸۱	فصل ۲: اعداد ممیز شناور و خطاها در محاسبات.....
۸۱	۲-۱ دستگاه اعداد ممیز شناور.....

۸۴	خطاهای گرد کردن	۲-۲
۸۸	قوانین حساب ممیز شناور	۳-۲
۹۳	جمع n عدد ممیز شناور	۴-۲
۹۴	ضرب n عدد ممیز شناور	۵-۲
۹۵	محاسبه حاصلضرب داخلی	۶-۲
۹۸	کرانه‌های خطا برای محاسبات ماتریسی ممیز شناور	۷-۲
۱۰۰	خطاهای گرد کردن ناشی از حذف و محاسبات بازگشتی	۸-۲
۱۰۵	مرور و خلاصه	۹-۲
۱۰۷	تمرینها	
۱۱۱	پیشنهادها برای مطالعه بیشتر	
۱۱۳	فصل ۳: پایداری الگوریتمها و وضعیت مسائل	
۱۱۳	برخی الگوریتمهای اساسی	۱-۳
۱۲۳	تعریفها و مفاهیم پایداری	۲-۳
۱۲۸	وضعیت مسأله و تحلیل اختلال	۳-۳
۱۲۹	وضعیت مسأله، پایداری الگوریتم، و دقت جواب	۴-۳
۱۳۲	چند جمله‌ای و یلکینسون	۵-۳
۱۳۴	یک دستگاه خطی بد - وضع	۶-۳
۱۳۵	مثالهایی از مسائل مقادیر ویژه بد - وضع	۷-۳
۱۳۷	پایداری قوی، ضعیف و معتدل	۸-۳
۱۴۰	مرور و خلاصه	۹-۳
۱۴۱	تمرینها	
۱۴۳	پیشنهادها برای مطالعه بیشتر	
۱۴۵	فصل ۴: الگوریتمهای عددی کارا و نرم‌افزارهای ریاضی	
۱۴۵	تعاریف و مثالها	۱-۴
	شمارش اعمال ممیز شناور (flops) و ملاحظات ذخیره‌سازی برای برخی	۲-۴
۱۴۹	الگوریتمهای مقدماتی	
۱۵۷	چند نرم‌افزار ریاضی با کیفیت بالا و موجود برای مسائل جبر خطی	۴-۳

۱۶۴	مرور و خلاصه
۱۶۵	تمرینها
۱۶۷	برنامه‌های MATLAB و MATCOM و مسائل
۱۶۹	پیشنهادها برای مطالعه بیشتر
۱۷۳	فصل ۵: برخی تبدیلات مفید در جبر خطی و کاربردهای آن
۱۷۳	۱-۵ یک راه کار محاسباتی در جبرخطی عددی
۱۷۴	۲-۵ ماتریسهای مقدماتی و تجزیه LU
۲۰۲	۳-۵ پایداری روش حذفی گاوس
۲۰۵	۴-۵ تبدیلات هاوس هولدر و کاربردها برای تجزیه QR و کاهش هسبرگی
۲۳۲	۵-۵ ماتریسهای گیونز و کاربردها برای تجزیه QR و کاهش هسبرگی
۲۴۹	۶-۵ پایه‌های یکا متعامد و تصویرهای متعامد
۲۵۴	۷-۵ تجزیه QR با محورگیری ستونی
۲۵۹	۸-۵ اصلاح یک تجزیه QR
۲۶۲	۹-۵ خلاصه و جدول مقایسه‌ها
۲۶۵	تمرینها
۲۷۵	برنامه‌های MATLAB و MATCOM و مسائل
۲۷۸	پیشنهادها برای مطالعه بیشتر
۲۷۹	فصل ۶: جوابهای عددی دستگاههای خطی
۲۷۹	۱-۶ مقدمه
۲۸۱	۲-۶ نتایج اساسی در مورد وجود، منحصر به فرد بودن، و پایایی جوابها
۲۸۳	۳-۶ برخی کاربردهای به وجود آورنده مسائل دستگاههای خطی
۳۰۵	۴-۶ روشهای مستقیم
۳۴۵	۵-۶ معکوسها و دترمینانها
۳۵۳	۶-۶ تحلیل اختلال مسأله دستگاه خطی
۳۶۱	۷-۶ عدد وضعیت و دقت جواب
۳۷۴	۸-۶ اختلالهای مؤلفه‌ای و خطاها
۳۷۶	۹-۶ تصفیه تکراری

۳۸۰	۱۰-۶	روشهای تکراری
۴۱۷	۱۱-۶	مرور و خلاصه
۴۲۳		تمرینها
۴۳۸		برنامه‌های MATLAB و MATCOM و مسائل
۴۴۵		پیشنهادها برای مطالعه بیشتر
۴۴۹		فصل ۷: جوابهای کم‌ترین توانهای دوم برای دستگاههای خطی
۴۴۹	۱-۷	مقدمه
۴۵۱	۲-۷	یک کاربرد ساده و منجر به یک دستگاه فرامعین
۴۵۲	۳-۷	وجود و یکتایی جواب
۴۵۵	۴-۷	تعبیر هندسی مسأله کم‌ترین توانهای دوم
۴۵۵	۵-۷	معادلات نرمال و برازش چندجمله‌ای
۴۶۰	۶-۷	شبه معکوس و مسأله کم‌ترین توانهای دوم
۴۶۱	۷-۷	حساسیت مسأله کم‌ترین توانهای دوم
۴۶۹	۸-۷	روشهای محاسباتی برای مسائل کم‌ترین توانهای دوم فرامعین
۴۹۳	۹-۷	دستگاههای فرومعین
۵۰۰	۱۰-۷	تصفیه تکراری
۵۰۵	۱۱-۷	محاسبه ماتریس وارینانس و کوواریانس $(A^T A)^{-1}$
۵۰۸	۱۲-۷	مرور و خلاصه
۵۱۲		تمرینها
۵۱۸		برنامه‌های MATLAB و مسائل
۵۲۴		پیشنهادها برای مطالعه بیشتر
۵۲۷		فصل ۸: مسائل عددی مقادیر ویژه یک ماتریس
۵۲۷	۱-۸	مقدمه
۵۲۹	۲-۸	برخی نتایج اساسی در مورد مقادیر ویژه و بردارهای ویژه
۵۳۹	۳-۸	مسائل مقدار ویژه مطرح در کاربردهای عملی
۵۶۳	۴-۸	محدودسازی مقادیر ویژه
۵۷۰	۵-۸	محاسبه مقادیر ویژه و بردارهای ویژه انتخابی

۵۹۰	تبدیل‌های تشابهی و محاسبات مقدار ویژه	۶-۸
۵۹۵	حساسیت مقدار ویژه	۷-۸
۶۰۲	حساسیت بردار ویژه	۸-۸
۶۰۳	شکل شور حقیقی و تکرارهای QR	۹-۸
۶۳۱	محاسبه بردارهای ویژه	۱۰-۸
۶۳۵	مسأله مقدار ویژه متقارن	۱۱-۸
۶۵۲	الگوریتم لانزوس برای ماتریسهای متقارن	۱۲-۸
۶۶۰	مرور و خلاصه	۱۳-۸
۶۶۵	تمرینها	
۶۷۶	برنامه‌های MATLAB و مسائل	
۶۸۴	پیشنهادها برای مطالعه بیشتر	
۶۸۹	فصل ۹: مسأله مقدار ویژه تعمیم یافته	
۶۸۹	مقدمه	۱-۹
۶۹۲	تجزیه شور تعمیم یافته	۲-۹
۶۹۳	الگوریتم QZ	۳-۹
۷۰۴	تکرار معکوس برای بردارهای ویژه مداد (A, B)	۴-۹
۷۰۵	مسأله مقدار ویژه تعمیم یافته معین متقارن	۵-۹
۷۱۳	مسائل مقدار ویژه تعمیم یافته معین متقارن مطرح در ارتعاشهای سازه‌ها	۶-۹
۷۲۳	کاربردها برای جداسازی و کاهش مدل	۷-۹
۷۳۷	مسأله مقدار ویژه درجه دو	۸-۹
۷۴۱	مسائل مقدار ویژه تعمیم یافته متقارن برای ماتریسهای ساخت یافته و بزرگ	۹-۹
۷۴۶	روش لانزوس برای مقادیر ویژه مداد $A - \lambda B$ در یک بازه	۱۰-۹
۷۴۸	مرور و خلاصه مطالب	۱۱-۹
۷۵۱	تمرینها	
۷۵۷	برنامه‌های MATLAB و مسائل	
۷۶۱	پیشنهادها برای مطالعه بیشتر	

۷۶۵	فصل ۱۰: تجزیه مقدار تکین (SVD)
۷۶۵	۱-۱۰ مقدمه
۷۶۷	۲-۱۰ قضیه تجزیه مقدار تکین
۷۷۱	۳-۱۰ یک رابطه بین تجزیه مقدار تکین و تجزیه‌های ویژه
۷۷۴	۴-۱۰ کاربردهای SVD
۷۷۷	۵-۱۰ حساسیت مقادیر تکین
۷۷۹	۶-۱۰ تجزیه مقدار تکین و ساختار یک ماتریس
۷۹۰	۷-۱۰ محاسبه ماتریس واریانس - کوواریانس
۷۹۱	۸-۱۰ تجزیه مقدار تکین، مسأله کمترین توانهای دوم، و شبه معکوس
۸۰۰	۹-۱۰ محاسبه تجزیه مقدار تکین
۸۱۷	۱۰-۱۰ SVD تعمیم یافته
۸۱۸	۱۱-۱۰ مرور و خلاصه
۸۱۹	تمرینها
۸۲۵	برنامه‌های MATLAB و مسائل
۸۳۰	پیشنهادها برای مطالعه بیشتر
۸۳۳	فصل ۱۱: نمونه کوچکی از تحلیل خطای گرد کردن
۸۳۳	۱-۱۱ قوانین اساسی حساب ممیز شناور
۸۳۴	۲-۱۱ تحلیل خطای پسرو برای طرح حذفی پیشرو و جایگزینی پسرو
۸۳۷	۳-۱۱ آنالیز خطای پسرو برای مثلثی سازی با استفاده از روش حذفی گاوس
۸۴۴	۴-۱۱ تحلیل خطای پسرو برای حل $Ax = b$
۸۴۶	۴-۱۱ مرور و خلاصه
۸۴۷	تمرینها
۸۴۹	پیشنهادها برای مطالعه بیشتر
۸۵۱	ضمیمه A: مقدمه مختصری برای MATLAB
۸۵۱	A - ۱ برخی اطلاعات اساسی در مورد MATLAB
۸۵۷	A - ۲ نوشتن برنامه‌های خودتان با استفاده از دستورهای MATLAB

۸۶۳	ضمیمه B : MATCOM و برنامه‌های انتخاب شده از MATCOM
۸۶۳	B - ۱ MATCOM چیست؟
۸۶۴	B - ۲ استفاده از MATCOM
۸۶۵	B - ۳ فهرست فصل به فصل برنامه‌های MATCOM
۸۶۹	B - ۴ برخی برنامه‌های انتخاب شده MATCOM
۸۷۷	ضمیمه C : یک فهرست از روالهای LAPACK که به طور متداول مورد استفاده‌اند
۸۸۱	جملات کلیدی در جبر خطی عددی
۸۹۱	جملات کلیدی در جبر خطی عددی (انگلیسی به فارسی)
۸۹۵	بخشی از راه حلها و جوابها برای برخی مسائل انتخاب شده
۸۹۵	فصل ۱
۸۹۸	فصل ۲
۸۹۹	فصل ۳
۸۹۹	فصل ۴
۹۰۱	فصل ۵
۹۰۲	فصل ۶
۹۰۵	فصل ۷
۹۰۷	فصل ۸
۹۰۸	فصل ۹
۹۰۹	فصل ۱۰
۹۱۰	فصل ۱۱
۹۱۳	منابع
۹۲۹	جدول نمادها
۹۳۳	فهرست الگوریتمها
۹۳۷	فهرست راهنما

در مورد نویسنده

بیسوا نات داتا^(۱) پرفسور علوم ریاضی و رئیس فعلی واحد درگیر^(۲) کاربردی برنامه دکترا در علوم ریاضی در دانشگاه ایلینویز شمالی می باشد. او همچنین استاد مدعو در دانشگاه ایلینویز در اربانا - چامپاین^(۳)، دانشگاه ایالتی پنسیلوانیا، دانشگاه ایلینویز جنوبی در کارین دیل^(۴)، دانشگاه کالیفرنیا در سان دیاگو، دانشگاه ایالتی کامپیناز^(۵)، برزیل، همچنین دانشگاههای بسیاری در سراسر جهان بوده است. او نویسنده تقریباً شصت و پنج مقاله میان رشته‌ای مرکب از جبر خطی، جبر خطی عددی و نظریه کنترل است. عضو هیأت تحریریه مجله‌های اولیه جبر خطی نظیر مجلات SIAM زیر بوده است *Matrix Analysis and applications, Linear algebra and its applications.*

و اخیراً مشاور علمی مجله زیر می باشد.

Journal of Mathematical systems, Estimation, and control.

داتا با عنوانهای نایب رئیس گروه فعال جبر خطی SIAM، عضو کمیته کارشناسی ارشد و نماینده آزاد در انجمن بین المللی جبر خطی خدمت کرده است. او همچنین چندین کنفرانس خیلی موفق میان رشته‌ای AMS و SIAM در جبر خطی، جبر خطی عددی، و کنترل، دستگاهها، و پردازش علامت را برگزار نموده است.

1. Biswa Nath Datta
3. Urbana-Champaign
5. Campinas

2. Application Involvement Component (AIC)
4. Carbondale

پیشگفتار مؤلف

جبر خطی عددی دیگر یک زیر مجموعه از آنالیز عددی نمی باشد و به صورت یک موضوع مستقل برای تحقیق و تدریس رشد کرده است. چون جبر خطی عددی برای محاسبات علمی لازم است (یک مؤلفه مهم از تحقیقات مهندسی و کاربردی می باشد) این موضوع برای رشته های ریاضی، علوم کامپیوتر و مهندسی به صورت کلاسهای کامل برگزار می گردد.

جبر خطی عددی و کاربردها می تواند هم در دوره های کارشناسی و هم در ابتدای کارشناسی ارشد و همچنین در کارهای دانشمندان و مهندسين برای مطالعات فردی و مرجع استفاده شود. من و سایر همکاران سر فصلها را به طور آزمایشی دسته بندی کرده ایم. مواد بیشتر از آنچه که یک ترم را بتواند پوشاند ارائه شده اند، بنابراین مدرسان می توانند مواد را مطابق با کلاسهای خاص خود تنظیم کنند و به راحتی سرفصل دوره را تدوین نمایند.

هدف اصلی من ترویج تعلیم جبر خطی عددی در دوره کارشناسی است. من مایلم که دانشجویان جدید که از این کتاب استفاده می کنند یک درک عمیق از مفاهیم خطاهای گرد کردن، پایداری، وضعیت و دقت داشته باشند، آنها همچنین باید قدر الگوریتمهای اصلی، سودمندی و پیاده سازی آنها را بدانند، برای این مقصود من ترکیبهای خاص **فهرت** را جهت بهره برداری آماده نموده ام.

ترکیبهای خاص

کاربردها. هرگاه ممکن باشد مسائل محاسباتی با مثالهای زندگی واقعی که از موضوعهای بشماري نظیر انتقال حرارت، دینامیک سیالات، پردازش سیگنال، مهندسی زیست پزشکی، آمار، بازرگانی، علوم

زیستی و نظریه کنترل، مهندسی ارتعاش، سرچشمه می‌گیرند، همراه خواهند شد. کاربردها همچنین کتاب را برای دوره‌هایی در جبرخطی کاربردی مناسب می‌سازند.

تحقیقات جدید. دانشجویان سالهای بالا بویژه از بحثهای به روز شده در مورد مفاهیم زیر بهره‌مند خواهند شد: پایداری، قوی، ضعیف و معتدل، مثالهای نقض از یک حدس ثابت قدیمی بر روی عامل رشد، تحلیل اختلال مؤلفه به مؤلفه مسائل دستگاه خطی، روشهای زیر فضای کريلف برای دستگاههای خطی و مسائل مقادیر ویژه، معادل بودن فرایند هاوس هولدر و گرام اشمیت برای مسائل کمترین توانهای دوم، استفاده از فرایند گرام اشمیت اصلاح شده برای حل دستگاههای فرو معین به طریق پایدار عددی، الگوریتم تقسیم و تسخیر برای مسائل مقدار ویژه متقارن و الگوریتم انتقال صفر برای مسئله SVD.

توضیحاتی روشن از مفاهیم اساسی. در کتاب ما خیلی زود وارد مفاهیم وضعیت مسائل و پایداری یک الگوریتم از طریق تحلیل پسرو خطای گرد کردن می‌شویم. مثالهای ساده ترغیب برانگیز همچنین ارائه می‌شوند. تاثیر این مفاهیم بر روی نتایج خاص به طور واضح نشان داده خواهند شد.

مسائل جبر خطی اساسی. در فصل صفر مسائل جبر خطی محاسباتی اساسی همراه با ذکر خلاصه‌ای از اهمیت هر یک بیان می‌شوند. سپس مشکلات حل این مسائل در پیاده‌سازی محاسباتی با استفاده از راههای روشن توضیح داده می‌شوند. این فصل به عنوان یک فصل ترغیب‌برانگیز برای کل کتاب به کار می‌رود.

بحث جامع ابزار اساسی. در سراسر کتاب ما نشان می‌دهیم که چگونه الگوریتمها می‌توانند با استفاده از ساختارهای ماتریسهای ابتدایی، هاوس هولدر و گینز کم هزینه و کارا گردند. فصل ۵ به آنها اختصاص داده شده است. ایجاد صفر در مکانهای خاص یک بردار یا ماتریس ساده است اما خاصیت خیلی قوی از این ماتریسها است. به طور واضح و روشن توضیح داده می‌شود که چگونه می‌توان از آنها برای تدوین الگوریتمها، برای تجزیه‌های ماتریسی مفید نظیر LU و QR و برای کاهش یک ماتریس به شکل ساده‌تر نظیر هسنبرگی استفاده کرد.

توضیح گام به گام الگوریتمها. قبل از این که یک الگوریتم معرفی شود، ما در مورد ایده پایه‌ای آن و اصول زیربنایی آن بحث خواهیم کرد. ابتدا چند گام اول را به صورت مقدماتی و با جزئیات توصیف می‌کنیم و سپس نشان می‌دهیم که چگونه گام k ام الگوی تثبیت شده را دنبال می‌کند. سپس دانشجویان قادر خواهند بود که برخی الگوریتمهای معروف را حتی قبل از یادگیری در کلاس دوباره تولید کنند.

نرم‌افزار و الگوریتمهای مؤثر عددی. در فصل ۴ توصیف می‌کنیم که چه عواملی الگوریتمها و نرم‌افزار وابسته را به طور عددی مؤثر می‌سازند: کارایی، پایداری عددی، صرفه‌جویی در ذخیره‌سازی

و غیره. اغلب دانشجویان اشتباه می‌کنند و یک الگوریتم را که فقط کارا است به عنوان الگوریتم خوب در نظر می‌گیرند، و درک نمی‌کنند که ممکن است خیلی ناپایدار باشد (برای مثال روش حذفی گاوس بدون محورگیری). در اینجا و سایر فصلها آنها یاد می‌گیرند که نه فقط چگونه بین الگوریتمها تمایز قائل شوند بلکه همچنین یک الگوریتم بد را به خوب در صورت امکان تبدیل کنند.

MATLAB و MATCOM. ما باور داریم که مدرسین جبرخطی عددی در صورتی که بتوانند باید از MATLAB استفاده کنند. این نرم‌افزار در همهٔ زمینه‌های محاسبات علمی و مهندسی به طور روز افزون عمومی‌تر می‌شود. ما برنامه‌هایی از MATLAB را برای برخی از الگوریتمهای پایه‌ای فراهم کرده‌ایم تا دانشجویان بتوانند از آنها به عنوان الگو برای نوشتن برنامه‌ها برای الگوریتمهای پیشرفته‌تر استفاده کنند. بسته ابزار MATLAB و MATCOM، دانشجویان را قادر می‌سازد که الگوریتمهای مختلف را برای یک مسأله مطالعه کنند و کارایی، پایداری و دقت را مقایسه نمایند (صفحه اطلاعات را در انتهای کتاب ملاحظه کنید).

حل و جواب مسائل انتخاب شده. بویژه برای استفاده در دورهٔ کارشناسی، برخی از راه‌حلها و جوابها را برای مسائل مهم، با تاکید بر آنهایی که نیاز به اثبات دارند، ارائه داده‌ایم. فهرست اصطلاحهای کلیدی. اصطلاحهای متداول در جبرخطی برای ارجاع سریع فهرست شده‌اند.

قدردانی

این کتاب به شکل کنونی‌اش بدون کمک، راهنمایی، پیشنهادهای، و تشویقهای بسیاری از دوستان و همکارانم در سراسر دنیا، وجود نمی‌داشت.

عمیق‌ترین سپاسهای خود را به یاری‌رسانان تقدیم می‌دارم. پرفسورها جیمز آر. بانچ^(۱) از دانشگاه کالیفرنیا، سن دیاگو؛ ام. بن بورنین^(۲)، وای پی. هونگ^(۳)، ای. گوپتا^(۴)، پرادپ مژومدار^(۵) از دانشگاه ایلینویز شمالی، میکائیل بری^(۶) از دانشگاه تنسی، زون گانگ زنگ^(۷) از دانشگاه ایلینویز شمال شرقی، و دکتر دانیل پیرس^(۸) از سرویسهای محاسبهٔ بویینگ. پرفسورها بانچ و زنگ و دکتر پیرس نسخه‌های مختلف دستی را بارها خوانده‌اند و انتقادهای سازنده و پیشنهادهای مفیدی برای بهبود کتاب ارائه داده و گاه‌گاهی در نگارش اثر یاری نموده‌اند. بعلاوه پرفسور بانچ یک دوره جبر خطی عددی را در دانشگاه کالیفرنیا در سن دیاگو از روی نسخه نیمه‌نهایی کتاب تدریس کرده است و بازتاب ارزشمند آن منجر

1. James. R. Bunch
3. Y. P. Hong
5. Pradip Majumdar
7. Zhonggang Zeng

2. M. Benbourenane
4. A. Gupta
6. Michael Berry
8. Daniel Pierce

به اصلاح و بهبود گردیده است. من عمیقاً سپاسگزار بانچ برای این یاری‌ها هستم. پرفسور هونگ نیز با حوصله بسیار، بخش قابل ملاحظه‌ای از قضایای جبر خطی نظری کتاب را بررسی نموده است و همکاران من پرفسورها بن بورنین، گوپتا و مؤومدار نه فقط در نوشتن بخشهای مربوط به کاربردهای مهندسی کمک نمودند بلکه نسخه‌های مختلف را غلط‌گیری کرده‌اند.

افراد ذیل برخی از فصلها را با دقت و کامل خوانده، اشتباهها را تذکر داده و توصیه‌ها و پیشنهادهایی برای بهبود ارائه نموده‌اند، من از یکایک آنها قدردانی می‌کنم. پرفسورها مارک آرنولد^(۱) از دانشگاه آرکانزاس، زد. بای^(۲) از دانشگاه کنتوکی، میکائیل بری از دانشگاه تنسی، اک بیورک^(۳) از دانشگاه لینک آپینگ سوئد، کی. بی. داتا^(۴) از انستیتوی تکنولوژی هند، خاکپور، هند؛ کارابی داتا^(۵) از دانشگاه ایلینویز شمالی، بارت دی مور^(۶) از دانشکده کاتولیک، لوون، بلژیک؛ دانیل اینمن^(۷) از پلی تکنیک ویرجینیا و دانشکده ایالتی؛ مورو لئون چینی^(۸) از دانشگاه پیزا، ایتالیا؛ برس فورد پارلت^(۹) از دانشگاه کالیفرنیا، برکلی، ایتشک رام^(۱۰) از دانشگاه آدلد، استرالیا، آکسل روه^(۱۱) از دانشگاه تکنولوژی شارلمرس، سوئد، یوسف سد^(۱۲) از دانشگاه مینسوتا؛ ژیلبرت استرانگ^(۱۳) از انستیتوی تکنولوژی ماساچوست؛ و پل ون دورن^(۱۴) از مرکز مهندسی دستگاهها و مکانیک کاربردی بلژیک. پرفسور فزال سعید^(۱۵) از دانشگاه ایلینویز، در سخنرانیش در مورد آموزش جبرخطی در مجمع کاربردی SIAM 1994 در استنبرد، اوتا، تفسیرهایی در مورد این کتاب نمود که قویاً برای من محرک و شوق‌آفرین بود.

بعلاوه از پرفسور بانچ و زنگ و دکتر پیرس به خاطر انتقادهای سازنده و راهنماییهای بسیار مفیدشان تشکر و قدردانی می‌نمایم؛ همچنین از پرفسورها دانیل بولی^(۱۶) از دانشگاه مینسوتا؛ آلن ادلمن^(۱۷) از انستیتوی تکنولوژی ماساچوست، فلویید هانسون^(۱۸) از دانشگاه ایلینویز شیکاگو؛ راج نیکانت پاتل^(۱۹) از دانشگاه کون کوردیا، مونترال؛ و سیمون تاوونر^(۲۰) از دانشگاه ایالتی پنسیلوانیا تشکر می‌کنم.

1. Mark Arnold
3. Ake Bjorck
5. Karabi Datta
7. Daniel Inman
9. Beresford Parlett
11. Axel Ruhe
13. Gilbert Strang
15. Faisal Saied
17. Alan Edelman
19. Rajnikant Patel

2. Z. Bai
4. K. B. Datta
6. Bart De Moor
8. Mauro Leoncini
10. Yitshak Ram
12. Youcef saad
14. Paul Van Dooren
16. Daniel Boley
18. Floyd Hanson
20. Simon Tavener

دانشجویان کنونی و سابق من، مارک آنولد^(۱)، سامر چودهاری^(۲)، کنراد فرناندز^(۳)، ویاو. هی^(۴)، سی. هتی^(۵)، ایشیتا ماجومدار^(۶)، و آویجیت پورکایاستا^(۷) که همگی به نوعی از غلط‌گیری گرفته تا حل مسأله و تایپ نقشی در این پروژه داشته‌اند؛ بویژه از کنراد فرناندز و ویاو. هی تشکر می‌نمایم. کنراد M-فایلها را برای نرم‌افزار MATCOM بر اساس MATLAB نوشته است، و یا او هر فصل را با دقت غلط‌گیری کرده و برخی اشکالات در الگوریتمها را تصحیح نموده است و در تهیه قسمتی از جوابهای مسائل انتخاب شده نیز کمک نموده است.

از دانشجویان دوره کارشناسی و کارشناسی ارشد دانشگاه ایلینویز شمالی و دانشگاه کالیفرنیا در سن دیاگو که نسخه‌های چاپ نشده کتاب را در اولین دوره‌شان در جبر خطی عددی استفاده کرده‌اند، بسیار تشکر می‌نمایم. انتقادهای سازنده آنها به من در تجدید نظر در برخی مواد کمک کرده است. از همکاران خودم پرفسورها ویلیام بل ایر^(۸)، جان بیچی^(۹)، لیندا سانز^(۱۰)، و جیندریچ نکاس^(۱۱) که خودشان نیز نویسندگان کتابهای درسی هستند، و گاه و بی‌گاه به پرسشهای من با شکیب بسیار پاسخ گفته و تجربیات خود را در اختیارم قرار داده‌اند و مرا از نظر روحی حمایت و تشویق نموده‌اند، تشکر و قدردانی می‌نمایم.

از کارکنان فعلی و سابق گروه علوم ریاضی در دانشگاه ایلینویز شمالی به خاطر کمکهای زیادشان در خصوص تایپ نسخه‌های مختلف کتاب و مساعدتهای به موقع‌شان برای نسخه‌های دیگر دست نویس تشکر می‌کنم. بویژه از لیزا آلیسون^(۱۲)، الیزابت بورک^(۱۳)، سارا کلایتون^(۱۴) و ماری - آن هارتینگ^(۱۵)، جنیفر جاکبسون^(۱۶)، پولا پروپست^(۱۷)، و میکائیل اسمیت^(۱۸) تشکر می‌نمایم. سارا کلایتون و ماری - آن هارتینگ بیشتر تایپ را انجام داده‌اند. مهارت و ابتکار ماری - آن را در LATEX می‌توان در همه جای کتاب مشاهده کرد؛ او معمار دست نویس بوده است. از آقای اریک بهر^(۱۹)، مدیر دستگاههای کامپیوتری به خاطر کمکهای فراوانش تشکر می‌نمایم.

- | | |
|-----------------------|-----------------------|
| 1. Mark Arnold | 2. Samar Choudhary |
| 3. Conrad Fernandes | 4. Wei-yao He, |
| 5. C. Hetti | 6. Ishita Majumdar |
| 7. Avijit Purkayastha | 8. William Blair |
| 9. John Beachy | 10. Linda Sons |
| 11. Jindrich Necas | 12. Lisa Allison |
| 13. Elizabeth Burk | 14. Sara Clayton |
| 15. Mari-Anne Hartig | 16. Jennifer Jacobson |
| 17. Paula Propst | 18. Michael Smith |
| 19. Eric. Behr | |

این کتاب بدون حمایت معنوی و درک و تشویق اعضای خانواده‌ام هرگز کامل نمی‌شد، کارابی^(۱)، همسر، که او هم یک ریاضیدان است؛ راجارشی^(۲)، پسر، که اکنون یک همکار جوان در دانشگاه ایلینویز در اوربانا - چامپاین^(۳) است؛ راخی^(۴)، دخترم که دانش‌آموز دبیرستانی در دکالب^(۵) است، پسر خواهرم سودیپ^(۶)، که اکنون در دانشگاه تکزاس در استون دانشجو است، پدر و مادر همسر آقای آر. ال. سارکار^(۷) و خانم بلارانی سارکار^(۸) که در هند زندگی می‌کنند اما هر هفته در مورد پیشرفت من پرس و جو می‌کردند و مرا تلفنی تشویق می‌نمودند. واضح است که من وظایف خود را به‌عنوان یک پدر و یک همسر کامل انجام ندهادم؛ بعلاوه همسر و فرزندانم با کوتاهی‌های گاه به‌گاه من در طی سالهای گذشته مدارا نموده‌اند، من خوشحالم که چنین خانواده مهربانی دارم.

همچنین از کارمندان شرکت انتشارات Brooks/Cole تشکر می‌نمایم. گاری اوستد^(۹) و یراستار اکتسابی، کارول بندیکت^(۱۰)، دستیار و یراستار؛ نانسی شامس^(۱۱) و یراستار تولید؛ جان کیمل^(۱۲) (اکنون در Chapman & Hall) که ابتدا کتاب را حروف چینی کرد. بعلاوه از لیندا تامسون^(۱۳) و باربارا کیمل^(۱۴)، و یراستارهای دوم، و آدرین کارتر^(۱۵)، هماهنگ‌کننده تبلیغاتی، به خاطر علاقمندی‌شان و مساعدتهای همیشگی‌شان تشکر و قدردانی می‌نمایم. کار کردن با آنها یک سرگرمی بود.

سرانجام این که مدیون نویسنده‌های چند کتاب ارزشمند در زمینه جبر خطی عددی هستم. بویژه پرفسورها. آک بیورک^(۱۶)، ژرژ فورسیت^(۱۷)، ژن. اچ. گولوب^(۱۸)، آر. جی. هانسون^(۱۹)، سی. ال. لاسون^(۲۰)، ام. ای. مال کولم^(۲۱)، کلو مولر^(۲۲)، جیمز ارتوگا^(۲۳)، برس فورد پارلت^(۲۴)، جی. دبلیو استوارت^(۲۵)، چارلز ون لون^(۲۶)، ریچارد وارگا^(۲۷)،

- | | |
|---------------------|-----------------------|
| 1. Karabi | 2. Rajarshi |
| 3. Urbana-Champaign | 4. Rakhi |
| 5. Dekalb | 6. Sudeep |
| 7. R. L. Sarkar | 8. Belarani Sarkar |
| 9. Gary Ostedt | 10. Carol Benedict |
| 11. Nancy Shammas | 12. John Kimmel |
| 13. Linda Thompson | 14. Barbara Kimmel |
| 15. Adrienne Carter | 16. Ake Bjorck |
| 17. George Forsyth | 18. Gene. H. Golub |
| 19. R. J. Hanson | 20. C. L. Lawson |
| 21. M. A. Malcolm | 22. Cleve Moler |
| 23. James Ortega | 24. Beresford Parlett |
| 25. G. W. Stewart | 26. Charles Van Loan |
| 27. Richard Varga | |

پیشگفتار مترجم

در دو دهه اخیر کارهای محاسباتی از همان اهمیت کارهای نظری و تجربی در بسیاری از شاخه‌های علوم و مهندسی برخوردار گشته‌اند. جبر خطی عددی اغلب در دل بسیاری از علوم محاسباتی و مسائل مهندسی نظیر انتقال حرارت، دینامیک سیالات، پردازش سیگنال، مهندسی زیست‌شناسی، آمار، نظریه کنترل، و غیره قرار دارد. چون نتایج یک چنین کارهای محاسباتی شدیداً بستگی به الگوریتمهای پیشرفته برای حل مسائل جبر خطی عددی گوناگون دارد، از این رو نیاز به کتابهای جبر خطی عددی کاملاً محسوس بوده و وجود آنها برای دانشجویان دوره کارشناسی و کارشناسی ارشد ریاضی، علوم کامپیوتر، مهندسی و دانشمندان سایر علوم و مهندسان امری اجتناب‌ناپذیر است. هدف اصلی من از ترجمه کتاب جبر خطی عددی و کاربردها این بوده است که دانشجویان با مطالعه آن یک درک عمیق از مفاهیم جبر خطی عددی نظیر خطاهای گرد کردن، پایداری، کارایی، وضعیت یک ماتریس داشته باشند و با مسائل جبر خطی و مشکلات حل آنها در پیاده‌سازی با استفاده از راههای روشن و واضح آشنا شوند و یاد بگیرند که چگونه با استفاده از الگوریتمهای کارا و مؤثر عددی به حل مسائل جبر خطی عددی بپردازند.

در این کتاب هر مسأله محاسباتی جبر خطی عددی با مثالهایی که از علوم دیگر سرچشمه می‌گیرد، مطرح می‌شود. سپس الگوریتمهای گوناگون برای حل آن ارائه می‌گردد. برای معرفی هر الگوریتم ابتدا در مورد ایده پایه‌ای آن و اصول زیربنایی آن بحث می‌شود و سپس الگوریتم به طور دقیق توضیح داده می‌شود و در پایان در مورد عواملی نظیر کارایی، پایداری عددی، صرفه‌جویی در ذخیره‌سازی که

الگوریتم را مؤثر عددی می‌سازند، بحث خواهد شد. مثالها عددی ارائه شده نیز موجب می‌شود که هیچ‌گونه ابهامی برای خواننده باقی نماند و بتواند به طور صحیح از الگوریتمها استفاده نماید. تمرینهای متنوع هر بخش موجب گسترش فهم جبر خطی عددی دانشجویان می‌شوند و تبحر آنها را در این زمینه افزایش می‌دهند. در این کتاب برنامه‌هایی از MATLAB برای برخی از الگوریتمهای پایه‌ای فراهم شده است تا دانشجویان بتوانند از آنها به عنوان الگو برای نوشتن کدها برای الگوریتمهای پیشرفته‌تر استفاده نمایند. بدین نحو دانشجویان با استفاده از MATLAB و MATCOM قادر خواهند بود که الگوریتمهای مختلف را برای یک مسأله مطالعه کنند و کارایی و پایداری و دقت آنها را مقایسه نمایند. این کتاب اصولاً اساس و بنیاد الگوریتمهای جبر خطی عددی و مزایا و معایب آنها را ارائه می‌دهد و برای دانشجویان دوره کارشناسی و کارشناسی ارشد ریاضی و علوم کامپیوتر و مهندسی پیش‌بینی شده است. ضمناً این کتاب می‌تواند به عنوان یک مرجع مورد استفاده همه افراد جامعه علمی و صنعتی قرار گیرد. امیدوارم ترجمه این کتاب مورد استفاده همه رشته‌های کاربردی و افراد جامعه علمی و صنعتی کشورمان قرار گیرد و در رسیدن آنها به نتایج علمی رضایت‌بخش مفید واقع گردد.

در اینجا لازم است به اطلاع خوانندگان گرامی رسانیده شود که چون ترجمه این کتاب بسیار حجیم بوده است، لذا به صورت دو جلد (اول و دوم) ارائه گردیده است. در جلد اول ترجمه این کتاب خوانندگان گرامی هشت فصل اول را ملاحظه می‌نمایند و در جلد دوم ترجمه این کتاب چهار فصل آخر را ملاحظه خواهند نمود. ضمناً فهرست همه فصلها، ضمائهم، مراجع و فهرست راهنما که برای همه فصلها مورد نیاز هستند، در هر دو جلد ارائه گردیده‌اند.

در اینجا بر خود لازم می‌دانم که صمیمانه‌ترین تشکرهای قلبی خود را به آقای دکتر کرایه‌چیان و آقای دکتر قدیریان که ویراستاری علمی و ادبی کتاب را انجام داده‌اند، تقدیم دارم و از زحمات بی‌شائبه خانم صدیقی و آقای افشاران که در تایپ و صفحه‌آرایی این کتاب کوشیده‌اند صمیمانه تشکر نمایم. همچنین از مسؤولین محترم معاونت پژوهشی و انتشارات دانشگاه فردوسی که نهایت همکاری را در جهت انتشار این کتاب نموده‌اند، سپاسگزاری می‌نمایم.

در خاتمه تذکرات خوانندگان گرامی و علاقه‌مند در مورد لغزشهای کتاب موجب امتنان خواهد بود، و از دریافت نظرات و پیشنهادات سازنده صاحب‌نظران سپاسگزار خواهم شد.

فائزه توتونیان

استاد گروه ریاضی دانشگاه فردوسی مشهد

مسائل جبر خطی عددی: اهمیت و مشکلات محاسباتی آنها

۱-۰ مقدمه

هدفهای اصلی این فصل عبارتند از: بیان مسائل خطی عددی، سپس ارائه خلاصه‌ای از اهمیت آنها و بالاخره ذکر مشکلاتی که در اجرای محاسبات به هنگام حل مسائل با استفاده از راههای واضح و روشن با آنها برخورد می‌نماییم.

۲-۰ مسائل جبرخطی عددی و اهمیت آنها

مسائل جبرخطی عددی در کاربرهای زیر نشان داده شده‌اند.

مسئله دستگاه خطی

با مفروض بودن یک ماتریس نامتفرد $n \times n$ مانند A و یک بردار n بعدی b ، مسئله پیدا کردن یک بردار n بعدی x می‌باشد به طوری که $Ax = b$.

یک شکل عملی از مسأله نیاز به جوابهای چندین دستگاه خطی با یک ماتریس A در طرف چپ دارد. یعنی، مسأله پیدا کردن یک ماتریس $X = (x_1, x_2, \dots, x_m)$ است به طوری که:

$$AX = B$$

که در آن $B = (b_1, b_2, \dots, b_m)$ یک ماتریس $n \times m$ است.

مسائل وابسته به مسائل دستگاه خطی عبارتند از: پیدا کردن معکوس یک ماتریس، رتبه، دترمینان، کدهای اصلی پیشرو، یک پایه یکا متعامد برای برد و فضای پوچ A ، و ماتریسهای تصویری گوناگون وابسته به A . جوابهای برخی از این مسائل اخیر به تجزیه‌های ماتریس نیاز دارند و مسأله تجزیه‌های ماتریس و مسائل دستگاههای خطی بسیار به یکدیگر وابسته‌اند.

شاید اغراق نباشد که بگوییم که مسأله دستگاه خطی در اغلب شاخه‌های علوم و مهندسی: ریاضیات کاربردی، زیست‌شناسی، شیمی، فیزیک، برق، مکانیک، راه و ساختمان، ارتعاشات و غیره مطرح می‌باشد.

مشترک‌ترین منبع مسائل فوق جوابهای عددی معادلات دیفرانسیل هستند. بسیاری از مدل‌های فیزیکی و دستگاههای مهندسی، دستگاههای معادلات دیفرانسیل هم معمولی و هم جزئی هستند. یک دستگاه معادلات معمولاً به طور عددی توسط گسسته سازی دستگاه از طریق روشهای تفاضلات متناهی یا عناصر متناهی حل می‌شود. فرایند گسسته سازی، در حالت کلی، به یک دستگاه خطی منجر می‌شود که جواب آن یک جواب تقریبی برای معادلات دیفرانسیل می‌باشد. (فصل ۶ را برای جزئیات بیشتر ملاحظه کنید).

مسأله کمترین توانهای دوم

با مفروض بودن یک ماتریس $n \times m$ مانند A ، مسأله کمترین توانهای دوم پیدا کردن یک بردار n بعدی x است، به طوری که نرم بردار مانده، $\|Ax - b\|_2$ ، تا حد ممکن کوچک باشد.

مسائل کمترین توانهای دوم در کاربردهای آماری و هندسی که نیاز به برازش دادن یک چندجمله‌ای یا یک منحنی به داده‌های تجربی دارند و همچنین در کاربردهای مهندسی نظیر پردازش سیگنال

و تصویر، مطرح می‌شوند. فصل ۷ را برای برخی کاربردهای خاص مسائل کمترین توانهای دوم ملاحظه نمایید. در اینجا لازم است متذکر شویم که روشها برای حل عددی مسائل کمترین توانهای دوم به جوابهای مسائل دستگاههای خطی منجر می‌شوند (مجدداً برای جزئیات فصل ۷ را ملاحظه کنید).

مسأله مقدار ویژه

با مفروض بودن یک ماتریس $n \times n$ مانند A ، مسأله پیدا کردن n اسکالر λ و بردارهای مخالف صفر x است، به طوری که

$$Ax = \lambda x$$

مسأله مقدار ویژه نوعاً در جواب صریح و تحلیل پایداری یک دستگاه همگن از معادلات دیفرانسیل مرتبه اول مطرح می‌شود. تحلیل پایداری فقط به شناخت ضمنی مقادیر ویژه نیاز دارد در صورتی که جواب صریح به مقادیر ویژه و بردارهای ویژه، به طور صریح نیاز دارد. کاربردهایی نظیر مسائل کمانش، تحلیل بورس کالاها، مختلف و مطالعه رفتار دستگاههای پویا به محاسبه فقط تعداد کمی از مقادیر ویژه، معمولاً تعداد کمی از بزرگترین یا کوچکترین مقادیر ویژه نیاز دارند.

در مثالهای کاربردی بسیاری، ماتریس A متقارن می‌باشد و در نتیجه مسأله مقدار ویژه، به یک مسأله مقدار ویژه متقارن تبدیل می‌شود. برای جزئیات بیشتر برخی کاربردهای خاص فصل ۸ را ملاحظه کنید. یک تعداد زیادی از مسائل مقادیر ویژه که در کاربردهای مهندسی مطرح می‌شوند مسائل مقدار ویژه تعمیم یافته هستند که به صورت زیر بیان می‌شوند.

مسأله مقدار ویژه تعمیم یافته

با مفروض بودن ماتریسهای مربعی A و B ، مسأله پیدا کردن اسکالرهایی λ و بردارهای مخالف صفر x است، به طوری که

$$Ax = \lambda Bx$$

مسأله مقدار ویژه تعمیم یافته نوعاً در تحلیل ارتعاشات ساختارهای نامیرا که توسط یک دستگاه معادلات دیفرانسیل مرتبه دوم و به شکل زیر مدل بندی می شوند، مطرح می شود.

$$B \frac{d^2 z}{dt^2} + Az = 0$$

به طور مشابه دستگاه میرای

$$A \frac{d^2 z}{dt^2} + B \frac{dz}{dt} + C = 0$$

به مسأله مقدار ویژه درجه دوم منجر می شود.

مسأله مقدار ویژه درجه دو

با مفروض بودن ماتریسهای مربعی A ، B و C ، مسأله پیدا کردن اسکالرهایی λ و بردارهای مخالف صفر x است، به طوری که

$$(\lambda^2 A + \lambda B + C) x = 0$$

برای جزئیات این که چگونه مسأله مقدار ویژه تعمیم یافته و مسأله مقدار ویژه درجه دو در کاربردهای خاص مطرح می شوند، فصلهای ۸ و ۹ را ملاحظه کنید.

مسأله تجزیه مقدار تکین

با مفروض بودن یک ماتریس $m \times n$ مانند A ، مسأله پیدا کردن ماتریسهای متعامد U و V و یک ماتریس قطری Σ است، به طوری که

$$A = U(\Sigma)V^T$$

این تجزیه به عنوان تجزیه مقدار تکین A معروف است. عناصر Σ مقادیر تکین هستند. بردارهای ستونی U و V بردارهای تکین نامیده می شوند.

زمینه‌های بسیاری از مهندسی، مانند نظریه کنترل و سیستمها، مهندسی طب زیستی، پردازش سیگنال و تصویر و کاربردهای آماری به مسأله تجزیه مقدار تکین منجر می‌شوند. این کاربردها نوعاً به رتبه ماتریس A ، یک پایه یکا متعامد، تصویرها، فاصله یک ماتریس از ماتریس دیگر با رتبه پایین‌تر، و غیره، به هنگام وجود ناخالصیها خاص (معروف به نوفه‌ها) در داده‌ها، نیاز دارند. مقادیر تکین و بردارهای تکین از لحاظ عددی ابزارهای قابل اطمینانی برای پیدا کردن این عناصر هستند. تجزیه مقدار تکین نیز کاراترین راه عددی برای حل مسأله کمترین توانهای دوم، بویژه در حالت رتبه ناقص است.

۰-۳ حل مسائل جبرخطی عددی با استفاده از شیوه‌های متداول

در این بخش مایلیم متذکر شویم که چگونه شیوه‌های متداول برای حل برخی از مسائل ذکر شده قبلی می‌توانند در پیاده سازی محاسباتی غیر عملی باشند و می‌توانند نتایج غیر دقیق تولید کنند.

● حل یک دستگاه خطی توسط قاعده کرامر: قاعده کرامر که در جبرخطی دوره کارشناسی تدریس می‌شود، البته از نظر تئوری و تاریخی اهمیت زیادی دارد (برای توضیحی از این روش فصل ۶ را ملاحظه کنید). متأسفانه این روش نمی‌تواند به عنوان یک روش محاسباتی عملی توصیه شود.

برای حل یک دستگاه خطی 20×20 ، حتی بر روی یک کامپیوتر مدرن و سریع امروزی باید با این قاعده، با استفاده از تعریف معمول دترمینان یک ماتریس، بیش از یک میلیون سال وقت صرف کرد.

● محاسبه جواب منحصر به فرد یک دستگاه خطی به وسیله معکوس ماتریس: جواب منحصر به فرد یک دستگاه خطی نامنفرد می‌تواند به صورت $x = A^{-1}b$ نوشته شود.

بدبختانه، محاسبه یک جواب برای یک دستگاه خطی با ابتدا محاسبه کردن صریح معکوس ماتریس غیر عملی می‌باشد.

محاسبه معکوس ماتریس تقریباً سه برابر پرهزینه‌تر از حل خود دستگاه خطی با استفاده از یک روش حذفی استاندارد می‌باشد (فصل ۶ را ملاحظه کنید)، و اغلب به دقت کمتری منجر می‌شود. یک مثال بدیهی را ملاحظه کنید: $27 = 3x$ را حل کنید. یک روش حذفی $9 = x$ را ارائه خواهد داد و فقط به یک تقسیم نیاز دارد. از طرف دیگر حل معادله با استفاده از معکوس ماتریس به شکل $x = \frac{1}{3} \times 27$ در خواهد آمد، که جواب $27 = \frac{1}{33333} \times x$ را (در حساب چهاررقمی) ارائه می‌دهد که به یک تقسیم و یک ضرب نیاز خواهد داشت.

توجه کنید که زمان کامپیوتری مصرفی توسط یک الگوریتم از لحاظ نظری توسط تعداد اعمال حسابی لازم برای انجام الگوریتم اندازه‌گیری می‌شود.

● حل یک مسأله کمترین توانهای دوم با استفاده از معادلات نرمال: اگر ماتریس $n \times m$ بعدی

A از رتبه کامل بوده، و m بزرگ‌تر یا مساوی n باشد، آنگاه مسأله کمترین توانهای دوم دارای یک جواب منحصر به فرد خواهد بود و این جواب از لحاظ نظری توسط جواب x دستگاه معادلات خطی زیر ارائه می‌شود.

$$A^T A x = A^T b$$

این معادلات معروف به معادلات نرمال هستند. بدبختانه این روش دارای برخی محدودیتهای عددی جدی است. اولاً در حساب دقت‌متناهی، در طی تشکیل صریح $A^T A$ ، برخی اطلاعات اساسی ممکن است از بین بروند. ثانیاً معادلات نرمال در برابر اختلالها بیشتر از دستگاه معادلات خطی معمولی $Ax = b$ حساس هستند و این حساسیت در موارد خاصی دقت جواب محاسبه شده کمترین توانهای دوم را بیش از آنچه که داده‌ها موجب می‌شوند، خراب می‌کند (فصل ۷ را برای جزئیات بیشتر ملاحظه کنید).

● محاسبه مقادیر ویژه یک ماتریس توسط پیدا کردن ریشه‌های چندجمله‌ای مشخصه‌اش:

مقادیر ویژه یک ماتریس A صفرهای چندجمله‌ای مشخصه‌اش هستند. بنابراین یک روش ساده برای پیدا کردن مقادیر ویژه محاسبه چندجمله‌ای مشخصه A و سپس پیدا کردن صفرهای آن به وسیله یک روش استاندارد ریشه‌یابی مطمئن است. متأسفانه این یک راه مناسب عددی نمی‌باشد. خطاهای تولید شده ناشی از گرد کردن در طی یک فرایند پیدا کردن چندجمله‌ای مشخصه، با احتمال خیلی زیاد برخی اختلالهای کوچک در ضرایب محاسبه شده تولید خواهند کرد. این خطاهای کوچک در ضرایب می‌توانند خطاهای بزرگی را در مقادیر ویژه در حالت‌های خاص تولید کنند. صفرهای برخی چندجمله‌ایها در برابر اختلالهای کوچک در ضرایب بسیار حساس می‌باشند. یک مثال کلاسیک از این نوع، چندجمله‌ای ویلکینسون است (فصل ۳ را ملاحظه کنید). ویلکینسون یک چندجمله‌ای درجه ۲۰ با ریشه‌های متمایز ۱ تا ۲۰ اختیار نمود و در ضریب x^{19} به وسیله یک مقدار کوچک معنی‌دار اختلال ایجاد کرد. سپس صفرهای این چندجمله‌ای کمی اختلال یافته توسط یک روش ریشه‌یابی مطمئن محاسبه شدند و نتیجه این بود که برخی ریشه‌ها کاملاً تغییر کردند و حتی برخی به صورت مختلط درآمدند.

● حل مسأله مقدار ویژه تعمیم یافته و مسائل مقدار ویژه درجه دوم توسط معکوس ماتریس:

مسأله مقدار ویژه تعمیم یافته در حالتی که B نامفرد است،

$$Ax = \lambda Bx$$

از لحاظ نظری معادل مسأله مقدار ویژه معمولی زیر است:

$$B^{-1}Ax = \lambda x$$

ولیکن، اگر ماتریس B نزدیک به نامنفرد باشد، آنگاه تشکیل ماتریس طرف چپ با محاسبه صریح معکوس B منجر به بی دقتی می شود که به نوبه خود، به محاسبات غیر دقیق مقادیر ویژه منجر می گردد. نتایج مشابهی برای مسأله مقدار ویژه درجه دوم برقرار است. در بیشتر کاربردهای مهندسی، نظیر مهندسی ارتعاش، ماتریس A معین مثبت متقارن است و در نتیجه نامنفرد است. در این حالت مسأله مقدار ویژه درجه دوم معادل با مسأله مقدار ویژه $Eu = \lambda u$ است که در آن

$$E = \begin{bmatrix} \cdot & I \\ -A^{-1}C & -A^{-1}B \end{bmatrix}$$

اما از لحاظ عددی حل مسأله مقدار ویژه درجه دوم توسط محاسبه ماتریس E توصیه نمی شود. اگر A تقریباً منفرد باشد، ماتریس E نمی تواند به طور دقیق تشکیل شود و در نتیجه مقادیر ویژه محاسبه شده نادقیق خواهند بود.

• پیدا کردن مقادیر تکین توسط مقادیر ویژه $A^T A$: هنگامی که $m \geq n$ ، مقادیر تکین A ریشه های دوم نامنفی مقادیر ویژه $A^T A$ هستند. ولیکن پیدا کردن مقادیر تکین به این طریق توصیه نمی شود. مجدداً، تشکیل ماتریس حاصلضرب به طور صریح، ممکن است منجر به از دست دادن اطلاعات معنی دار شود. یک مثال نسبتاً بدیهی را ملاحظه کنید:

$$A = \begin{bmatrix} 1 & 1 \\ \varepsilon & \cdot \\ \cdot & \cdot \end{bmatrix}$$

که در آن ε به گونه ای است که در محاسبات دقت متناهی $1 + \varepsilon^2 = 1$. در نتیجه پس از محاسبه داریم:

$$A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

مقادیر ویژه ۲ و ۰ هستند. بنابراین مقادیر تکین محاسبه شده به صورت $\sqrt{2}$ و ۰ ارائه می شوند. ولیکن مقادیر تکین دقیق عبارتند از $\sqrt{2}$ و $\varepsilon\sqrt{2}$. (فصل ۱۰ را برای جزئیات بیشتر ملاحظه کنید).

نتیجه گیری

در این فصل اشاره نمودیم که چگونه برخی از شیوه های نظری متداول برای مسائل جبرخطی ممکن

است به مشکلات محاسباتی و بی‌دقتی در نتایج محاسبه شده منجر شوند. جبرخطی عددی به تحلیل عمیق چنین مشکلاتی، به بررسی این که چگونه این مشکلات می‌توانند در برخی مثالها برطرف شوند، و به فرمول‌بندی و پیاده‌سازی الگوریتمهای عددی مناسب برای استفاده در علوم و مهندسی مربوط می‌شود.

مروری بر مفاهیم جبر خطی مورد نیاز

۱-۱ مقدمه

اگر چه یک دوره مقدماتی در جبر خطی یک پیشنیاز برای این کتاب است، برای کامل کردن آن ما برخی نمادگذاریها را متذکر می‌شویم و خیلی سریع تعریفهای اساسی و مفاهیم ماتریسها و بردارها را در این فصل مرور می‌کنیم. سپس ما با جزئیات بیشتر مفاهیم و نتایج اساسی از نرم برداری و نرم ماتریسی و کاربردهای آنها را برای مطالعه ماتریسهای همگرا بحث خواهیم کرد. این نتایج اغلب در فصلهای بعدی این کتاب مورد استفاده قرار می‌گیرند.

۲-۱ بردارها

یک مجموعه مرتب از اعداد حقیقی یک بردار نامیده می‌شود. خود اعداد مؤلفه‌های بردار نامیده می‌شوند، یک حرف کوچک معمولاً برای نشان دادن یک بردار استفاده می‌شود. یک بردار n مؤلفه‌ای v دارای شکل زیر است

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

برداری به این شکل یک بردار ستونی نامیده می‌شود، و ترانهاده آن یک بردار سطری است. مجموعه همه بردارهای n بعدی (یعنی هر بردار دارای n مؤلفه) توسط \mathbb{R}^n نمایش داده می‌شود که فضای برداری \mathbb{R}^n نامیده می‌شود. ترانهاده یک بردار v با v^T نمایش داده می‌شود. یک بردار ستونی به طور ساده یک بردار نامیده خواهد شد. مگر ذکر شود.

اگر u و v دو بردار در \mathbb{R}^n باشند، آنگاه مجموع آنها $u + v$ به صورت زیر تعریف می‌شود.

$$u + v = (u_1 + v_1, u_2 + v_2, \dots, u_n + v_n)$$

اگر c یک اسکالر باشد، آنگاه $cu = (cu_1, cu_2, \dots, cu_n)^T$. حاصلضرب داخلی دو بردار u و v یک اسکالر است که توسط رابطه زیر ارائه می‌شود.

$$u^T v = (u_1 v_1 + u_2 v_2 + \dots + u_n v_n)$$

طول یک بردار v که توسط $\|v\|$ نمایش داده می‌شود، برابر $\sqrt{v^T v}$ می‌باشد، یعنی طول بردار v (یا طول اقلیدسی v) برابر $\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$ می‌باشد.

گفته می‌شود که یک مجموعه بردار $\{m_1, \dots, m_k\}$ در \mathbb{R}^n وابسته خطی است اگر اسکالرهایی c_1, c_2, \dots, c_k نه همگی صفر، یافت شوند به طوری که $c_1 m_1 + c_2 m_2 + \dots + c_k m_k = 0$ (بردار صفر، یعنی برداری که همه مؤلفه‌هایش برابر صفر هستند). در غیر این صورت، مجموعه مستقل خطی نامیده می‌شود. مجموعه همه ترکیبهای خطی m_1, m_2, \dots, m_k با $\text{span}\{m_1, \dots, m_k\}$ نمایش داده خواهد شد.

مثال ۱-۲-۱ مجموعه بردارهای

$$e_i = (0, 0, \dots, 0, 1, 0, \dots, 0)^T, \quad i = 1, \dots, n$$



مؤلفه i ام

مستقل خطی است، زیرا رابطه

$$c_1 e_1 + c_2 e_2 + \dots + c_n e_n = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = 0$$

درست است اگر و فقط اگر

$$c_1 = c_2 = \dots = c_n = 0$$

مثال ۱-۲-۲ بردارهای $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ و $\begin{bmatrix} -3 \\ 6 \end{bmatrix}$ وابسته خطی هستند، زیرا

$$3 \begin{bmatrix} 1 \\ -2 \end{bmatrix} + \begin{bmatrix} -3 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

در اینجا $c_1 = 3$ و $c_2 = 1$.

۱-۲-۱ متعامد بودن بردارها و زیر فضاها

زاویه θ بین دو بردار u و v به صورت زیر ارائه می شود

$$\cos(\theta) = \frac{u^T v}{\|u\| \|v\|}$$

دو بردار u و v متعامد هستند اگر $\theta = 90^\circ$ ، یعنی $u^T v = 0$. نماد \perp برای نمایش متعامد بودن استفاده می شود.

فرض کنید S یک مجموعه از بردارهای واقع در \mathbb{R}^n باشد. آنگاه S یک زیر فضا از \mathbb{R}^n نامیده می شود اگر $s_1, s_2 \in S$ ایجاب کند $c_1 s_1 + c_2 s_2 \in S$ ، که در آن c_1 و c_2 اسکالرهایی دلخواه هستند. یعنی S یک زیر فضا است اگر هر ترکیب خطی از دو بردار در S نیز در S واقع باشد. توجه کنید که فضای \mathbb{R}^n خودش یک زیر فضا از \mathbb{R}^n است. برای هر زیر فضا یک کوچک ترین عدد صحیح و منحصر به فرد مانند r یافت می شود به قسمی که هر بردار در زیر فضا بتواند به صورت یک ترکیب خطی از حداکثر r بردار در زیر فضا بیان شود؛ r بعد زیر فضا نامیده می شود و با $\dim[S]$ نمایش داده می شود. گفته می شود که هر مجموعه از r بردار مستقل خطی از S با بعد $\dim[S] = r$ یک پایه از زیر فضا تشکیل می دهد.

دو زیر فضای S_1 و S_2 از \mathbb{R}^n گفته می شود که متعامد هستند اگر برای هر بردار $s_1 \in S_1$ و هر بردار $s_2 \in S_2$ رابطه $s_2^T s_1 = 0$ برقرار باشد. دو زیر فضای متعامد S_1 و S_2 با $S_1 \perp S_2$ نمایش داده می شوند.

۱-۳-۱ ماتریسها

۱-۳-۱ مفاهیم اساسی

گردایه ای از n بردار در \mathbb{R}^n و مرتب شده در یک آرایه مستطیلی m سطری و n ستونی یک ماتریس نامیده می شود. بنابراین یک ماتریس A دارای شکل زیر است.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

این ماتریس با $A = (a_{ij})_{m \times n}$ ، یا به طور ساده $A = (a_{ij})$ نمایش داده می شود، که از آن استنباط می شود که $i = 1, 2, \dots, m$ و $j = 1, 2, \dots, n$. گفته می شود که A از مرتبه $m \times n$ می باشد. مجموعه همه ماتریسهای $m \times n$ توسط $\mathbb{R}^{m \times n}$ نمایش داده می شود.

یک ماتریس A که دارای تعداد یکسان سطر و ستون باشد یک ماتریس مربعی نامیده می شود. ماتریس مربعی با ۱ در امتداد قطر اصلی و صفر در سایر جاها ماتریس واحد (همانی) نامیده می شود و توسط I نمایش داده می شود. در متن برخی اوقات یک ماتریس واحد از مرتبه $p \times p$ توسط $I_{p \times p}$ نمایش داده می شود.

مجموع دو ماتریس $A = (a_{ij})$ و $B = (b_{ij})$ در $\mathbb{R}^{m \times n}$ ماتریسی از همان مرتبه A و B می باشد و توسط رابطه زیر ارائه می شود.

$$A + B = (a_{ij} + b_{ij})$$

اگر c یک اسکالر باشد، آنگاه cA ماتریسی است که به صورت زیر ارائه می شود.

$$cA = (ca_{ij})$$

فرض کنید A از مرتبه $m \times n$ و B از مرتبه $n \times p$ باشند. آنگاه حاصلضرب آنها AB یک ماتریس از مرتبه $m \times p$ ارائه شده به صورت زیر می باشد.

$$AB = \left(\sum_{k=1}^n a_{ik} b_{kj} \right); \quad i = 1, 2, \dots, m, j = 1, 2, \dots, p$$

توجه کنید که اگر b یک بردار ستونی باشد، آنگاه Ab یک بردار ستونی است. از طرف دیگر، اگر a یک بردار ستونی و b^T یک بردار سطری باشد، آنگاه ab^T یک ماتریس است، و به حاصلضرب خارجی دو بردار a و b معروف است بنابراین،

$$ab^T = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} (b_1 \ b_2 \ \dots \ b_m) = \begin{pmatrix} a_1 b_1 & \dots & a_1 b_m \\ a_2 b_1 & \dots & a_2 b_m \\ \vdots & \ddots & \vdots \\ a_n b_1 & \dots & a_n b_m \end{pmatrix}$$

مثال ۱-۳-۱ فرض کنید

$$a = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad b^T = (2 \quad 3 \quad 4)$$

حاصلضرب خارجی ماتریس $ab^T = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 6 & 8 \\ 6 & 9 & 12 \end{bmatrix}$ است.

حاصلضرب داخلی برابر اسکالر ۲۰ است. $a^T b = (1 \quad 2 \quad 3) \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = 20$

ترانهاده یک ماتریس A از مرتبه $m \times n$ با A^T نمایش داده می شود، و ماتریسی از مرتبه $n \times m$ است که با تعویض سطرها و ستونهای A به دست می آید

$$A^T = (a_{ji}), \quad i = 1, \dots, n, \quad j = 1, \dots, m$$

توجه کنید که ضرب ماتریسی جابجایی نیست، یعنی در حالت کلی داریم:

$$AB \neq BA$$

همچنین،

$$(AB)^T = B^T A^T$$

تعریف ۱-۳-۱ یک ماتریس A متقارن نامیده می شود اگر $A^T = A$.

یک راه دیگر نوشتن ماتریس حاصلضرب این است که بنویسیم $B = (b_1, \dots, b_p)$ ، که در آن b_i ، i امین ستون B می باشد. در این صورت ماتریس حاصلضرب AB می تواند به صورت زیر نوشته شود.

$$AB = (Ab_1, \dots, Ab_p)$$

به طور مشابه، اگر a_i^T ، i امین سطر A باشد،

$$AB = \begin{bmatrix} a_1^T B \\ a_2^T B \\ \vdots \\ a_m^T B \end{bmatrix}$$

ماتریسهای بلوکی

اگر ماتریسهای A و B بتوانند به صورت زیر افراز شوند:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

آنگاه با در نظر گرفتن هر بلوک به عنوان یک عنصر ماتریس، ما می توانیم جمع، ضرب اسکالر و ضرب ماتریسی را به طریق معمول تشکیل دهیم. بنابراین،

$$A+B = \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} \\ A_{21}+B_{21} & A_{22}+B_{22} \end{bmatrix}$$

و

$$AB = \begin{bmatrix} A_{11}B_{11}+A_{12}B_{21} & A_{11}B_{12}+A_{12}B_{22} \\ A_{21}B_{11}+A_{22}B_{21} & A_{21}B_{12}+A_{22}B_{22} \end{bmatrix}$$

با فرض این که افراز به صورت مناسبی انجام شده باشد که ضربهای ماتریسی متناظر ممکن باشند مفهوم افراز دو بلوکی می تواند به سادگی تعمیم داده شود.

بنابراین اگر $A = (A_{ij})$ و $B = (B_{ij})$ دو ماتریس بلوکی باشند، $C = AB$ به صورت

$$C = (C_{ij}) = \left[\sum_{k=1}^n A_{ik} B_{kj} \right]$$

ارائه می شود که در آن هر A_{ij} ، B_{ij} و C_{ij} یک ماتریس بلوکی است.

یک ماتریس قطری بلوکی یک ماتریس $n \times n$ است که در آن هر عنصر قطری یک ماتریس مربعی است و سایر عناصر همگی برابر صفرند. برای راحتی به صورت زیر نوشته می‌شود.

$$A = \text{diag}(A_{11}, \dots, A_{kk})$$

که در آن A_{ii} ها ماتریسهای مربعی هستند. مجموع مرتبه‌های این ماتریسها برابر n است.

دترمینان یک ماتریس

برای هر ماتریس مربعی A ، یک اسکالر منحصر به فرد وابسته به ماتریس یافت می‌شود که دترمینان A نامیده می‌شود، و توسط $\det(A)$ نمایش داده می‌شود. برای یک ماتریس A از مرتبه 2×2 ، $\det(A) = a_{11}a_{22} - a_{12}a_{21}$ ؛ برای یک ماتریس $A = (a_{ij})$ از مرتبه 3×3 ، $\det(A) = a_{11} \cdot \det(A_{11}) - a_{12} \cdot \det(A_{12}) + a_{13} \cdot \det(A_{13})$ ، که در آن A_{1i} یک زیر ماتریس 2×2 است که از حذف سطر اول و ستون i ام به دست می‌آید. این تعریف می‌تواند به آسانی تعمیم داده شود. بنابراین برای یک ماتریس $A = (a_{ij})$ از مرتبه $n \times n$ داریم:

$$\det(A) = (-1)^{i+1} a_{i1} \det(A_{i1}) + (-1)^{i+2} a_{i2} \det(A_{i2}) + \dots + (-1)^{i+n} a_{in} \det(A_{in})$$

که در آن A_{ij} یک زیر ماتریس A از مرتبه $(n-1)$ می‌باشد که از حذف سطر i ام و ستون j ام به دست می‌آید.

مثال ۱-۳-۲

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

با قرار دادن $i = 1$ داریم:

$$\begin{aligned} \det(A) &= 1 \cdot \det \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix} - 2 \cdot \det \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix} + 3 \cdot \det \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix} \\ &= 1(-3) - 2(-6) + 3(-3) = 0 \end{aligned}$$

قضیه ۱-۳-۱ خواص ساده زیر برای $\det(A)$ برقرار هستند.

$$. \det(A) = \det(A^T) \quad .1$$

۲. $\det(\alpha A) = \alpha^n \det(A)$ ، که در آن α یک اسکالراست.

۳. $\det(AB) = \det(A) \cdot \det(B)$.

۴. اگر دو سطر یا دو ستون A یکسان باشند، آنگاه $\det(A) = 0$.

۵. اگر B یک ماتریس به دست آمده از A با تعویض دو سطر یا دو ستون باشد، آنگاه $\det(B) = -\det(A)$.

۶. دترمینان یک ماتریس مثلثی برابر حاصلضرب عناصر قطری آن می‌باشد.

(یک ماتریس مربعی A مثلثی است اگر همه عناصر زیر یا بالای قطر آن برابر صفر باشند.)

چند جمله‌ای مشخصه، مقادیر ویژه، و بردارهای ویژه یک ماتریس

فرض کنید A یک ماتریس $n \times n$ باشد. آنگاه چندجمله‌ای $P_n(\lambda) = \det(A - \lambda I)$ چندجمله‌ای مشخصه نامیده می‌شود. صفرهای چندجمله‌ای مشخصه مقادیر ویژه A نامیده می‌شوند. توجه کنید که این تعریف معادل زیر است: λ یک مقدار ویژه A است اگر و فقط اگر یک بردار مخالف صفر x یافت شود به طوری که $Ax = \lambda x$. بردار x یک بردار ویژه راست (یا فقط یک بردار ویژه) نامیده می‌شود.

تعریف ۱-۳-۲ یک ماتریس A از مرتبه $n \times n$ که کمتر از n بردار ویژه مستقل خطی داشته باشد یک ماتریس ناقص نامیده می‌شود.

مثال ۱-۳-۳ ماتریس $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ ناقص است. این ماتریس فقط یک بردار ویژه $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ دارد.

دترمینان یک ماتریس مثلثی بلوکی

فرض کنید

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

که در آن A_{11} و A_{22} ماتریسهای مربعی هستند. آنگاه $\det(A) = \det(A_{11}) \cdot \det(A_{22})$.

۱-۳-۲ برد و فضای پوچ

برای هر ماتریس A از مرتبه $m \times n$ دو زیر فضای وابسته مهم وجود دارد: برد A ، که توسط $R(A)$ ، و فضای پوچ A که توسط $N(A)$ نمایش داده می‌شوند:

$$R(A) = \{b \in \mathbb{R}^m \mid b = Ax, x \in \mathbb{R}^n \text{ مانند برداری}\}$$

$$N(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$$

فرض کنید S یک زیر فضا از \mathbb{R}^m باشد. آنگاه زیر فضای S^\perp که به صورت

$$S^\perp = \{y \in \mathbb{R}^m \mid y^T x = 0, x \in S\}$$

به ازای همه بردارهای $x \in S$

تعریف می شود متمم متعامد S نامیده می شود. می توان نشان داد (تمرینها را ملاحظه کنید) که

$$1. \quad N(A) = R(A^T)^\perp$$

$$2. \quad R(A)^\perp = N(A^T)$$

بعد $N(A)$ پوچی A نامیده می شود و توسط $\text{null}(A)$ نمایش داده می شود.

۱-۳-۳ رتبه یک ماتریس

فرض کنید A یک ماتریس از مرتبه $m \times n$ باشد. آنگاه زیر فضای به وجود آمده توسط بردارهای سطری A فضای سطری A نامیده می شود. زیر فضای به وجود آمده توسط ستونهای A فضای ستونی A نامیده شود. برد A همان فضای ستونی A می باشد.

رتبه یک ماتریس A بعد فضای ستونی A می باشد و به $\text{rank}(A)$ نمایش داده می شود. یک ماتریس مربعی $A \in \mathbb{R}^{n \times n}$ نامنفرد نامیده می شود اگر $\text{rank}(A) = n$ باشد. در غیر این صورت منفرد می باشد.

گفته می شود که یک ماتریس A از مرتبه $m \times n$ ، $A \in \mathbb{R}^{m \times n}$ دارای رتبه ستونی کامل است اگر ستونهای آن مستقل خطی باشند. رتبه سطری کامل به طور مشابه تعریف می شود. گفته می شود یک ماتریس A دارای رتبه کامل است اگر دارای رتبه سطری کامل یا رتبه ستونی کامل باشد. اگر دارای رتبه کامل نباشد، ناقص رتبه است.

مثال ۱-۳-۴ ماتریس

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

دارای رتبه کامل است؛ $\text{rank}(A) = 2$ (دارای رتبه ستونی کامل است)؛ $\text{null}(A) = 0$.

مثال ۱-۳-۵ ماتریس

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 0 & 0 \end{bmatrix}$$

دارای رتبه ناقص است؛ $\text{rank}(A) = 1$ ؛ $\text{null}(A) = 1$.

برخی از خواص رتبه

فرض کنید A یک ماتریس $m \times n$ باشد. آنگاه

۱. $\text{rank}(A) = \text{rank}(A^T)$;
۲. $\text{rank}(A) + \text{null}(A) = n$;
۳. $\text{rank}(AB) \geq \text{rank}(A) + \text{rank}(B) - n$; که در آن B از مرتبه $n \times p$ است
۴. $\text{rank}(BA) = \text{rank}(A) = \text{rank}(AC)$; که در آن B و C نامنفرد و از مرتبه‌های مناسب هستند
۵. $\text{rank}(AB) \leq \min \{ \text{rank}(A), \text{rank}(B) \}$.
۶. $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$.

۱-۳-۲ معکوس یک ماتریس

گفته می‌شود که یک ماتریس A از مرتبه $n \times n$ معکوس پذیر است اگر یک ماتریس B از مرتبه $n \times n$ وجود داشته باشد به طوری که

$$AB = BA = I$$

معکوس A توسط A^{-1} نمایش داده می‌شود. معکوس منحصر به فرد است. یک ماتریس معکوس پذیر اغلب یک ماتریس نامنفرد نامیده می‌شود.

یک خاصیت جالب از معکوس حاصلضرب دو ماتریس معکوس پذیر این است که

$$(AB)^{-1} = B^{-1}A^{-1}$$

قضیه ۱-۳-۲ برای هر ماتریس از مرتبه $n \times n$ ، گزاره‌های زیر معادلند:

۱. A نامنفرد است.
۲. $\det(A)$ مخالف صفر است.
۳. $\text{rank}(A) = \text{rank}(A^T) = n$.
۴. $N(A) = \{0\}$.
۵. A^{-1} وجود دارد.
۶. A دارای سطرها و ستونهای مستقل خطی است.
۷. مقادیر ویژه A مخالف صفر هستند.

۱-۳-۵ ماتریسهای مشابه

دو ماتریس A و B مشابه نامیده می‌شوند اگر یک ماتریس نامنفرد T وجود داشته باشد به قسمی که

$$T^{-1}AT = B$$

یک خاصیت مهم ماتریسهای مشابه این است که دو ماتریس مشابه دارای مقادیر ویژه یکسان هستند (برای اثبات بخش ۸-۲ را ملاحظه کنید).

۱-۳-۶ متعامد بودن و تصویرها

مجموعه بردارهای $\{v_1, \dots, v_m\}$ در \mathbb{R}^n متعامد است اگر

$$v_i^T v_j = 0, \quad i \neq j$$

بعلاوه اگر $v_i^T v_i = 1$ ، به ازای هر i ، آنگاه آنها یکا متعامد هستند. یک پایه برای یک زیرفضا که یکا متعامد نیز باشد، یک پایه یکا متعامد برای زیرفضا نامیده می‌شود.

مثال ۱-۳-۶ برای ماتریس

$$A = \begin{bmatrix} 0 & 0 \\ \frac{1}{2} & 1 \\ \frac{1}{2} & 1 \end{bmatrix}$$

بردار

$$\begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

یک پایه یکا متعامد برای $R(A)$ تشکیل می‌دهد (بخش ۵-۶ را ملاحظه کنید).

مثال ۱-۳-۷ برای ماتریس

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

بردارهای

$$\begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

یک پایه یکا متعامد برای $R(A)$ تشکیل می‌دهند (بخش ۵-۶-۱ را ملاحظه کنید).

تصویر قائم

فرض کنید S یک زیرفضا از \mathbb{R}^n باشد. آنگاه یک ماتریس P از مرتبه $n \times n$ و دارای خواص

۱. $R(P) = S$

۲. $P^T = P$ (متقارن است)،

۳. $P^2 = P$ (خود توان است).

تصویر قائم بر روی S یا، به طور ساده، ماتریس تصویر نامیده می‌شود. ما تصویر قائم P را بر روی S توسط P_S نمایش می‌دهیم. تصویر قائم بر روی یک زیرفضا منحصر به فرد است. توجه کنید که یک تصویر قائم لزوماً متعامد نیست.

فرض کنید $V = (v_1, \dots, v_k)$ ، که در آن $\{v_1, \dots, v_k\}$ یک پایه یکا متعامد برای یک زیرفضای S باشد. آنگاه

$$P_S = VV^T$$

تصویر قائم منحصر به فرد بر روی S است. توجه کنید که V منحصر به فرد نیست، اما P_S منحصر به فرد است.

یک رابطه بین P_{S^\perp} و P_S

اگر P_S تصویر قائم بر روی S باشد، آنگاه $I - P_S$ ، که در آن I ماتریس واحد از همان مرتبه P_S است، تصویر قائم بر روی S^\perp می‌باشد. [تمرین ۱۴ (الف) را ملاحظه کنید].

تصویرهای قائم بر روی $R(A)$ و $N(A^T)$

هنگامی که S برابر $R(A)$ یا $N(A^T)$ ، وابسته به A باشد، تصویرهای قائم منحصر به فرد بر روی $R(A)$ و $N(A^T)$ را به ترتیب توسط P_A و P_N نمایش می‌دهیم.

می‌توان نشان داد [تمرین ۱۴ (ب)] که اگر A از مرتبه $m \times n$ ($m \geq n$) دارای رتبه کامل باشد، آنگاه

$$P_A = A(A^T A)^{-1} A^T$$

$$P_N = I - A(A^T A)^{-1} A^T$$

مثال ۱-۳-۸

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}, \quad (A^T A)^{-1} = \begin{bmatrix} \frac{5}{6} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

$$P_A = \begin{bmatrix} \frac{5}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{5}{6} \end{bmatrix}$$

۷-۳-۱ تصویر یک بردار بر روی برد و فضای پوچ یک ماتریس فرض کنید S یک زیرفضا از \mathbb{R}^n باشد. آنگاه هر بردار b را می توان به صورت

$$b = b_S + b_{S^\perp}$$

نوشت که در آن $b_S \in S$ و $b_{S^\perp} \in S^\perp$. فرض کنید S برابر $R(A)$ یک ماتریس A باشد. آنگاه $b_S \in R(A)$ و $b_{S^\perp} \in N(A^T)$. بنابراین b_S را به b_R و b_{S^\perp} را به b_N نمایش خواهیم داد، به این معنا که b_R در برد A و b_N در فضای پوچ A^T هستند.

می توان نشان داد [تمرین ۱۴ (ج)] که

$$b_R = P_A b \quad \text{و} \quad b_N = P_N b$$

در اینجا، b_R و b_N به ترتیب تصویر قائم b بر روی $R(A)$ و تصویر قائم b بر روی $N(A^T)$ هستند.

از مطالب قبل به آسانی ملاحظه می کنیم که

$$b_R^T b_N = 0$$

$$A = \begin{bmatrix} \cdot & \cdot \\ \frac{1}{\sqrt{2}} & 1 \\ \frac{1}{\sqrt{2}} & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$V = \text{یک بردار یکا متعامد} = \begin{bmatrix} \cdot \\ -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$$P_A = VV^T = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \frac{1}{2} & \frac{1}{2} \\ \cdot & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$b_R = P_A b = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \frac{1}{2} & \frac{1}{2} \\ \cdot & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \cdot \\ 1 \\ 1 \end{bmatrix}$$

$$P_N = (I - P_A) = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & \frac{1}{2} & -\frac{1}{2} \\ \cdot & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$b_N = P_N b = \begin{bmatrix} 1 \\ \cdot \\ \cdot \end{bmatrix}$$

توجه کنید که $b = b_R + b_N$

۴-۱ برخی ماتریسهای خاص

۱-۴-۱ ماتریسهای قطری و مثلثی

یک ماتریس $A = (a_{ij})$ از مرتبه $m \times n$ یک ماتریس قطری است اگر $a_{ij} = 0$ به ازای $i \neq j$ می نویسیم $A = \text{diag}(a_{11}, \dots, a_{ss})$ ، که در آن $s = \min(m, n)$.

یک ماتریس مربعی $A = (a_{ij})$ یک ماتریس بالا مثلثی است اگر $a_{ij} = 0$ به ازای $i > j$.
 ترانهاده یک ماتریس بالا مثلثی پائین مثلثی است، یعنی $A = (a_{ij})$ پایین مثلثی است اگر $a_{ij} = 0$ به ازای $i < j$.

$$\begin{bmatrix} \diagup & 0 \\ * & \diagdown \end{bmatrix} \quad \text{پایین مثلثی} \qquad \begin{bmatrix} \diagup & * \\ 0 & \diagdown \end{bmatrix} \quad \text{بالا مثلثی}$$

برخی خواص مفید ماتریسهای مثلثی

خواص زیر از ماتریسهای مثلثی مفید هستند.

۱. حاصلضرب دو ماتریس بالا (پایین) مثلثی یک ماتریس بالا (پایین) مثلثی است. عناصر قطری ماتریس حاصلضرب دقیقاً برابر حاصلضرب عناصر قطری دو ماتریس است [تمرین ۱۹(الف)].

۲. معکوس یک ماتریس بالا (پایین) مثلثی نامنفرد یک ماتریس بالا (پایین) مثلثی است. عناصر قطری معکوس، معکوس عناصر قطری ماتریس اصلی هستند [تمرین ۱۹(ب)].

۳. مقادیر ویژه یک ماتریس مثلثی عناصر قطری آن هستند [تمرین ۱۹(د)].

۴. دترمینان یک ماتریس مثلثی برابر حاصلضرب عناصر قطری آن است [تمرین ۱۹(ج)].
 بنابراین یک ماتریس مثلثی نامنفرد است اگر و فقط اگر همه عناصر قطری آن مخالف صفر باشند.

۲-۴-۱ ماتریس متعامد

یک ماتریس مربعی U متعامد است اگر

$$U^T U = U U^T = I$$

ماتریسهای متعامد یک نقش مهم در محاسبات ماتریسی عددی بازی می کنند.

- دو خاصیت زیر از ماتریسهای متعامد برای محاسبات عددی جذاب هستند.
۱. معکوس یک ماتریس متعامد O دقیقاً ترانواده آن است: $O^{-1} = O^T$.
 ۲. حاصلضرب دو ماتریس متعامد یک ماتریس متعامد است.

۳-۴-۱ ماتریس جایگشت

یک ماتریس مربعی مخالف صفر مانند P یک ماتریس جایگشت نامیده می‌شود اگر تنها یک عنصر مخالف صفر که ۱ است در هر سطر و هر ستون یافت شود و بقیه همگی برابر صفر باشند. بنابراین اگر $(\alpha_1, \dots, \alpha_n)$ یک جایگشت از $(1, 2, \dots, n)$ باشد، آنگاه

$$P = \begin{bmatrix} e_{\alpha_1}^T \\ e_{\alpha_2}^T \\ \vdots \\ e_{\alpha_n}^T \end{bmatrix}$$

که در آن e_i ، i امین ستون ماتریس واحد I از مرتبه $n \times n$ می‌باشد. به طور مشابه ماتریس

$$P = (e_{\alpha_1}, e_{\alpha_2}, \dots, e_{\alpha_n})$$

که در آن e_i ، i امین ستون I می‌باشد، یک ماتریس جایگشت است.

مثال ۱-۴-۱ ماتریسهای

$$P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

همگی ماتریسهای جایگشت هستند.

اثرهای پیش ضرب و پس ضرب توسط یک ماتریس جایگشت

اگر

$$P_1 = \begin{bmatrix} e_{\alpha_1}^T \\ \vdots \\ e_{\alpha_n}^T \end{bmatrix}$$

آنگاه

$$P_1 A = \begin{pmatrix} \text{سطر } \alpha_1 \text{ از } A \\ \text{سطر } \alpha_2 \text{ از } A \\ \vdots \\ \text{سطر } \alpha_n \text{ از } A \end{pmatrix}$$

و به طور مشابه، اگر $P_2 = (e_{\alpha_1}, e_{\alpha_2}, \dots, e_{\alpha_n})$ ، آنگاه

$$AP_2 = (\text{ستون } \alpha_n \text{ از } A, \dots, \text{ستون } \alpha_2 \text{ از } A, \text{ستون } \alpha_1 \text{ از } A)$$

مثال ۱-۲-۲

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad P_1 = \begin{pmatrix} \cdot & 1 & \cdot \\ \cdot & \cdot & 1 \\ 1 & \cdot & \cdot \end{pmatrix} = \begin{pmatrix} e_3^T \\ e_2^T \\ e_1^T \end{pmatrix} \quad .1$$

$$P_1 A = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \end{pmatrix} = \begin{pmatrix} \text{سطر دوم } A \\ \text{سطر سوم } A \\ \text{سطر اول } A \end{pmatrix}$$

$$P_2 = \begin{pmatrix} \cdot & 1 & \cdot \\ \cdot & \cdot & 1 \\ 1 & \cdot & \cdot \end{pmatrix} = [e_3, e_1, e_2] \quad .2$$

$$AP_2 = \begin{pmatrix} a_{13} & a_{11} & a_{12} \\ a_{23} & a_{21} & a_{22} \\ a_{33} & a_{31} & a_{32} \end{pmatrix} = (\text{ستون دوم } A, \text{ستون اول } A, \text{ستون سوم } A)$$

یک خاصیت مهم یک ماتریس جایگشت این است که یک ماتریس جایگشت متعامد است بنابراین مطالب زیر درست است.

۱. معکوس یک ماتریس جایگشت P ترانهاده آن است، و ترانهاده آن نیز یک ماتریس جایگشت است.

۲. حاصلضرب دو ماتریس جایگشت یک ماتریس جایگشت است، و در نتیجه متعامد است.

۴-۴-۱ ماتریس هسنبرگی (تقریباً مثلثی)

یک ماتریس مربعی A بالا هسنبرگی است اگر $a_{ij} = 0$ ، به ازای $i > j + 1$. ترانهاده یک ماتریس بالا هسنبرگی یک ماتریس پایین هسنبرگی است، یعنی یک ماتریس $A = (a_{ij})$ پایین هسنبرگی است اگر $a_{ij} = 0$ به ازای $j > i + 1$. یک ماتریس مربعی که هم بالا هسنبرگی و هم پایین هسنبرگی باشد سه قطری است.

$$\begin{pmatrix} * & * & & 0 \\ \vdots & & \ddots & \\ * & & & * \\ * & * & \dots & * \end{pmatrix} \quad \begin{pmatrix} * & \dots & * & * \\ * & \dots & * & * \\ & \ddots & \vdots & \vdots \\ 0 & & * & * \end{pmatrix}$$

پایین هسنبرگی

بالا هسنبرگی

یک ماتریس بالا هسنبرگی $A = (a_{ij})$ کاهش ناپذیر است اگر

$$a_{i, i-1} \neq 0 \quad i = 2, 3, \dots, n \quad \text{به ازای}$$

به طور مشابه، یک ماتریس پایین هسنبرگی $A = (a_{ij})$ کاهش ناپذیر است اگر

$$a_{i, i+1} \neq 0 \quad i = 1, 2, \dots, n-1 \quad \text{به ازای}$$

مثال ۴-۴-۱

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 3 & 4 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{یک ماتریس پایین هسنبرگی کاهش ناپذیر است.}$$

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 2 & 3 \end{pmatrix} \quad \text{یک ماتریس بالا هسنبرگی کاهش ناپذیر است.}$$

برخی خواص مفید

۱. هر ماتریس مربعی A می‌تواند توسط یک تبدیل تشابهی متعامد به یک ماتریس بالا (یا پایین) هسنبرگی تبدیل شود، یعنی با مفروض بودن یک ماتریس حقیقی A ، یک ماتریس متعامد U وجود دارد به طوری که

$$UAU^T = H$$

که در آن H یک ماتریس بالا هسنبرگی است (یک اثبات ساختنی در فصل ۵ ارائه شده است).
 ۲. اگر A متقارن باشد، آنگاه ماتریس هسنبرگی تبدیل یافته حاصل از A سه قطری متقارن است.
 ۳. یک ماتریس هسنبرگی دلخواه را همواره می‌توان به بلوک‌هایی افزاز کرد که هر بلوک قطری یک ماتریس هسنبرگی کاهش‌ناپذیر باشد.

مثال ۱-۴-۴

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$A = \begin{bmatrix} A_1 & A_2 \\ 0 & A_3 \end{bmatrix}$$

توجه کنید که $A_1 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ و $A_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ ماتریسهای هسنبرگی کاهش‌ناپذیر هستند.

۱-۴-۵ ماتریس همراه

یک ماتریس بالا هسنبرگی کاهش‌ناپذیر به شکل

$$C = \begin{bmatrix} 0 & 0 & \dots & \dots & a_1 \\ 1 & 0 & \dots & \dots & a_2 \\ 0 & 1 & \dots & \dots & a_3 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & 1 & a_n \end{bmatrix}$$

یک ماتریس بالا همراه نامیده می‌شود. ترانزاده یک ماتریس بالا همراه یک ماتریس پایین همراه است. معادله مشخصه یک ماتریس همراه به سادگی تعیین می‌شود.

$$\det(C - \lambda I) = \det(C^T - \lambda I) = (-1)^n (\lambda^n - a_n \lambda^{n-1} - a_{n-1} \lambda^{n-2} - \dots - a_2 \lambda - a_1)$$

۶-۴-۱ ماتریس ناموهن

یک ماتریس A ناموهن است اگر A مشابه با یک ماتریس همراه باشد، یعنی A ناموهن است اگر یک ماتریس نامنفرد T وجود داشته باشد به طوری که TAT^{-1} یک ماتریس همراه باشد.

تبصره: یک ماتریس هسبرگی کاهش‌ناپذیر ناموهن است، اما عکس آن درست نیست.

یک ماتریس که ناموهن نیست موهن نامیده می‌شود. یک ماتریس موهن A مشابه با یک مجموع مستقیم از یک تعداد ماتریسهای همراه است. یعنی A با

$$\begin{bmatrix} C_1 & & & \\ & \ddots & & \\ & & O & \\ & & & \ddots \\ & O & & & C_k \end{bmatrix}$$

مشابه است که در آن هر C_i یک ماتریس همراه می‌باشد، $k > 1$ ، و چند جمله‌ای مشخصه هر C_i بر چند جمله‌ایهای مشخصه همه C_i های قبلی بخش‌پذیر است. این شکل همچنین به عنوان شکل کانونی فروبنیوس معروف است.

۷-۴-۱ ماتریس غالب قطری

یک ماتریس $A = (a_{ij})$ غالب قطری سطری است اگر

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| \quad \text{به ازای همه مقادیر } i$$

یک ماتریس غالب قطری ستونی به طور مشابه تعریف می‌شود. ماتریس

$$A = \begin{bmatrix} 1.0 & 1 & 1 \\ 1 & 1.0 & 1 \\ 1 & 1 & 1.0 \end{bmatrix}$$

هم غالب قطری سطری و هم ستونی است.

تذکر: برخی اوقات در ادبیات جبر خطی، یک ماتریس با این خواص یک ماتریس غالب قطری اکید نامیده می شود.

۱-۴-۸ ماتریس معین مثبت

یک ماتریس معین مثبت نامیده می شود اگر برای هر بردار مخالف صفر x داشته باشیم،

$$x^T A x > 0$$

فرض کنید $x = (x_1, x_2, \dots, x_n)^T$. آنگاه $x^T A x = \sum_{i,j=1}^n a_{ij} x_i x_j$ شکل درجه دوم وابسته به A نامیده می شود.

مثال ۱-۴-۵

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix},$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$x^T A x = (x_1 \ x_2) \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = (2x_1 + x_2, x_1 + 5x_2) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= 2x_1^2 + 2x_1x_2 + 5x_2^2 = 2(x_1^2 + x_1x_2 + \frac{1}{4}x_2^2) + \frac{9}{4}x_2^2$$

$$= 2(x_1 + \frac{1}{4}x_2)^2 + \frac{9}{4}x_2^2 > 0$$

یک ماتریس نیمه معین مثبت به طور مشابه تعریف می شود. یک ماتریس متقارن A نیمه معین مثبت است اگر $x^T A x \geq 0$ به ازای هر x .

یک نماد که معمولاً برای یک ماتریس متقارن معین مثبت (نیمه معین مثبت) استفاده می شود $A > 0$ ($A \geq 0$) می باشد.

برخی مشخصه ها و خاصیت های ماتریس های معین مثبت

۱. یک ماتریس متقارن A معین مثبت است اگر و فقط اگر همه مقادیر ویژه آن مثبت باشند. توجه

کنید که در مثال ۱-۴-۵ مقادیر ویژه عبارتند از $1/6972$ و $5/3028$.

۲. یک ماتریس متقارن A معین مثبت است اگر و فقط اگر کهادهای اصلی پیشرو آن همگی مثبت

باشند. n کهاد اصلی پیشرو برای یک ماتریس A از مرتبه $n \times n$ وجود دارد. i امین کهاد اصلی

پیشرو دترمینان زیر ماتریسی از A است که از i سطر و i ستون اول تشکیل می‌شود.
برای مثال اگر

$$A = \begin{bmatrix} 10 & 1 & 1 \\ 1 & 10 & 1 \\ 1 & 1 & 10 \end{bmatrix}$$

آنگاه

$$10 = \text{اولین کهاد اصلی پیشرو}$$

$$99 = \det \begin{bmatrix} 10 & 1 \\ 1 & 10 \end{bmatrix} = \text{دومین کهاد اصلی پیشرو}$$

$$972 = \det A = \text{سومین کهاد اصلی پیشرو}$$

۳. مجموع دو ماتریس معین مثبت یک ماتریس معین مثبت است.

۴. اگر $A = (a_{ij})$ معین مثبت باشد، آنگاه $a_{ii} > 0$ به ازای همه مقادیر i .

۵. اگر $A = (a_{ij})$ معین مثبت باشد، آنگاه بزرگترین عنصر (از لحاظ اندازه) همه ماتریس باید بر روی قطر آن قرار داشته باشد.

۶. یک ماتریس متقارن غالب قطری با عناصر قطری مثبت، معین مثبت است. توجه کنید که ماتریس A ، در مثال قبل، غالب قطری است.

تبصره: توجه کنید که (۴) و (۵) فقط شرایط لازم برای یک ماتریس متقارن هستند که معین مثبت باشد. آنها نمی‌توانند فقط به عنوان آزمونهای اولیه برای معین مثبت بودن به کار برده شوند. برای مثال ماتریسهای

$$A = \begin{bmatrix} 4 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 \\ 1 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 20 & 12 & 25 \\ 12 & 15 & 2 \\ 25 & 2 & 5 \end{bmatrix}$$

نمی‌توانند معین مثبت باشد، زیرا در ماتریس A یک عنصر صفر بر روی قطر و در ماتریس B بزرگترین عنصر، ۲۵، بر روی قطر قرار ندارد.

۵-۱ قضیه کیلی-هامیلتون

قضیه ۱-۵-۱ یک ماتریس مربعی A در معادله مشخصه خود صدق می‌کند؛ یعنی اگر $A = (a_{ij})$ یک ماتریس $n \times n$ باشد و $P_n(\lambda)$ چندجمله‌ای مشخصه A باشد آنگاه، $P_n(A)$ یک ماتریس صفر است.

اثبات: کتاب نظریه ماتریس نوشته فرانکلین (*Matrix theory by Franklin* , pp. 113-114) را ملاحظه کنید.

مثال ۱-۵-۱ فرض کنید

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$$

$$P_2(\lambda) = \lambda^2 - 2\lambda - 1$$

$$\begin{aligned} P_2(A) &= \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} - 2 \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

۶-۱ مقادیر تکین

فرض کنید A یک ماتریس $m \times n$ ($m \geq n$) باشد. آنگاه مقادیر ویژه و ویژه ماتریس از مرتبه $n \times n$ و مقارن $A^T A$ ، حقیقی و نامنفی هستند. فرض کنید این مقادیر ویژه، توسط σ_i^2 نمایش داده شوند و $\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_n^2$. آنگاه $\sigma_1, \sigma_2, \dots, \sigma_n$ مقادیر تکین A نامیده می‌شوند، هر ماتریس A از مرتبه $m \times n$ می‌تواند به صورت

$$A = U \Sigma V^T$$

تجزیه شود که در آن $U_{m \times m}$ و $V_{n \times n}$ ماتریسهای متعامد هستند و Σ یک ماتریس «شبه قطری» $m \times n$ است. این تجزیه، تجزیه مقدار تکین یا SVD نامیده می‌شود. مقادیر تکین σ_i ، $i = 1, 2, \dots, n$ عناصر قطری Σ هستند. تعداد مقادیر تکین مخالف صفر برابر رتبه ماتریس A است (بخش ۱۰-۳ را ملاحظه کنید).

مثال ۱-۶-۱ فرض کنید

$$A = \begin{bmatrix} 0 & 1 \\ 2 & 2 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 4 & 4 \\ 4 & 5 \end{bmatrix}$$

مقادیر ویژه $A^T A$ عبارتند از $\left[\frac{9 \pm \sqrt{65}}{2} \right]$:

$$\sigma_1 = \sqrt{\left[\frac{9 + \sqrt{65}}{2} \right]},$$

$$\sigma_2 = \sqrt{\left[\frac{9 - \sqrt{65}}{2} \right]}$$

۷-۱ نرمهای برداری و ماتریسی

۱-۷-۱ نرمهای برداری

فرض کنید

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

یک بردار n بعدی در \mathbb{R}^n باشد. آنگاه یک نرم برداری که توسط نماد $\|x\|$ نمایش داده می‌شود یک تابع پیوسته از مؤلفه‌های x_1, x_2, \dots, x_n از x ، با مقدار حقیقی تعریف شده بر روی \mathbb{R}^n می‌باشد که دارای خواص زیر است.

۱. $\|x\| > 0$ به ازای هر بردار مخالف x ، $\|x\| = 0$ اگر x یک بردار صفر باشد.

۲. $\|\alpha x\| = |\alpha| \|x\|$ به ازای هر بردار x در \mathbb{R}^n و به ازای همه اسکالرها α .

۳. $\|x + y\| \leq \|x\| + \|y\|$ به ازای هر بردار x و y در \mathbb{R}^n .

خاصیت ۳ معروف به نامساوی مثلثی است.

تذکر:

$$\|-x\| = \|x\|$$

$$|\|x\| - \|y\|| \leq \|(x - y)\|$$

به سادگی می توان تحقیق نمود که توابع زیرنرمهای برداری هستند.

$$1. \|x\|_1 = |x_1| + |x_2| + \dots + |x_n| \text{ (نرم مجموع یا نرم یک)}$$

$$2. \|x\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \text{ (نرم اقلیدسی یا نرم دو)}$$

$$3. \|x\|_\infty = \max_i |x_i| \text{ (نرم بی نهایت یا نرم ماکزیمم)}$$

در حالت کلی، اگر p یک عدد حقیقی بزرگتر یا مساوی ۱ باشد، p -نرم، یا نرم هولدر، به صورت زیر تعریف می شود.

$$\|x\|_p = \left[|x_1|^p + \dots + |x_n|^p \right]^{\frac{1}{p}}$$

منظور ما از $\|x\|$ نرم $\|x\|_2$ خواهد، مگر متذکر شویم.

مثال ۱-۷-۱ فرض کنید $x = (1, 1, -2)^T$ ، آنگاه

$$\|x\|_1 = 4$$

$$\|x\|_2 = \sqrt{1^2 + 1^2 + (-2)^2} = \sqrt{6}$$

$$\|x\|_\infty = 2$$

یک خاصیت مهم نرم هولدر نامساوی هولدر است.

$$|x^T y| \leq \|x\|_p \|y\|_q$$

که در آن

$$\frac{1}{p} + \frac{1}{q} = 1$$

یک حالت خاص نامساوی هولدر نامساوی کوشی - شوارتز است .

$$|x^T y| \leq \|x\|_2 \|y\|_2$$

یعنی:

$$\left| \sum_{j=1}^n x_j y_j \right| \leq \sqrt{\sum_{j=1}^n x_j^2} \sqrt{\sum_{j=1}^n y_j^2}$$

خاصیت هم ارزی نرمهای برداری

همه نرمهای برداری به این مفهوم هم ارز هستند که ثابتهایی مانند α و β وجود دارند به قسمی که به ازای هر بردار x داریم:

$$\alpha \|x\|_\mu \leq \|x\|_\nu \leq \beta \|x\|_\mu$$

برای نرمهای ۱، ۲، ∞ می توانیم α و β را به سادگی محاسبه کنیم:

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \|x\|_2$$

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n} \|x\|_\infty$$

$$\|x\|_\infty \leq \|x\|_1 \leq n \|x\|_\infty$$

۱-۷-۲ نرمهای ماتریسی

فرض کنید A یک ماتریس $m \times n$ باشد، آنگاه مشابه بانرم برداری، یک نرم ماتریسی $\|A\|$ با خواص زیر تعریف می کنیم.

$$1. \|A\| \geq 0, \|A\| = 0 \text{ اگر و فقط اگر } A \text{ یک ماتریس صفر باشد.}$$

$$2. \|\alpha A\| = |\alpha| \|A\| \text{ به ازای هر اسکالر } \alpha.$$

$$3. \|A + B\| \leq \|A\| + \|B\|$$

نرمهای ماتریسی وابسته

با مفروض بودن یک ماتریس A و یک نرم برداری $\|\cdot\|$ ، یک عدد نامنفی که به صورت

$$\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

تعریف می‌شود در همه خواص یک نرم ماتریسی صدق می‌کند. این نرم یک نرم ماتریسی وابسته به نرم برداری است. یک خاصیت خیلی مفید و اغلب مورد استفاده از یک نرم ماتریسی وابسته (برخی اوقات ما آن را p -نرم یک ماتریس A خواهیم نامید) عبارت است از:

$$\|Ax\|_p \leq \|A\|_p \|x\|_p$$

این خاصیت به سادگی از تعریف p -نرمها نتیجه می‌شود. توجه کنید که برای هر بردار خاص و مخالف صفر داریم:

$$\|A\|_p \geq \frac{\|Ax\|_p}{\|x\|_p}$$

از ضرب طرفین این رابطه در $\|x\|_p$ نامساوی اصلی حاصل می‌گردد.

دو p -نرم که از لحاظ محاسبه ساده‌ترین هستند، عبارتند از:

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}| \quad (\text{نرم مجموع ستونی ماکزیمم})$$

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}| \quad (\text{نرم مجموع سطری ماکزیمم})$$

مثال ۱-۷-۲

$$A = \begin{pmatrix} 1 & -2 \\ 3 & 4 \\ -5 & 6 \end{pmatrix}$$

$$\|A\|_1 = 12, \quad \|A\|_\infty = 11$$

یک p -نرم مفید دیگر، نرم طیفی است.

تعریف ۱-۷-۱ نرم طیفی عبارت است از

$$\max_{x \neq 0} \|Ax\|_2 = \frac{\|Ax\|_2}{\|x\|_2}$$

می‌توان نشان داد که

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$$

ماکزیمم مقدار ویژه $\lambda_{\max}(A^T A)$

(توجه کنید که مقایره ویژه $A^T A$ حقیقی و نامنفی هستند.)

مثال ۱-۷-۳

$$A = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 5 & 13 \\ 13 & 34 \end{bmatrix}$$

مقادیر ویژه $A^T A$ عبارتند از: 0.257 و 38.9743 .

$$\|A\|_F = \sqrt{38.9743} = 6.2429$$

نرم فروبنیوس

یک نرم ماتریسی مهم که با نرم برداری $\|x\|_2$ سازگار است نرم فروبنیوس می باشد:

$$\|A\|_F = \left[\sum_{j=1}^n \sum_{i=1}^m |a_{ij}|^2 \right]^{\frac{1}{2}}$$

یک نرم ماتریسی $\| \cdot \|_M$ و یک نرم برداری $\| \cdot \|_V$ سازگارند اگر

$$\|Ax\|_V \leq \|A\|_M \|x\|_V$$

برای هر نرم ماتریسی وابسته و نرم فروبنیوس می توان ثابت کرد که

$$\|AB\| \leq \|A\| \|B\|$$

تذکر:

۱. برای ماتریس واحد I ،

$$\|I\|_F = \sqrt{n}$$

در حالی که $\|I\|_1 = \|I\|_2 = \|I\|_\infty = 1$.

۲. $\|A\|_F^2 = \text{trace}(A^T A)$ ، که در آن $\text{trace}(A)$ به صورت مجموع عناصر قطری A تعریف می‌شود، یعنی اگر $A = (a_{ij})$ ، آنگاه $\text{trace}(A) = a_{11} + a_{22} + \dots + a_{nn}$.

خاصیت هم‌ارزی نرم‌های ماتریسی

همانند حالت نرم‌های برداری، نرم‌های ماتریسی نیز مرتبط هستند. اسکالرهایی α و β وجود دارند به قسمی که

$$\alpha \|A\|_\mu \leq \|A\|_\nu \leq \beta \|A\|_\mu$$

بروژه نامساویهای زیر که نرم‌های مختلف را به هم مرتبط می‌سازند درست هستند و اغلب در عمل بسیار مورد استفاده قرار می‌گیرند.

قضیه ۱-۷-۱ فرض کنید A یک ماتریس از مرتبه $m \times n$ باشد،

$$1. \frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\|_2 \leq \sqrt{m} \|A\|_\infty$$

$$2. \|A\|_2 \leq \|A\|_F \leq \sqrt{n} \|A\|_2$$

$$3. \frac{1}{\sqrt{m}} \|A\|_1 \leq \|A\|_2 \leq \sqrt{n} \|A\|_1$$

$$4. \|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}$$

نامساویهای ۱ و ۲ را اثبات می‌کنیم و بقیه را به عنوان تمرین باقی می‌گذاریم.

اثبات (۱): برطبق تعریف داریم:

$$\|A\|_\infty = \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty}$$

از خاصیت هم‌ارزی نرم‌های برداری داریم:

$$\|Ax\|_\infty \leq \|Ax\|_2 \quad \text{و} \quad \|x\|_2 \leq \sqrt{n} \|x\|_\infty$$

از نامساوی دوم به دست می‌آوریم:

$$\frac{1}{\|x\|_\infty} \leq \frac{\sqrt{n}}{\|x\|_2}$$

بنابراین نتیجه می شود که :

$$\frac{\|Ax\|_{\infty}}{\|x\|_{\infty}} \leq \sqrt{n} \frac{\|Ax\|_p}{\|x\|_p}$$

یا

$$\max_{x \neq 0} \frac{\|Ax\|_{\infty}}{\|x\|_{\infty}} \leq \sqrt{n} \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \sqrt{n} \|A\|_p$$

یعنی،

$$\frac{1}{\sqrt{n}} \|A\|_{\infty} \leq \|A\|_p$$

قسمت اول ثابت شد. برای اثبات قسمت دوم، دوباره از تعریف $\|A\|_p$ و خاصیت هم‌ارزی نرم‌های برداری استفاده می‌کنیم:

$$\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

$$\|Ax\|_p \leq \sqrt{m} \|Ax\|_{\infty}$$

$$\|x\|_{\infty} \leq \|x\|_p$$

بنابراین:

$$\frac{\|Ax\|_p}{\|x\|_p} \leq \sqrt{m} \frac{\|Ax\|_{\infty}}{\|x\|_{\infty}}$$

در نتیجه

$$\max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} \leq \sqrt{m} \max_{x \neq 0} \frac{\|Ax\|_{\infty}}{\|x\|_{\infty}} \quad \text{یا} \quad \|A\|_p \leq \sqrt{m} \|A\|_{\infty}$$

اثبات (۱) اکنون کامل است.

اثبات (۲): رابطه (۲) را با استفاده از یک روش متفاوت ثابت می‌کنیم. یادآور می‌شویم که

$$\|A\|_F^2 = \text{trace}(A^T A)$$

چون $A^T A$ متقارن است، یک ماتریس متعامد O وجود دارد به طوری که

$$O^T (A^T A) O = D = \text{diag}(d_1, \dots, d_n)$$

(قضیه ۸-۲-۸ را ملاحظه کنید.)

اثر یک ماتریس تحت یک تبدیل تشابه تغییرناپذیر است (تمرین ۳۵ (د) را ملاحظه کنید). بنابراین داریم

$$\text{trace}(A^T A) = \text{trace}(D) = d_1 + \dots + d_n$$

فرض کنید $d_k = \max_i (d_i)$ ، آنگاه چون d_1, \dots, d_n مقادیر ویژه نیز هستند، داریم

$$\|A\|_F^2 = d_k$$

بنابراین:

$$\|A\|_F^2 = \text{trace}(A^T A) = d_1 + \dots + d_n \geq d_k = \|A\|_F^2$$

برای اثبات قسمت دیگر، توجه می‌کنیم که:

$$\|A\|_F^2 = d_1 + \dots + d_n \leq d_k + d_k + \dots + d_k = n d_k$$

□ یعنی $\|A\|_F^2 \leq n d_k = n \|A\|_F^2$. بنابراین $\|A\|_F \leq \sqrt{n} \|A\|_F$.

۳-۷-۱ همگرایی یک دنباله ماتریس و ماتریسهای همگرا

گفته می‌شود یک دنباله از بردارهای $v^{(1)}, v^{(2)}, \dots$ به بردار v همگرا است اگر

$$\lim_{k \rightarrow \infty} v_i^{(k)} = u_i \quad i = 1, \dots, n$$

گفته می‌شود یک دنباله از ماتریسهای $A^{(1)}, A^{(2)}, \dots$ به ماتریس $A = (a_{ij})$ همگرا است اگر

$$\lim_{k \rightarrow \infty} a_{ij}^{(k)} = a_{ij}, \quad i, j = 1, 2, \dots, n$$

اگر دنباله $\{A^{(k)}\}$ به A همگرا باشد، آنگاه می‌نویسیم

$$\lim_{k \rightarrow \infty} A^{(k)} = A$$

اکنون بدون اثبات، شرایط لازم و کافی را برای همگرایی دنباله‌های برداری و ماتریسی بیان می‌کنیم. اثباتها می‌توانند به آسانی انجام شوند.

قضیه ۲-۷-۱ دنباله $v^{(1)}, v^{(2)}, \dots$ به v همگراست اگر و فقط اگر

$$\lim_{k \rightarrow \infty} \|v^{(k)} - v\| = 0$$

یک قضیه مشابه برای یک دنباله ماتریس نیز برقرار است.

قضیه ۳-۷-۱ دنباله $A^{(1)}, A^{(2)}, \dots$ به ماتریس A همگراست اگر و فقط اگر

$$\|A - A^{(r)}\| \rightarrow 0$$

که در آن $\|\cdot\|$ یک نرم دلخواه می باشد.

اکنون نتیجه‌ای در مورد همگرایی دنباله توانهای یک ماتریس به ماتریس صفر را بیان و اثبات می‌کنیم.

قضیه ۴-۷-۱ دنباله A^1, A^2, \dots از توانهای ماتریس A به ماتریس صفر همگراست اگر و فقط اگر $|\lambda_i| < 1$ به ازای هر مقدار ویژه λ_i از A .

اثبات: برای هر ماتریس A از مرتبه $n \times n$ یک ماتریس نامنفرد T وجود دارد به قسمی که

$$T^{-1}AT = J = \begin{bmatrix} J_1 & & \cdot \\ & \ddots & \\ \cdot & & J_r \end{bmatrix}$$

که در آن J_i دارای شکل زیر است.

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \cdot \\ & \lambda_i & 1 & \cdot \\ & & \ddots & \ddots \\ \cdot & & & \ddots \\ & & & & \lambda_i \end{bmatrix}$$

با یک محاسبه ساده می توان نشان داد که

$$J_i^k = \begin{bmatrix} \lambda_i^k & k\lambda_i^{k-1} & \binom{k}{2} \lambda_i^{k-2} & \dots & \binom{k}{n-1} \lambda_i^{k-n+1} \\ & \lambda_i^k & k\lambda_i^{k-1} & \dots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & k\lambda_i^{k-1} \\ & & & & \lambda_i^k \end{bmatrix}$$

از این رابطه مشاهده می کنیم که $J_i^{(k)} \rightarrow 0$ اگر و فقط اگر $|\lambda_i| < 1$. این بدین معناست که $\lim_{k \rightarrow \infty} A^k = 0$ اگر و فقط اگر $|\lambda_i| < 1$ به ازای هر i .

شکل فوق شکل متعارف جردن A نامیده می شود و ماتریسهای بلوکی قطری ماتریسهای جردن نامیده می شوند.

تعریف ۱-۷-۲ یک ماتریس A یک ماتریس همگرا نامیده می شود اگر $A^k \rightarrow 0$ هنگامی که $k \rightarrow \infty$. اکنون یک شرط کافی برای همگرایی یک ماتریس A برحسب نرمی از ماتریس بیان می کنیم. ابتدا نتیجه زیر را ثابت می کنیم.

یک رابطه بین نرمها و مقادیر ویژه

قضیه ۱-۷-۵ فرض کنید λ یک مقدار ویژه یک ماتریس A باشد. آنگاه برای هر نرم ماتریسی وابسته داریم:

$$|\lambda| \leq \|A\|$$

اثبات: برطبق تعریف یک بردار مخالف صفر x وجود دارد به قسمی که

$$Ax = \lambda x$$

با گرفتن نرم از طرفین این رابطه داریم:

$$\|Ax\| = \|\lambda x\| = |\lambda| \|x\|$$

ولیکن، $\|Ax\| \leq \|A\| \|x\|$ ، بنابراین $\|x\| \leq \|A\| \|x\|$ در نتیجه خواهیم داشت

$$|\lambda| \leq \|A\|$$

تعریف ۱-۷-۳ مقدار $\rho(A)$ که به صورت

$$\rho(A) = \max_i |\lambda_i|$$

تعریف می شود که در آن $\lambda_1, \dots, \lambda_n$ مقادیر ویژه A هستند، شعاع طیفی A نامیده می شود. به عنوان یک حالت خاص از قضیه ۱-۷-۵ داریم:

$$\rho(A) \leq \|A\|$$

با در نظر گرفتن این نتیجه، اکنون می توانیم نتیجه زیر را بیان کنیم

نتیجه ۱-۷-۱ یک ماتریس A همگراست اگر $\|A\| < 1$ ، که در آن $\| \cdot \|$ یک نرم ماتریسی وابسته است.

همگرایی یک سری ماتریسی نامتناهی

قضیه ۱-۷-۶ سری ماتریسی

$$I + A + A^2 + \dots$$

به $(I - A)^{-1}$ همگراست اگر A یک ماتریس همگرا باشد.

اثبات: چون A یک ماتریس همگراست، $A^k \rightarrow 0$ هنگامی که $k \rightarrow \infty$. پس از قضیه ۱-۷-۴ باید به ازای هر مقدار ویژه λ_i از A داشته باشیم $|\lambda_i| < 1$. این بدان معناست که ماتریس $(I - A)$ نامنفرد است، زیرا مقادیر ویژه $I - A$ برابر $1 - \lambda_1, 1 - \lambda_2, \dots, 1 - \lambda_n$ هستند و یک ماتریس A نامنفرد است اگر و فقط اگر مقادیر ویژه اش مخالف صفر باشند. چون مقادیر ویژه $(I - A)$ مقادیر $1 - \lambda_1, 1 - \lambda_2, \dots, 1 - \lambda_n$ هستند، رابطه $|\lambda_i| < 1$ ایجاب می کند که هیچ کدام از آنها صفر نباشند. بنابراین از تساوی

$$(I - A)(I + A + A^2 + \dots + A^k) = I - A^{k+1}$$

داریم:

$$(I + A + A^2 + \dots + A^k) = (I - A)^{-1} - (I - A)^{-1} A^{k+1}$$

چون A یک ماتریس همگرا است، $\rightarrow 0$ A^{k+1} هنگامی که $k \rightarrow \infty$. بنابراین هنگامی که $k \rightarrow \infty$ ،

$$I + A + A^2 + \dots \rightarrow (I - A)^{-1}$$

۴-۷-۱ فرمها و معکوسها

هنگامی که خطاها در یک الگوریتم مورد تجزیه و تحلیل قرار می‌گیرند، برخی اوقات لازم است بدانیم که یک ماتریس نامنفرد A به چه اندازه می‌تواند اختلال یابد به قسمی که ماتریس اختلال یافته $A + E$ نامنفرد باشد و چگونه می‌توان خطا را در معکوس ماتریس اختلال یافته برآورد نمود. با ماتریس واحد شروع می‌کنیم. در روابط زیر $\| \cdot \|$ یک نرم ماتریسی وابسته است که برای آن $\|I\| = 1$ می‌باشد.

قضیه ۷-۷-۱ فرض کنید که $\|E\| < 1$ ، آنگاه $(I - E)$ نامنفرد است و

$$\|(I - E)^{-1}\| \leq (1 - \|E\|)^{-1}$$

اثبات: فرض کنید $\lambda_1, \dots, \lambda_n$ مقادیر ویژه E باشند، آنگاه مقادیر ویژه $I - E$ عبارتند از:

$$1 - \lambda_1, 1 - \lambda_2, \dots, 1 - \lambda_n$$

چون $\|E\| < 1$ ، $|\lambda_i| < 1$ به ازای هر i . بنابراین هیچ کدام از مقادیر $1 - \lambda_1, 1 - \lambda_2, \dots, 1 - \lambda_n$ صفر نیستند. این ثابت می‌کند که $I - E$ نامنفرد است.

برای اثبات قسمت دوم، توجه داریم که چون $\|E\| < 1$ ، برطبق قضیه ۶-۷-۱ داریم:

$$(I - E)^{-1} = I + E + E^2 + \dots = \lim_{k \rightarrow \infty} \left(\sum_{i=0}^k E^i \right)$$

چون $\|E^i\| \leq \|E\|^i$ ، با گرفتن نرم از هر دو طرف رابطه فوق داریم:

$$\|(I - E)^{-1}\| \leq \sum_{i=0}^{\infty} \|E\|^i = (1 - \|E\|)^{-1} \quad (\|I\| = 1 \text{ زیرا})$$

(توجه کنید که سری نامتناهی $\dots + x + x^2 + \dots$ به $\frac{1}{1-x}$ همگرا است اگر و فقط اگر $|x| < 1$)

قضیه ۸-۷-۱ فرض کنید که $\|E\| < 1$ ، آنگاه،

$$\|(I - E)^{-1} - I\| \leq \frac{\|E\|}{1 - \|E\|}$$

اثبات: برای هر دو ماتریس نامنفرد A و B می‌توانیم بنویسیم.

$$A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}$$

در این رابطه با جایگزین کردن $B = I$ و $A = I - E$ خواهیم داشت

$$(I-E)^{-1} - I = (I-E)^{-1}E$$

(توجه کنید که $I^{-1} = I$). گرفتن نرم از طرفین این رابطه نتیجه می‌دهد.

$$\|(I-E)^{-1} - I\| \leq \|I-E\|^{-1} \|E\|$$

از قضیه ۱-۷-۷ می‌دانیم که:

$$\|(I-E)^{-1}\| \leq (1 - \|E\|)^{-1}$$

بنابراین نتیجه مورد نظر به دست می‌آید.

□

پيامد نتیجه

اگر ماتریس E خیلی کوچک باشد، آنگاه $\|E\| - 1$ نزدیک به واحد است. بنابراین نتیجه قبلی ایجاب می‌کند که اگر یک ماتریس واحد کمی اختلال یافته را معکوس کنیم، آنگاه خطا در معکوس ماتریس اختلال یافته از نرم ماتریس اختلال تجاوز ننماید.

قضیه ۱-۷-۹ فرض کنید A نامنفرد و $\|A^{-1}E\| < 1$ باشد. آنگاه $A - E$ نامنفرد است و

$$\frac{\|A^{-1} - (A-E)^{-1}\|}{\|A^{-1}\|} \leq \frac{\|A^{-1}E\|}{1 - \|A^{-1}E\|}$$

اثبات: می‌توانیم بنویسیم

$$A - E = A(I - A^{-1}E)$$

چون $\|A^{-1}E\| < 1$ ، از قضیه ۱-۷-۷ نتیجه می‌گیریم که $I - A^{-1}E$ نامنفرد است. بنابراین $A - E$ که

حاصلضرب ماتریسهای نامنفرد A و $I - A^{-1}E$ می باشد، همچنین یک ماتریس نامنفرد است. برای اثبات قسمت دوم، دوباره از تساوی

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$$

استفاده می کنیم. با جایگزینی $B = A - E$ ، سپس داریم

$$A^{-1} - (A - E)^{-1} = -A^{-1}E(A - E)^{-1}$$

با گرفتن نرم از طرفین رابطه فوق داریم:

$$\|A^{-1} - (A - E)^{-1}\| \leq \|A^{-1}E\| \|(A - E)^{-1}\|$$

چون

$$B = A - (A - B)$$

$$= A[I - A^{-1}(A - B)]$$

$$B^{-1} = [I - A^{-1}(A - B)]^{-1}A^{-1}$$

(توجه کنید که $(XY)^{-1} = Y^{-1}X^{-1}$). اگر $B = A - E$ را جایگزین کنیم، آنگاه خواهیم داشت

$$(A - E)^{-1} = [I - A^{-1}E]^{-1}A^{-1}$$

با گرفتن نرم، به دست می آوریم

$$\|(A - E)^{-1}\| \leq \|A^{-1}\| \|(I - A^{-1}E)^{-1}\|$$

اما از قضیه ۷-۷-۱ می دانیم که

$$\|(I - A^{-1}E)^{-1}\| \leq (1 - \|A^{-1}E\|)^{-1}$$

بنابراین داریم

$$\|(A - E)^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}E\|}$$

در نتیجه

$$\|A^{-1} - (A - E)^{-1}\| \leq \frac{\|A^{-1}E\| \cdot \|A^{-1}\|}{1 - \|A^{-1}E\|}$$

یا

$$\square \quad \frac{\|A^{-1} - (A - E)^{-1}\|}{\|A^{-1}\|} \leq \frac{\|A^{-1}E\|}{1 - \|A^{-1}E\|}$$

تذکره: قضیه ۱-۷-۸ در مفهوم برآورد بزرگ‌ترین خطا در معکوس یک ماتریس که عناصرش دقیقاً معلوم نیستند، اهمیت دارد.

۱-۸ خواص تغییرناپذیری نرم ماتریسهای متعامد

این فصل را با فهرست کردن برخی خواص خیلی مفید نرم ماتریسهای متعامد که اغلب در عمل مورد استفاده قرار می‌گیرند، به پایان می‌رسانیم.

قضیه ۱-۸-۱ فرض کنید O یک ماتریس متعامد باشد، آنگاه

$$\|O\|_2 = 1$$

اثبات:

$$\square \quad \|O\|_2 = \sqrt{O^T O} = \sqrt{\rho(O^T O)} = \sqrt{\rho(I)} = 1$$

قضیه ۲-۸-۱

$$\|AO\|_2 = \|A\|_2$$

اثبات:

$$\square \quad \|AO\|_2 = \sqrt{\rho(O^T A^T A O)} = \sqrt{\rho(A^T A)} = \|A\|_2$$

(توجه کنید که شعاع طیفی تحت تبدیل تشابهی تغییرناپذیر است. فصل ۸ را ملاحظه کنید.)

قضیه ۳-۸-۱

$$\|AO\|_F = \|A\|_F$$

اثبات:

$$\square \quad \|AO\|_F^2 = \text{trace}(O^T A^T A O) = \text{trace}(A^T A) = \|A\|_F^2$$

۹-۱ مرور و خلاصه

مفاهیم اساسی که برای فهم بقیه کتاب لازم می‌باشد به طور خلاصه در این فصل بیان شده‌اند. مهمترین آنها عبارتند از:

۱. ماتریسهای خاص: ماتریسهای قطری، مثلثی، متعامد، جایگشت، هسبرگی، سه قطری، غالب قطری و معین مثبت تعریف شدند و خواص آنها مورد بحث قرار گرفتند.

۲. نرمهای برداری و ماتریسی: برخی نرمهای ماتریسی مهم عبارتند از نرم مجموع-سطری، نرم مجموع-ستونی، نرم فروبنیوس و نرم طیفی. رابطه بین نرمهای ماتریسی مختلف در قضیه ۱-۷-۱ بیان و اثبات گردید.

خاصیت نرم ماتریسهای متعامد دارای اهمیت ویژه است. سه نتیجه ساده اما مهم در بخش ۸-۱ بیان و اثبات شدند (قضیه‌های ۱-۸-۱، ۲-۸-۱، ۳-۸-۱). این نتایج بیان می‌کنند که (الف) نرم طیفی یک ماتریس متعامد برابر ۱ است، و

(ب) نرمهای طیفی و فروبنیوس تحت ضرب ماتریسی تغییرناپذیر باقی می‌مانند.

۳. همگرایی یک دنباله ماتریسی: نظریه همگرایی دنباله توانهای ماتریسی $\{A^k\}$ در مطالعه همگرایی روشهای تکراری برای دستگاههای خطی مهم است. مهمترین نتایج در این رابطه عبارتند از:

(الف) دنباله $\{A^k\}$ به ماتریس صفر همگراست اگر و فقط اگر $\|A\| < 1$ به ازای هر مقدار ویژه λ_i از A (قضیه ۱-۷-۴).

(ب) دنباله $\{A^k\}$ به ماتریس صفر همگراست اگر $\|A\| < 1$ (نتیجه قضیه ۱-۷-۵).

۴. نرمها و معکوسها: اگر یک ماتریس نامنفرد A توسط یک ماتریس E اختلال یابد، بعضی اوقات برای ما جالب است بدانیم که آیا ماتریس اختلال یافته $A + E$ نامنفرد باقی می‌ماند و چگونه می‌توان خطا را در معکوس $A + E$ برآورد کرد. سه قضیه (قضیه‌های ۱-۷-۷، ۱-۷-۸ و ۱-۷-۹) در این زمینه در بخش ۱-۷-۴ ثابت شده‌اند. این نتایج یک نقش مهم در تجزیه و تحلیل اختلال دستگاههای خطی بازی می‌کنند (فصل ۶).

تمرینها

بخشهای ۱-۲ و ۱-۳

۱. هر کدام از گزاره‌های زیر را ثابت کنید.

(الف) یک مجموعه از n بردار مستقل خطی در \mathbb{R}^n یک پایه برای \mathbb{R}^n است.

- (ب) مجموعه $\{e_1, e_2, \dots, e_n\}$ یک پایه برای \mathbf{R}^n است.
- (ج) یک مجموعه از m بردار در \mathbf{R}^n ، که در آن $m > n$ ، وابسته خطی است.
- (د) هر دو پایه در \mathbf{R}^n ، دارای تعداد یکسان بردار هستند.
- (ه) $\text{span}\{v_1, \dots, v_k\}$ یک زیر فضا از \mathbf{R}^n است، که در آن $\text{span}\{v_1, \dots, v_k\}$ مجموع ترکیبهای خطی از k بردار v_1, \dots, v_k از یک فضای برداری \mathbf{R}^n می باشد.
- (و) $\text{span}\{v_1, \dots, v_k\}$ کوچکترین زیرفضایی از \mathbf{R}^n است که v_1, \dots, v_k را در بردارد.
۲. ثابت کنید که اگر $S = \{s_1, \dots, s_k\}$ یک مجموعه متعامد از بردارهای مخالف صفر باشد آنگاه S مستقل خطی است.
۳. فرض کنید S یک زیر فضای m بعدی از \mathbf{R}^n باشد. آنگاه ثابت کنید که S دارای یک پایه یکا متعامد است. (راهنمایی: فرض کنید $\{v_1, \dots, v_m\}$ یک پایه مفروض از S باشد. یک مجموعه بردار $\{u_k\}$ به صورت

$$u_1 = \frac{v_1}{\|v_1\|}$$

$$u_{k+1} = \frac{v'_{k+1}}{\|v'_{k+1}\|}$$

تعریف کنید که در آن

$$v'_{k+1} = v_{k+1} - (v_{k+1}^T u_1) u_1 - (v_{k+1}^T u_2) u_2 - \dots - (v_{k+1}^T u_k) u_k, \quad k = 1, 2, \dots, m-1$$

سپس نشان دهید که $\{u_1, \dots, u_m\}$ یک پایه یکا متعامد از S است. این فرایند کلاسیک گرام - اشمیت است.

۴. با استفاده از فرایند گرام - اشمیت یک پایه یکا متعامد از \mathbf{R}^3 بسازید.

۵. یک پایه یکا متعامد از $R(A)$ بسازید، که در آن

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 4 & 5 \end{bmatrix}$$

۶. فرض کنید S_1 و S_2 دو زیر فضا از \mathbf{R}^n باشند، ثابت کنید که

$$\dim(S_1 + S_2) = \dim(S_1) + \dim(S_2) - \dim(S_1 \cap S_2)$$

۷. قضیه ۱-۳-۱ را در مورد خواص دترمینان یک ماتریس ثابت کنید.

۸. هر یک از گزاره‌های زیر را ثابت کنید.

(الف) $\text{null}(A) = 0$ اگر و فقط اگر A دارای ستونهای مستقل خطی باشد.

(ب) $\text{rank}(A) = \text{rank}(A^T)$

(ج) $\text{rank}(A) + \text{null}(A) = n$

(د) اگر A یک ماتریس از مرتبه $m \times n$ باشد و $m < n$ ، آنگاه $\text{rank}(A) \leq m$.

(ه) اگر A و B ماتریسهای از مرتبه $m \times n$ و $n \times q$ باشند، آنگاه

$$\text{rank}(AB) \leq \min \{ \text{rank}(A), \text{rank}(B) \}$$

(و) هنگامی که ماتریس در یک ماتریس معکوس پذیر ضرب می‌شود، رتبه ماتریس بدون

تغییر باقی می‌ماند

(ز) اگر $B = UAV$ ، که در آن U و V معکوس پذیرند، آنگاه $\text{rank}(B) = \text{rank}(A)$.

(ح) $N(A) = R(A^T)^\perp$ و $R(A)^\perp = N(A^T)$

۹. فرض کنید A یک ماتریس $m \times n$ باشد، آنگاه A دارای رتبه ۱ است اگر و فقط اگر A را بتوان به

صورت $A = ab^T$ نوشت که در آن a و b بردارهای ستونی هستند.

۱۰. درستی روابط اساسی زیر را در مورد نامنفرد بودن و معکوس A ثابت کنید.

(الف) $(A^{-1})^{-1} = A$

(ب) $(A^T)^{-1} = (A^{-1})^T$

(ج) $(cA)^{-1} = \frac{1}{c}A^{-1}$ ، که در آن c یک اسکالر مخالف صفر است.

(د) $(AB)^{-1} = B^{-1}A^{-1}$

۱۱. فرض کنید که A را بتوان به صورت

$$A = LU$$

نوشت، که در آن L یک ماتریس پایین مثلثی با ۱ در امتداد قطر و $U = (u_{ij})$ یک ماتریس بالا

مثلثی است. ثابت کنید که

$$\det(A) = \prod_{i=1}^n u_{ii}$$

۱۲. فرض کنید $A = \begin{bmatrix} A_1 & 0 \\ 0 & A_p \end{bmatrix}$ ، که در آن A_1 و A_p ماتریسهای مربعی هستند. ثابت کنید که

$$\det(A) = \det(A_1)\det(A_p)$$

۱۳. فرض کنید که A را بتوان به صورت

$$A = LDL^T$$

نوشت که در آن L یک ماتریس پایین مثلثی با عناصر قطری ۱ و $D = \text{diag}(d_{11}, \dots, d_{nn})$ یک ماتریس قطری است. ثابت کنید که کهادهای اصلی پیشرو (دترمینانهای) A عبارتند از: $d_{11}, \dots, d_{11}d_{22}, \dots, d_{11} \dots d_{nn}$

$$d_{11} \dots d_{nn}, \dots, d_{11}d_{22}$$

۱۴. (الف) نشان دهید که اگر P_S یک عملگر تصویر متعامد بر روی S باشد، آنگاه $I - P_S$ یک عملگر تصویر متعامد بر روی S^\perp خواهد بود.

(ب) هر یک از گزاره‌های زیر را ثابت کنید.

$$P_A = A(A^T A)^{-1} A^T \quad (\text{i})$$

$$P_N = I - A(A^T A)^{-1} A^T \quad (\text{ii})$$

(ج) هر یک از گزاره‌های زیر را ثابت کنید.

$$b_R = P_A b \quad (\text{i})$$

$$b_N = P_N b \quad (\text{ii})$$

۱۵. (الف) P_A و P_N را برای ماتریسهای زیر پیدا کنید.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 \\ 10^{-4} & 0 \\ 0 & 10^{-4} \end{bmatrix}$$

(ب) به ازای بردار $b = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ ، b_R و b_N را برای هر یک از ماتریسهای قبل بیابید.

(ج) یک پایه متعامد برای هر یک از ماتریسهای قبل با استفاده از فرایند گرام - اشمیت به دست آورید و سپس P_A ، P_N ، b_R و b_N را پیدا کنید، برای توصیفی از الگوریتم گرام - اشمیت فصل ۷ یا تمرین ۳ این فصل را ملاحظه کنید.

۱۶. فرض کنید A یک ماتریس $m \times n$ و از رتبه r باشد. تجزیه مقدار تکین A را ملاحظه کنید:

$$A = U \Sigma V^T = (U_r, \hat{U}_r) \Sigma (V_r, \hat{V}_r)^T$$

که در آن $U_r = (u_1, \dots, u_r)$ ، $V_r = (v_1, \dots, v_r)$ و Σ یک ماتریس قطری $m \times n$ باشد. آنگاه

گزاره‌های زیر را ثابت کنید.

(الف) $V_r V_r^T$ تصویر قائم بر روی $N(A)^\perp = R(A^T)$ است.

(ب) $U_r U_r^T$ تصویر قائم بر روی $R(A)$ است.

(ج) $(\hat{U}_r)(\hat{U}_r)^T$ تصویر قائم بر روی $N(A^T) = R(A)^\perp$ است.

(د) $(\hat{V}_r)(\hat{V}_r)^T$ تصویر قائم بر روی $N(A)$ است.

۱۷. (فاصله دو زیر فضا). فرض کنید S_1 و S_2 دو زیر فضا از R^n باشند به قسمی که

$\dim(S_1) = \dim(S_2)$. فرض کنید P_1 و P_2 به ترتیب تصویرهای قائم بر روی S_1 و S_2 باشند

آنگاه $\|P_1 - P_2\|_F$ برابر فاصله بین S_1 و S_2 تعریف می‌شود. ثابت کنید که فاصله S_1 و S_2

برابر $\text{dist}(S_1, S_2) = \sin(\theta)$ می‌باشد، که در آن θ زاویه بین S_1 و S_2 است.

($\|A\|_F = \sqrt{A^T A}$ مقدار ویژه A)

۱۸. ثابت کنید که اگر P_S تصویر قائم بر روی S باشد، آنگاه $I - 2P_S$ متعامد است.

بخشهای ۱-۴ تا ۱-۶

۱۹. گزاره‌های زیر را ثابت کنید.

(الف) حاصلضرب دو ماتریس بالا (پایین) مثلثی یک ماتریس بالا (پایین) مثلثی است. در

حالت کلی، اگر $A = (a_{ij})$ یک ماتریس بالا مثلثی باشد، آنگاه $p(A)$ ، که در آن $p(x)$ یک

چندجمله‌ای است، یک ماتریس بالا (پایین) مثلثی است که عناصر قطری آن $p(a_{ii})$ ،

$i = 1, \dots, n$ هستند.

(ب) معکوس یک ماتریس پایین (بالا) مثلثی، ماتریس پایین (بالا) مثلثی دیگری است که

عناصر قطری آن معکوس عناصر قطری ماتریس مثلثی هستند.

(ج) دترمینان یک ماتریس مثلثی حاصلضرب عناصر قطری آن است.

(د) مقادیر ویژه یک ماتریس مثلثی عناصر قطری آن هستند.

(ه) اگر $A \in R^{n \times n}$ بالا مثلثی اکید باشد، آنگاه $A^n = 0$. $A = (a_{ij})$ بالا مثلثی اکید

است اگر بالا مثلثی باشد و $a_{ii} = 0$ به ازای هر i .

(و) معکوس یک ماتریس نامنفرد A را می‌توان به صورت یک چندجمله‌ای برحسب A نوشت

(از قضیه کیلی - هامیلتون استفاده کنید).

۲۰. ثابت کنید که حاصلضرب یک ماتریس بالا هسنبرگی و یک ماتریس بالا مثلثی یک ماتریس

بالا هسنبرگی است.

۲۱. ثابت کنید که یک ماتریس هسنبرگ متقارن سه قطری متقارن است.

۲۲. یک ماتریس مربعی $A = (a_{ij})$ یک ماتریس نواری با عرض نواری $1 + 2k$ است اگر $|i - j| > k$ ایجاب کند $a_{ij} = 0$. عرضهای نواری ماتریسهای سه قطری و پنج قطری چند هستند؟ آیا حاصلضرب دو ماتریس نواری دارای عرض نواری یکسان یک ماتریس نواری با همان عرض نواری است؟ دلایلی برای جواب خود ارائه دهید.

۲۳. (الف) نشان دهید که ماتریس

$$H = I - 2 \frac{uu^T}{u^T u}$$

که در آن u یک بردار ستونی است، متعامد است (ماتریس H یک ماتریس هاوس هولدر نامیده می شود).

(ب) نشان دهید که ماتریس $J = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}$ ، که در آن $c^2 + s^2 = 1$ ، متعامد است. (ماتریس J یک ماتریس گیونز نامیده می شود).

(ج) ثابت کنید که حاصلضرب دو ماتریس متعامد یک ماتریس متعامد است.

(د) ثابت کنید که یک ماتریس مثلثی که متعامد است قطری می باشد.

۲۴. فرض کنید A و B دو ماتریس متقارن باشند.

(الف) ثابت کنید که $(A + B)$ متقارن است.

(ب) ثابت کنید که AB لزوماً متقارن نیست. شرطی به دست آورید که تحت آن AB متقارن باشد.

(ج) اگر A و B معین مثبت متقارن باشند، ثابت کنید که $(A + B)$ معین مثبت است. آیا AB

معین مثبت است؟ دلایلی برای جواب خود ارائه دهید. چه موقع $(A - B)$ معین مثبت متقارن است؟

۲۵. فرض کنید $A = (a_{ij})$ یک ماتریس معین مثبت متقارن از مرتبه $n \times n$ باشد. گزاره های زیر را ثابت کنید.

(الف) هر عنصر قطری A باید مثبت باشد.

(ب) A نامنفرد است.

(ج) $a_{ii}a_{jj} < (a_{ij})^2$ به ازای $i, j = 1, 2, \dots, n$ ، $i \neq j$.

(د) بزرگترین عنصر باید بر روی قطر قرار داشته باشد.

۲۶. فرض کنید A یک ماتریس معین مثبت متقارن و x یک بردار n بعدی مخالف صفر باشد. ثابت

کنید که $A + xx^T$ معین مثبت است.

۲۷. ثابت کنید که یک ماتریس غالب قطری نامنفرد است، و یک ماتریس غالب قطری متقارن با عناصر قطری مثبت یک ماتریس معین مثبت است.

۲۸. ثابت کنید که اگر مقادیر ویژه یک ماتریس همگی متمایز باشند، آنگاه A ناموهن است.

۲۹. ثابت کنید که یک ماتریس متقارن A معین مثبت است اگر و فقط اگر A^{-1} وجود داشته و معین مثبت باشد.

۳۰. فرض کنید A یک ماتریس $m \times n$ ($m \geq n$) با رتبه کامل باشد. آنگاه $A^T A$ معین مثبت است.

۳۱. حقیقتهای اساسی زیر را در مورد مقادیر ویژه، و بردارهای ویژه ثابت کنید.

(الف) یک ماتریس A نامنفرد است اگر و فقط اگر A دارای یک مقدار ویژه صفر نباشد

$$(\text{راهنمایی: } \det(A) = \lambda_1 \lambda_2 \dots \lambda_n)$$

(ب) مقادیر ویژه A^T و A یکسان هستند.

(ج) اگر دو ماتریس دارای مقادیر ویژه یکسان باشند، آنها لازم نیست مشابه باشند (با مثالی که خودتان می‌سازید، این مطلب را نشان دهید).

(د) یک ماتریس متقارن معین مثبت است اگر و فقط اگر همه مقادیر ویژه آن مثبت باشند.

(ه) مقادیر ویژه یک ماتریس مثلثی عناصر قطری آن هستند.

(و) مقادیر ویژه یک ماتریس متعامد دارای مدول ۱ هستند.

(ز) فرض کنید A یک ماتریس متقارن باشد و فرض کنید Q متعامد بوده و $Q^T A Q$ قطری باشد. آنگاه نشان دهید که ستونهای Q بردارهای ویژه A هستند.

(ر) بردارهای ویژه یک ماتریس متقارن می‌توانند طوری انتخاب شوند که یکا متعامد باشند.

۳۲. فرض کنید A یک ماتریس متقارن با مقادیر ویژه $\lambda_1, \lambda_2, \dots, \lambda_n$ و بردارهای ویژه یکا متعامد v_1, v_2, \dots, v_n باشد. آنگاه نشان دهید که:

$$A = \lambda_1 v_1 v_1^T + \lambda_2 v_2 v_2^T + \dots + \lambda_n v_n v_n^T$$

۳۳. فرض کنید H یک ماتریس بالا هسنبرگی کاهش‌ناپذیر باشد. فرض کنید $(x_1, \dots, x_n) = X$ به صورت زیر تعریف شود.

$$x_1 = e_1 = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}, \quad x_{i+1} = Hx_i, \quad i = 1, 2, \dots, n-1$$

آنگاه ثابت کنید که X نامنفرد است و $X^{-1}HX$ یک ماتریس همراه است (ماتریس همراه بالا هسنبرگی).

۳۴. مقادیر تکین یک ماتریس متقارن چه هستند؟ مقادیر تکین یک ماتریس معین مثبت متقارن چه هستند؟ ثابت کنید که یک ماتریس مربعی A نامنفرد است اگر و فقط اگر دارای مقادیر تکین صفر نباشد.

۳۵. فرض کنید $\text{trace}(A) = \sum_{i=1}^n a_{ii}$. آنگاه گزاره‌های زیر را ثابت کنید.

$$\text{trace}(AB) = \text{trace}(BA) \quad (\text{الف})$$

$$\text{trace}(AA^T) = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \quad (\text{ب})$$

که در آن $A = (a_{ij})$ یک ماتریس $m \times n$ است.

$$\text{trace}(A+B) = \text{trace}(A) + \text{trace}(B) \quad (\text{ج})$$

$$\text{trace}(TAT^{-1}) = \text{trace}(A) \quad (\text{د})$$

بخشهای ۷-۱ و ۸-۱

۳۶. نشان دهید که $\|x\|_1, \|x\|_2, \|x\|_\infty$ (به صورت تعریف شده در بخش ۷-۱) نرمهای برداری هستند.

۳۷. نشان دهید که اگر x و y دو بردار باشند، آنگاه

$$|\|x\| - \|y\|| \leq \|x - y\| \leq \|x\| + \|y\|$$

۳۸. اگر x و y دو بردار n بعدی باشند، آنگاه ثابت کنید که

$$|x^T y| \leq \|x\|_2 \|y\|_2 \quad (\text{الف}) \quad (\text{نامساوی کوشی - شوارتز})$$

$$\|xy^T\|_2 = \|x\|_2 \|y\|_2 \quad (\text{ب})$$

۳۹. فرض کنید x و y دو بردار متعامد باشند، آنگاه نشان دهید که:

$$\|x+y\|_2^2 = \|x\|_2^2 + \|y\|_2^2$$

۴۰. ثابت کنید که برای هر بردار x داریم:

$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$$

۴۱. ثابت کنید که $\|A\|_1, \|A\|_\infty, \|A\|_2$ نرمهای ماتریسی هستند.

۴۲. فرض کنید $A = (a_{ij})$ یک ماتریس $m \times n$ باشد. تعریف کنید $A_1 = \max_{ij} |a_{ij}|$. آیا A_1 یک نرم ماتریسی است؟ دلایلی برای جواب خودتان ارائه دهید.

۴۳. (الف) ثابت کنید که طول بردار توسط ضرب ماتریس متعامد حفظ می‌شود. یعنی اگر $x \in \mathbb{R}^n$ و $Q \in \mathbb{R}^{n \times n}$ متعامد باشد، آنگاه $\|Qx\|_2 = \|x\|_2$ (لم ایزومتري).

(ب) اگر نرمهای ۱ و بی نهایت استفاده شوند آیا مطلب ذکر شده در قسمت (الف) درست است؟ دلایلی ارائه دهید. اگر نرم فروبنیوس استفاده شود آیا مطلب مذکور درست است.

۴۴. ثابت کنید که $\|I\|_2 = 1$. ثابت کنید که $\|I\|_F = \sqrt{n}$

۴۵. ثابت کنید که اگر Q و P ماتریسهای متعامد باشند، آنگاه

$$\|QAP\|_F = \|A\|_F \quad (\text{الف})$$

$$\|QAP\|_2 = \|A\|_2 \quad (\text{ب})$$

۴۶. ثابت کنید که نرم طیفی یک ماتریس متقارن همان شعاع طیفی آن می‌باشد.

۴۷. فرض کنید $A \in \mathbb{R}^{n \times n}$ و x, y و z بردارهای n بعدی باشند به قسمی که $Ax = b$ و

$$Ay = b + z \quad \text{آنگاه نشان دهید که: } \|x - y\|_2 \leq \|A^{-1}\|_2 \|z\|_2 \quad (\text{با فرض}$$

این که A^{-1} وجود دارد).

۴۸. ثابت کنید که $\|A\|_2 \leq \|A\|_F$ (از نامساوی کوشی - شوارتز استفاده کنید).

۴۹. ثابت کنید که $\|A\|_2$ دقیقاً بزرگترین مقدار تکین A است. چگونه $\|A^{-1}\|_2$ به یک مقدار تکین A وابسته است.

$$\|A^T A\|_2 = \|A\|_2^2 \quad \text{ثابت کنید}$$

۵۰. ثابت کنید

$$\|AB\|_F \leq \|A\|_F \|B\|_F$$

$$\|AB\|_F \leq \|A\|_2 \|B\|_F$$

۵۲. فرض کنید $A = (a_1, \dots, a_n)$ ، که در آن a_j ، زمین ستون A است. آنگاه ثابت کنید که

$$\|A\|_F^2 = \sum_{i=1}^n \|a_i\|_2^2$$

۵۳. ثابت کنید که اگر A و $A + E$ هر دو نامنفرد باشند، آنگاه

$$\|(A + E)^{-1} - A^{-1}\| \leq \|E\| \|A^{-1}\| \|(A + E)^{-1}\|$$

(لم با ناخ). پیامد این نتیجه چیست؟

۵۴. فرض کنید $A \in \mathbb{R}^{m \times n}$ دارای رتبه کامل باشد. ثابت کنید که $A + E$ همچنین دارای رتبه

کامل است اگر E به قسمی باشد که، $\|E\|_2 < \frac{1}{\|A^+\|_2}$ ، که در آن $A^+ = (A^T A)^{-1} A^T$.

۵۵. نشان دهید که ماتریسهای

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 4 & 5 & 6 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 4 \\ 0 & 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}$$

ماتریسهای همگرا نیستند.

۵۶. یک مثال ساده بسازید که در آن آزمون نرم برای ماتریسهای همگرا باشکست مواجه گردد، اما

ماتریس باز هم همگرا باشد.

۵۷. ثابت کنید که سری $(I + A + A^2 + \dots)$ همگرا است اگر $\|B\| < 1$ ، که در آن $B = PAP^{-1}$.

پیامد این نتیجه چیست؟ مثال ساده‌ای بسازید که فایده این نتیجه در محاسبات عملی دیده

شود (برای جزئیات صفحه ۶۰ کتاب Wilkinson, The Algebraic Eigenvalue Problem

را ملاحظه کنید).

پیشنهادها برای مطالعه بیشتر

مواد ارائه شده در این فصل را می‌توان در هر کتاب استاندارد در مورد جبر خطی و

نظریه ماتریس ملاحظه کرد، بویژه کتابهای زیر را برای مطالعه بیشتر پیشنهاد

می‌کنیم.

Franklin, Joel N. *Matrix Theory*. Englewood Cliffs, N.J.: Prentice Hall, 1968.

Hill, Richard O., Jr. *Elementary Linear Algebra with Applications*. Orlando, Fla.:

Harcourt Brace Jovanovich, 1991.

- Horn, Roger and Charles Johnson. *Matrix Analysis*. New York: Cambridge University Press, 1985.
- Householder, A. S. *The Theory of Matrices in Numerical Analysis*. New York: Dover Publications, 1964.
- Lancaster, Peter. *The Theory of Matrices*. New York: Academic Press, 1969.
- Lancaster, Peter and M. Tismenetsky. *The Theory of Matrices with Applications*. 2d ed. Orlando, Fla.: Academic Press, 1985.
- Lay, David C. *Linear Algebra and Its Applications*. Reading, Mass.: Addison-Wesley, 1994.
- Leon, Steven J. *Linear Algebra with Applications*. New York: Macmillan, 1994.
- Noble, Ben and James Daniel. *Applied Linear Algebra*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1988.
- Schneider, Hans and George Philip Barker. *Matrices and Linear Algebra*. New York: Dover Publications, 1989.
- Strang, Gilbert. *Introduction to Linear Algebra*. Wellesley, Mass.: Wellesley-Cambridge Press, 1993.
- Strang, Gilbert. *Linear Algebra and Its Applications*. 3rd ed. San Diego: Harcourt Brace Jovanovich, 1988.
- Wilkinson, James H. *The Algebraic Eigenvalue Problem*. Oxford: Clarendon Press, 1965, Chapter 1.

اعداد ممیز شناور و خطاها در محاسبات

۲-۱ دستگاه اعداد ممیز شناور

به علت گنجایش محدود ذخیره سازی، یک عدد حقیقی ممکن است که به طور دقیق بر روی یک کامپیوتر نمایش داده شود یا این کار ممکن نباشد. بنابراین، به هنگام استفاده از یک کامپیوتر، ما با تقریبهای دستگاه اعداد حقیقی با استفاده از نمایشهای کامپیوتری متناهی سر و کار داریم. این فصل به مطالعه چنین اعداد تقریبی محدود خواهد بود. بویژه IEEE استاندارد برای حساب ممیز شناور دودویی (IEEE, 1985) را که به طور وسیعی استفاده می شود و مورد قبول ایستگاههای کاری (نظیر IBM، SUN، کامپیوترهای شخصی و غیره) می باشد، امتحان خواهیم کرد.

یک عدد ممیز شناور نرمال شده مخالف صفر x در مبنای ۲ دارای شکل زیر است.

$$(-1)^s d_1 d_2 d_3 \dots d_t 2^e$$

$$x = \pm d_1 d_2 \dots d_t 2^e$$

یا

$$x = \pm r 2^e$$

که در آن e نما و r ماتیس نامیده می شوند و t دقت، $(-1)^s$ علامت و $(d_1 d_2 \dots d_t)$ قسمت کسری

عدد هستند. (توجه کنید که t متناهی است).

$$d_1 = 1$$

$$d_i = 0 \text{ یا } 1, \quad 2 \leq i \leq t$$

سه پارامتر همه مقادیر عددی را که می‌توانند نمایش داده شوند، مشخص می‌کنند. این سه پارامتر عبارتند از دقت، t ، و می‌نیمم و ماکزیمم نماها، L و U . اعداد L و U در کامپیوترها تغییر می‌کنند حتی آنهایی که از IEEE استاندارد تبعیت می‌کنند، زیرا استاندارد فقط می‌نیممها را توصیه می‌کند. به عنوان مثال، استاندارد برای دقت ساده $t = 24$ ، $L = -126$ و $U = 127$ را توصیه می‌کند. توصیه برای دقت مضاعف $t = 53$ ، $L = -1022$ و $U = 1023$ است. برای مثال یک کلمه ۳۲ بیتی را در نظر بگیرید.

۱	۸	۲۳
s	e	f

در اینجا s علامت عدد، e دامنه برای نما و f قسمت کسری است.

توجه کنید که برای اعداد ممیز شناور نرمال شده در مبنای ۲، $d_1 = 1$ معلوم می‌باشد و در نتیجه d_1 می‌تواند به طور ضمنی ذخیره شود.

ذخیره‌سازی واقعی نما توسط ذخیره سازی نمای درست بعلاوه یک انحراف انجام می‌شود. انحراف به نحوی انتخاب می‌شود که e همواره نامنفی باشد. IEEE استاندارد نیز نیاز به این دارد که نمای انحراف نیافته دارای دو مقدار نادیده $1 - L$ و $1 + U$ باشد. $L - 1$ برای کد کردن $0 \pm$ و اعداد نرمال نشده (یعنی آن اعدادی که برای آنها $d_1 \neq 1$) استفاده می‌شود. $1 + U$ برای کد کردن $\pm \infty$ و غیر اعداد، نظیر $(-\infty) + (+\infty)$ ، که توسط NaN نمایش داده می‌شوند به کار می‌رود.

توجه کنید که برای مثال دقت ساده که اکنون ارائه شد، انحراف ۱۲۷ است. بنابراین اگر نمای انحراف یافته ۲۵۵ باشد آنگاه $\pm \infty$ (یا NaN) استنتاج می‌شود. همین‌طور اگر نمای انحراف یافته برابر 0 باشد آنگاه $0 \pm$ (یا عدد نرمال نشده) استنتاج می‌شود. استاندارد مشخص می‌کند که چگونه باید حالت‌های مختلف را برای وضعیت‌های خاص تعیین کرد. در اینجا لازم نیست که وارد چنین جزئیاتی شویم، خواننده‌های کنجکاو می‌توانند به IEEE(1985) مراجعه نمایند.

از بحث بالا ملاحظه می‌کنیم که IEEE استاندارد برای دقت ساده تقریباً دقتی با ۷ رقم اعشار فراهم می‌سازد زیرا $10^{-7} \times 1/2 \cong 2^{-23}$. به طور مشابه دقت مضاعف تقریباً دقتی با ۱۶ رقم اعشار فراهم

می سازد ($2^{-52} \cong 2/2 \times 10^{-16}$).

همچنین یک IEEE استاندارد برای اعداد ممیز شناوری که لزوماً از مبنای ۲ نیستند، وجود دارد. با اجازه دادن به انتخاب یک پایه، مثلاً β ، ملاحظه می‌کنیم که مجموعه همهٔ اعداد ممیز شناور، که دستگاه اعداد ممیز شناور نامیده می‌شود، توسط چهار پارامتر مشخص می‌گردد:

β عدد پایه

t دقت

L و U حدهای پایین و بالای نما

مجموعه اعداد ممیز شناور نرمال شده با دقت t را توسط F_t نمایش می‌دهیم. در F_t ، $d_1 \neq 0$. مجموعه F_t تحت اعمال حسابی بسته نیست؛ یعنی مجموع، تفاضل، ضرب، یا خارج قسمت دو عدد ممیز شناور در F_t لزوماً یک عدد در F_t نیست. برای دیدن این مطلب، مثال ساده‌ای را در دستگاه ممیز شناور با $\beta = 10$ ، $t = 3$ ، $L = -1$ و $U = 2$ ملاحظه می‌کنیم:

$$a = 11/2 = 1/12 \times 10^1$$

$$b = 1/13 = 1/13 \times 10^0$$

حاصلضرب $10^1 \times 1/2656 = 12/656 = c$ در F_t نیست. این مثال نشان می‌دهد که در طول یک محاسبه، یک عدد محاسبه شده ممکن است در خارج F_t قرار گیرد.

البته دواره وجود دارد که یک عدد بتواند خارج از دامنه F_t قرار گیرد؛ اولاً، نمای عدد ممکن است در خارج از بازه $[L, U]$ افتد، و ثانیاً قسمت کسری ممکن است بیش از $(t - 1)$ رقم داشته باشد (این دقیقاً حالتی است که در این مثال اتفاق افتاد).

اگر محاسبات در یک کامپیوتر یک نمای خیلی بزرگ (خیلی کوچک) تولید کنند آنگاه وضعیت سرریز (یا پاریز) نامیده می‌شود. سرریز یک مسأله جدی است، برای اغلب دستگاهها نتیجه یک سرریز $\pm \infty$ است. پاریز معمولاً خیلی جدی نیست. نتیجه یک پاریز ممکن است برابر قرار دادن یک مقدار با صفر، یا یک عدد نرمال نشده، یا $\pm 2^L$ باشد.

مثال ۱-۱-۲ سرریز و پاریز

۱. فرض کنید $\beta = 10$ ، $t = 3$ ، $L = -3$ ، $U = 3$.

$$a = 1/11 \times 10^3$$

$$b = 1/20 \times 10^3$$

$$c = a \times b = 1/33 \times 10^6$$

این حاصلضرب یک سرریز نتیجه می دهد، زیرا نما ۶ خیلی بزرگ است.

$$۲. \text{ فرض کنید } U = ۳, L = -۲, t = ۳, \beta = ۱۰$$

$$a = ۱/۰ \times ۱۰^{-۲}$$

$$b = ۲/۰ \times ۱۰^{-۲}$$

$$c = ab = ۲ \times ۱۰^{-۴}$$

این حاصلضرب یک پاریز نتیجه می دهد.

برخی محاسبات ریاضی نظیر پیدا کردن یک ریشه، پیدا کردن توانی از یک عدد، یا محاسبه کسرها می توانند سرریز نتیجه دهند. برای مثال، محاسبه $c = \sqrt{a^2 + b^2}$ را در نظر بگیرید. اگر a یا b خیلی بزرگ باشند آنگاه به هنگام محاسبه $a^2 + b^2$ سرریز خواهیم داشت.

IEEE استاندارد نتایج عملیات را با بینهایت ها و NaN ها در معرض دید قرار می دهد. همه اعمال با بی نهایت ها با حالت حدی در آنالیز حقیقی متناظر هستند. نتیجه وضعیتهای مبهمی نظیر $0 \cdot \infty$ ، NaN می باشد، و نتیجه همه اعمال دودویی با یک یا دو NaN، یک NaN می باشد.

گاهی اوقات با انجام محاسبات به گونه ای متفاوت، می توان از سرریزی و پاریزی پرهیز نمود. برای مثال عمل محاسبه طول یک بردار n بعدی x را با مؤلفه های x_1, \dots, x_n در نظر بگیرید.

$$\|x\|_2^2 = x_1^2 + x_2^2 + \dots + x_n^2$$

اگر x_i خیلی کوچک یا خیلی بزرگ باشد ما می توانیم سرریز یا پاریز از طریق معمول محاسبه $\|x\|_2$ داشته باشیم. ولیکن اگر هر مؤلفه را با تقسیم آن بر $m = \max(|x_1|, \dots, |x_n|)$ نرمال کنیم و سپس مجذور نموده و جمع نماییم، آنگاه می توانیم از مشکلات سرریز پرهیز نماییم. بنابراین یک راه بهتر برای محاسبه $\|x\|_2$ به صورت زیر است.

$$۱. m = \max(|x_1|, \dots, |x_n|)$$

$$۲. y_i = x_i/m, i = 1, \dots, n$$

$$۳. \|x\|_2 = m \sqrt{y_1^2 + y_2^2 + \dots + y_n^2}$$

۲-۲ خطاهای گرد کردن

اگر یک نتیجه محاسبه شده از یک عدد حقیقی مفروض قابل نمایش در ماشین نباشد، آنگاه به دو طریق

می تواند در ماشین نمایش داده شود. عدد

$$\pm 0.d_1 d_2 \dots d_t d_{t+1} \dots$$

را ملاحظه کنید. اولین روش، برش زدن، روشی است که در آن ارقام بعد از d_t حذف می شوند. دومین روش گرد کردن است که در آن ارقام بعد از d_t نه فقط حذف می شوند بلکه d_t به طرف بالا یا پایین گرد می شود، بسته به این که $d_{t+1} \geq \beta/2$ یا $d_{t+1} < \beta/2$.

فرض کنید $\text{fl}(x)$ نمایش ممیز شناور یک عدد حقیقی x را نمایش دهد.

مثال ۲-۲-۱ گرد کردن: مبنای ۱۰ را در نظر بگیرید. فرض کنید $x = 3/141596$,

$$\begin{array}{ccc} t = 2 & t = 3 & t = 4 \\ \text{fl}(x) = 3/1 & \text{fl}(x) = 3/14 & \text{fl}(x) = 3/142 \end{array}$$

اکنون عبارتی برای اندازه گیری خطای به وجود آمده برای نمایش یک عدد حقیقی x بر روی کامپیوتر ارائه می دهیم و سپس نشان می دهیم که چگونه این اندازه می تواند برای ارائه کرانهایی برای خطاها در سایر محاسبات ممیز شناور مورد استفاده قرار گیرد.

تعریف ۲-۲-۱ فرض کنید \hat{x} تقریبی از x را نمایش دهد؛ آنگاه دو راه وجود دارد که می توانیم خطا را اندازه گیری کنیم.

$$\begin{aligned} \text{خطای مطلق} &= |\hat{x} - x| \\ \text{خطای نسبی} &= \frac{|\hat{x} - x|}{|x|}, \quad x \neq 0 \end{aligned}$$

توجه کنید که خطای نسبی معنای بیشتری از خطای مطلق دارد. مثال ساده زیر این مطلب را نشان می دهد.

مثال ۲-۲-۲ مقادیر

$$\begin{aligned} x_1 &= 1/31, & \hat{x}_1 &= 1/30 \\ x_2 &= 0/12, & \hat{x}_2 &= 0/11 \end{aligned}$$

را در نظر بگیرید. خطای مطلق در هر دو حالت یکسان هستند:

$$|\hat{x}_1 - x_1| = |\hat{x}_2 - x_2| = 0.01$$

از طرف دیگر، خطای نسبی در حالت اول عبارت است از:

$$\frac{|\hat{x}_1 - x_1|}{|x_1|} = 0.0076$$

و خطای نسبی در حالت دوم عبارت از:

$$\frac{|\hat{x}_2 - x_2|}{|x_2|} = 0.0833$$

بنابراین خطاهای نسبی نشان می‌دهند که \hat{x}_1 به x_1 نزدیک‌تر از \hat{x}_2 به x_2 است، در حالی که خطاهای مطلق هیچ نشانی در این زمینه ارائه نمی‌دهند.

خطای نسبی نشانی از تعداد ارقام بامعنی در یک جواب تقریبی ارائه می‌دهد. به طور مشخص‌تر تعریف زیر را داریم.

تعریف ۲-۲-۲ گفته می‌شود \hat{x} تقریبی برای x با s رقم بامعنی است اگر s بزرگ‌ترین عدد صحیح نامنفی باشد که به ازای آن خطای نسبی در رابطه $(10^{-s}) < |x - \hat{x}| / |x|$ صدق نماید.

بنابراین در مثالهای فوق \hat{x}_1 و x_1 با دو رقم بامعنی با یکدیگر مطابقت می‌نمایند، در حالی که \hat{x}_2 و x_2 در حدود یک رقم بامعنی با یکدیگر مطابقت می‌نمایند.

اکنون عبارتی برای خطای نسبی نمایش یک عدد حقیقی x توسط نمایش ممیز شناورش $\text{fl}(x)$ ارائه می‌دهیم.

قضیه ۱-۲-۲ فرض کنید $\text{fl}(x)$ نمایش ممیز شناور یک عدد حقیقی x را نشان دهد. آنگاه

$$\frac{|\text{fl}(x) - x|}{|x|} \leq \mu = \begin{cases} \beta^{1-t} & \text{برای برش زدن} \\ \frac{1}{\beta} \beta^{1-t} & \text{برای گرد کردن} \end{cases} \quad (1-2-2)$$

اثبات: ما کران را برای گرد کردن ثابت می‌کنیم و قسمت دیگر را به عنوان تمرین ۱ ب باقی می‌گذاریم. فرض کنید x به صورت

$$x = (0.d_1 d_2 \dots d_t d_{t+1} \dots) \times \beta^e$$

نمایش داده شود که در آن $d_1 \neq 0$ و $d_i < \beta$ و $0 \leq d_i$. هنگامی که x را گرد می‌کنیم، یکی از اعداد ممیز شناور زیر را به دست می‌آوریم.

$$x' = (0.d_1 d_2 \dots d_t) \times \beta^e$$

$$x'' = [(0.d_1 d_2 \dots d_t) + \beta^{-t}] \times \beta^e$$

واضح است که داریم $x \in (x', x'')$. فرض کنید، بدون از دست دادن کلیات، که x به x' نزدیکتر باشد. پس داریم

$$|x - x'| \leq \frac{1}{2} |x' - x''| = \frac{1}{2} \beta^{e-t}$$

بنابراین خطای نسبی عبارت است از:

$$\frac{|x - x'|}{|x|} \leq \frac{1}{2} \frac{\beta^{-t}}{0.d_1 d_2 \dots d_t \dots}$$

$$\leq \frac{1}{2} \frac{\beta^{-t}}{1/\beta} \quad (\text{چون } d_i < \beta)$$

$$= \frac{1}{2} \beta^{1-t}$$

مثال ۲-۲-۳ نمایش سه رقمی عدد اعشاری $x = 0.2346$ را ملاحظه کنید ($\beta = 10, t = 3$). اگر گرد کردن مورد استفاده قرار گیرد، داریم:

$$\text{fl}(x) = 0.235$$

$$\text{خطای نسبی} \approx 0.001705 < \frac{1}{2} 10^{-2}$$

اگر برش زدن استفاده شود، داریم:

$$\text{fl}(x) = 0.234$$

$$\text{خطای نسبی} \approx 0.0025575 < 10^{-2}$$

تعریف ۲-۲-۳ عدد μ در $(1 - 2^{-t}, 1)$ دقت ماشین یا واحد خطای گرد کردن نامیده می‌شود. این کوچکترین عدد ممیز شناور مثبتی است که

$$\text{fl}(1 + \mu) > 1$$

μ معمولاً از مرتبه 10^{-16} و 10^{-6} (برروی بیشتر ماشینها) به ترتیب برای دقت مضاعف و ساده است. برای IBM ۳۶۰ و IBM ۳۷۰، $\beta = 16$ و $t = 6$ و $\mu = 4/77 \times 10^{-7}$ است.

دقت ماشین در محاسبات علمی خیلی مهم است. اگر مقادیر β و t ، L و U برای یک کامپیوتر معلوم نباشند، برنامه فرترن زیر می تواند برای برآورد μ برای آن کامپیوتر مورد استفاده قرار گیرد. (Forsythe, Malcolm, and Moler 1977, 14)

```
Real MEU, MEU_1, SAVE
MEU = 1.0
MEU_1 = 2.0
DO WHILE (MEU_1 .GT. 1.0)
    SAVE = MEU
    MEU = 0.5*MEU
    MEU_1 = MEU+ 1.0
END DO
MEU = SAVE
PRLINT *, ' MACHINE EPSLION IS ', MEU
```

این برنامه فرترن تقریبی از μ را محاسبه می کند که حداکثر با مضربی از ۲ با μ تفاوت دارد. این تقریب کاملاً قابل قبول است، زیرا یک مقدار دقیق از μ چندان مهم نیست و بندرت مورد نیاز است.

کتاب Forsythe, Malcolm, and Moler (1977) همچنین حاوی یک فهرست وسیع از U و L برای کامپیوترهای مختلف می باشد.

۳-۲ قوانین حساب ممیز شناور

فرمول

$$\frac{|fl(x)-x|}{|x|} \leq \mu = \begin{cases} \beta^{1-t} & \text{برای برش زدن} \\ \frac{1}{4} \beta^{1-t} & \text{برای گرد کردن} \end{cases}$$

را می توان به صورت زیر نوشت:

$$fl(x) = x(1 + \delta) \quad |\delta| \leq \mu$$

با فرض این که IEEE استاندارد در دست باشد، به آسانی می توان قوانین ساده حساب ممیز شناور زیر را به دست آورد.

قضیه ۲-۳-۱ فرض کنید x و y دو عدد ممیز شناور باشند و فرض کنید $\text{fl}(x+y)$ ، $\text{fl}(x-y)$ ، $\text{fl}(xy)$ و $\text{fl}(x/y)$ به ترتیب مجموع، تفاضل، حاصلضرب و خارج قسمت محاسبه شده را نمایش دهند. آنگاه

$$1. \quad |\delta| \leq \mu \text{ که در آن } \text{fl}(x \pm y) = (x \pm y)(1 + \delta)$$

$$2. \quad |\delta| \leq \mu \text{ که در آن } \text{fl}(xy) = (xy)(1 + \delta)$$

$$3. \quad \text{اگر } y \neq 0 \text{ آنگاه } \text{fl}(x/y) = (x/y)(1 + \delta) \text{ که در آن } |\delta| \leq \mu$$

بر روی کامپیوترهایی که از IEEE استاندارد استفاده نمی کنند قانون جمع ممیز شناور زیر باید برقرار باشد.

$$4. \quad |\delta_1| \leq \mu \text{ و } |\delta_2| \leq \mu \text{ که در آن } \text{fl}(x+y) = x(1 + \delta_1) + y(1 + \delta_2)$$

مثال ۲-۳-۱ اعمال ممیز شناور با گرد کردن

فرض کنید که در قسمتهای (الف) تا (ج) $\beta = 10$ و $t = 3$ باشند.

(الف) برای $x = 0.999 \times 10^2$ و $y = 0.111 \times 10^0$ داریم

$$x + y = 100.0110 = 0.100011 \times 10^3$$

$$\text{fl}(x + y) = 0.100 \times 10^3$$

بنابراین $\text{fl}(x + y) = (x + y)(1 + \delta)$ که در آن $\delta = -1/0.999 \times 10^{-4}$ ، $|\delta| \leq \frac{1}{4}(10^{-2})$.

(ب) برای $x = 0.999 \times 10^2$ و $y = 0.111 \times 10^0$ داریم

$$xy = 110.889$$

$$\text{fl}(xy) = 0.111 \times 10^2$$

بنابراین $\text{fl}(xy) = (xy)(1 + \delta)$ که در آن $\delta = 1/0.0100 \times 10^{-3}$ ، $|\delta| \leq \frac{1}{4}(10^{-1-3})$.

(ج) برای $x = 0.999 \times 10^2$ و $y = 0.111 \times 10^0$ داریم

$$\frac{x}{y} = 900$$

$$\text{fl}\left(\frac{x}{y}\right) = 0/900 \times 10^3$$

$$\delta = 0$$

(د) فرض کنید $t = 4$, $\beta = 10$, $x = 0/1112$ و $y = 0/2245 \times 10^5$ آنگاه

$$xy = 0/249644 \times 10^4$$

$$\text{fl}(xy) = 0/2496 \times 10^4$$

$$\text{بنابراین } |\text{fl}(xy) - xy| = 0/44 \text{ و } \frac{1}{4} \times 10^{-3} < 1/7625 \times 10^{-4} = |\delta|$$

محاسبات بدون یک رقم ضامن

قضیه ۱-۳-۲ و مثال ۱-۳-۲ نشان می‌دهند که خطاهای نسبی در محاسبه مجموع، تفاضل، حاصلضرب و خارج قسمت در حساب ممیز شناور کوچک هستند. ولیکن کامپیوترهای بدون رقم ضامن وجود دارند که در آنها جمعها و تفریقها ممکن است دقیق نباشند. ما این وضعیت را با جزئیات بیشتر بیان می‌کنیم.

یک رقم ضامن یک رقم اضافی بر روی انتهای پایینی ثبات حسابی است که هدفش نگهداری رقم پایین‌ترین مرتبه است که در غیر این صورت هنگامی که ممیزهای اعشاری تنظیم می‌شوند، از بین می‌رود. مثال زیر تفاوت بین دو مدل را نشان می‌دهد. فرض کنید

$$\beta = 10, \quad t = 3, \quad \mu = 0/005$$

مثال ۲-۳-۲ جمع با یک رقم ضامن

$$x = 0/101 \times 10^2, \quad y = -0/994 \times 10^1$$

گام ۱: در یک ردیف قرار دادن دو عدد

	رقم ضامن
$x = 0/101$	0×10^2
$y = -0/099$	4×10^2

گام ۲: جمع کردن (با یک رقم اضافی)

$$\begin{array}{r} 0/1010 \times 10^2 \\ -0/0994 \times 10^2 \\ \hline \text{fl}(x+y) = 0/0016 \times 10^2 \end{array}$$

گام ۳: نرمال کردن

$$\text{fl}(x+y) = 1/60 \times 10^{-1}$$

نتیجه: $\text{fl}(x+y) = (x+y)(1+\delta)$ با $\delta = 0$

مثال ۲-۳-۳ جمع بدون یک رقم ضامن

$$x = 0/101 \times 10^2, \quad y = -0/994 \times 10^1$$

گام ۱: در یک ردیف قرار دادن دو عدد

$$x = 0/101 \times 10^2$$

$$y = -0/099[4] \times 10^2$$

رقم پایین ترین مرتبه [۴] خارج می شود

گام ۲: جمع کردن

$$\begin{array}{r} 0/101 \times 10^2 \\ -0/099 \times 10^2 \\ \hline \text{fl}(x+y) = 0/002 \times 10^2 \end{array}$$

گام ۳: نرمال کردن

$$\text{fl}(x+y) = 2/00 \times 10^{-1}$$

نتیجه: $\text{fl}(x+y) = (x+y)(1+\sigma)$ با $\sigma = 0/25 = 50\mu$

بنابراین، تکرار می‌کنیم که برای کامپیوترهای با یک رقم ضامن داریم:

$$\text{fl}(x \pm y) = (x \pm y)(1 + \delta), \quad |\delta| \leq \mu$$

ولیکن، برای آن کامپیوترهایی که بدون یک رقم ضامن کار می‌کنند داریم:

$$\text{fl}(x \pm y) = x(1 + \delta_1) \pm y(1 + \delta_2)$$

$$|\delta_1| \leq \mu, \quad |\delta_2| \leq \mu$$

یک تذکر نهایی: در تمام این کتاب، فرض خواهیم کرد که محاسبات با یک رقم ضامن انجام شده باشند، به همان صورتی که بر روی تقریباً همه ماشینهای موجود انجام می‌شود. ما نتایج ۱ تا ۳ قضیه ۲-۳-۱ را همراه با رابطه (۲-۲-۱) از قضیه ۲-۲-۱ قانونهای اساسی ممیز شناور می‌نامیم. این قوانین اساسی پایه و اساس برقرارسازی کرانها برای محاسبات ممیز شناور هستند.

برای مثال، محاسبه ممیز شناور $x(y+z)$ را در نظر بگیرید.

$$\begin{aligned} \text{fl}(x(y+z)) &= [x \cdot \text{fl}(y+z)](1 + \delta_1) \\ &= x(y+z)(1 + \delta_2)(1 + \delta_1) \\ &= x(y+z)(1 + \delta_1\delta_2 + \delta_1 + \delta_2) \\ &= x(y+z)(1 + \delta_3) \end{aligned}$$

که در آن $\delta_3 = \delta_1 + \delta_2$ ؛ به دلیل آن δ_1 و δ_2 کوچک هستند، از حاصلضرب آنها چشم‌پوشی می‌شود. اکنون به آسانی می‌توانیم کران δ_3 را برقرار سازیم. فرض کنید $\beta = 10$ و گرد کردن به کار می‌رود آنگاه

$$\begin{aligned} |\delta_3| &= |\delta_1 + \delta_2| \leq |\delta_1| + |\delta_2| \\ &\leq \frac{1}{\beta} \times 10^{1-t} + \frac{1}{\beta} \times 10^{1-t} = 10^{1-t} \end{aligned}$$

بنابراین خطای نسبی ناشی از گرد کردن در محاسبه $\text{fl}(x(y+z))$ در حدود 10^{1-t} در بدترین حالت می‌باشد.

۲-۲ جمع n عدد ممیز شناور

جمع n عدد ممیز شناور x_1, x_2, \dots, x_n را با گرد کردن در نظر بگیریم. تعریف می‌کنیم $s_r = fl(x_1 + x_r)$.
 آنگاه

$$s_r = fl(x_1 + x_r) = (x_1 + x_r)(1 + \delta_r)$$

که در آن $\mu = \frac{1}{4}\beta^{1-t}$ یعنی $|\delta_r| \leq \mu$ ، یعنی $(x_1 + x_r) = \delta_r(x_1 + x_r) + s_r$. s_2, s_3, \dots, s_n را به طور بازگشتی توسط

$$s_{i+1} = fl(s_i + x_{i+1}), \quad i = 2, 3, \dots, n-1$$

تعریف می‌کنیم. آنگاه $s_r = fl(s_r + x_r) = (s_r + x_r)(1 + \delta_r)$ یعنی

$$\begin{aligned} s_3 - (x_1 + x_2 + x_3) &= (x_1 + x_2) \delta_2 + (x_1 + x_2)(1 + \delta_2)\delta_3 + x_3\delta_3 \\ &\approx (x_1 + x_2) \delta_2 + (x_1 + x_2 + x_3) \delta_3 \end{aligned}$$

(با چشم پوشی از جمله $\delta_2\delta_3$ ، که کوچک است، و غیره). بنابراین با استقرا می‌توان نشان داد که:

$$\begin{aligned} s_n - (x_1 + x_2 + \dots + x_n) &\approx (x_1 + x_2) \delta_2 + (x_1 + x_2 + x_3) \delta_3 \\ &\quad + \dots + (x_1 + x_2 + \dots + x_n) \delta_n \end{aligned}$$

(مجدداً با چشم پوشی از جمله $\delta_i\delta_j$ ، که کوچک هستند).

این نتیجه می‌تواند به صورت

$$\begin{aligned} s_n - (x_1 + x_2 + \dots + x_n) &\approx x_1(\delta_1 + \delta_2 + \dots + \delta_n) \\ &\quad + x_2(\delta_2 + \dots + \delta_n) + x_3(\delta_3 + \dots + \delta_n) \\ &\quad + \dots + x_n \delta_n \end{aligned}$$

نوشته شود که در آن $\mu = \frac{1}{4}\beta^{1-t}$ با تعریف $\delta_1 = 0$ می‌توانیم بنویسیم:

قضیه ۲-۴-۱ فرض کنید x_1, x_2, \dots, x_n عدد ممیز شناور باشند. آنگاه

$$\begin{aligned} fl(x_1 + x_2 + \dots + x_n) - (x_1 + x_2 + \dots + x_n) \\ \approx x_1(\delta_1 + \delta_2 + \dots + \delta_n) + x_2(\delta_2 + \dots + \delta_n) + \dots + x_n \delta_n \end{aligned}$$

که در آن هر $i = 1, 2, \dots, n$ ، $|\delta_i| \leq \mu$

تبصره: از فرمول قبل ملاحظه می‌کنیم که می‌باید در حالت کلی انتظار خطای کمتری را داشته باشیم اگر n عدد ممیز شناور را به ترتیب صعودی اندازه هایشان

$$|x_1| \leq |x_2| \leq \dots \leq |x_n|$$

جمع کنیم. اگر اعداد به ترتیب صعودی اندازه هایشان مرتب شوند، آنگاه خطاهای بزرگ‌تر مربوط به اعداد کوچک‌تر خواهند بود.

۲-۵ ضرب n عدد ممیز شناور

اگر همانند حالت جمع n عدد ممیز شناور که در بخش آخر ملاحظه نمودیم، پیش رویم، می‌توان نشان داد که قضایای زیر برقرار هستند.

قضیه ۲-۵-۱

$$\prod_{i=1}^n (x_i) \approx (1 + \varepsilon) \prod_{i=1}^n x_i$$

که در آن $\varepsilon = |(1 + \delta_1)(1 + \delta_2) \dots (1 + \delta_n) - 1|$ و $|\delta_i| \leq \mu$ ، $i = 1, 2, \dots, n$.

یک کران بالا برای ε

با فرض این که $(n - 1)\mu < 0.1$ ، ثابت خواهیم کرد که $\varepsilon < 1/0.6(n - 1)\mu$. (این فرض کاملاً واقع‌بینانه است و بر روی بیشتر ماشینها برای مقادیر n نسبتاً بزرگ برقرار خواهد بود).

چون $|\delta_i| \leq \mu$ ، داریم، $\varepsilon \leq (1 + \mu)^{n-1} - 1$. دوباره چون

$$\ln(1 + \mu)^{n-1} = (n - 1)\ln(1 + \mu) < (n - 1)\mu$$

داریم:

$$(1 + \mu)^{n-1} < e^{(n-1)\mu}$$

بنابراین

$$\begin{aligned} (1 + \mu)^{n-1} - 1 &< e^{(n-1)\mu} - 1 = (n - 1)\mu + \frac{((n - 1)\mu)^2}{2} + \dots \\ &= (n - 1)\mu \left[1 + \frac{(n - 1)\mu}{2} + \frac{((n - 1)\mu)^2}{6} + \dots \right] \\ &< (n - 1)\mu \left[1 + \frac{0.05}{1 - 0.05} \right] \end{aligned}$$

(با استفاده از سری هندسی). (توجه کنید که $0/1 < \mu(n-1)$).

بنابراین

$$\varepsilon \leq (1 + \mu)^{n-1} - 1 < (n-1)\mu \left[1 + \frac{0/05}{1 - 0/05} \right] < 1/06(n-1)\mu \quad (1-5-2)$$

بنابراین با ترکیب قضیه ۲-۵-۱ و رابطه (۲-۵-۱)، می‌توانیم قضیه زیر را بنویسیم.

قضیه ۲-۵-۲ خطای نسبی در محاسبه حاصلضرب n عدد ممیز شناور حداکثر برابر $1/06(n-1)\mu$ است، با فرض این که $0/1 < \mu(n-1)$.

۲-۶ محاسبه حاصلضرب داخلی

یک امر مهم محاسباتی که اغلب در جبر خطی عددی رخ می‌دهد محاسبه حاصلضرب داخلی دو بردار n بعدی x و y است:

$$x^T y = x_1 y_1 + x_2 y_2 + \dots + x_n y_n \quad (1-6-2)$$

که در آن x_i و y_i ، $i = 1, \dots, n$ ، مؤلفه‌های x و y هستند.

فرض کنید x_i و y_i ، $i = 1, \dots, n$ ، اعداد ممیز شناور باشند تعریف می‌کنیم

$$S_1 = \text{fl}(x_1 y_1) \quad (2-6-2)$$

$$S_2 = \text{fl}(S_1 + \text{fl}(x_2 y_2)) \quad (3-6-2)$$

⋮

$$S_k = \text{fl}(S_{k-1} + \text{fl}(x_k y_k)), \quad k = 2, 3, \dots, n \quad (4-6-2)$$

سپس با استفاده از قضیه ۱-۳-۲ داریم:

$$S_1 = x_1 y_1 (1 + \delta_1) \quad (5-6-2)$$

$$S_2 = [S_1 + x_2 y_2 (1 + \delta_2)] (1 + \eta_2) \quad (6-6-2)$$

⋮

$$S_n = [S_{n-1} + x_n y_n (1 + \delta_n)] (1 + \eta_n) \quad (7-6-2)$$

که در آن هر $|\delta_i| \leq \mu$ و $|\eta_i| \leq \mu$. با جایگذاری مقادیر S_1 تا S_{n-1} در S_n و انجام برخی مرتب

سازیه‌ها، می‌توان نوشت:

$$S_n = \sum_{i=1}^n x_i y_i (1 + \varepsilon_i) \quad (۸-۶-۲)$$

که در آن

$$1 + \varepsilon_i = (1 + \delta_i)(1 + \eta_i)(1 + \eta_{i+1}) \cdots (1 + \eta_n) \\ \approx 1 + \delta_i + \eta_i + \eta_{i+1} + \cdots + \eta_n \quad (\eta_1 = 0) \quad (۹-۶-۲)$$

(با چشم‌پوشی از حاصلضربهای $\delta_i \eta_j$ و $\eta_i \eta_k$ که کوچک هستند).

برای مثال، هنگامی که $n = ۲$ ، به آسانی می‌توان امتحان کرد که:

$$S_2 = x_1 y_1 (1 + \varepsilon_1) + x_2 y_2 (1 + \varepsilon_2) \quad (۱۰-۶-۲)$$

که در آن $1 + \varepsilon_2 \approx 1 + \delta_2 + \eta_2$ ، $1 + \varepsilon_1 \approx 1 + \delta_1 + \eta_2$ (با چشم‌پوشی از حاصلضربهای $\delta_1 \eta_2$ و $\delta_2 \eta_2$ که کوچک هستند).

همانند بخش قبل، می‌توان نشان داد که اگر $n\mu < ۰/۱$ آنگاه

$$|\varepsilon_i| \leq 1/06(n + 2 - i)\mu, \quad i = 1, 2, \dots, n \quad (۱۱-۶-۲)$$

از (۸-۶-۲) و (۱۱-۶-۲) می‌توان نوشت

$$|\mathbb{1}(x^T y) - x^T y| \leq \sum_{i=1}^n |x_i| |y_i| |\varepsilon_i| \\ \leq \phi(n)\mu |x|^T |y| \\ \leq \phi(n)\mu \|x\|_2 \|y\|_2 \quad (\text{بخش ۷-۱})$$

که در آن $|\alpha| = (|\alpha_1|, |\alpha_2|, \dots, |\alpha_n|)^T$ یک تابع کوچک از n است.

قضیه ۱-۶-۲

$$|\mathbb{1}(x^T y) - x^T y| \leq \phi(n)\mu |x|^T |y| \leq \phi(n)\mu \|x\|_2 \|y\|_2$$

محاسبه حاصلضرب داخلی در دقت مضاعف

وقتی در مورد محاسبه حاصلضرب داخلی صحبت می‌کنیم، مهم است توجه کنیم که چون در اغلب کامپیوترها محاسبات با دقت مضاعف امکان‌پذیر است، توصیه می‌شود که حاصلضرب داخلی در دقت مضاعف (با استفاده از ۲۴ رقم حسابی) محاسبه شود تا دقت بیشتری را حفظ نماید. توجه آن این است که اگر ما از دقت ساده برای محاسبه $x^T y$ استفاده کنیم $(1 - 2^n)$ خطای گرد کردن دقت ساده (یکی برای هر ضرب و هر جمع) وجود خواهد داشت. یک استراتژی بهتر این است که هر x و y را با توسعه مانتیسهای آنها با صفر به دقت مضاعف تبدیل کنیم و ضرب و جمع آنها را در دقت مضاعف انجام دهیم و نتیجه نهایی را در دقت ساده ارائه دهیم. این فرایند معروف به ذخیره حاصلضرب داخلی در دقت مضاعف (یا دقت توسعه یافته) می‌باشد. این فرایند را به صورت زیر خلاصه می‌کنیم.

۱. محاسبه حاصلضربهای فردی $x_i y_i$ در دقت مضاعف

۲. محاسبه مجموع $\sum_{i=1}^n x_i y_i$ در دقت مضاعف به ترتیب صعودی

۳. گرد کردن مجموع در دقت ساده

این فرایند خطای گرد کردن پایینی را با حداقل هزینه اضافی ارائه می‌دهد. می‌توان نشان داد (Wilkinson, 1963, 117-118) که خطا در این حالت اساساً مستقل از n است. بویژه می‌توان نشان داد که اگر حاصلضرب داخلی در دقت مضاعف ذخیره شود و $fl_p(x^T y)$ نتیجه چنین محاسباتی را نمایش دهد، آنگاه قضیه زیر برقرار است.

قضیه ۲-۶-۲ رابطه

$$|fl_p(x^T y) - x^T y| \leq c\mu |x^T y|$$

برقرار است مگر حذفهای جدی در هر جمله از $x^T y$ اتفاق افتد. در اینجا $0/1 < n\mu < c$ و یک عدد کوچک است.

تبصره: جمله آخر در قضیه ۲-۶-۲ مهم است. می‌توانیم یک مثال ساده (تمرین ۶(ب)) را برای دیدن این که اگر حذف رخ دهد، نتیجه قضیه ۲-۶-۲ برقرار نخواهد بود، بسازیم. پدیده حذف فاجعه‌آمیز در بخش بعد مورد بحث قرار می‌گیرد.

۲-۲ کرانه‌های خطا برای محاسبات ماتریسی ممیز شناور

قضیه ۲-۷-۱ فرض کنید $|M| = (|m_{ij}|)$. فرض کنید A و B دو ماتریس ممیز شناور و c یک عدد ممیز شناور باشند. آنگاه

$$|E| \leq \mu |cA|, \text{fl}(cA) = cA + E \quad ۱.$$

$$|E| \leq \mu(|A+B|), \text{fl}(A+B) = (A+B) + E \quad ۲.$$

اگر A و B دو ماتریس سازگار برای ضرب ماتریسی باشند، آنگاه از (۲-۶-۴) می‌توان نشان داد که:

$$|E| \leq n\mu |A| |B| + O(\mu^2), \text{fl}(AB) = AB + E \quad ۳.$$

معنای $O(\mu^2)$

در قضیه ۲-۷-۱ نماد $O(\mu^2)$ نشانی است برای یک عبارت پیچیده که توسط $c\mu^2$ کراندار می‌باشد. در اینجا c یک ثابت است و به مسأله بستگی دارد. عبارت $O(\mu^2)$ به طور مکرر در این کتاب استفاده می‌شود.

تبصره: نتیجه آخر نشان می‌دهد که ضرب ماتریسی در ممیز شناور می‌تواند خیلی نادقیق باشد، زیرا $|A||B|$ ممکن است بزرگ‌تر از خود $|AB|$ باشد (تمرین ۹). به این دلیل، هرگاه ممکن باشد باید، هنگام محاسبه ضربهای ماتریسی در ماتریس یا ماتریس در بردار، از ذخیره سازی ضربهای داخلی در دقت مضاعف استفاده شود، زیرا در این حالت می‌توان نشان داد که عناصر ماتریس خطا به طور قاطع توسط عناصر $|AB|$ به جای عناصر $|A||B|$ کراندار می‌باشند، مرجع Wilkinson (1965, 118) را ملاحظه کنید.

کرانه‌های خطا بر حسب نرمها

به طور مرسوم برای محاسبات ماتریسی کرانه‌های ماتریسهای خطا بر حسب نرمهای ماتریسها (به جای مقادیر مطلق ماتریسها، که قبلاً ارائه شدند) ارائه می‌شوند. در اینجا دوباره کران را برای ماتریسهای خطا برای ضربهای ماتریسی با استفاده از نرمها برای سادگی ارجاعهای بعدی در کتاب، می‌نویسیم. لیکن توجه کنید که کرانه‌های خطاهای عنصر به عنصر از خطاهای به صورت نرم دارای معنای بیشتری هستند (تبصره‌های بخش ۳-۲ را ملاحظه کنید). دوباره رابطه زیر را ملاحظه کنید.

$$\text{fl}(AB) = AB + E, \quad |E| \leq n\mu |A| |B| + O(\mu^2)$$

چون $\|E\| \leq \| |E| \|$ ، می توان این رابطه را دوباره به صورت

$$\text{fl}(AB) = AB + E$$

نوشت که در آن $\|E\| \leq \| |E| \| \leq n\mu \|A\| \|B\| + O(\mu^2)$.

بویژه برای $\| \cdot \|_2$ و $\| \cdot \|_\infty$ ، داریم:

$$\|E\|_\infty \leq n\mu \|A\|_\infty \|B\|_\infty + O(\mu^2)$$

$$\|E\|_2 \leq n^2\mu \|A\|_2 \|B\|_2 + O(\mu^2)$$

قضیه ۲-۷-۲ $\text{fl}(AB) = AB + E$ که در آن $\|E\|_2 \leq n^2\mu \|A\|_2 \|B\|_2 + O(\mu^2)$.

دو حالت خاص وجود دارند که باید ملاحظه کنیم.

نتیجه ۱-۷-۲ ضرب ماتریس - بردار . اگر b یک بردار باشد، آنگاه از قبل داریم :

$$\text{fl}(Ab) = Ab + e$$

که در آن $\|e\|_2 \leq n^2\mu \|A\|_2 \|b\|_2 + O(\mu^2)$.

(بعلاوه مسأله ۱۱ و تبصره های ارائه شده در آنجا را ملاحظه کنید).

نتیجه ۲-۷-۲ ضرب ماتریس در یک ماتریس متعامد . فرض کنید $A \in \mathbb{R}^{n \times n}$ و $Q \in \mathbb{R}^{n \times n}$

متعامد باشد آنگاه

$$\text{fl}(QA) = Q(A + E)$$

که در آن $\|E\|_2 \leq O(\mu) \|A\|_2$

نتیجه ۲-۷-۲ می گوید، با این که ضرب ماتریس می تواند در حالت کلی نادقیق باشد، اگر یکی از ماتریسها متعامد باشد، آنگاه ضرب ماتریسی ممیز شناور فقط یک خطای کوچک و قابل پذیرش ارائه می دهد. همچنان که در فصلهای بعد خواهیم دید، این نتیجه پایه و اساس بسیاری از الگوریتمهای مناسب عددی مورد بحث در این کتاب است.

برای مثال، نتیجه زیر که اغلب در این کتاب استفاده می‌شود، یکی از پیامدهای نتیجه قبلی است.

نتیجه ۲-۷-۳ فرض کنید P یک ماتریس متعامد تعریف شده توسط رابطه

$$P = I - \gamma \frac{uu^T}{u^T u}$$

باشد، که در آن u یک بردار ستونی است. فرض کنید \hat{P} گونه محاسبه شده P در حساب ممیز شناور باشد. آنگاه

$$\| \hat{P}A - P(A + E) \|$$

که در آن $\gamma \|A\| \leq O(\mu) \|E\|$ و c یک ثابت از مرتبه واحد است.

اثبات: مرجع (Wilkinson (1965, 152-160 را ملاحظه کنید.

۲-۸ خطاهای گرد کردن ناشی از حذف و محاسبات بازگشتی

به طور شهودی، واضح است که اگر تعداد زیادی محاسبات ممیز شناور انجام شود، آنگاه خطای انباشته شده می‌تواند کاملاً بزرگ باشد؛ لیکن خطای گرد کردن می‌تواند حتی در یک تک گام از محاسبات خطرناک باشد. برای مثال، دو عدد زیر را ملاحظه کنید.

$$x = 0.54617$$

$$y = 0.54601$$

مقدار دقیق عبارت است از:

$$d = x - y = 0.00016$$

فرض کنید اکنون از حساب چهار رقمی با گرد کردن استفاده می‌کنیم. آنگاه داریم.

$$\hat{x} = 0.5462 \quad (\text{درست با چهار رقم با معنی})$$

$$\hat{y} = 0.5460 \quad (\text{درست با چهار رقم با معنی})$$

$$\hat{d} = \hat{x} - \hat{y} = 0.0002$$

در حساب هشت رقمی داریم $\sqrt{10^{10} - 4} = 10^5$.

جلوگیری از حذف

اگر از یک جفت فرمول معادل استفاده شود، از حذف می‌توان جلوگیری نمود

$$x_1 = \frac{-b - \text{sign}(b)\sqrt{b^2 - 4ac}}{2a}$$

$$x_2 = \frac{c}{ax_1}$$

که در آن $\text{sign}(b)$ علامت b است. با استفاده از این فرمولها، به آسانی ملاحظه می‌کنیم که:

$$x_1 = 1000000/00$$

$$x_2 = \frac{1/00000000}{1000000/00}$$

$$= 0/00000100000000$$

مثال ۲-۸-۲ باز هم با مثالی دیگر ملاحظه می‌کنیم که چگونه حذف شدن می‌تواند منجر به بی‌دقتی شود. مسأله محاسبه $1 - x - e^x = f(x)$ را در $x = 0/01$ در نظر بگیرید. با استفاده از حساب پنج رقمی جواب درست برابر $0/000050167$ می‌باشد. اگر $f(x)$ مستقیماً از عبارت داده شده محاسبه شود، داریم

$$f(0/01) = 1/0101 - (0/01) - 1 = 0/0001$$

$$\text{خطای نسبی} = \frac{0/0001 - 0/000050167}{0/000050167} = 0/99 \times 10^0$$

که اشاره می‌کند به این که نمی‌توان حتی به اولین رقم با معنی اعتماد کرد. خوشبختانه می‌توان با استفاده از سری همگرای زیر دوباره از حذف شدن جلوگیری کرد.

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

در این حالت داریم:

$$e^x - x - 1 = (1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots) - x - 1$$

$$= \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

برای $x = 0/01$ ، این فرمول نتیجه می‌دهد.

$$\begin{aligned} & \frac{(0/01)^2}{2!} + \frac{0/01)^3}{3!} + \frac{(0/01)^4}{4!} + \dots \\ & = 0/00005 + 0/000000166666 + 0/0000000004166 + \dots \\ & = 0/000050167 \quad (\text{درست تا پنج رقم با معنی}) \end{aligned}$$

تبصره: توجه کنید که اگر x منفی می‌بود، آنگاه استفاده از سری همگرا برای e^x نمی‌توانست کمک کند. برای محاسبه e^x به ازای یک مقدار منفی x می‌توان با استفاده از فرمول زیر از حذف شدن پرهیز کرد.

$$e^{-x} = \frac{1}{e^x} = \frac{1}{1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\dots}$$

تبصره: توجه کنید که اگر چه حذف شدن منجر به بی‌دقتی می‌شود، ولیکن نباید برای نتایج غلط مقصر باشد. سرچشمه می‌تواند جای دیگر، مثلاً در محاسبات قبلی باشد. برای نمونه، در مثال ۲-۸-۱، سرچشمه محاسبه $4 - 10^{10}$ در حساب ۸ رقمی است.

محاسبات بازگشتی

در مثالهای قبل، ملاحظه کردیم که چگونه حذفهای تفریقی می‌توانند جوابهای غیر دقیق ارائه دهند. ولیکن محاسبات متداول دیگری نیز وجود دارند که باعث افزایش خطا می‌شوند. برای مثال، محاسبات بازگشتی را ملاحظه کنید. این محاسبات، محاسباتی هستند که به طور بازگشتی انجام می‌شوند؛ به نحوی که محاسبه یک گام به محاسبه گامهای قبلی نیاز دارد. در چنین حالتی حتی اگر خطای ایجاد شده در گام اول ناچیز باشد، بر اثر تجمع و بزرگ شدن خطا در هر گام، خطای نهایی می‌تواند کاملاً بزرگ باشد و یک جواب کاملاً غلط ارائه دهد.

برخی فرمولهای بازگشتی به طریقه‌های ناخوش آیند خطاها را تکثیر می‌کنند. یک مثال خیلی زیبا و حاوی خطاهای بازگشتی را دوباره از کتاب Forsythe, Malcolm, and Moler (1977, 16-17) ملاحظه کنید.

مثال ۲-۸-۳ فرض کنید به محاسبه انتگرال

$$E_n = \int_0^1 x^n e^{x-1} dx$$

به ازای مقادیر مختلف n نیاز داشته باشیم. با انتگرالگیری جزء به جزء خواهیم داشت،

$$E_n = \int_0^1 x^n e^{x-1} dx = (x^n e^{x-1}) \Big|_0^1 - \int_0^1 n x^{n-1} e^{x-1} dx$$

یا

$$E_n = 1 - nE_{n-1}, \quad n = 2, 3, \dots$$

بنابراین اگر E_1 معلوم باشد، آنگاه برای مقادیر مختلف n ، E_n می‌تواند با استفاده از فرمول بازگشتی فوق محاسبه شود.

در واقع، با $\beta = 10$ و $t = 6$ و شروع با $E_1 = 0/367879$ به عنوان یک تقریب ۶ رقمی برای $E_1 = \frac{1}{e}$ داریم.

$$E_1 = 0/367879$$

$$E_2 = 0/264242$$

$$E_3 = 0/207274$$

$$E_4 = 0/170904$$

⋮

$$E_9 = -0/0684800$$

اگرچه انتگرالده در بازه $(0, 1)$ مثبت است، مقدار محاسبه شده E_9 منفی می‌باشد. این پدیده را می‌توان به صورت زیر توضیح داد.

خطا در محاسبه E_2 ، -2 برابر خطا در محاسبه E_1 است؛ خطا در محاسبه E_3 ، -3 برابر خطا در محاسبه E_2 است (بنابراین خطا در این گام ۶ برابر خطا در E_1 می‌باشد). از اینرو خطا در محاسبه E_9 ، $9! = (-9) \dots (-4)(-3)(-2)$ برابر خطا در E_1 است. خطا در E_1 ناشی از گرد کردن $\frac{1}{e}$ با استفاده از شش رقم با معنی بوده؛ که در حدود $10^{-7} \times 4/412 \times 10^{-7}$ می‌باشد. ولیکن این خطای کوچک در ۹! ضرب می‌شود و مقدار $0/1601 \approx 10^{-7} \times 4/412 \times 9!$ را ارائه می‌دهد که کاملاً بزرگ است.

آرایش مجدد فرمول بازگشتی

برای مثال ۲-۸-۳ دوباره به این نتیجه می‌رسیم که می‌توان یک نتیجه بهتر با آرایش مجدد فرمول بازگشتی به دست آورد به نحوی که خطا در هر گام به جای بزرگ شدن، کاهش یابد. در واقع اگر فرمول بازگشتی را به صورت:

$$E_{n-1} = \frac{1-E_n}{n}, \quad n = \dots, 3, 2$$

بنویسیم، آنگاه خطا در هر گام توسط عامل $\frac{1}{n}$ کاهش می‌یابد. بنابراین از یک مقدار بزرگ n و کارکردن به صورت پسرو، ملاحظه می‌کنیم که E_q با دقت شش رقم کامل دقیق خواهد بود. برای به دست آوردن یک مقدار اولیه، توجه می‌کنیم به این‌که:

$$E_n = \int_0^1 x^n e^{x-1} dx \leq \int_0^1 x^n dx = \frac{1}{n+1}$$

با $n = 20$ ، $E_{20} \leq \frac{1}{21}$ ، اجازه دهید اختیار کنیم $E_{20} = 0$. سپس با شروع با مقدار $E_{20} = 0$ می‌توان نشان داد (Forsythe, Malcolm, and Moler (1977, 17)) که $E_q = 0.0916123$ ، که با دقت ۶ رقم کامل درست است.

دلیل به دست آوردن این دقت این است که خطا در E_{20} حداکثر $\frac{1}{21}$ بود؛ این خطا در محاسبه E_{19} در $\frac{1}{20}$ ضرب شده و در محاسبه E_{19} خطایی حداکثر برابر $0.0024 = \frac{1}{20} \times \frac{1}{21}$ ارائه می‌دهد، و غیره.

۹-۲ مرور و خلاصه

در این فصل مفاهیم اعداد ممیز شناور و خطاهای گردکردن معرفی شده و مورد بحث قرار گرفته‌اند.

۱. اعداد ممیز شناور: یک عدد ممیز شناور نرمال شده دارای شکل

$$x = \pm r \beta^e$$

است که در آن e نما، r مانیتیس و β پایه دستگاه نامیده می‌شوند. دستگاه اعداد ممیز شناور

توسط چهار پارامتر زیر مشخص می‌شود:

β پایه

t دقت

U و L حد پایین و بالای نما

۲. خطاها: خطاها در یک محاسبه توسط خطای مطلق یا خطای نسبی اندازه‌گیری می‌شوند.

خطاهای نسبی معنای بیشتری از خطاهای مطلق دارند. خطای نسبی نشانه‌ای از ارقام با معنی در

یک جواب تقریبی ارائه می‌دهد. خطای نسبی در نمایش یک عدد حقیقی x توسط نمایش ممیز

شناورش $fl(x)$ توسط یک عدد μ ، که دقت ماشین نامیده می‌شود، کراندار می‌شود

(قضیه ۲-۲-۱).

۳. قوانین حساب ممیز شناور:

$$\text{fl}(x * y) = (x * y)(1 + \delta)$$

که در آن * هر یک از چهار عمل اصلی +، -، x، یا ÷ را نشان می‌دهد، و $|\delta| \leq \mu$.

۴. محاسبات جمع، ضرب، ضرب داخلی: نتایج جمع و ضرب n عدد ممیز شناور به ترتیب در قضیه‌های ۲-۴-۱ و ۲-۵-۱ ارائه شده‌اند.

(الف) n عدد ممیز شناور را به ترتیب صعودی اندازه‌هایشان جمع کنید.

(ب) به هنگام محاسبه ضرب داخلی دو بردار، حاصلضرب داخلی را در دقت مضاعف ذخیره می‌کنیم هرگاه ممکن باشد.

۵. ضربهای ماتریسی ممیز شناور: کرانه‌های خطا به صورت عنصر به عنصر و به صورت نرم برای ضرب ماتریسی دو ماتریس ممیز شناور به ترتیب در قضیه‌های ۲-۷-۱ و ۲-۷-۲ ارائه شده‌اند.

(الف) ضرب ماتریسی در حساب ممیز شناور می‌تواند خیلی نادقیق باشد مگر یکی از ماتریسها متعامد باشد. ذخیره‌سازی حاصلضرب داخلی، هرگاه ممکن باشد، در محاسبه ضرب ماتریس در ماتریس یا ماتریس در بردار پیشنهاد می‌شود.

(ب) دقت بالا در یک محاسبه ضرب ماتریسی که حاوی یک ماتریس متعامد باشد، استفاده از ماتریس متعامد را در محاسبات ماتریسی خیلی جذاب می‌سازد.

۶. خطاهای گرد کردن ناشی از حذف و محاسبه بازگشتی: دو محاسبه مهم که خطاهای گرد کردن را به وجود می‌آورند عبارتند از تفاضل دو عدد نزدیک به هم و محاسبات بازگشتی. این دو به طور نسبتاً مفصل در بخش ۲-۸ مورد بحث قرار گرفتند. مثالهایی برای نشان دادن چگونه وارد شدن این خطاها در بسیاری از محاسبات اساسی ارائه گردیدند. پیام امید بخش در اینجا این است که در بسیاری از مثالها، محاسبات می‌توانند دوباره به‌نحوی مرتب شوند که بتوان از حذف جلوگیری نمود و خطا در هر گام از محاسبات بازگشتی بتواند تقلیل یابد.

۱. (الف) نشان دهید که

$$\frac{|\text{fl}(x) - x|}{|x|} \leq \mu = \begin{cases} \beta^{1-t} & \text{برای برش زدن} \\ \frac{1}{3}\beta^{1-t} & \text{برای گرد کردن} \end{cases}$$

(ب) نشان دهید که (الف) را می توان به شکل زیر نوشت:

$$\text{fl}(x) = x(1 + \delta), \quad |\delta| \leq \mu$$

۲. فرض کنید x یک عدد ممیز شناور و k یک عدد صحیح مثبت باشد. آنگاه نشان دهید که

$$\text{fl}\left(\frac{x^k}{k!}\right) = \frac{x^k}{k!}(1 + e_k)$$

$$\text{که در آن } |e_k| \leq 2k\mu + O(\mu^2)$$

۳. برای نشان دادن این که قانون توزیعی برای جمع و ضرب ممیز شناور برقرار نیست، مثالی بسازید. چه می توانید درباره جابجایی و شرکت پذیری این اعمال بگویید؟ برای جوابهای خود دلائلی ارائه دهید.

۴. فرض کنید x_1, x_2, \dots, x_n عدد ممیز شناور باشند. تعریف کنید.

$$s_2 = \text{fl}(x_1 + x_2), \quad s_k = \text{fl}(s_{k-1} + x_k), \quad k = 3, \dots, n$$

سپس با استفاده از قضیه ۲-۴-۱ نشان دهید که

$$S_n = \text{fl}(x_1 + x_2 + \dots + x_n) = x_1(1 + v_1) + x_2(1 + v_2) + \dots + x_n(1 + v_n)$$

کرانی برای هر $i = 1, \dots, n$ ، v_i ارائه دهید.

۵- (الف) مثالی بزنید که نشان دهد که به هنگام جمع یک فهرست از اعداد ممیز شناور، خطای گرد کردن می تواند کوچکتر باشد اگر اعداد به ترتیب صعودی اندازه هایشان جمع شوند.

(ب) مثال دیگری پیدا کنید که نشان دهید که این مطلب همواره درست نیست.

۶. (الف) ثابت کنید که خطا در محاسبه یک ضرب داخلی با ذخیره سازی در دقت مضاعف اساساً مستقل از n است. یعنی نشان دهید که اگر $\text{fl}_\gamma(x^T y)$ حاصل ضرب داخلی با ذخیره سازی را در دقت مضاعف نشان دهد، آنگاه به جز در حالت هایی که حذف های جدی رخ می دهد، داریم:

$$|\text{fl}_\gamma(x^T y) - x^T y| \leq c\mu |x^T y| + O(\mu^2)$$

(قضیه ۲-۶-۲)

(ب) با یک مثال ساده نشان دهید که اگر حذف وجود داشته باشد، آنگاه $\text{fl}_r(x^T y)$ می‌تواند به‌طور با معنایی با $x^T y$ تفاوت داشته باشد (اختیار کنید $t = 3$)
 (ج) اگر s یک اسکالر باشد، آنگاه ثابت کنید که

$$\text{fl}_r\left(\frac{x^T y}{s}\right) = \frac{x_1 y_1 (1 + \delta_1) + \dots + x_n y_n (1 + \delta_n)}{s / (1 + \delta)}$$

کرانی برای δ و δ_i ، $i = 1, \dots, n$ ، بیابید.

۷. نشان دهید که

$$|E| \leq \mu |cA|, \text{fl}(cA) = cA + E \quad (\text{الف})$$

$$|E| \leq \mu(|A| + |B|), \text{fl}(A + B) = (A + B) + E \quad (\text{ب})$$

$$|E| \leq n\mu |A| |B| + O(\mu^2), \text{fl}(AB) = AB + E \quad (\text{ج})$$

۸. یک مثال ساده بسازید که نشان دهد ضرب ماتریسی در حساب ممیز شناور لازم نیست دقیق باشد. با استفاده از ذخیره سازی حاصلضرب داخلی در دقت مضاعف، ضرب داخلی را دوباره با مثال خود انجام دهید.

۹. فرض کنید A و B ماتریسهای $n \times n$ باشند، آنگاه نشان دهید که

$$\text{fl}(AB) = AB + E$$

که در آن

$$|e_{ij}| \leq n^2 |f_{ij}| \mu + O(\mu^2)$$

حاصلضرب داخلی i امین سطر A و j امین ستون B

۱۰. ثابت کنید که اگر Q متعامد باشد، آنگاه

$$\text{fl}(QA) = Q(A + E)$$

که در آن $\|E\|_r \leq n^2 \mu \|A\|_r + O(\mu^2)$

۱۱. فرض کنید b یک بردار ستونی و $x = Ab$. علاوه فرض کنید $\hat{x} = \text{fl}(x)$. آنگاه نشان دهید

که

$$\frac{\|\hat{x} - x\|}{\|x\|} \leq p(n) \mu \|A^{-1}\| \|A\|$$

که در آن $p(n)$ یک چندجمله‌ای بر حسب n و از درجه پایین است. (عدد $\|A\| \|A^{-1}\|$ عدد وضعیت A نامیده می‌شود. ماتریسهایی یافت می‌شوند که برای آنها این عدد می‌تواند خیلی بزرگ باشد. سپس برای آن ماتریسها نتیجه می‌گیریم که خطای نسبی در یک ضرب ماتریس - بردار می‌تواند کاملاً بزرگ باشد).

۱۲. با استفاده از قضیه ۲-۷-۱ ثابت کنید که اگر B نامنفرد باشد، آنگاه

$$\frac{\|fl(AB) - AB\|_F}{\|AB\|_F} \leq n\mu \|B\|_F \|B^{-1}\|_F + O(\mu^2)$$

۱۳. فرض کنید y_1, y_2, \dots, y_n بردار ستونی تعریف شده به صورت بازگشتی زیر باشند

$$y_{i+1} = Ay_i, \quad i = 1, 2, \dots, n-1$$

فرض $\hat{y}_i = fl(y_i)$ یک کران برای خطای نسبی در محاسبه هر y_i ، $i = 1, 2, \dots, n$ ، بیابید.

۱۴. فرض کنید $\beta = 10$ ، $\epsilon = 4$. $fl(A^T A)$ را محاسبه کنید که در آن

$$A = \begin{bmatrix} 1 & 1 \\ 10^{-\epsilon} & 0 \\ 0 & 10^{-\epsilon} \end{bmatrix}$$

محاسبه خود را با $9 = \epsilon$ تکرار کنید. نتایج را با هم مقایسه کنید.

۱۵. نشان دهید که چگونه باید محاسبات را برای هر یک از موارد زیر مرتب نمود تا بتوان از دست

دادن ارقام بامعنی جلوگیری کرد. در هر حالت با یک مثال عددی جواب خود را تایید نمایید.

(الف) $e^x - x - 1$ ، برای مقایسه منفی از x

(ب) $\sqrt{x^2 + 1} - x^2$ ، برای مقادیر بزرگ x

(ج) $\frac{1}{x} - \frac{1}{x+1}$ ، برای مقادیر بزرگ x

(د) $x - \sin x$ ، برای مقادیر x نزدیک به صفر

(ه) $1 - \cos x$ ، برای مقادیر x نزدیک به صفر

۱۶. خطاهای نسبی و مطلق را برای تقریب هر یک از مقادیر زیر بیابید.

(الف) π توسط $\frac{22}{7}$

(ب) $\frac{1}{3}$ توسط 0.3333

(ج) $\frac{1}{6}$ توسط 0.166

چند رقم با معنی در هر تقریب وجود دارد.

۱۷. فرض کنید $\beta = 10$ و $t = 4$. محاسبه زیر را در نظر بگیرید.

$$a = \frac{(\frac{1}{6} - 0.1666)}{0.1666}$$

چند رقم با معنی درست از جواب واقعی به دست خواهید آورد.

۱۸. محاسبه $e = \sqrt{a^2 + b^2}$ را در نظر بگیرید. چگونه می توان محاسبات را تنظیم نمود تا بتوان

از سرریز در محاسبه $a^2 + b^2$ برای مقادیر بزرگ a یا b پرهیز کرد.

۱۹. چه جوابهایی به دست خواهید آورد اگر اعداد زیر را بر روی ماشین حساب یا کامپیوتر خود محاسبه کنید.

(الف) $\sqrt{10^8 - 1}$

(ب) $\sqrt{10^{-20} - 1}$

(ج) $10^{16} - 50$

خطاهای مطلق و نسبی را در هر حالت حساب کنید.

۲۰. با استفاده از فرمول معروف $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ در حل معادلات درجه دوم زیر چه مشکلاتی را

پیش بینی می کنید.

(الف) $x^2 - 10^6x + 1 = 0$

(ب) $10^{-10}x^2 - 10^{10}x + 10^{10} = 0$

چه چاره ای را پیشنهاد می کنید؟ اکنون با راه پیشنهاد شده خود معادلات را حل کنید.

۲۱. نشان دهید که انتگرال

$$y_i = \int_0^1 \frac{x^i}{x + \delta} dx$$

می تواند با استفاده از فرمول

$$y_i = \frac{1}{i} - \delta y_{i-1}$$

محاسبه شود. y_1, y_2, \dots, y_n را با استفاده از این فرمول و با اختیار کردن

$$y_0 = \ln(x + \delta) \Big|_{x=0}^1 = \ln 6 - \ln 5 = \ln(1/2)$$

محاسبه کنید. چه چیز غیر عادی در این محاسبه مشاهده می‌نمایید؟ توضیح دهید که چه اتفاقی افتاده است. اکنون فرمول بازگشتی را به نحوی مرتب کنید که مقادیر x را بتواند با دقت بیشتری محاسبه شوند.

۲۲. فرض کنید که x مقادیر 10^4 ، 50000 و 55596 را با پنج رقم با معنی تقریب زند. بزرگ‌ترین بازه‌ای را بیابید که در هر حالت x را در برداشته باشد.

پیشنهادها برای مطالعه بیشتر

برای جزئیات IEEE استاندارد، ویژه نگاشت

"An American National Standard: IEEE Standard for Binary Floating-Point Arithmetic" (Piscataway, N.J.:IEEE, 1985).

و Goldberg (1991) را ملاحظه کنید.

برای نتایجی در مورد کرانه‌های خطا برای اعمال ماتریسی ممیز شناور پایه‌ای کتاب

James H. Wilkinson, *Rounding Errors in Algebraic Processes* (Englewood Cliffs, N.J.: Prentice - Hall, 1963).

یک مرجع بسیار مفید و با ارزش می‌باشد.

بحث مربوط به اعمال ممیز شناور مقدماتی و خطاهای گرد کردن ناشی از حذف شدن و محاسبات بازگشتی امروزه در اغلب کتابهای آنالیز عددی مقدماتی ارائه می‌شوند. تعداد کمی از آنها را که استفاده کرده‌ایم و مفید یافته‌ایم، نام می‌بریم:

Atkinson, Kendall. *Elementary Numerical Analysis*. New York: John Wiley, 1993.

Cheney, Ward, and David Kincaid. *Numerical Mathematics and Computing*, 3d ed., Pacific Grove, Calif.: Brooks/Cole, 1994.

Conte, S. D., and C. de Boor. *Elementary Numerical Analysis: An Algorithmic Approach*, 3d ed. New York: McGraw-Hill, 1980.

Dahlquist. G., and A. Bjorck. *Numerical Methods*. Englewood Cliffs, N.J.: Prentice-Hall, 1974.

Faires, J. D., and R. L. Burden. *Numerical Methods*. Boston: PWS-Kent, 1993.

Forsythe, George E., Michael A. Malcolm, and Cleve B. Moler. *Computer Methods for Mathematical Computations*. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Golub, Gene H., and James M. Ortega. *Scientific Computing and Differential Equations: An Introduction to Numerical Methods*. San Diego: Academic Press,

- 1992.
- Johnston, R, L. *Numerical Methods: A Software Approach*. Toronto: John Wiley, 1982.
- Kahaner, D., C. B. Moler, and S. Nash. *Numerical Methods and Software*. Englewood Cliffs, N.J.: Printice-Hall, 1988.
- Ortega, James M., and William G. Poole, Jr. *Numerical Methods for Differential Equations*. Mass.: Pitman Publishing, 1981.
- Rice, John R. *Numerical Methods, Software and Analysis*. New York: McGraw-Hill, 1983.
- Skeel, Robert D., and J. B. Keiper. *Elementary Numerical Computing with Mathematica*. New York: McGraw-Hill, 1993.

پایداری الگوریتمها و وضعیت مسائل

۱-۳ برخی الگوریتمهای اساسی

با یک تعریف شروع می‌کنیم

تعریف ۱-۱-۳ یک الگوریتم، یک مجموعه مرتب از اعمال منطقی و حسابی است که هنگامی که در مورد یک مسأله محاسباتی تعریف شده توسط یک مجموعه داده (که داده‌های ورودی نامیده می‌شوند) به کار می‌رود یک جواب برای مسأله تولید می‌کند. یک جواب شامل یک مجموعه داده است که داده خروجی نامیده می‌شود.

در این کتاب برای سهولت بیشتر اغلب الگوریتمها را به صورت شبه کدها بیان خواهیم کرد که می‌توانند به سادگی به کدهای کامپیوتری تبدیل شوند. توصیف الگوریتمها به وسیله شبه کدها توسط استوارت (Stewart (1973)) رایج گردیده است.

۱-۱-۳ محاسبه نرم یک بردار

با مفروض بودن $x = (x_1, \dots, x_n)^T$ ، $\|x\|_2$ را محاسبه کنید.

الگوریتم ۱-۱-۳ محاسبه نرم یک بردار

داده‌های ورودی: n, x_1, \dots, x_n

گام ۱: $r = \max(|x_1|, \dots, |x_n|)$ را محاسبه کنید.

- گام ۲: $i = 1, \dots, n, y_i = x_i / r$ را محاسبه کنید.
 گام ۳: $\|x\|_2 = r\sqrt{y_1^2 + \dots + y_n^2}$ را محاسبه کنید.
 داده‌های خروجی: s

Pseudocodes (شبه کدها)

$$r = \max(|x_1|, \dots, |x_n|)$$

$$s = 0$$

For $i = 1$ to n do

$$y_i = x_i / r$$

$$s = s + y_i^2$$

$$s = r(s)^{1/2}$$

یک تذکر الگوریتمی: برای پرهیز از سرریز، هر عنصر x قبل از استفاده از فرمول زیر نرمال شده است.

$$\|x\|_2 = \sqrt{x_1^2 + \dots + x_n^2}$$

۲-۱-۳ محاسبه ضرب داخلی دو بردار

با مفروض بودن دو بردار n بعدی x و y ، ضرب داخلی زیر را محاسبه کنید.

$$x^T y = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$$

الگوریتم ۲-۱-۳ محاسبه ضرب داخلی دو بردار

داده‌های ورودی: $n, x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$

گام ۱: حاصلضربهای جزئی را محاسبه کنید: $s_i = x_i y_i, i = 1, \dots, n$

گام ۲: حاصلضربهای جزئی را جمع کنید: $\sum_{i=1}^n s_i$

داده‌های خروجی: Sum

Pseudocodes (شبه کدها)

$$\text{Sum} = 0$$

For $i = 1, \dots, n$ do

$$\text{Sum} = \text{Sum} + x_i y_i$$

۳-۱-۳ جواب یک دستگاه بالا مثلثی

دستگاه $Ty = b$ را در نظر بگیرید که در آن $T = (t_{ij})$ یک ماتریس بالا مثلثی نامنفرد است و $y = (y_1, y_2, \dots, y_n)^T$ بویژه

$$t_{11}y_1 + t_{12}y_2 + \dots + t_{1n}y_n = b_1$$

$$t_{22}y_2 + \dots + t_{2n}y_n = b_2$$

$$t_{33}y_3 + \dots + t_{3n}y_n = b_3$$

⋮

$$t_{n-1,n-1}y_{n-1} + t_{n-1,n}y_n = b_{n-1}$$

$$t_{nn}y_n = b_n$$

که در آن هر $t_{ii} \neq 0$ به ازای $i = 1, 2, \dots, n$.

ابتدا معادله آخر را برای به دست آوردن y_n حل می‌کنیم؛ سپس این مقدار را در معادله ما قبل آخر قرار داده و y_{n-1} را به دست می‌آوریم، و غیره. این فرایند معروف به جایگزینی پسرو می‌باشد. این الگوریتم می‌تواند به آسانی به صورت زیر نوشته شود.

الگوریتم ۳-۱-۳ جایگزینی پسرو

داده‌های ورودی: $T = (t_{ij})$ ، یک ماتریس بالا مثلثی نامنفرد $n \times n$ بوده، و b یک بردار n بعدی است.

$$\text{گام ۱: } y_n = \frac{b_n}{t_{nn}}$$

گام ۲: y_{n-1} تا y_1 را به طور متوالی محاسبه کنید.

$$y_i = \frac{1}{t_{ii}} \left(b_i - \sum_{j=i+1}^n t_{ij}y_j \right), \quad i = n-1, \dots, 2, 1$$

داده‌های خروجی: $y = (y_1, \dots, y_n)^T$

Pseudocodes (شبه کدها)

For $i = n, n-1, \dots, 3, 2, 1$ do

$$y_i = \frac{1}{t_{ii}} \left(b_i - \sum_{j=i+1}^n t_{ij}y_j \right)$$

تذکر: هنگامی که $i = n$ مجموع (Σ) انجام نمی‌شود.

تذکر MATCOM: این الگوریتم در برنامه BACKSUB در MATCOM پیاده‌سازی شده است. فصل ۴ را برای توصیف MATCOM ملاحظه کنید.

۳-۱-۴ محاسبه معکوس یک ماتریس بالا مثلثی

پیدا کردن معکوس یک ماتریس A از مرتبه $n \times n$ معادل پیدا کردن یک ماتریس X است به قسمی که

$$AX = I$$

فرض کنید $X = (x_1, \dots, x_n)$ و $I = (e_1, \dots, e_n)$ ، که در آن x_i ستون i ام ماتریس و e_i ستون i ام ماتریس I هستند. آنگاه حل معادله ماتریسی $AX = I$ معادل n دستگاه خطی زیر است.

$$Ax_i = e_i, \quad i = 1, \dots, n$$

بویژه هنگامی که A یک ماتریس مثلثی است کار ساده است. فرض کنید T یک ماتریس بالا مثلثی باشد، آنگاه پیدا کردن معکوس آن $S = (s_{11}, \dots, s_{nn})$ معادل حل n دستگاه خطی بالا مثلثی زیر است.

$$Tx_i = e_i, \quad i = 1, \dots, n$$

قرار دهید $s_i = (s_{1i}, s_{2i}, \dots, s_{ni})^T$

برای $i = 1$: $Ts_1 = e_1$ نتیجه می‌دهد

$$s_{11} = \frac{1}{t_{11}}$$

(عناصر s_{21} تا s_{n1} همگی برابر صفرند.)

برای $i = 2$: $Ts_2 = e_2$ نتیجه می‌دهد

$$s_{22} = \frac{1}{t_{22}}, \quad s_{12} = -\frac{1}{t_{11}} (t_{12}s_{22})$$

(عناصر s_{32} تا s_{n2} همگی برابر صفرند.)

برای $i = k$: $Ts_k = e_k$ نتیجه می‌دهد

$$s_{kk} = \frac{1}{t_{kk}}, \quad s_{ik} = -\frac{1}{t_{ii}} (t_{i,i+1}s_{i+1,k} + \dots + t_{ik}s_{kk}), \quad i = k-1, k-2, \dots, 1$$

(سایر عناصر ستون s_k همگی برابر صفرند.) شبه کدهای الگوریتم اکنون می‌توانند به سادگی به صورت

زیر نوشته شوند.

یک قرارداد

از هم اکنون، از قالب زیر برای توصیف الگوریتمها استفاده خواهیم کرد.

الگوریتم ۳-۱-۴ معکوس یک ماتریس بالا مثلثی. فرض کنید T یک ماتریس بالا مثلثی نامنفرد $n \times n$ باشد. الگوریتم زیر ماتریس S ، معکوس T را محاسبه می‌کند.

For $k = n, n - 1, \dots, 1$ do

$$1: s_{kk} = \frac{1}{t_{kk}}$$

$$2: s_{ik} = -t_{ii}^{-1} \sum_{j=i+1}^k t_{ij} s_{jk} \quad (i = k-1, k-2, \dots, 1)$$

مثال ۳-۱-۱ S را برای ماتریس T ارائه شده در زیر حساب کنید.

$$T = \begin{bmatrix} 5 & 2 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 4 \end{bmatrix}$$

$k = 3:$

$$s_{33} = \frac{1}{4}$$

$$s_{23} = -\frac{1}{t_{22}}(t_{23}s_{33})$$

$$= -\frac{1}{2}\left(1 \times \frac{1}{4}\right) = -\frac{1}{8}$$

$$s_{13} = -\frac{1}{t_{11}}(t_{13}s_{33} + t_{12}s_{23})$$

$$= -\frac{1}{5}\left(2\left(-\frac{1}{8}\right) + 3 \times \frac{1}{4}\right)$$

$$= -\frac{1}{10}$$

$k = ۲:$

$$s_{۲۲} = \frac{1}{t_{۲۲}} = \frac{1}{۲}$$

$$s_{۱۲} = -\frac{1}{t_{۱۱}}(t_{۱۲}s_{۲۲})$$

$$= -\frac{1}{۵}(۲ \times \frac{1}{۲}) = -\frac{1}{۵}$$

$k = ۱:$

$$s_{۱۱} = \frac{1}{t_{۱۱}} = \frac{1}{۵}$$

$$T^{-1} = S = \begin{bmatrix} \frac{1}{۵} & -\frac{1}{۵} & -\frac{1}{۱۰} \\ ۰ & \frac{1}{۲} & -\frac{1}{۸} \\ ۰ & ۰ & \frac{1}{۴} \end{bmatrix}$$

۵-۱-۳ روش حذفی گاوس برای حل دستگاه $Ax = b$

مسأله حل دستگاه خطی n معادله و n مجهول زیر را در نظر بگیرید.

$$a_{۱۱}x_1 + a_{۱۲}x_2 + \dots + a_{۱n}x_n = b_1$$

$$a_{۲۱}x_1 + a_{۲۲}x_2 + \dots + a_{۲n}x_n = b_2$$

⋮

$$a_{n۱}x_1 + a_{n۲}x_2 + \dots + a_{nn}x_n = b_n$$

یا در نمادگذاری ماتریسی،

$$Ax = b$$

که در آن $b = (b_1, \dots, b_n)^T$ و $A = (a_{ij})$

یک شیوه آشنا برای حل این مسأله طرح حذفی کلاسیک معروف به حذفی گاوس است. توصیف مفصل و نحوه توسعه این الگوریتم تاریخی و تغییرات عملی مهم آن در فصل ۵ و ۶ ارائه شده‌اند. ولیکن برای فهم بهتر از برخی مواد ارائه شده در این فصل، فقط یک توصیف خلاصه از طرح مقدماتی حذفی گاوس را ارائه می‌دهیم.

ایده مقدماتی کاهش دستگاه به یک دستگاه بالا مثلثی معادل است تا دستگاه بتواند به سادگی با استفاده از الگوریتم جایگزینی پسرو (الگوریتم ۳-۱-۳) حل شود.

فرایند کاهش از $n - 1$ گام تشکیل می‌شود. در توصیف الگوریتم از نمادگذارهای زیر استفاده می‌کنیم:

$$b^{(k)} = \begin{bmatrix} b_1^{(k)} \\ b_2^{(k)} \\ \vdots \\ b_n^{(k)} \end{bmatrix}, \quad A^{(k)} = [a_{ij}^{(k)}],$$

$$A^{(0)} = A, \quad b^{(0)} = b$$

گام ۱: در گام ۱، مجهول x_1 از معادله دوم تا معادله n ام حذف می‌شود. این کار توسط ضرب معادله اول در

$$\left(-\frac{a_{21}}{a_{11}}, -\frac{a_{31}}{a_{11}}, \dots, -\frac{a_{n1}}{a_{11}}\right)$$

و اضافه کردن آن به معادلات دوم تا n ام انجام می‌شود. مقادیر

$$m_{i1} = -\frac{a_{i1}}{a_{11}}, \quad i = 2, \dots, n$$

ضریب‌ها نامیده می‌شوند. در انتهای گام ۱، دستگاه $Ax = b$ به $A^{(1)}x = b^{(1)}$ تبدیل می‌شود، که در آن عناصر $A^{(1)} = (a_{ij}^{(1)})$ و $b^{(1)}$ به عناصر A و b به صورت زیر وابسته هستند.

$$a_{ij}^{(1)} = a_{ij} + m_{i1} a_{1j} \quad (i = 2, \dots, n; j = 2, \dots, n)$$

$$b_i^{(1)} = b_i + m_{i1} b_1 \quad (i = 2, \dots, n)$$

(توجه کنید که $(a_{11}^{(1)}, \dots, a_{1n}^{(1)})$ همگی برابر صفرند).

گام ۲: در گام ۲، x_2 از معادلات سوم تا n ام دستگاه $A^{(1)}x = b^{(1)}$ توسط ضرب معادله دوم در $A^{(1)}x = b^{(1)}$ در ضریب‌های

$$m_{i2} = -\frac{a_{i2}^{(1)}}{a_{22}^{(1)}}, \quad i = 3, \dots, n$$

و اضافه کردن نتیجه به معادلات سوم تا n ام، حذف می شود. اکنون دستگاه به شکل $A^{(2)}x = b^{(2)}$ می باشد، که عناصر آن به صورت زیر ارائه می شوند.

$$a_{ij}^{(2)} = a_{ij}^{(1)} + m_{i2} a_{2j}^{(1)} \quad (i = 3, \dots, n; j = 3, \dots, n)$$

$$b_i^{(2)} = b_i^{(1)} + m_{i2} b_2^{(1)} \quad (i = 3, \dots, n)$$

$$a_{i1}^{(2)} = 0 \quad (i = 3, \dots, n)$$

عناصر دیگر $A^{(2)}$ و $b^{(2)}$ به ترتیب همان عناصر $A^{(1)}$ و $b^{(1)}$ هستند.

گام k : در گام k ام، $(n - k)$ ضربگر

$$m_{ik} = -\frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \quad i = k+1, \dots, n$$

تشکیل می شوند؛ با استفاده از آنها، از معادلات $(k + 1)$ ام تا n ام دستگاه $A^{(k-1)}x = b^{(k-1)}$ حذف می شود. عناصر $A^{(k)}$ و $b^{(k)}$ توسط روابط زیر ارائه می شوند.

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} + m_{ik} a_{kj}^{(k-1)} \quad (i = k+1, \dots, n; j = k+1, \dots, n)$$

$$b_i^{(k)} = b_i^{(k-1)} + m_{ik} b_k^{(k-1)} \quad (i = k+1, \dots, n)$$

$$a_{ik}^{(k)} = 0 \quad (i = k+1, \dots, n)$$

عناصر دیگر $A^{(k)}$ و $b^{(k)}$ به ترتیب همان عناصر $A^{(k-1)}$ و $b^{(k-1)}$ هستند.

گام $n-1$: در انتهای گام $(n - 1)$ ام، ماتریس کاهش یافته $A^{(n-1)}$ بالا مثلثی است و بردار اصلی b به $b^{(n-1)}$ تبدیل می شود.

اکنون برای نوشتن شبه کدهای طرح حذفی گاوس آماده هستیم.

ملاحظات زیر به نوشتن شبه کدها کمک خواهند کرد:

۱. $(n-1)$ گام وجود دارند؛ $k = 1, 2, \dots, n-1$

۲. برای هر مقدار k ، $n-k$ ضربگر وجود دارند: m_{ik} ، $(i = k+1, \dots, n)$

۳. برای هر مقدار k ، فقط $(n-k)^2$ عنصر $(a_{ij}^{(k)})$ تغییر می کنند

$(i = k+1, \dots, n; j = k+1, \dots, n)$. $n-k$ عنصر زیر عنصر (k, k) ستون k ام صفر

هستند و سایر عناصر که تغییر نمی یابند همان عناصر متناظر با عناصر $A^{(k-1)}$ هستند.

الگوریتم ۳-۱-۵ الگوریتم حذفی گاوس پایه ای

For $k = 1, 2, \dots, n-1$ do

For $i = k+1, \dots, n$ do

$$m_{i,k} = -\frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} \quad (\text{با فرض این که } a_{kk}^{(k-1)} \neq 0)$$

$$b_i^{(k)} = b_i^{(k-1)} + m_{i,k} b_k^{(k-1)}$$

For $j = k+1, \dots, n$ do

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} + m_{i,k} a_{kj}^{(k-1)}$$

(توجه کنید که $(a_{ij}^{(k)}) = (a_{ij}^{(k-1)})$ و $(b = b^{(k)})$)

تبصره ها:

۱. الگوریتم حذفی گاوس پایه ای معمولاً به عنوان الگوریتم حذفی گاوس بدون تعویض سطرها

یا الگوریتم حذفی گاوس بدون محورگیری معروف است. دلیل داشتن چنین نامی در بحث این

الگوریتم در فصل ۵ روشن خواهد شد.

۲. الگوریتم حذفی گاوس پایه ای به صورتی که در اینجا ارائه شده است، معمولاً در عمل مورد

استفاده قرار نمی گیرد. دو گونه عملی این الگوریتم معروف به حذفی گاوس با محورگیری جزئی

و کامل در فصل ۵ و ۶ توصیف می شوند.

۳. فرض کرده ایم که مقادیر $a_{11}, \dots, a_{11}^{(n-1)}, \dots, a_{nn}^{(n-1)}$ مخالف صفر باشند. اگر هر یک از آنها صفر

باشد، الگوریتم متوقف خواهد شد.

مثال ۳-۱-۲ دستگاه زیر را حل کنید.

$$5x_1 + x_2 + x_3 = 7$$

$$x_1 + x_2 + x_3 = 3$$

$$2x_1 + x_2 + 3x_3 = 6$$

یا

$$\begin{pmatrix} 5 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 7 \\ 3 \\ 6 \end{pmatrix}$$

$$Ax = b$$

گام ۱:

$$k = 1$$

$$i = 2, 3$$

$$m_{21} = -\frac{a_{21}}{a_{11}} = -\frac{1}{5}$$

$$b_2^{(1)} = b_2 + m_{21}b_1 = \frac{1}{5}$$

$$j = 2, 3$$

$$i = 2, j = 2: \quad a_{22}^{(1)} = a_{22} + m_{21}a_{12} = \frac{4}{5}$$

$$i = 2, j = 3: \quad a_{23}^{(1)} = a_{23} + m_{21}a_{13} = \frac{4}{5}$$

$$i = 3, j = 2: \quad a_{32}^{(1)} = a_{32} + m_{31}a_{12} = \frac{3}{5}$$

$$m_{31} = -\frac{a_{31}}{a_{11}} = -\frac{2}{5}$$

$$i = 3, j = 3: \quad a_{33}^{(1)} = a_{33} + m_{31}a_{13} = \frac{13}{5}$$

$$b_3^{(1)} = b_3 + m_{31}b_1 = \frac{16}{5}$$

(توجه کنید که: $b_1^{(1)} = b_1$, $a_{31}^{(1)} = a_{31} = 0$, $a_{11}^{(1)} = a_{11}$, $a_{12}^{(1)} = a_{12}$, $a_{13}^{(1)} = a_{13}$.)

$$\begin{bmatrix} 5 & 1 & 1 \\ 0 & \frac{4}{5} & \frac{4}{5} \\ 0 & \frac{2}{5} & \frac{13}{5} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ \frac{8}{5} \\ \frac{16}{5} \end{bmatrix}$$

$$A^{(1)}x = b^{(1)}$$

$$k = 2$$

گام ۲:

$$i = 3$$

$$m_{32} = -\frac{a_{32}^{(1)}}{a_{22}^{(1)}} = -\frac{2}{4}$$

$$i = 3, j = 3: a_{33}^{(2)} = a_{33}^{(1)} + m_{32}a_{23}^{(1)} = 2$$

$$b_3^{(2)} = b_3^{(1)} + m_{32}b_2^{(1)} = 2$$

$$\begin{bmatrix} 5 & 1 & 1 \\ 0 & \frac{4}{5} & \frac{4}{5} \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ \frac{8}{5} \\ 2 \end{bmatrix}$$

$$A^{(2)}x = b^{(2)}$$

توجه کنید که $A^{(2)}$ بالا مثلثی است.

جایگزینی پسرو: این دستگاه مثلثی به آسانی با استفاده از جایگزینی پسرو حل می‌شود:

$$2x_3 = 2 \Rightarrow x_3 = 1$$

$$\frac{4}{5}x_2 + \frac{4}{5}x_3 = \frac{8}{5} \Rightarrow x_2 = 1$$

$$5x_1 + x_2 + x_3 = 7 \Rightarrow x_1 = 1$$

۲-۳ تعریفها و مفاهیم پایداری

مثالهای مربوط به حذفهای فاجعه‌آمیز و محاسبات بازگشتی در فصل ۲ یک وجه مشترک داشتند: دقیق نبودن نتیجه محاسبه شده در هر حالت کاملاً ناشی از الگوریتم استفاده شده بود، زیرا به محض آن که الگوریتم عوض شد یا دوباره مرتب شد و در مورد مسأله با همان داده‌ها به کار برده شد، نتیجه خیلی

رضایت بخش شد. بنابراین ما از دو نوع متفاوت الگوریتم برای یک مسأله مفروض صحبت می‌کنیم. الگوریتمهای نوع اول که نتایج نادقیق ارائه می‌دهند - نمونه هایی از الگوریتمهای ناپایدارند، در حالی که الگوریتمهای نوع دوم که نتایج رضایت بخش ارائه می‌دهند - الگوریتمهای پایدار هستند.

مطالعه پایداری بسیار مهم است. این مطالعه از طریق تحلیل خطای گرد کردن انجام می‌شود. دو نوع وجود دارد: تحلیل خطای پیشرو و تحلیل خطای پسرو.

در تحلیل پیشرو کوششی انجام می‌شود تا ملاحظه شود که چه مقدار جواب محاسبه شده حاصل از الگوریتم از جواب واقعی مبتنی بر همان داده‌ها تفاوت دارد.

تعریف ۱-۲-۳ یک الگوریتم پایدار پیشرو نامیده می‌شود اگر جواب محاسبه شده x از جهتی نزدیک به جواب واقعی x باشد.

کرانه‌های خطای گرد کردن به دست آمده در فصل ۲ برای اعمال حسابی مختلف نتیجه تحلیلهای پیشرو خطا هستند، بنابراین از قضیه ۲-۳-۱ بلافاصله نتیجه می‌گیریم که اعمال حساب ممیز شناور (جمع، تفریق، ضرب، تقسیم) پایدار پیشرو هستند.

از طرف دیگر تحلیل پسرو، خطا را به داده‌های مسأله به جای جواب مسأله وابسته می‌سازد. بنابراین ما پایدار پسرو را به صورت زیر تعریف می‌کنیم.

تعریف ۲-۲-۳ یک الگوریتم پایدار پسرو نامیده می‌شود اگر یک جواب واقعی برای یک مسأله نزدیک به مسأله مورد بحث تولید کند.

تحلیل پسرو خطا که توسط ویلکینسون (۱۹۸۶-۱۹۱۹) در نوشته‌ها متداول گردید، اکنون به طور وسیعی در محاسبات ماتریسی استفاده می‌شود؛ با استفاده از این تحلیل، پایداری (یا ناپایداری) بسیاری از الگوریتمها در جبر خطی عددی در سالهای اخیر بنا نهاده شده‌اند. در این کتاب، هدف ما از «پایداری» پایداری پسرو خواهد بود، مگر ذکر شود.

به عنوان یک مثال ساده از پایداری پسرو، مجدداً حالت مجموع دو عدد ممیز شناور x و y را در نظر بگیرید. قبلاً ملاحظه کردیم که:

$$\begin{aligned} fl(x + y) &= (x + y)(1 + \delta) \\ &= x(1 + \delta) + y(1 + \delta) \\ &= x' + y' \end{aligned}$$

بنابراین مجموع محاسبه شده دو عدد ممیز شناور x و y مجموع واقعی دو عدد ممیز شناور x' و y'

است. چون

$$|\delta| \leq \mu$$

' x و y هر دو به ترتیب به x و y نزدیک هستند. بنابراین نتیجه می‌گیریم که عمل جمع دو عدد ممیز شناور پایدار پسرو است. البته عبارتهای مشابه برای اعمال دیگر حساب ممیز شناور وجود دارد. برای مثال، جمع n عدد ممیز شناور و محاسبه ضرب داخلی دو بردار ممیز شناور پایدار پسرو هستند. برای مثالی از نوع دیگر، مسأله حل دستگاه خطی $Ax = b$ را در نظر بگیرید.

تعریف ۳-۲-۳ یک الگوریتم برای حل $Ax = b$ پایدار نامیده خواهد شد اگر جواب محاسبه شده \bar{x} به قسمی باشد که

$$(A + E)\bar{x} = b + \delta b$$

با E و δb کوچک.

اندازه‌گیری کوچک بودن

«کوچک بودن» یک ماتریس یا یک بردار یا با نگاه کردن به عناصر ماتریس یا با محاسبه نرم آن اندازه‌گیری می‌شود.

خطاها در نرم ماتریسی در برابر خطاهای عناصر

اگر چه اندازه‌گیری خطاها در محاسبات با استفاده از نرمها در محاسبات ماتریسی مرسوم است، و لیکن اهمیت اندازه خطاها مؤلفه به مؤلفه به طور روز افزون افزایش می‌یابد. در حقیقت مفهوم بیشتری را می‌رساند.

یک ماتریس A از مرتبه $n \times n$ دارای n^2 عنصر است اما نرم A یک عدد تک است. بنابراین کوچک بودن یا بزرگ بودن نرم یک ماتریس خطا مانند E به درستی کوچک بودن یا بزرگ بودن عناصر فردی E را منعکس نمی‌نماید. برای مثال اگر $E = (10, 0/00001, 1)^T$ آنگاه $\|E\| = 10/0499$. بنابراین عنصر کوچک $0/00001$ در اندازه نرم منعکس نمی‌شود.

مثالهایی از الگوریتمهای پایدار و ناپایدار توسط تحلیل خطای پسرو

مثال ۳-۲-۱ یک الگوریتم پایدار - حل یک دستگاه بالا مثلثی با استفاده از جایگزینی پسرو. الگوریتم ۳-۱-۳ (روش جایگزینی پسرو) را در نظر بگیرید. فرض کنید الگوریتم با استفاده از

ذخیره‌سازی حاصلضرب داخلی در دقت مضاعف پیاده‌سازی شود. آنگاه می‌توان نشان داد (فصل ۱۱ ملاحظه شود) که جواب محاسبه شده \hat{x} در رابطه

$$(T + E)\hat{x} = b$$

صدق می‌کند که در آن عناصر ماتریس خطا E کاملاً کوچک هستند. در حقیقت، اگر $E = (e_{ij})$ و $T = (t_{ij})$ آنگاه

$$|e_{ij}| \leq c\mu |t_{ij}|, \quad i, j = 1, \dots, n$$

و c یک ثابت از مرتبه واحد است. بنابراین فرایند جایگزینی پسر و برای حل یک دستگاه بالا مثلثی پایدار است.

مثال ۳-۲-۲ یک الگوریتم ناپایدار - حذفی گاوس بدون محورگیری

مسئله حل دستگاه خطی نامنفرد $Ax = b$ را با استفاده از روش حذفی گاوس (الگوریتم ۳-۱-۵) در نظر بگیرید.

توسط ویلکینسون (فصل ۱۱ این کتاب را ملاحظه کنید) نشان داده شده است که اگر فرایند با شکست مواجه نشود جواب محاسبه شده \hat{x} در رابطه

$$(A + E)\hat{x} = b$$

با

$$\|E\|_{\infty} \leq cn^2 \rho \|A\|_{\infty} \mu + o(\mu^2)$$

صدق می‌کند که در آن

$$A^{(k)} = (a_{ij}^{(k)})$$

ماتریسهای کاهش یافته در فرایند حذفی هستند و ρ معروف به عامل رشد توسط رابطه زیر ارائه می‌شود.

$$\rho = \frac{\max_k \max_{i,j} |a_{ij}^{(k)}|}{\max_{i,j} |a_{ij}|}$$

بویژه اگر $\alpha = \max_{ij} |a_{ij}|$ و $\alpha_k = \max_{ij} |a_{ij}^{(k)}|$ ، آنگاه عامل رشد ρ به صورت زیر ارائه می‌شود.

$$\rho = \frac{\max(\alpha, \alpha_1, \dots, \alpha_{n-1})}{\alpha}$$

برای یک ماتریس دلخواه A ، ρ می‌تواند کاملاً بزرگ باشد، زیرا عناصر ماتریسهای کاهش یافته $A^{(k)}$ می‌توانند به طور دلخواه بزرگ شوند. برای دیدن این مطلب، مثال ساده زیر را ملاحظه کنید.

$$A = \begin{bmatrix} 10^{-10} & 1 \\ 1 & 2 \end{bmatrix}$$

یک گام از روش حذفی گاوس با استفاده حساب ممیز شناور با ۹ رقم اعشار ماتریس کاهش یافته زیر را نتیجه خواهد داد.

$$A^{(1)} = \begin{bmatrix} 10^{-10} & 1 \\ 0 & 2 - 10^{-10} \end{bmatrix} = \begin{bmatrix} 10^{-10} & 1 \\ 0 & -10^{-10} \end{bmatrix}$$

بنابراین عامل رشد برای این مسأله عبارت است از:

$$\rho = \frac{\max(\alpha, \alpha_1)}{\alpha} = \frac{\max(2, 10^{-10})}{2} = \frac{10^{-10}}{2}$$

که کاملاً بزرگ است. بنابراین اگر یک دستگاه خطی با این ماتریس بالا مثلثی کاهش یافته را حل کنیم، نمی‌توانیم انتظار یک ماتریس خطای کوچک E داشته باشیم. در واقع اگر ما بخواهیم دستگاه

$$10^{-10}x_1 + x_2 = 1$$

$$x_1 + 2x_2 = 3$$

را با استفاده از ماتریس $A^{(1)}$ قبلی حل کنیم، جواب محاسبه شده $x_2 = 1$ ، $x_1 = 0$ خواهد بود، در حالی که جواب واقعی $x_1 = x_2 = 1$ است. توجه کنید که در حل دستگاه بالا با استفاده از روش حذفی گاوس بردار

$$b = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

به بردار

$$b^{(1)} = \begin{bmatrix} 1 \\ 3 - 10^{-10} \end{bmatrix} = \begin{bmatrix} 1 \\ -10^{-10} \end{bmatrix}$$

(در حساب ۹ رقمی) تغییر خواهد کرد. این نشان می‌دهد که روش حذفی گاوس برای یک دستگاه دلخواه ناپایدار است.

تذکره: روش حذفی گاوس بدون محورگیری برای همه ماتریسها ناپایدار نیست. رده‌های خاصی از ماتریسها وجود دارند، نظیر ماتریسهای متقارن معین مثبت، که برای آنها روش حذفی گاوس پایدار است. این دستگاه خاص را در فصل ۶ با جزئیات بیشتر مورد بحث قرار خواهیم داد.

اگر یک الگوریتم برای یک ماتریس مفروض پایدار باشد، آنگاه خوب است که بدانیم که آیا الگوریتم برای هر ماتریس A در یک رده مفروض پایدار است.

تعریف ۳-۲-۴ یک الگوریتم برای یک رده از ماتریسهای C پایدار است اگر برای هر ماتریس A در C ، جواب محاسبه شده توسط الگوریتم جواب واقعی یک مسأله نزدیک به مسأله مورد بحث باشد. بنابراین برای دستگاه خطی

$$Ax = b$$

یک الگوریتم برای یک رده از ماتریسهای C پایدار است اگر برای هر ماتریس $A \in C$ و برای هر b ، یک جواب محاسبه شده \hat{x} تولید کند که در رابطه

$$(A + E)\hat{x} = b + \delta b$$

به ازای ماتریسی مانند E و برداری مانند δb صدق کند که در آن $A + E$ نزدیک به A و $b + \delta b$ نزدیک به b هستند.

۳-۳ وضعیت مسأله و تحلیل اختلال

از بحث قبل نباید این نظریه را برداشت نماییم که اگر یک الگوریتم پایدار برای یک مسأله استفاده شود، آنگاه جواب محاسبه شده دقیق خواهد بود. یک خاصیت مسأله که وضعیت نامیده می‌شود همچنین در دقیق بودن یا دقیق نبودن نتیجه محاسبه شده سهم دارد.

وضعیت یک مسأله یک خاصیت از خود مسأله است. این مربوط می‌شود به این که چگونه جواب مسأله عوض خواهد شد اگر داده‌های ورودی حاوی برخی ناخالصیها باشند. این ناخالصیها از این حقیقت ناشی می‌شوند که در کاربردهای عملی اغلب داده‌ها از برخی مشاهدات تجربی به دست می‌آیند که در آنها اندازه‌ها می‌توانند تحت تأثیر اختلالها در داده‌ها قرار گیرند. منابع دیگری از خطا نیز، نظیر خطاهای گرد کردن (بحث شده در فصل ۱۱) خطاهای گسسته‌سازی، و غیره وجود دارند. بنابراین

هنگامی که تحلیل‌گر عددی یک مسئله برای حل در اختیار دارد باید مسئله را نه تنها با داده‌های اصلی، بلکه با داده‌های اختلال یافته حل کند. سؤالی که طبیعتاً مطرح می‌شود این است که چه اثرهایی این اختلالها بر روی جواب دارند؟

یک مطالعه نظری انجام یافته توسط تحلیل‌گرهای عددی برای بررسی این اثرها، که مستقل از الگوریتم خاص استفاده شده برای حل مسئله می‌باشد، تحلیل اختلال نامیده می‌شود. این مطالعه به ما کمک می‌کند که کشف کنیم که آیا مسئله داده شده «بد» یا «خوب» است در این جهت که آیا اختلالهای کوچک در داده‌ها یک تغییر بزرگ یا کوچک در جواب ایجاد می‌کنند. بویژه ما از تعریف استاندارد زیر استفاده می‌کنیم.

تعریف ۳-۳-۱ یک مسئله (نسبت به یک مجموعه مفروض از داده‌ها) یک مسئله بد وضع نامیده می‌شود اگر صرفنظر از روش حل یک خطای نسبی کوچک در داده‌ها موجب یک خطای نسبی بزرگ در جواب شود. در غیر این صورت خوش وضع نامیده می‌شود.

فرض کنید یک مسئله P با یک داده ورودی c حل شود. فرض کنید $P(c)$ مقدار مسئله را با ورودی c نمایش دهد. فرض کنید δc اختلال در c را نمایش دهد. آنگاه گفته می‌شود P برای داده ورودی c بد-وضع است اگر خطای نسبی در جواب،

$$\frac{|P(c + \delta c) - P(c)|}{|P(c)|}$$

بزرگتر از خطای نسبی در داده، یعنی $\frac{|\delta c|}{|c|}$ باشد.

تذکر: تعریف وضعیت بستگی به داده دارد. بنابراین یک مسئله که برای یک مجموعه از داده بد-وضع است می‌توند برای مجموعه دیگری خوش-وضع باشد.

۳-۴ وضعیت مسئله، پایداری الگوریتم، ودقت جواب

همان‌گونه که در بخش قبل بیان شد، وضعیت یک مسئله یک خاصیت از خود مسئله است و نه از الگوریتم استفاده شده برای حل مسئله. البته برای یک استفاده کننده دقت جواب محاسبه شده اهمیت زیادی دارد. ولیکن دقت یک جواب محاسبه شده توسط یک الگوریتم مستقیماً هم به پایداری الگوریتم و هم به وضعیت مسئله وابسته است. اگر مسئله بد-وضع باشد، هر طور که پایداری الگوریتم باشد، دقت جواب محاسبه شده نمی‌تواند تضمین شود.

پایداری پرسو و دقت

توجه کنید که تعریف پایداری پرسو بیان نمی‌کند که جواب محاسبه شده \hat{x} توسط یک الگوریتم پایدار پرسو نزدیک به جواب واقعی مسأله اصلی خواهد بود. با وجود این هنگامی که یک الگوریتم پایدار در مورد یک مسأله خوش - وضع به کار برده می‌شود، جواب محاسبه شده باید نزدیک به جواب واقعی باشد، زیرا در این حالت پایداری جواب واقعی یک مسأله نزدیک را تضمین خواهد کرد و خوش وضعیت تضمین خواهد کرد که جواب برای مسأله اصلی و جواب برای مسأله نزدیک به هم نزدیک می‌باشند.

بد - وضعی یک مسأله، جواب محاسبه شده را، حتی با استفاده از یک الگوریتم پایدار، آلوده می‌سازد، بنابراین یک جواب غیر قابل پذیرش حاصل می‌گردد. هنگامی که یک جواب محاسبه شده قانع‌کننده نیست، برخی استفاده‌کننده‌ها (که معمولاً به وضعیت اهمیت نمی‌دهند) تمایل دارند که الگوریتم را برای بی‌دقتی مقصر بدانند، این نادرست است. با وجود این اگر یک الگوریتم «پایدار» در مورد یک مسأله بد-وضع به کار برده شود، خطای بیشتری از آنچه که داده‌ها موجب می‌شوند، وارد نخواهد کرد.

از بحث قبل، کاملاً واضح است که بررسی وضعیت یک مسأله خیلی مهم است.

عدد وضعیت یک مسأله

تحلیل گران عددی معمولاً سعی می‌کنند یک عدد که عدد وضعیت نامیده می‌شود به یک مسأله وابسته سازند. عدد وضعیت مشخص می‌کند که آیا مسأله بد - وضع یا خوش - وضع است. به ویژه عدد وضعیت یک کران برای خطای نسبی جواب هنگامی که یک اختلال کوچک در مورد داده‌ها اعمال می‌شود، ارائه می‌دهد.

در جبر خطی عددی، اعداد وضعیت برای بسیاری (نه همه) مسائل مشخص شده‌اند. متأسفانه محاسبه عدد وضعیت اغلب پیچیده‌تر از حل خود مسأله بوده و به صرف وقت بیشتری نیاز دارد. برای مثال (همان‌گونه که در فصل ۶ ملاحظه خواهیم کرد) برای مسأله دستگاه خطی $Ax = b$ ، عدد وضعیت

عبارت است از:

$$\text{Cond}(A) = \|A\| \|A^{-1}\|$$

بنابراین محاسبه عدد وضعیت در این حالت محاسبه معکوس A را در بر دارد؛ که گران‌تر از حل دستگاه $Ax = b$ است. در فصل ۶ روشهای برآورد $\text{Cond}(A)$ را بدون محاسبه صریح A^{-1} مورد بحث قرار خواهیم داد.

ما وضعیت هر مسأله را به طور مفصل در فصلهای مربوطه توضیح خواهیم داد؛ ولیکن قبل از تمام کردن این بخش اجازه دهید چند مثال معروف از مسائل بد - وضع را بیان نماییم.

یک تفریق بد - وضع

تفریق $c = a - b$ را در نظر بگیرید:

$$a = 12354101$$

$$b = 12345678$$

$$c = a - b = 8423$$

اکنون در a در رقم ششم اختلال ایجاد می‌کنیم.

$$\hat{a} = 12354001$$

$$\hat{c} = c + \delta c = \hat{a} - b = 8323$$

بنابراین، یک اختلال در ششمین رقم مقدار ورودی موجب یک تغییر در رقم دوم جواب شده است. توجه کنید که خطای نسبی در داده‌ها عبارت است از:

$$\frac{a - \hat{a}}{a} = 0.000008$$

در حالی که خطای نسبی در جواب محاسبه شده به صورت زیر است.

$$\frac{c - \hat{c}}{c} = 0.1187722$$

یک مسأله ریشه‌یابی بد - وضع

حل معادله درجه دوم ساده‌ی زیر را در نظر بگیرید.

$$f(x) = x^2 - 2x + 1$$

ریشه‌ها عبارتند از $x = 1, 1$. اکنون در ضریب ۲ به اندازه $1/0.00001$ اختلال ایجاد می‌کنیم. ریشه‌های محاسبه شده چندجمله‌ای اختلال یافته $f(x) = x^2 - 2/0.00001x + 1$ عبارتند از $x_1 = 1/0.032$ و $x_2 = 0.99968$. خطاهای نسبی در x_1 و x_2 برابر 0.00032 هستند. خطای نسبی در داده‌ها برابر $10^{-6} \times 5$ می‌باشد.

۳-۵ چندجمله‌ای ویلکینسون

مثال قبل حاوی ریشه‌های مضاعف بود. ریشه‌های چندگانه و یا ریشه‌های نزدیک به یکدیگر مسأله ریشه‌یابی را همواره بد - وضع می‌سازند؛ با وجود این مسأله می‌تواند حتی به هنگامی که ریشه‌ها کاملاً مجزا هستند، نیز بد - وضع باشد. مثال معروف زیر را که توسط ویلکینسون ارائه شده است، در نظر بگیرید (مرجع Forsythe, Malcolm, and Moler (1977, 18-19) را ملاحظه کنید).

$$P(x) = (x - 1)(x - 2) \dots (x - 20) \\ = x^{20} - 210x^{19} + \dots$$

صفرهای $P(x)$ مقادیر ۱، ۲، ...، ۲۰ بوده و متمایز می‌باشند. اکنون ضریب x^{19} را از -210 به $-210 + 2 - 23$ تغییر داده و بقیه ضرایب را بدون تغییر باقی می‌گذاریم. ویلکینسون از یک کامپیوتر دودویی با $t = 30$ استفاده کرد. بنابراین این تغییر در مبنای ۲ موجب یک تغییر در رقم بامعنای 30 ام گردید. ریشه‌های چندجمله‌ای اختلال یافته که به دقت توسط ویلکینسون محاسبه شده‌اند، عبارتند از (از کتاب Forsythe, Malcolm, and Moler (1977, 18) اقتباس شده‌اند):

1/0.00000	0.0000	10/0.9526	6145	±	0/64350	0.904i
2/0.00000	0.0000	11/79363	3881	±	1/65232	9728i
3/0.00000	0.0000	13/99235	8137	±	2/51883	0.070i
4/0.00000	0.0000	16/730.73	7466	±	2/81262	4894i
4/999999	99928	19/50.243	9400	±	1/940.33	0.347i
6/0.00000	6944					
6/999969	7234					
8/0.00726	7603					
8/91725	0.249					
20/84690	8101					

جدول فوق نشان می‌دهد که برخی صفرها نسبت به اختلال حساس‌تر از بقیه هستند. تحلیل زیر، که توسط ویلکینسون انجام شده است (کتاب Forsythe, Malcolm, and Moler (1977, 19) را ملاحظه کنید)، کوششی برای توصیف این پدیده است.

فرض کنید چند جمله‌ای اختلال یافته به صورت زیر باشد.

$$P(x, \alpha) = x^{20} - \alpha x^{19} + \dots$$

آنگاه عدد (وضعیت)

$$\left. \frac{\partial x}{\partial \alpha} \right|_{x=i}$$

حساس بودن ریشه $x = i$ ، $i = 1, 2, \dots, 20$ ، را اندازه‌گیری می‌کند. برای محاسبه این عدد، از معادله $P(x, \alpha) = 0$ نسبت به α مشتق می‌گیریم.

$$\begin{aligned} \frac{\partial x}{\partial \alpha} &= \frac{-\partial P / \partial \alpha}{\partial P / \partial x} \\ &= \frac{x^{19}}{\sum_{i=1}^{20} \sum_{\substack{j=1 \\ j \neq i}}^{20} (x-j)} \end{aligned}$$

مقادیر $\left. \frac{\partial x}{\partial \alpha} \right|_{x=i}$ به ازای $i = 1, 2, \dots, 20$ در جدول زیر فهرست شده‌اند. (کتاب Forsythe, Malcolm, and Moler (1977, 19) را ملاحظه کنید).

ریشه	$\left. \frac{\partial x}{\partial \alpha} \right _{x=i}$	ریشه	$\left. \frac{\partial x}{\partial \alpha} \right _{x=i}$
۱	$-8/2 \times 10^{-18}$	۱۱	$-4/6 \times 10^7$
۲	$8/2 \times 10^{-11}$	۱۲	$2/0 \times 10^8$
۳	$-1/6 \times 10^{-6}$	۱۳	$-6/1 \times 10^8$
۴	$2/2 \times 10^{-3}$	۱۴	$1/3 \times 10^9$
۵	$-6/1 \times 10^{-1}$	۱۵	$-2/1 \times 10^9$
۶	$5/8 \times 10^1$	۱۶	$2/4 \times 10^9$
۷	$-2/5 \times 10^3$	۱۷	$-1/9 \times 10^9$
۸	$6/0 \times 10^4$	۱۸	$1/0 \times 10^9$
۹	$-8/3 \times 10^5$	۱۹	$-3/1 \times 10^8$
۱۰	$7/6 \times 10^6$	۲۰	$4/3 \times 10^7$

ریشه یابی و محاسبه مقادیر ویژه

مثالهای قبل به ما درسهای خیلی مهمی را می آموزند. ایده خوبی نیست که برای محاسبه مقادیر ویژه یک ماتریس ضرایب چندجمله‌ای مشخصه به صورت صریح محاسبه شوند و صفرهای آن به دست آیند، زیرا خطاهای گرد کردن در محاسبات به طور باثباتی برخی اختلالهای کوچک در ضرایب محاسبه شده چندجمله‌ای مشخصه ایجاد خواهند کرد، و این اختلالهای کوچک در ضرایب ممکن است موجب تغییرات بزرگ در صفرها شوند، و در نتیجه مقادیر ویژه به طور غیر دقیق محاسبه شوند.

۳-۶ یک دستگاه خطی بد-وضع

ماتریس

$$H = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \dots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \dots & \frac{1}{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} & \dots & \frac{1}{2n-1} \end{pmatrix}$$

به نام ریاضیدان معروف دیوید هیلبرت، ماتریس هیلبرت نامیده می شود. یک دستگاه خطی، حتی با ماتریس هیلبرت از مرتبه پایین، بد-وضع است. برای مثال $n = 5$ اختیار کنید و حل دستگاه

$$Hx = b$$

را با $b \approx (2/2833, 1/4500, 1/0929, 0/8845, 0/7456)^T$ در نظر بگیرید. جواب واقعی برابر $x = (1, 1, 1, 1, 1)^T$ است. اکنون در عنصر $(5, 1)$ ماتریس H در رقم پنجم اختلال ایجاد کرده و به دست می آوریم $0/20001$. جواب محاسبه شده با این ماتریس اختلال یافته بسیار جزئی، برابر $(0/9937, 1/1252, 0/4365, 1/8765, 0/5618)^T$ است. توجه کنید که $Cond(H) = O(10^5)$. برای مثالهای بیشتری از مسائل دستگاههای خطی بد-وضع فصل ۶ را ملاحظه کنید.

۷-۳ مثالهایی از مسائل مقادیر ویژه بد - وضع

مثال ۷-۳-۱ ماتریس 10×10 زیر را در نظر بگیرید.

$$A = \begin{bmatrix} 1 & & & & & & & & & & \\ & 1 & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & \ddots & & & & & & & \\ & & & & \ddots & & & & & & \\ & & 0 & & & \ddots & & & & & \\ & & & & & & \ddots & & & & \\ & & & & & & & \ddots & & & \\ & & & & & & & & \ddots & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{bmatrix}$$

مقادیر ویژه A همگی برابر ۱ هستند. اکنون اگر در عنصر $(1, 10)$ ماتریس A یک اختلال به اندازه مقدار کوچک $\epsilon = 10^{-10}$ ایجاد کنیم. آنگاه برای مقادیر ویژه ماتریس اختلال یافته با استفاده از نرم افزار *MATLAB* (توصیف شده در فصل بعد) که از یک الگوریتم کارای عددی برای محاسبه مقدار ویژه استفاده می کند، مقادیر زیر را پیدا خواهیم کرد:

$$1/0184 + 0/0980i$$

$$0/9506 + 0/0876i$$

$$1/0764 + 0/0632i$$

$$0/9051 + 0/0350i$$

$$1/0999 + 0/00i$$

$$1/0764 - 0/0632i$$

$$0/9051 - 0/0350i$$

$$1/0184 - 0/0980i$$

$$0/9506 - 0/0876i$$

(به تغییر در مقادیر ویژه توجه کنید).

مثال ۷-۳-۲ ماتریس دو قطری ویلکینسون. دوباره نباید تصور شود که هر مسأله مقدار ویژه فقط می تواند هنگامی بد - وضع باشد که مقادیر ویژه چندگانه و یا نزدیک به یکدیگر باشند. یک مسأله مقدار ویژه یا مقادیر ویژه کاملاً متمایز نیز می تواند بد - وضع باشد. ماتریس مثلثی 20×20 زیر (معروف به ماتریس دو قطری ویلکینسون) را در نظر بگیرید.

$$A = \begin{bmatrix} 20 & & & & \\ & 20 & & & \\ & 19 & 20 & & 0 \\ & & \ddots & \ddots & \\ 0 & & & \ddots & 20 \\ & & & & & 1 \end{bmatrix}$$

مقادیر ویژه A عبارتند از ۱، ۲، ...، ۲۰. اکنون در عنصر $(20, 1)$ به اندازه $\epsilon = 10^{-10}$ اختلال ایجاد می‌کنیم. اگر مقادیر ویژه این ماتریس کمی اختلال یافته با استفاده از یک الگوریتم پایدار (نظیر روش تکراری QR که در فصل ۸ بحث خواهد شد) محاسبه شوند، ملاحظه می‌گردد که برخی از آنها به طور جدی تغییر خواهند کرد؛ حتی مختلط خواهند شد.

در این حالت نیز، برخی از مقادیر ویژه بدوضع‌تر از سایرین هستند. برای توضیح این مطلب، ویلکینسون عدد وضعیت هر مقدار ویژه را حساب کرد. عدد وضعیت مقدار ویژه λ_i یک ماتریس A غیرناقص به صورت زیر تعریف می‌شود. (فصل ۸ ملاحظه شود).

$$\text{Cond}(\lambda_i) = \frac{1}{|y_i^T x_i|}$$

که در آن y_i و x_i به ترتیب بردارهای ویژه واحد چپ و راست A متناظر با مقدار ویژه λ_i هستند (یادآور می‌شویم که x یک بردار ویژه راست A وابسته به یک مقدار ویژه λ است اگر $Ax = \lambda x$ ، $x \neq 0$ ؛ به طور مشابه، y یک بردار ویژه چپ وابسته به λ است اگر $y^T A = \lambda y^T$). برای مثال فوق بردار ویژه x_r راست متناظر با $\lambda_r = r$ دارای مؤلفه‌های زیر است (کتاب Wilkinson (1965, 91) را ملاحظه کنید):

$$\left[1, \frac{20-r}{-20}, \frac{(20-r)(19-r)}{(-20)^2}, \dots, \frac{(20-r)!}{(-20)^{20-r}}, 0, \dots, 0 \right]$$

و مؤلفه‌های y_r عبارتند از

$$\left[0, \dots, 0, \frac{(r-1)!}{20^{r-1}}, \dots, \frac{(r-1)(r-2)}{20^2}, \frac{(r-1)}{20}, 1 \right]$$

این بردارها کاملاً در نرم ۲ نرمال نشده‌اند، با وجود این معکوس حاصلضرب آنها یک برآوردی از اعداد وضعیت ارائه می‌دهند. در حقیقت K_r ، عدد وضعیت مقدار ویژه $\lambda = r$ برابر است با

$$|K_r| = \frac{1}{|y_r^T x_r|} = \frac{2 \cdot 10^9}{(20-r)!(r-1)!}$$

عدد K_r برای همه مقادیر r بزرگ است. کوچک‌ترین K_i ها برای ماتریس دو قطری ویلکینسون عبارتند از $K_1 = K_{20} \approx 4/31 \times 10^7$ و بزرگ‌ترین آنها عبارتند از $K_{11} = K_1 \approx 3/98 \times 10^{12}$.

مثال ۳-۷-۳ (Wilkinson (1965, 92))

$$A = \begin{bmatrix} n & (n-1) & (n-2) & \dots & 3 & 2 & 1 \\ (n-1) & (n-1) & (n-2) & \dots & 3 & 2 & 1 \\ & (n-2) & (n-2) & \dots & 3 & 2 & 1 \\ & & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & & 2 & 2 & 2 \\ & & & & & 1 & 1 \end{bmatrix}$$

هنگامی که n افزایش می‌یابد، کوچکترین مقادیر ویژه به تدریج بدوضع می‌شوند. برای مثال، برای $n = 12$ اعداد وضعیت چند مقدار ویژه اول از مرتبه واحد هستند، در حالی که اعداد وضعیت متناظر با سه مقدار ویژه آخر از مرتبه 10^7 می‌باشند.

۸-۳ پایداری قوی، ضعیف و معتدل

هنگامی که پایداری یک الگوریتم را توسط تحلیل خطای پسرو بررسی می‌کنیم، برخی اوقات بیشتر از آنچه که تعریف قبلی پایداری مطرح می‌کند، به دست می‌آوریم. برای مثال، می‌توان نشان داد که اگر روش حذفی گاوس با محورگیری جزئی و کامل (برای بحثهای مربوط به محورگیری جزئی و کامل فصل ۵ را ملاحظه کنید) برای حل یک دستگاه نامنفرد $Ax = b$ به کار رود، جواب محاسبه شده \bar{x} نه تنها در رابطه

$$(A + E)\bar{x} = b$$

با ماتریس خطای کوچک E صدق می‌کند، بلکه $(A + E)$ حاصل نیز نامنفرد می‌باشد. تعریف استاندارد پایداری البته به این نیازی ندارد.

از طرف دیگر، اگر روش حذفی گاوس بدون محورگیری در مورد یک دستگاه معین مثبت متقارن به کار رود، آنگاه تحلیل خطای پسرو استاندارد نشان می‌دهد که جواب محاسبه شده \tilde{x} در رابطه $(A + E)\tilde{x} = b$ با ماتریس خطای کوچک E صدق می‌کند، اما نشان نمی‌دهد که $(A + E)$ معین مثبت متقارن است.

بنابراین در باره دو نوع الگوریتم پایدار می‌توانیم صحبت کنیم. یک نوع آن نه تنها یک ماتریس خطای کوچک ارائه می‌دهد، بلکه ماتریس $(A + E)$ اختلال یافته نیز متعلق به همان رده ماتریس A می‌باشد، و نوع دیگر آن فقط یک ماتریس خطای کوچک با هیچ قیدی بر روی $(A + E)$ ارائه می‌دهد. برای تمایز گذاشتن بین این دو نوع پایداری، اخیراً بانچ ((bunch (1987)) مفهوم پایداری قوی را مطرح کرده است، در پیروی از بانچ ما تعریف زیر را ارائه می‌دهیم.

تعریف ۱-۸-۳ یک الگوریتم برای حل مسأله دستگاه خطی $Ax = b$ برای یک رده از ماتریسهای C پایدار قوی است اگر برای هر A در C ، جواب محاسبه شده، جواب واقعی یک مسأله نزدیک آن باشد، و ماتریس $(A + E)$ نیز به C متعلق باشد.

مثالها (بانچ)

۱. روش حذفی گاوس با محورگیری بر روی رده ماتریسهای نامنفرد پایدار قوی است.
 ۲. الگوریتم چولسکی برای محاسبه تجزیه یک ماتریس معین مثبت متقارن به شکل $A = HH^T$ بر روی رده ماتریسهای معین مثبت متقارن پایدار قوی است.
 ۳. روش حذفی گاوس بدون محورگیری بر روی رده ماتریسهای غالب قطری نامنفرد پایدار قوی است.
- مشابه با تعریف پایداری قوی، بانچ همچنین مفهوم پایداری ضعیف را که به وضعیت مسأله بستگی دارد، مطرح کرد.

تعریف ۲-۸-۳ یک الگوریتم برای یک رده از ماتریسهای C پایدار ضعیف است اگر الگوریتم برای هر ماتریس خوش وضع در C یک جواب دقیق قابل پذیرش تولید کند.

بنابراین یک الگوریتم برای حل دستگاه خطی $Ax = b$ برای یک رده از ماتریسهای C پایدار ضعیف است اگر برای هر ماتریس خوش-وضع در C و به ازای هر b ، جواب محاسبه شده \tilde{x} برای $Ax = b$ به قسمی باشد که $\frac{\|x - \tilde{x}\|}{\|x\|}$ کوچک باشد.

انگیزه بانچ در مطرح کردن این تعریف این بود که به این مطلب اشاره کند که الگوریتم معروف

لوینسون^(۱) برای حل دستگاههای خطی حاوی ماتریسهای توپلیتز^(۲) (یک ماتریس $T = (t_{ij})$ توپلیتز است اگر عناصر در امتداد هر قطر آن یکسان باشند) بر روی رده ماتریسهای توپلیتز معین مثبت متقارن پایدار ضعیف است. این نتیجه بسیار مهم و قابل توجه توسط سی بنکو (Cybenko 1980) ثابت گردید. نتیجه مهم بود زیرا انجمن پردازش سیگنال به طور عادی از الگوریتم لوینسون برای سالها بدون بررسی رفتار پایداری این الگوریتم مهم، استفاده می کرد. (اخیراً در مرجع Bojanczyk, Brent, de Hoog, and Sweet (1994) نشان داده شده است که الگوریتم ارائه شده برای ماتریسهای توپلیتز، معروف به الگوریتم باریس^(۳)، دارای خواص عددی بهتری از الگوریتم لوینسون است).

تبصره‌ها در مورد پایداری، پایداری قوی و پایداری ضعیف:

۱. اگر یک الگوریتم پایدار قوی باشد، لزوماً پایدار است.
۲. توجه کنید که پایداری، پایداری ضعیف را ایجاب می کند. پایداری ضعیف برای اغلب استفاده کننده‌ها قابل قبول است.
۳. اگر یک تحلیل‌گر عددی بتواند ثابت کند که الگوریتم خاصی به طور ضعیف پایدار نیست، آنگاه نتیجه می شود که الگوریتم پایدار نیست زیرا پایدار نبودن ضعیف پایدار نبودن را ایجاب می کند.

پایداری معتدل: ما تعریف کرده‌ایم که الگوریتم پایدار پسرو می باشد، اگر الگوریتم جوابی تولید کند که جواب واقعی یک مسأله نزدیک باشد. اما احتمال زیاد وجود دارد که الگوریتم جوابی تولید کند که فقط نزدیک به جواب واقعی یک مسأله نزدیک باشد. این چنین الگوریتمی را چه می نامیم؟ وان دورن (Van Dooren (1991))، در پیروی از جونگ (Jong (1977)) یک چنین الگوریتمی را الگوریتم پایدار آمیخته نامیده است. ما تصور می کنیم که بهتر است چنین پایداری را پایداری معتدل بنامیم. بالاخره یک چنین الگوریتمی پایدار به مفهوم معتدل است.

تعریف ۳-۸-۳ یک الگوریتم به طور معتدل پایدار است اگر جوابی تولید کند که نزدیک به جواب واقعی یک مسأله نزدیک باشد.

مثال ۳-۸-۱

۱. الگوریتم QR برای مسائل کمترین توانهای دوم با رتبه ناقص به طور معتدل پایدار است (مرجع Lawson and Hanson (1974, 95) و فصل ۷ این کتاب را ملاحظه کنید).
۲. الگوریتم QR برای مسأله دستگاه خطی فرو معین با رتبه کامل به طور معتدل پایدار است (مرجع Lawson and Hanson (1974, 93) و فصل ۷ این کتاب را ملاحظه کنید).

۹-۳ مرور و خلاصه

در این بخش دو تا از مهمترین مفاهیم جبرخطی عددی، یعنی وضعیت مسأله و پایداری الگوریتم را مطرح کرده‌ایم و در مورد این که چگونه آنها بر دقت جواب اثر می‌گذارند، بحث کرده‌ایم.

۱. وضعیت مسأله: وضعیت مسأله یک خاصیت مسأله است. یک مسأله بدوضع گفته می‌شود، اگر یک تغییر کوچک در داده‌ها موجب یک تغییر بزرگ در جواب شود. در غیر این صورت خوش وضع است. وضعیت یک مسأله وابسته به داده‌ها است. یک مسأله می‌تواند نسبت به یک مجموعه از داده‌ها بدوضع باشد، اما نسبت به مجموعه دیگری کاملاً خوش وضع باشد. بدوضع‌ی یک ماتریس معمولاً توسط یک عدد که عدد وضعیت نامیده می‌شود، اندازه‌گیری می‌گردد. عدد وضعیت برای دستگاه خطی $Ax = b$ عبارت است از $\|A^{-1}\| \|A\|$.

مثالهای معروف از مسائل بدوضع عبارتند از چندجمله‌ای ویلکینسون برای مسأله ریشه‌یابی، ماتریس دو قطری ویلکینسون برای مسأله مقدار ویژه و ماتریس هیلبرت برای مسأله دستگاه خطی جبری.

۲. پایداری یک الگوریتم: گفته می‌شود یک الگوریتم یک الگوریتم پایدار پسرو است اگر جواب واقعی یک مسأله نزدیک را محاسبه کند. برخی از مثالهای الگوریتمهای پایدار عبارتند از جایگزینی پسرو و حذفی پیشرو برای دستگاههای مثلثی، روش حذفی گاوس با محورگیری برای دستگاه خطی، تجزیه QR با استفاده از تبدیلهای هاوس هولدر، گونوز و الگوریتم تکراری QR برای محاسبه مقادیر ویژه.

الگوریتم حذفی گاوس بدون تعویض سطر برای ماتریسهای دلخواه ناپایدار است.

۳. آثار وضعیت و پایداری بر روی دقت جواب: وضعیت مسأله و پایداری الگوریتم هر دو بر دقت جواب محاسبه شده توسط الگوریتم اثر می‌گذارند.

اگر یک الگوریتم پایدار در مورد یک مسأله خوش وضع به کار برده شود، یک جواب دقیق محاسبه خواهد کرد. از طرف دیگر اگر یک الگوریتم پایدار در مورد یک مسأله بد وضع به

کار برده شود، تضمینی وجود ندارد که جواب محاسبه شده دقیق باشد. تعریف پایداری پسرو آن را ایجاب نمی‌کند. با وجود این اگر یک الگوریتم پایدار در مورد یک مسأله بد وضع به کار برده شود خطاهای بیشتری از آنچه را که داده‌ها موجب می‌شوند، به وجود نخواهد آورد.

تمرینها

۱. (الف) نشان دهید که محاسبات ممیز شناور جمع، ضرب و تقسیم دو عدد پایدار پسرو و پایدار پیشرو هستند.

(ب) آیا محاسبات ممیز شناور ضرب داخلی و خارجی دو بردار پایدار پسرو هستند؟ برای جواب خود دلایلی ارائه دهید.

۲. ضریب رشد روش حذفی گاوس بدون محورگیری را برای ماتریسهای زیر پیدا کنید.

$$\begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0/\dots\dots\dots 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0/\dots\dots\dots 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 0/9 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 1 & 0/99 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 1 & 0/999 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 1 & 1/0001 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0/9 & 0/81 \\ 1 & 1/9 & 3/61 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 1 \\ 0/9 & 0/9 & 0/9 \\ 1 & 1/9 & 3/61 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}$$

۳. عدد وضعیت هر یک از ماتریسهای تمرین ۲ را به دست آورید.

۴. (الف) نشان دهید که به ازای همه اسکالرهایی c مخالف صفر داریم: $\text{Cond}(cA) = \text{Cond}(A)$.

(ب) نشان دهید که اگر $\|I\| \geq 1$ ، آنگاه $\text{Cond}(A) \geq 1$.

۵. ثابت کنید

$$\frac{\|fl(AB) - AB\|_F}{\|AB\|_F} \leq n\mu \text{Cond}_F(B) + O(\mu^2)$$

که در آن A و B ماتریس بوده و B نامنفرد است. $(\text{Cond}_F(B) = \|B\|_F \|B^{-1}\|_F)$
 ۶. آیا محاسبات ممیز شناور زیر پایدار پرسرو هستند؟ دلایلی برای جواب خود در هر حالت ارائه دهید.

(الف) $\text{fl}(x(y+z))$

(ب) $\text{fl}(x_1+x_2+\dots+x_n)$

(ج) $\text{fl}(x_1 x_2 \dots x_n)$

(د) $\text{fl}(x^T y/c)$ ، که در آن x و y بردار بوده و c یک اسکالر است.

(هـ) $\text{fl}(\sqrt{x_1^2+x_2^2+\dots+x_n^2})$

۷. عامل رشد روش حذفی گاوس را برای هر کدام از ماتریسهای زیر پیدا کنید و نتیجه بگیرید که روش حذفی گاوس برای دستگاههای خطی با این ماتریسها پایدار پرسرو است.

(الف)
$$\begin{bmatrix} 10 & 1 & 1 \\ 1 & 10 & 1 \\ 1 & 1 & 10 \end{bmatrix}$$

(ب)
$$\begin{bmatrix} 4 & 0 & 2 \\ 0 & 4 & 0 \\ 2 & 0 & 5 \end{bmatrix}$$

(ج)
$$\begin{bmatrix} 10 & 1 & 1 \\ 1 & 15 & 5 \\ 5 & 5 & 14 \end{bmatrix}$$

۸. نشان دهید که روش حذفی گاوس بدون محورگیری برای ماتریس زیر به طور قوی پایدار است.

$$\begin{bmatrix} 10 & 1 & 1 \\ 1 & 10 & 1 \\ 1 & 1 & 10 \end{bmatrix}$$

۹. فرض کنید H یک ماتریس بالا هسنبرگی کاهش ناپذیر باشد. یک ماتریس قطری D به قسمی پیدا کنید که $D^{-1}HD$ یک ماتریس بالا هسنبرگی نرمال شده باشد (یعنی همه عناصر زیر قطر آن برابر ۱ باشند). نشان دهید که ماتریس تبدیل D باید بد وضع باشد اگر یک یا چند عنصر زیر قطر H خیلی کوچک باشند. یک مثال عددی از مرتبه ۵ ارائه دهید که در این خاصیت صدق

کند.

۱۰. نشان دهید که ریشه‌های چندجمله‌ایهای زیر بد وضع هستند

$$(الف) \quad x^3 - 3x^2 + 3x - 1$$

$$(ب) \quad (x - 1)^3(x - 2)$$

$$(ج) \quad (x - 1)(x - 0.999)(x - 2)$$

۱۱. یک نتیجه ریاضی بیان کنید که نشان دهد ضرب ماتریس - بردار با یک ماتریس بد وضع ممکن است یک خطای نسبی بزرگ در نتیجه محاسبه شده ایجاد کند. یک مثال 2×2 برای دیدن این مطلب بسازید.

۱۲. (الف) با استفاده از روش حذفی گاوس بدون محورگیری، دستگاه خطی $Ax = b$ را با هر کدام از ماتریسهای تمرین ۲ حل کنید، و b را در هر حالت به نحوی انتخاب کنید که هر مؤلفه جواب واقعی برابر ۱ باشد

(ب) برای هر دستگاه در ۱۲ (الف) خطای نسبی $\frac{\|x - \hat{x}\|}{\|x\|}$ را حساب کنید که در آن \hat{x} جواب محاسبه شده می‌باشد.

(ج) در صورت بزرگ بودن خطای نسبی برای هر جواب، توضیح دهید که چه اتفاقی افتاده است.

۱۳. با استفاده از تمرین ۵، ثابت کنید که خطا در ضرب ماتریسی $B_1 B_2 \dots B_N A C_1 C_2 \dots C_N$ حداکثر به اندازه $\prod_{i=1}^N \text{Cond}(B_i) \text{Cond}(C_i)$ بزرگ خواهد شد.

پیشنهادها برای مطالعه بیشتر

برای کدهای کامپیوتری محاسبات ماتریسی استاندارد کتاب راهنما
T. Coleman and Charles Van Loan, *Matrix Computations*, (Philadelphia: SIAM, 1988)

را ملاحظه کنید. مفاهیم پایداری و وضعیت به طور کامل در کتاب
G.W. Stewart, *An Introduction to Matrix Computations* New York: Academic Press,
1973, Chapter 2

بحث شده است. توجه کنید که تعریف پایداری پرسو استوارت کمی با تعریف معمول پایداری که توسط ویلکینسون ارائه شده است، تفاوت دارد. بعلاوه ما قویاً به خواننده‌ها توصیه می‌کنیم که مقاله روشن‌کننده زیر که در زمینه پایداری می‌باشد را مطالعه نمایند.

James R. Bunch, *The Weak and Strong Stability of Algorithms in Numerical Linear Algebra*, *Lin. Alg. Appl.* 88-89(1987): 46-66.

بعلاوه مقاله Bunch, Demmel, and Van Loan (1989) را ملاحظه کنید.
 کتاب Wilkinson, *The Algebraic Eigenvalue Problem* یک منبع غنی از اطلاعات و
 نتایج در مورد پایداری پسر و برای الگوریتمهای ماتریسی می باشد.
 یک مقاله مهم که در مورد نظریه و مفاهیم انواع مختلف پایداری در حالت کلی بحث کرده
 است مقاله L. S. de Jong (1977) می باشد.
 یک کتاب جدید نوشته شده توسط استوارت و سان (Stewart and Sun (1990)) به نظریه
 اختلالهای مسائل ماتریسی مختلف اختصاص داده شده است و یک منبع غنی از اطلاعات در این زمینه
 می باشد. برای خواننده های با اطلاعات پیشرفته تر، مطالعه مقاله اصلی Rice (1966) و مقاله معروف
 Kahan (1966) را قویاً توصیه می کنیم.

الگوریتمهای عددی کارا و نرم افزارهای ریاضی

۴-۱ تعاریف و مثالها

- حل یک مسأله بر روی یک کامپیوتر شامل گامهای زیر است که باید به ترتیب انجام شوند:
۱. ساختن یک مدل ریاضی از مسأله - یعنی، تبدیل مسأله به زبان ریاضیات. برای مثال مدل‌های ریاضی بسیاری از مسائل مهندسی مجموعه‌هایی از معادلات دیفرانسیل معمولی و جزئی هستند.
 ۲. پیدا کردن یا توسعه روشهای سازنده برای حل مدل ریاضی. این گام معمولاً از یک جست‌وجو در میان نوشته‌ها برای پیدا کردن روشهایی که برای حل این مسأله موجود هستند، تشکیل می‌شود.
 ۳. شناسایی بهترین روش از تقطه نظر عددی (بهترین ممکن است ترکیبی از چند تای دیگر باشد). این روش را روش مؤثر (قابل اجرا) عددی می‌نامیم.
 ۴. بالاخره، پیاده‌سازی روش مؤثر عددی مشخص شده در گام ۳. این گام مشتمل بر نوشتن و اجرای یک برنامه کامپیوتری کارا و قابل اطمینان مبتنی بر روش مؤثر عددی مشخص شده است و ممکن است بلافاصله نیاز به بهره‌برداری از ساختار کامپیوتری خاص باشد.
- منظور از ایجاد نرم‌افزار ریاضی این است که برای یک دانشمند یا مهندس برنامه‌های کامپیوتری

فراهم کرد که او بتواند با اطمینان یک مسأله را حل کند. بنابراین نرم افزار ریاضی باید از کیفیت بالایی برخوردار باشد.

اجازه دهید مشخص کنیم که وقتی ما یک نرم افزار را یک نرم افزار ریاضی با کیفیت بالا می نامیم، منظورمان چیست. نرم افزار ریاضی با کیفیت بالا باید به طور ایده آل دارای خصوصیات زیر باشد (Rice, (1981):

۱. توانایی و انعطاف پذیری: نرم افزار بتواند برای حل تغییرات گوناگون و متعددی از مسأله اصلی و مسائل وابسته و نزدیک به آن، استفاده شود. برای مثال، مسائل وابسته و نزدیک به مسأله دستگاه خطی $Ax = b$ عبارتند از:

(الف) محاسبه معکوس A - یعنی پیدا کردن یک ماتریس X به قسمی که $AX = I$. اگر چه پیدا کردن معکوس A و حل $AX = b$ مسائل معادل هستند، ولیکن جواب یک دستگاه خطی با استفاده از معکوس توصیه نمی شود. محاسبه معکوس به طور صریح پرهیز می شود، مگر این که کاربرد خاصی از آن مورد نیاز باشد.

(ب) پیدا کردن دترمینان و رتبه A .

(ج) حل $AX = B$ ، که در آن B یک ماتریس است.

همچنین یک مسأله ماتریسی ممکن است دارای ساختار خاصی باشد و ممکن است معین مثبت، نواری، توپلیتز، متراکم، تنک و غیره باشد. نرم افزار باید بیان کند که چه تغییراتی از مسأله را می تواند انجام دهد و آیا برای ساختار خاصی میزان شده است.

۲. خواندن و تغییر به نحوی ساده: نرم افزار باید خوب مستندسازی شده باشد. مستندسازی باید روشن بوده و برای خواندن ساده باشد، حتی برای یک استفاده کننده، تا اگر بعضی تغییرات لازم باشد، آنها بتوانند به سادگی انجام شوند. در این رابطه از یادداشت روی جلد کتاب Forsythe, Malcolm, and Moler (1977) عبارت زیر را نقل می کنیم:

... نوشتن دو زیر روال خوب از تصمیم گیری در مورد این که کدام یک بهتر هستند، ساده تر است. در انتخاب یک زیر روال از میان زیرروالهای موجود برای یک مسأله خاص، ما تأکید زیادی بر روشن بودن و نحوه برنامه نویسی داریم. اگر چندین زیر روال از نظر دقت، اعتبار و کارایی قابل مقایسه باشند، ما آن را انتخاب می کنیم که از جهت خواندن و استفاده کمترین پیچیدگی را داشته باشد.

۳. قابل حمل بودن: باید قابل اجرا بر روی کامپیوترهای مختلف با کمی تغییر یا بدون تغییرات باشد.

۴. توانمند بودن: باید قادر به مقابله با وضعیت غیر قابل انتظار در طی اجرا باشد.
 ۵. مبتنی بر یک الگوریتم مؤثر عددی: باید مبتنی بر الگوریتمی با خواص عددی جذاب باشد. ما چندین بار از عبارت مؤثر عددی بدون مشخص کردن آن استفاده کرده‌ایم. این مهم‌ترین مؤلفه از یک نرم‌افزار ریاضی با کیفیت بالا است. ما یک الگوریتم ماتریسی را مؤثر عددی می‌نامیم اگر دارای خواص زیر باشد:

۱. منظور کلی: الگوریتم برای یک رده بزرگی از ماتریسها کار کند.
۲. قابل اطمینان بودن: الگوریتم باید هرگاه که در معرض شکست ناشی از خطاهای گرد کردن بیش از اندازه قرار گیرد یا قادر نباشد تا برخی معیارهای همگرایی خاص را برآورده سازد، اخطار دهد. الگوریتمهایی وجود دارند که جوابهای کاملاً غلط تولید می‌کنند بدون این که هیچ اخطاری دهند. روش حذفی گاوس بدون محورگیری (برای دستگاه خطی یا مسأله معادل) یکی از این گونه الگوریتمها است. این روش قابل اطمینان نیست.
۳. پایدار: کل خطاهای گرد کردن الگوریتم نباید از خطاهایی که در مسأله اصلی ذاتی هستند، تجاوز کند (بخشهای ۲-۳ و ۳-۳ را ملاحظه کنید).
۴. کارایی: کارایی یک الگوریتم توسط مقدار زمان کامپیوتری مصرف شده در پیاده‌سازی آن، اندازه‌گیری می‌شود. از نظر تئوری تعداد اعمال ممیز شناور لازم برای پیاده‌سازی الگوریتم کارایی آن را نشان می‌دهد.

تعریف ۴-۱-۱ یک عمل ممیز شناور، یا فلاپ "flop"، زمان کامپیوتری لازم برای اجرای دستور فرترن زیر می‌باشد.

$$A(I, J) = A(I, J) + t * A(I, J)$$

یک flop شامل یک ضرب، یک جمع و مقداری کار جهت اندیسها است. به طور مشابه، یک تقسیم همراه با یک جمع و یا تفریق به عنوان یک flop محسوب می‌شود. از این تعریف برای flop در بسته نرم‌افزاری همگانی LINPACK استفاده شده است (این بسته نرم‌افزاری در بخش ۴-۳ به طور مفصل توصیف خواهد شد).

تذکری در مورد تعریف شمارش flop ها: با پیشرفت تکنولوژی ابرکامپیوترها، تمایلی برای محسوب کردن یک جمع یا یک تفریق نیز به عنوان یک flop وجود دارد. این تعریف در چاپ دوم کتاب Golub and Van Loan (1989) مورد استفاده قرار گرفته است. ولیکن ما تصمیم گرفته‌ایم که از

تعریف LINPACK اصلی استفاده کنیم. توجه کنید که اگر یک جمع یا تفریق به عنوان یک flop استفاده شود، آنگاه flop جدید دارای ارزشی دو برابر قدیم خواهد بود.

تعریف ۴-۱-۲ یک الگوریتم ماتریسی حاوی محاسبات با ماتریسهایی از مرتبه n یک الگوریتم کارا نامیده می‌شود اگر بیش از $O(n^3)$ عمل ممیز شناور (flop) اختیار نکند (بنابراین قاعده تاریخی کرامر برای حل یک دستگاه خطی کارا نیست، زیرا $O(n!)$ عمل ممیز شناور (flop) برای اجرا نیاز دارد (فصل ۶ را ملاحظه کنید).

یک نکته لازم است در اینجا ذکر شود. یک الگوریتم ممکن است کارا باشد ولیکن ناپایدار باشد. برای مثال روش حذفی گاوس بدون محورگیری تقریباً $\frac{n^3}{3}$ عمل ممیز شناور (flops) برای یک ماتریس $n \times n$ نیاز دارد. بنابراین اگرچه کارا است ولیکن برای یک ماتریس دلخواه غیر قابل اطمینان و ناپایدار است.

۵. صرفه‌جویی در استفاده از حافظه: معمولاً در حدود n^2 مکان از حافظه برای ذخیره‌سازی یک ماتریس متراکم از مرتبه n مورد نیاز است. بنابراین اگر الگوریتم ذخیره‌سازی چندین ماتریس را در طی اجرائش لازم داشته باشد، یک تعداد زیادی از مکانهای حافظه لازم خواهد بود حتی اگر n چندان بزرگ نباشد. مهم است به هنگام طراحی یک الگوریتم توجه خاصی به صرفه‌جویی حافظه گردد.

با آرایش مجدد و دقیق یک الگوریتم، می‌توان حافظه لازم را به طور زیادی کاهش داد. (مثالهایی در این زمینه بعداً ارائه خواهند شد). به طور کلی، اگر یک ماتریس تولید شده در طی اجرای الگوریتم برای استفاده‌های بعدی مورد نیاز نباشد، باید عنصر محاسبه شده دیگری بر روی آن نوشته شود.

نمادگذاری برای رویهم نوشتن و تعویض

از نماد

$$a \equiv b$$

برای نمایش این که b بر روی a نوشته می‌شود، استفاده خواهیم کرد. به طور مشابه تعویض

دو مقدار محاسبه شده a و b به طور نمادی به صورت زیر نوشته می‌شود.

$$a \Leftrightarrow b$$

۴-۲ شمارش اعمال ممیز شناور (flops) و ملاحظات ذخیره سازی برای برخی الگوریتمهای مقدماتی

در جبر خطی عددی با ماتریسهای مثلثی سروکار داریم که می توانند با استفاده از فقط $n(n+1)/2$ مکان به جای n^2 مکان ذخیره شوند. این حقیقت مفید باید به هنگام طراحی یک الگوریتم در جبر خطی عددی در نظر گرفته شود، در نتیجه فضای زیادی موجود می تواند برای مصارف دیگر مورد استفاده قرار گیرد.

ماتریسهای تنک، در حالت کلی، دارای عناصر صفر زیادی هستند. یک طرح مناسب برای ذخیره سازی یک ماتریس تنک، تنها شامل ذخیره سازی عناصر مخالف صفر است. در زیر ما شمارش اعمال ممیز شناور (flops) و طرح ذخیره سازی را برای برخی محاسبات مقدماتی ساده با ماتریسهای متراکم مورد بررسی قرار خواهیم داد.

در تعیین تعداد اعمال ممیز شناور (flops) یک الگوریتم، توجه می کنیم که تعداد ضربها و جمعها در یک الگوریتم ماتریسی تقریباً یکسان هستند. بنابراین فقط شمارش ضربها در یک الگوریتم ماتریسی یک ایده از کل تعداد اعمال ممیز شناور (flops) آن الگوریتم به ما ارائه خواهد داد. (نکات مشابه نیز برای تقسیمها و تفریقها برقرارند). بعلاوه اعمال حاوی عناصر صفر می توانند در شمارش محسوب نشوند.

مثال ۴-۲-۱ محاسبه حاصلضرب داخلی. فرض کنید

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{و} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

دو بردار n بعدی باشند. آنگاه حاصلضرب داخلی $z = x^T y = \sum_{i=1}^n x_i y_i$ می تواند به صورت زیر محاسبه شود (الگوریتم ۳-۱-۲)

$$z = 0$$

For $i = 1, 2, \dots, n$ do

$$z = z + x_i y_i$$

فقط یک عمل ممیز شناور (flop) برای هر i مورد نیاز است. پس در کل n عمل ممیز شناور (flops) برای اجرای الگوریتم مورد نیاز است.

مثال ۲-۲-۴ حاصلضرب خارجی. حاصلضرب خارجی xy^T یک ماتریس Z است که به صورت زیر تعریف می‌شود: مؤلفه (i, j) این ماتریس $x_i y_j$ است. چون n^2 مؤلفه وجود دارد و هر کدام نیاز به یک ضرب دارند، محاسبه حاصلضرب خارجی به n^2 عمل ممیز شناور (flops) و n^2 مکان از حافظه نیاز دارد. ولیکن اغلب به ماتریس حاصل از ضرب خارجی به طور صریح نیاز نداریم.

$$z = xy^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} (y_1 \ y_2 \ \dots \ y_n) = \begin{pmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \dots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \dots & x_n y_n \end{pmatrix}$$

مثال ۳-۲-۴ ضرب ماتریس - بردار. فرض کنید $A = (a_{ij})$ یک ماتریس $n \times n$ و

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

یک بردار n بعدی باشند. آنگاه

$$Ab = \begin{pmatrix} a_{11}b_1 + a_{12}b_2 + \dots + a_{1n}b_n \\ a_{21}b_1 + a_{22}b_2 + \dots + a_{2n}b_n \\ \vdots \\ a_{n1}b_1 + a_{n2}b_2 + \dots + a_{nn}b_n \end{pmatrix}$$

شمارش اعمال ممیز شناور: هر مؤلفه از بردار Ab نیاز به n ضرب و $(n - 1)$ جمع دارد. چون n مؤلفه وجود دارد، محاسبه Ab حدوداً به n^2 عمل ممیز شناور (flops) نیاز دارد.

مثال ۴-۲-۴ ضرب ماتریس - ماتریس با المثلثی. فرض کنید $U = (u_{ij})$ و $V = (v_{ij})$ دو ماتریس بالا مثلثی از مرتبه n باشند. آنگاه الگوریتم زیر حاصلضرب $C = UV$ را محاسبه می‌کند. الگوریتم حاصلضرب UV را بر روی V می‌نویسد.

For $i = 1, 2, \dots, n$ do

For $j = i, i + 1, \dots, n$ do

$$v_{ij} \equiv c_{ij} = \sum_{k=i}^j u_{ik} v_{kj}$$

توجه کنید که در این شبه کد، z حلقه داخلی و i حلقه خارجی را نشان می دهند. برای هر i از ۱ تا n ، z مقادیر i تا n را اختیار می کند.

شمارش اعمال ممیز شناور

۱. محاسبه c_{ij} به $i + 1 - z$ ضرب نیاز دارد.

۲. چون z از i تا n و i از ۱ تا n اجرا می شوند، تعداد کل ضرب عبارت است از:

$$\begin{aligned} \sum_{i=1}^n \sum_{j=i}^n (j-i+1) &= \sum_{i=1}^n (1+2+\dots+(n-i+1)) \\ &= \sum_{i=1}^n \frac{(n-i+1)(n-i+2)}{2} \\ &\approx \frac{n^3}{6} \quad (\text{برای } n \text{ بزرگ}) \end{aligned}$$

(یادآور می شویم که $1+2+3+\dots+r = r(r+1)/2$ و $1^2+2^2+3^2+\dots+r^2 = r(r+1)(2r+1)/6$.)

تعداد اعمال ممیز شناور برای ضرب دو ماتریس مثلثی

برای n بزرگ، ضرب دو ماتریس $n \times n$ حدوداً به $\frac{n^3}{6}$ عمل ممیز شناور (flops) نیاز دارد.

مثال ۴-۲-۵ ضرب ماتریس - ماتریس. فرض کنید A یک ماتریس $m \times n$ و B یک ماتریس $n \times p$ باشند. آنگاه الگوریتم زیر حاصل ضرب $C = AB$ را محاسبه می کند.

For $i = 1, 2, \dots, m$ do

For $j = 1, 2, \dots, p$ do

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

شمارش اعمال ممیز شناور: n ضرب در محاسبه هر c_{ij} وجود دارد. چون z از ۱ تا p و i از ۱ تا m اجرا می شوند، کل تعداد ضربها برابر mnp می باشد. بنابراین برای دو ماتریس A و B از مرتبه n ، تعداد

اعمال ممیز شناور برابر n^3 است.

تعداد اعمال ممیز شناور برای ضرب دو ماتریس

فرض کنید A یک ماتریس $m \times n$ و B یک ماتریس $n \times p$ باشند. آنگاه محاسبه $C = AB$ به mnp عمل ممیز شناور نیاز دارد. به ویژه برای محاسبه دو ماتریس مربعی $n \times n$ ، n^3 عمل ممیز شناور (flops) مورد نیاز می‌باشد.

الگوریتم قبل برای حاصلضرب ماتریس - ماتریس برای دو ماتریس $n \times n$ به طور واضح به n^2 مکان حافظه برای هر ماتریس نیاز خواهد داشت. ولیکن می‌توان به همان صورتی که در زیر نشان داده شده است، الگوریتم را دوباره به نحوی نوشت که یک صرفه‌جویی قابل ملاحظه در حافظه وجود داشته باشد.

الگوریتم زیر بر روی B حاصلضرب AB را می‌نویسد، با این فرض که یک ستون اضافی به B ضمیمه شده باشد (یا می‌توانیم از یک بردار کاری برای نگهداری مقادیر موقتی استفاده کنیم). A و B ماتریسهای $n \times n$ هستند.

مثال ۴-۲-۶ ضرب ماتریس - ماتریس با صرفه‌جویی در حافظه

For $j = 1, 2, \dots, n$ do

$$1. \quad h_i = \sum_{k=1}^n a_{ik} b_{kj} \quad (i = 1, 2, \dots, n)$$

$$2. \quad b_{ij} \equiv h_i \quad (1, 2, \dots, n)$$

(h یک بردار کاری موقت است).

مثال ۴-۲-۷ محاسبه $(I - \frac{\gamma uu^T}{u^T u})A$. در جبر خطی عددی اغلب به محاسبه

$$(I - \frac{\gamma uu^T}{u^T u})A$$

نیاز داریم که در آن I یک ماتریس واحد $m \times m$ ، u یک بردار m بعدی و A یک ماتریس $m \times n$ هستند. ماتریس $I - \gamma uu^T / u^T u$ یک ماتریس هاوس هولدر نامیده می‌شود (فصل ۵ را ملاحظه کنید). به صورت مقدماتی، ما باید ماتریس $(I - \gamma uu^T / u^T u)$ را از بردار u تشکیل دهیم و سپس یا ضرب در

A ، ماتریس حاصلضرب را به طور صریح بسازیم. این کار به $O(m^2n)$ عمل ممیز شناور نیاز دارد. اکنون نشان می دهیم که این ماتریس حاصلضرب را می توان به طور ضمنی با $O(mn)$ عمل ممیز شناور انجام داد. نکته کلیدی این است که ما به تشکیل صریح ماتریس $(I - \gamma uu^T/u^T u)$ نیاز نداریم.

الگوریتم زیر حاصلضرب را حساب می کند. الگوریتم بر روی A حاصلضرب را می نویسد. فرض کنید

$$\beta = \frac{\gamma}{u^T u}$$

آنگاه

$$(I - \frac{\gamma uu^T}{u^T u})A$$

به صورت $A - \beta uu^T A$ در می آید. فرض کنید.

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix}$$

آنگاه عنصر (i, j) ام از $(A - \beta uu^T A)$ برابر $(u_1 a_{1j} + u_2 a_{2j} + \dots + u_m a_{mj})u_i$ خواهد بود. بنابراین، الگوریتم زیر را داریم:

الگوریتم ۴-۲-۱ محاسبه $(I - \frac{\gamma uu^T}{u^T u})A$

Step 1: Compute $\beta = \frac{\gamma}{u^T u}$

Step 2: For $j = 1, 2, \dots, n$ do

$$\alpha = u_1 a_{1j} + u_2 a_{2j} + \dots + u_m a_{mj}$$

$$\alpha \equiv \beta \cdot \alpha$$

For $i = 1, 2, \dots, m$ do

$$a_{ij} \equiv a_{ij} - \alpha u_i$$

تذکر MATCOM: الگوریتم ۴-۲-۱ در برنامه PHOUSHMUL در بسته نرم افزار MATCOM پیاده سازی شده است. بخش ۴-۳-۸ و ضمیمه B را برای فراگیری بیشتر در مورد MATCOM ملاحظه کنید.

شمارش اعمال ممیز شناور

۱. $(m + 1)$ عمل ممیز شناور (flops) برای محاسبه β مورد نیاز است (m عمل ممیز شناور

(flops) برای محاسبه حاصلضرب داخلی و ۱ عمل ممیز شناور (flop) برای تقسیم ۲ بر حاصلضرب داخلی)

۲. n ضرب α وجود دارد، و هزینه محاسبه هر α ، $(m + 1)$ عمل ممیز شناور (flops) است. بنابراین $n(m + 1)$ عمل ممیز شناور (flops) برای محاسبه α ها نیاز داریم.

۳. mn عنصر a_{ij} برای محاسبه وجود دارد. هزینه محاسبه هر a_{ij} به محض آن که α ها محاسبه شوند، یک عمل ممیز شناور (flop) است.

کل تعداد اعمال ممیز شناور برابر $n + 2mn + (m + 1)$ خواهد بود.

اکنون این نتیجه مهم را (که به طور مکرر در این کتاب استفاده می شود) در زیر خلاصه می کنیم.

$$(I - \frac{2uu^T}{u^T u})A$$

تعداد اعمال ممیز شناور برای حاصلضرب A یک ماتریس $m \times n$ ($m \geq n$) و u یک بردار m بعدی باشند. آنگاه حاصلضرب قبل می تواند با فقط $n + 2mn + (m + 1)$ عمل ممیز شناور (flops) محاسبه شود. بویژه اگر $m = n$ باشد، آنگاه این محاسبه تقریباً به $2n^2$ عمل ممیز شناور (flops) نیاز دارد، در صورتی که اگر حاصلضرب به صورت معمول انجام شود، به n^3 عمل ممیز شناور (flops) نیاز خواهد داشت.

یک مثال عددی: فرض کنید

$$u = (1, 1, 1)^T, \quad A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 0 & 0 \end{bmatrix}$$

آنگاه

$$\beta = \frac{2}{3}$$

$j = 1$: محاسبه ستون اول حاصلضرب

$$\alpha = u_1 a_{11} + u_2 a_{21} + u_3 a_{31} = 1 + 2 = 3$$

$$\alpha \equiv \beta \alpha = \frac{2}{3} \times 3 = 2$$

$$a_{11} \equiv a_{11} - \alpha u_1 = 1 - 2 = -1$$

$$a_{21} \equiv a_{21} - \alpha u_2 = 2 - 2 \times 1 = 0$$

$$a_{31} \equiv a_{31} - \alpha u_3 = 0 - 2 \times 1 = -2$$

$z = 2$: محاسبه ستون دوم حاصلضرب

$$\alpha = u_1 a_{12} + u_2 a_{22} + u_3 a_{32} = 1 + 1 = 2$$

$$\alpha \equiv \beta \alpha = 2 \times \frac{2}{3} = \frac{4}{3}$$

$$a_{12} \equiv a_{12} - \alpha u_1 = 1 - \frac{4}{3} \times 1 = -\frac{1}{3}$$

$$a_{22} \equiv a_{22} - \alpha u_2 = 1 - \frac{4}{3} \times 1 = -\frac{1}{3}$$

$$a_{32} \equiv a_{32} - \alpha u_3 = 0 - \frac{4}{3} = -\frac{4}{3}$$

بنابراین

$$A \equiv \left(I - \frac{2uu^T}{u^T u} \right) A$$

$$= \begin{bmatrix} -1 & -\frac{1}{3} \\ 0 & -\frac{1}{3} \\ -2 & -\frac{4}{3} \end{bmatrix}$$

مثال ۴-۲-۸ شمارش تعداد اعمال ممیز شناور الگوریتم ۳-۱-۳ (فرایند جایگزینی پسر). از شبه کدهای این الگوریتم ملاحظه می‌کنیم که محاسبه γ_n نیاز به یک عمل ممیز شناور (flop)، محاسبه γ_{n-1} نیاز به ۲ عمل ممیز شناور (flops)، و غیره دارند. بنابراین برای محاسبه γ_1 تا γ_n ما به $(1 + 2 + 3 + \dots + n) = n(n+1)/2 \approx n^2/2$ عمل ممیز شناور (flops) نیاز داریم.

تعداد اعمال ممیز شناور برای فرایند جایگزینی پسر

برای حل یک دستگاه بالا مثلثی با استفاده از جایگزینی پسر تقریباً $n^2/2$ عمل ممیز شناور (flops) مورد نیاز است

مثال ۴-۲-۹ شمارش اعمال ممیز شناور و ملاحظات ذخیره‌سازی برای الگوریتم ۳-۱-۴ (معکوس یک ماتریس بالا مثلثی). اجازه دهید یک بار دیگر الگوریتم را در اینجا بیان کنیم:

For $k = n, n - 1, \dots, 2, 1$ do

$$s = \frac{1}{t_{kk}}$$

For $i = k - 1, k - 2, \dots, 1$

$$s_{ik} = -\frac{1}{t_{ik}} \left[\sum_{j=i+1}^k t_{ij} s_{jk} \right]$$

شمارش اعمال ممیز شناور: تعداد اعمال ممیز شناور تقریباً به صورت زیر است

$$k = 1 \quad 1 \text{ عمل ممیز شناور (flop)}$$

$$k = 2 \quad 3 \text{ عمل ممیز شناور (flops)}$$

$$k = 3 \quad 6 \text{ عمل ممیز شناور (flops)}$$

$$k = n \quad \frac{n(n-1)}{2} \text{ عمل ممیز شناور (flops)}$$

کل اعمال ممیز شناور (flops):

$$1 + 3 + 6 + \dots + \frac{n(n+1)}{2} = \sum_{r=1}^n \frac{r(r+1)}{2} = \sum_{r=1}^n \frac{r^2}{2} + \sum_{r=1}^n \frac{r}{2} \approx \frac{n^3}{6}$$

تعداد اعمال ممیز شناور (flops) برای محاسبه معکوس یک ماتریس بالا مثلثی

تعداد اعمال ممیز شناور (flops) مورد نیاز برای محاسبه معکوس یک ماتریس بالا مثلثی در

حدود $\frac{n^3}{6}$ می باشد.

ملاحظات ذخیره سازی. چون معکوس یک ماتریس بالا مثلثی یک ماتریس بالا مثلثی است و از الگوریتم واضح است که می توانیم s_{ik} را بر روی t_{ik} بنویسیم، الگوریتم می تواند به نحوی دوباره نوشته شود که معکوس k بر روی T نوشته شود. بنابراین می توانیم الگوریتم را به صورت زیر دوباره بنویسیم.

الگوریتم ۴-۲-۲ محاسبه معکوس یک ماتریس بالا مثلثی با ذخیره سازی اقتصادی

For $k = n, n - 1, \dots, 2, 1$ do

$$t_{kk} \equiv s_{kk} = \frac{1}{t_{kk}}$$

For $i = k - 1, k - 2, \dots, 1$ do

$$t_{ik} \equiv s_{ik} = -\frac{1}{t_{ik}} \left[\sum_{j=i+1}^k t_{ij} s_{jk} \right]$$

تذکر MATCOM : الگوریتم ۴-۲-۲ در برنامه INVUPTR در بسته نرم افزاری MATCOM پیاده سازی شده است.

مثال ۴-۲-۱۰ شمارش اعمال ممیز شناور و ملاحظات ذخیره سازی برای الگوریتم ۳-۱-۵ (روش حذفی گاوس). در فصل ۵ نشان داده شده است که الگوریتم حذفی گاوس حدوداً به $n^3/3$ عمل ممیز شناور (flops) نیاز دارد. الگوریتم می تواند بر روی A ، ماتریس بالا مثلثی $A^{(n-1)}$ را بنویسد؛ در حقیقت، الگوریتم می تواند بر روی A هر ماتریس $A^{(k)}$ را بنویسد. ضربگرهای m_{ik} می توانند در نیمه پایینی A به محض آن که محاسبه شوند نوشته شوند. هر $b^{(k)}$ می تواند بر روی b نوشته شود.

الگوریتم ۴-۲-۳ الگوریتم حذفی گاوس با ذخیره سازی اقتصادی

For $k = 1, 2, \dots, n - 1$ do

For $i = k + 1, \dots, n$ do

$$a_{ik} \equiv m_{ik} = -\frac{a_{ik}}{a_{kk}}$$

$$b_i \equiv b_i + m_{ik} b_k$$

For $j = k + 1, \dots, n$ do

$$a_{ij} \equiv a_{ij} + m_{ik} a_{kj}$$

تذکر MATCOM : الگوریتم ۴-۲-۳ در برنامه GAUSS در بسته نرم افزاری MATCOM پیاده سازی شده است.

۴-۳ چند نرم افزار ریاضی با کیفیت بالا و موجود برای مسائل جبر خطی

چندین بسته نرم افزاری ریاضی با کیفیت بالا برای انواع مختلف محاسبات ماتریسی وجود دارند. این نرم افزارها عبارتند از LINPACK، EISPACK، MATLAB، NETLIB، IMSL، NAG، و

تازه‌ترین نرم‌افزارهای منتشر شده عبارتند از LAPACK و TEMPLATES.

LINPACK ۱-۳-۴

برطبق راهنمای کاربر آن « LINPACK یک گردایه از زیر روالهای فرتن است که دستگاههای مختلف معادلات جبر خطی همزمان را تجزیه و تحلیل می‌کنند. زیر روالها طوری طراحی شده‌اند که کاملاً مستقل از ماشین بوده، تماماً قابل حمل، و با کارایی نزدیک به بهینه در بیشتر محیطهای عملی قابل اجرا باشند».

اگرچه هدف در اصل حل دستگاههای خطی بوده است، ولیکن بسته نرم‌افزاری حاوی زیر روالهایی برای تجزیه مقدار تکین (SVD) و مسائل وابسته به دستگاههای خطی، نظیر محاسبه معکوس، دترمینان، و مسأله کمترین توانهای دوم نیز می‌باشد. بیشتر زیر روالها برای ماتریسهای مربعی هستند، اما برخی از آنها در مورد ماتریسهای ضرایب مستطیلی وابسته به مسائل فرامعین یا فرومعین نیز عمل می‌کنند. زیر روالها برای مسائل کوچک و متراکم از مرتبه کمتر از چند صد و ماتریسهای نواری از مرتبه کمتر از چند هزار ساخته شده‌اند. برای روشهای تکراری زیر روالی وجود ندارد.

EISPACK ۲-۳-۴

EISPACK یک بسته نرم‌افزاری دستگاه ویژه است. بسته نرم‌افزاری در اصل برای محاسبه مقادیر ویژه و بردارهای ویژه یک ماتریس طراحی شده است، ولیکن حاوی زیر روالهایی برای مسأله مقدار ویژه تعمیم یافته به شکل $Ax = \lambda Bx$ و تجزیه مقدار تکین نیز می‌باشد.

مقادیر ویژه یک ماتریس دلخواه A به صورت دنباله‌ای در چندین فاز محاسبه می‌شوند، ابتدا ماتریس A متعادل می‌گردد. اگر ماتریس نامتقارن باشد، ماتریس تعادل یافته سپس به وسیله تشابه به یک ماتریس بالا هسنبرگی کاهش می‌یابد (اگر متقارن باشد به سه قطری متقارن کاهش می‌یابد). بالاخره مقادیر ویژه ماتریس بالا هسنبرگی و یا سه قطری متقارن تبدیل یافته با استفاده از تکرارهای QR ضمنی یا روش دنباله استورم محاسبه می‌شود. زیر روالهایی از EISPACK وجود دارند که همه این کارها را انجام می‌دهند. روش تکرار ضمنی QR و روش دنباله استورم در فصل ۸ توصیف می‌شوند.

LAPACK ۳-۳-۴

بنا بلوکهای جبر خطی عددی دارای سه سطح از BLAS^(۱) (زیر روالهای مقدماتی جبر خطی)

هستند:

- BLAS سطح ۱ برای اعمال بردار - بردار هستند. یک BLAS سطح ۱ نوعی به شکل $y \equiv \alpha x + y$ است که در آن x و y بردار هستند و α یک اسکالر است.
- BLAS سطح ۲ برای اعمال ماتریس - بردار هستند. یک BLAS سطح ۲ نوعی به شکل $y \equiv \alpha Ax + \beta y$ است.
- BLAS سطح ۳ برای اعمال ماتریس - ماتریس هستند. یک BLAS سطح ۳ نوعی به شکل $C \equiv \alpha AB + \beta C$ است.

BLAS سطح ۱ در LINPACK استفاده می شود. متأسفانه، الگوریتمهای مرکب از BLAS سطح ۱ برای رسیدن به کارایی بالا بر روی بیشتر ابرکامپیوترهای امروزی نظیر Cray Y-MP یا C-90 مناسب نیستند.

اگرچه BLAS سطح ۲ می تواند سرعت خوبی (گاهی اوقات تقریباً حداکثر) بر روی بیشتر کامپیوترهای برداری نظیر CRAY X-MP یا CRAY Y-MP ارائه دهند، اما آنها برای کارایی بالا بر روی ابرکامپیوترهای مدرن دیگر (مثلاً CRAY 2) مناسب نیستند.

BLAS سطح ۳ برای اغلب ابرکامپیوترهای امروزی ایده آل هستند. آنها می توانند $O(n^3)$ عمل ممیز شناور را بر روی $O(n^2)$ داده انجام دهند.

بنابراین در طی چندین سال گذشته، تلاش فراوان توسط دانشمندان جبرخطی عددی انجام شده است تا الگوریتمهای مبتنی بر اعمال BLAS-1 مرسوم به الگوریتمهای غنی و مبتنی بر اعمال BLAS-2 و BLAS-3 تبدیل شوند. به عنوان یک نتیجه، اکنون الگوریتمهایی از این نوع برای بیشتر محاسبات ماتریسی وجود دارند که الگوریتمهای بلوک شده نامیده می شوند. این الگوریتمها در یک بسته نرم افزاری به نام LAPACK پیاده سازی شده اند.

بر طبق راهنمای کاربر «LAPACK یک کتابخانه قابل حمل از زیر برنامه های فرترن ۷۷ برای حل متداول ترین مسائل جبر خطی عددی: دستگاههای معادلات خطی، مسائل کمترین توانهای دوم خطی، مسائل مقدار ویژه، و مسأله مقدار تکین است». این نرم افزار طوری طراحی شده است که بر روی دامنه وسیعی از کامپیوترهای مدرن با اجرای بالا کارا باشد.

LAPACK برای جایگزین شدن به جای LINPACK و EISPACK طراحی می شود و اصولاً این کار با دوباره تغییر دادن ساختار نرم افزار برای رسیدن به کارایی خیلی بیشتر بر روی پردازنده های برداری، ایستگاههای کاری ابر - اسکالر با اجرای بالا، و چند پردازنده های با حافظه تقسیم شده، انجام می شود.

LAPACK همچنین تابعهای دیگری اضافه می‌کند، از برخی الگوریتمهای جدید یا بهبود یافته استفاده می‌نماید و دو مجموعه الگوریتم را در یک بسته نرم‌افزاری کامل می‌کند. راهنمای کاربر LAPACK یک مقدمه غیر رسمی برای طراحی الگوریتم و نرم‌افزار ارائه می‌دهد، محتویات نرم‌افزار را خلاصه می‌کند، استفاده درست از نرم‌افزار و مستندات آن را توصیف می‌کند، و شامل مشخصات کاملی برای فراخوانی زیر روالها است.

۴-۳-۴ NETLIB

NETLIB مخفف کتابخانه شبکه‌ای است. زیر برنامه‌های LINPACK, EISPACK و LAPACK به صورت الکترونیکی از این کتابخانه همراه با انواع دیگری از نرم‌افزارها برای محاسبات ماتریسی در دسترس هستند. یک استفاده کننده می‌تواند نرم‌افزار را از این بسته‌های نرم‌افزاری با فرستادن پیام الکترونیکی به آدرس زیر به دست آورد.

`netlib@netlib.ornl.gov`

همچنین فایلها می‌توانند با دستور ناشناس ftp به یک فهرست راهنمای محلی انتقال یابند

`ftp netlib.att.com`

`ftp www.netlib.org`

برای پی بردن به چگونگی استفاده از netlib، دستور زیر را در سرآمد موضوعی^(۱) یا کالبد پیام الکترونیکی بگنجانید.

`send index`

اطلاعات در مورد زیر برنامه‌های موجود در یک نرم‌افزار مفروض می‌تواند با فرستادن پیام الکترونیکی زیر به دست آید:

`send index from {library}`

بنابراین برای به دست آوردن توصیفی از زیر برنامه‌ها برای LAPACK، پیام زیر را بفرستید:

`send index from LAPACK`

برای به دست آوردن یک قطعه یا قطعه‌های نرم‌افزاری از یک بسته نرم‌افزاری، پیام زیر را بفرستید:

`send {routines} from {library}`

بنابراین برای به دست آوردن یک زیر برنامه از LINPACK که SGECO نامیده می شود، پیام زیر را بفرستید:

send SGECO from LINPACK

(این پیام زیر روال SGECO و سایر زیر روالهایی را که SGECO نامیده می شود، برای شما خواهد فرستاد). Xnetlib، که از یک پنجره واسط X برای باز کردن مستقیم رایانه از راه دور، استفاده می کند، همچنین موجود است (و به محض نصب راحت تر می باشد).

اطلاعات بیشتر در مورد netlib می تواند توسط FTP ناشناس به هر یک از محل های زیر به دست آید:

netlib2.cs.utk.edu
netlib.att.com

NAG ۵-۳-۴

NAG مخفف *Numerical Algorithm Group* (گروه الگوریتم عددی) است. این گروه یک کتابخانه نرم افزاری بزرگ (که NAG نیز نامیده می شود) شامل زیر روالهایی برای اغلب مسائل محاسباتی شامل جبر خطی عددی، مسائل معادلات دیفرانسیل عددی (هم معمولی و هم جزئی)، مسائل بهینه سازی، مسائل معادلات انتگرالی، مسائل آماری و غیره تدوین نموده است.

IMSL ۶-۳-۴

IMSL مخفف *International Mathematical and Statistical Libraries* (کتابخانه های ریاضی و آمار بین المللی) است. همان گونه که عنوان آن بیان می کند، این کتابخانه شامل زیر روالهایی برای تقریباً همه محاسبات ریاضی و آماری است.

MATLAB ۷-۳-۴

نام MATLAB نشانه *MATrix LABORatory* (لابراتوار ماتریسی) است. این یک دستگاه محاسبه ای محاوره ای است که برای محاسبات ساده مسائل گوناگون علمی و مهندسی مبتنی بر ماتریس طراحی شده است. MATLAB دسترسی ساده به نرم افزار ماتریسی تدوین یافته توسط پروژه های نرم افزاری LINPACK و EISPACK را فراهم می کند.

MATLAB می تواند برای حل یک دستگاه خطی و مسائل مربوط به آن (نظیر معکوس کردن یک

ماتریس یا محاسبه رتبه و دترمینان یک ماتریس)، برای محاسبه مقادیر ویژه، بردارهای ویژه یک ماتریس، برای پیدا کردن تجزیه مقدار تکین یک ماتریس، برای محاسبه صفرهای یک چندجمله‌ای، برای محاسبه مقادیر ویژه و بردارهای ویژه تعمیم یافته، و غیره استفاده شود. MATLAB یک نرم‌افزار بی‌نهایت مفید و ارزشمند برای آموختن الگوریتمها برای مسائل کوچک و برای استفاده در کلاسهای درسی است. این نرم‌افزار یک وسیله ضروری برای تدریس جبر خطی عددی و کاربردی در کلاسهای درسی شده است. یک خصوصیت برجسته از MATLAB قابلیت‌های گرافیکی آن است (برای اطلاعات بیشتر در مورد MATLAB ضمیمه A را ملاحظه کنید).

یک ویرایش دانشجویی از MATLAB وجود دارد که توسط ناشر Prentice Hall در سال ۱۹۹۲ منتشر شده است. این ویرایش برای استفاده سهل و ساده در کلاس طراحی شده است.

۸-۳-۴ کدهای MATLAB و MATCOM

کدهای MATLAB برای الگوریتمهای انتخاب شده و توصیف شده در این کتاب در ضمیمه A برای دانشجویان مبتدی فراهم شده‌اند.

بعلاوه، یک بسته ابزار MATLAB محاوره‌ای که MATCOM نامیده می‌شود، و همه الگوریتمهای اصلی (که باید در دوره اول تدریس شوند) در آن پیاده‌سازی شده‌اند، می‌تواند توسط خواننده این کتاب به دست آید. بنابراین دانشجویان می‌توانند الگوریتمهای مختلف را برای یک مسأله خاص از لحاظ کارایی عددی، پایداری، دقت، و غیره مقایسه کنند (نرم‌افزار MATCOM از MathWorks در دسترس است. برگه اطلاعات را در انتهای کتاب ملاحظه کنید. گونه‌های Macintosh و PC موجود هستند).

۹-۳-۴ کتابخانه ACM

کتابخانه فراهم‌شده توسط انجمن محاسبات ماشینی (*Association for Computing Machinery* (ACM)) شامل زیر روالهایی برای اعمال ماتریس - بردار مقدماتی، دستگاههای خطی و مسائل وابسته، دستگاههای غیرخطی، صفرهای چندجمله‌ایها، و غیره است. مجله TOMS^(۱) (مجله انجمن محاسبات ماشینی (ACM) که در مورد نرم‌افزارهای ریاضی توضیح می‌دهد) این الگوریتمها را چاپ می‌کند.

۴-۳-۱۰ ITPACK (بسته نرم افزاری تکراری)

بسته نرم افزاری ITPACK برای حل دستگاه معادلات خطی $Ax = b$ توسط روشهای تکراری مختلف، که در آن A یک ماتریس بزرگ است، طراحی شده است. (این نرم افزار در مکان اینترنتی math.utexas.edu موجود است و توسط دستور ftp ناشناس قابل دسترسی است).

۴-۳-۱۱ بسته نرم افزاری TAEMPLATES (برای حل دستگاههای خطی: بنابلوکها برای روشهای تکراری)

TEMPLATES شامل تعداد زیادی برنامه است که به صورت بنا بلوکها برای تقریباً همه روشهای تکراری فن آوری روز برای حل دستگاههای خطی تنک و بزرگ به کار می روند. یک template (اسوه) توصیفی از یک الگوریتم کلی به جای object code (برنامه مقصود) یا source code (برنامه مبدأ) است که به طور متداول تر در یک کتابخانه نرم افزاری مرسوم پیدا می شوند.

این نرم افزار از netlib موجود است. یک تقاضا توسط پیام الکترونیکی به صورت زیر بفرستید:

`mail netlib@anl.gov`

یک یا چند تقاضا (یکی در هر خط) ممکن است بر روی خط پیام یا در بدنه به صورت زیر بفرستید:

`send index from linalg.`

`send sftemplates.shar from linalg.`

اولین تقاضا منجر به یک پیام الکترونیکی می شود که حاوی فهرست راهنما از کتابخانه `linalg` همراه با یک توصیف خلاصه از محتویات آن است. تقاضای دوم منجر به یک پیام الکترونیکی می شود که از یک فایل مشترک حاوی زیر روالهای فرتن با دقت ساده و یک فایل README تشکیل می شود.

۴-۳-۱۲ SPARSKIT : یک بسته ابزار پایه ای برای محاسبات ماتریسی تنک

SPARSKIT توسط یوسف سعد (Youcef Saad) از دانشگاه مینسوتا تدوین شده است. گونه دوم SPARSKIT اکنون موجود است و در یک فضای ftp ناشناس در ftp.cs.umn.edu

(128.101.230.9)، قرار گرفته است. برای دسترسی لازم است شما به صورت زیر عمل نمایید

• با login name ناشناس، و نام خودتان به عنوان کلمه عبور به آدرس ftp.cs.umn.edu

[128.101.230.9] ، FTP کنید.

• سپس با `cd` به `/pub/sparse` بروید. یک فایل `SPARSKIT2.tar.Z` و یک فهرست راهنما که

SPARSKIT2 نامیده می‌شود پیدا خواهید کرد. می‌توانید فایل فشرده SPARSKIT2.tar.Z را دریافت کنید (هنگامی که FTP می‌کنید از 'binary' قبل از 'get' استفاده کنید) یا می‌توانید فایل‌های خاصی را یا همه فهرست راهنما SPARSKIT2 را دریافت کنید.

- اگر فایل فشرده SPARSKIT2.tar.Z را دریافت می‌کنید شما لازم است دستور زیر را تایپ کنید تا زیر فهرست راهنمای SPARSKIT2 را در فهرست راهنمای خود ایجاد کنید.

`uncompress -c SPARSKIT2.tar.z | tar xvf -`

مرور و خلاصه

منظور از این فصل معرفی الگوریتم‌های مؤثر عددی و نرم‌افزارهای ریاضی با کیفیت بالا و وابسته بوده است.

۱. نوشتن کدها برای یک الگوریتم می‌تواند یک کار پیش پافتاده باشد. ولیکن همه نرم‌افزارهای نوشته شده نرم‌افزارهای با کیفیت بالا نیستند. نرم‌افزار ریاضی با کیفیت بالا نرم‌افزاری تعریف می‌کنیم که (۱) توانا و قابل انعطاف، (۲) ساده برای خواندن و تغییر، (۳) قابل حمل، (۴) قوی، و از همه مهم‌تر (۵) مبتنی بر یک الگوریتم مؤثر عددی باشد

۲. همانند نرم‌افزارها، ممکن است الگوریتم‌های زیادی برای یک مسأله وجود داشته باشد. ولیکن همه الگوریتم‌ها از نظر عددی مؤثر نمی‌باشند. ما یک الگوریتم مؤثر عددی را الگوریتمی تعریف می‌کنیم که از لحاظ کاری (۱) کلی، (۲) قابل اطمینان، (۳) پایدار، و (۴) نسبت به زمان و حافظه از کارایی برخوردار باشد.

۳. کارایی یک الگوریتم ماتریسی برحسب زمان کامپیوتری مصرف شده توسط الگوریتم اندازه‌گیری می‌شود. یک اندازه نظری، تعداد اعمال ممیز شناور (flops) لازم برای اجرای الگوریتم است. به طور تقریبی یک ضرب (یا یک تقسیم) همراه با یک جمع (یا یک تفریق) یک عمل ممیز شناور (flop) تعریف می‌شود.

یک الگوریتم ماتریسی حاوی ماتریسهای از مرتبه n که بیش از $O(n^3)$ عمل ممیز شناور (flops) نیاز نداشته باشد یک الگوریتم کارا تعریف می‌شود؛ پایداری یک الگوریتم در فصل ۳ تعریف شد. یک الگوریتم می‌تواند کارا باشد بدون آن که پایدار باشد. بنابراین یک الگوریتم کارا لازم نیست که یک الگوریتم مؤثر عددی باشد. الگوریتم‌هایی وجود دارند که سریع هستند ولیکن پایدار نمی‌باشند

۴. چندین مثال (بخش ۴-۲) ارائه شدند تا نشان دهند که چگونه برخی اعمال ماتریسی مقدماتی می توانند دوباره تنظیم شوند تا به طور نسبتاً ساده هم ذخیره سازی و هم زمان را بدون پیاده سازی طبیعی کارا سازند. مثال ۴-۲-۷ مهم ترین آنها در این زمینه است. در آن جا نشان دادیم که چگونه محاسبه حاصلضرب یک ماتریس $n \times n$ مانند A با ماتریس $H = (I - \gamma uu^T/uu^T)$ ، معروف به ماتریس هاوس هولدر، تنظیم می شود، به طوری که حاصلضرب می تواند با $O(n^2)$ عمل ممیز شناور (flops) به جای $O(n^3)$ عمل ممیز شناور (flops) انجام شود که لازم خواهد بود اگر محاسبه به طور ساده بدون در نظر گرفتن ساختار H انجام شود. این محاسبه برای بسیاری از محاسبات ماتریسی دیگر توصیف شده در این کتاب یک الگو تشکیل می دهد.

۵. خلاصه ای از نرم افزارهای ریاضی با کیفیت بالا نظیر LAPACK، EISPACK، LINPACK، MATLAB، IMSL، NAG، TEMPLATES، و SPARSKIT در بخش آخر فراهم شد.

تمرینها

۱. یک الگوریتم برای محاسبه حاصلضرب $C = AB$ در هر یک از حالت های زیر تدوین نمایید. الگوریتم باید مزایای ساختارهای خاص ماتریسها را در هر حالت در نظر بگیرد. تعداد اعمال ممیز شناور و حافظه لازم را در هر حالت مشخص کنید.

(الف) A و B هر دو ماتریسهای پایین مثلثی هستند.

(ب) A یک ماتریس دلخواه و B یک ماتریس پایین مثلثی است.

(ج) A و B هر دو سه قطری هستند.

(د) A دلخواه و B بالا هسنبرگی است.

(ه) A بالا هسنبرگی و B سه قطری است.

(و) A بالا هسنبرگی و B بالا مثلثی است.

۲. گفته می شود که یک ماتریس مربعی $A = (a_{ij})$ یک ماتریس نواری با طول نوار $1 + 2k$ است اگر برای $|i - j| > k$ داشته باشیم $a_{ij} = 0$. یک الگوریتم برای محاسبه حاصلضرب $C = AB$ تدوین نمایید که در آن A دلخواه و B یک ماتریس نواری با طول نوار ۳ باشد، مزایای ساختار ماتریس B را در نظر بگیرید. تعداد اعمال ممیز شناور را مشخص کنید.

۳. با استفاده از ایده های الگوریتم ۴-۲-۱ یک الگوریتم برای محاسبه حاصلضرب $A(I + xy^T)$ تدوین نمایید، که در آن A یک ماتریس $n \times n$ و x و y بردارهای n بعدی هستند. الگوریتم

شما باید تقریباً به $2n^2$ عمل ممیز شناور (flops) نیاز داشته باشید.

۴. الگوریتم تمرین ۱ را برای حالت‌های خاص هنگامی که ماتریس $I + xy^T$ به شکلهای زیر باشد، دوباره بنویسید

(الف) یک ماتریس مقدماتی: $I + m_k e_k^T$ ، $m_k = (0, 0, \dots, 0, m_{k+1, k}, \dots, m_{n, k})^T$ ، e_k^T ، k امین سطر I می‌باشد.

(ب) یک ماتریس هاوس هولدر: $I - \frac{yuu^T}{u^T u}$ ، که در آن u یک بردار n بعدی می‌باشد.

۵. فرض کنید A و B دو ماتریس متقارن و هم مرتبه باشند. یک الگوریتم برای محاسبه $C = A + B$ ، با در نظر گرفتن مزایای متقارن بودن هر دو ماتریس، تدوین نمایید. الگوریتم شما باید بر روی B ماتریس C را بنویسد. تعداد اعمال ممیز شناور را مشخص نمایید.

۶. فرض کنید $A = (a_{ij})$ یک ماتریس پایین هسنبرگی کاهش‌ناپذیر از مرتبه n باشد. سپس، با مفروض بودن اولین سطر r_1 ، می‌توان نشان داد ((Datta and Datta, 1976)) که سطرهای متوالی r_2 تا r_n از A^k (k یک عدد صحیح مثبت کوچک‌تر یا مساوی n است) می‌توانند به‌طور بازگشتی به صورت زیر محاسبه شوند.

$$r_{i+1} = \frac{\left[r_i B_i - \sum_{j=1}^{i-1} a_{ij} r_j \right]}{a_{i, i+1}}, \quad i = 1, 2, \dots, n-1,$$

که در آن $B_i = A - a_{ii}I$. یک الگوریتم برای محاسبه A^k تدوین نمایید. تعداد اعمال ممیز شناور را برای الگوریتم مشخص نمایید.

۷. فرض کنید a_r و b_r ، به ترتیب، ستون r ام و سطر r ام ماتریسهای A و B را نمایش دهند. آنگاه یک الگوریتم برای محاسبه حاصلضرب AB با استفاده از فرمول زیر تدوین نمایید.

$$AB = \sum_{i=1}^n a_i b_i$$

تعداد اعمال ممیز شناور و حافظه لازم را برای الگوریتم مشخص نمایید

۸. ماتریس 12×12 زیر را در نظر بگیرید.

$$A = \begin{bmatrix} 12 & 11 & 10 & \dots & 3 & 2 & 1 \\ 11 & 11 & 10 & \dots & 3 & 2 & 1 \\ 0 & 10 & 10 & \dots & 3 & 2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \dots & 2 & 2 & 1 \\ \vdots & \vdots & \vdots & \dots & 0 & 1 & 1 \end{bmatrix}$$

مقادیر ویژه این ماتریس را پیدا کنید. از MATLAB استفاده کنید.

اکنون در عنصر (۱ و ۱۲) به اندازه 10^{-9} اختلال ایجاد کنید و مقادیر ویژه این ماتریس اختلال یافته را محاسبه کنید. در مورد وضعیت مقادیر ویژه چه نتیجه‌ای می‌گیرید؟

برنامه‌های MATLAB و MATCOM و مسائل

تذکر: شما به برنامه phousmul از MATCOM نیاز خواهید داشت.

۱. با استفاده از تابع 'rand' از MATLAB یک ماتریس تصادفی 5×5 بسازید و سپس خروجیهای زیر را چاپ کنید.

$$A(2,:), A(:, 1), A(:, 5)$$

$$A(1, 1:2:5), A([1, 5]), A(4:-1:1, 5:-1:1)$$

۲. با استفاده از تابع 'for' یک برنامه MATLAB برای پیدا کردن حاصلضرب داخلی و حاصلضرب خارجی دو بردار n بعدی u و v بنویسید.

$$[s] = \text{inpro}(u, v)$$

$$[A] = \text{outpro}(u, v)$$

برنامه خود را با ایجاد دو بردار مختلف تصادفی u و v با استفاده از $\text{rand}(4, 1)$ بیازمایید.

۳. یاد بگیرید که چگونه از دستورهای MATLAB زیر برای ایجاد ماتریسهای خاص استفاده کنید

compan	ماتریس همراه
diag	ماتریس قطری
ones	ماتریس ثابت
zeros	ماتریس صفر
rand	ماتریس تصادفی
tril	قسمت پایین مثلثی
hankel	ماتریس هنکل
toeplitz	ماتریس توپلیتز
hilb	ماتریس هیلبرت
triu	ماتریس بالا مثلثی
vander	ماتریس واندرموند

۴. برنامه‌های MATLAB را برای ایجاد ماتریسهای معروف زیر بنویسید.

(الف) $[A] = \text{wilk}(n)$ برای ایجاد ماتریس دو قطری ویلکینسون (a_{ij}) از مرتبه n :

$$a_{ii} = n - i + 1, \quad i = 1, 2, \dots, n$$

$$a_{i-1,i} = n, \quad i = 2, 3, \dots, n$$

$$a_{ij} = 0, \quad \text{در غیر این صورت}$$

(ب) $[A] = \text{pie}(n, \alpha)$ برای ایجاد ماتریس پای (Pie) (a_{ij}) از مرتبه n :

$$a_{ii} = \alpha \quad \text{یک پارامتر نزدیک ۱ یا } n - 1 \text{ است}$$

$$a_{ij} = 1 \quad \text{به ازای } i \neq j$$

۵. با استفاده از دستورهای 'help' برای 'flops'، 'clock'، 'etim'، و غیره، یاد بگیرید که چگونه تعداد اعمال ممیز شناور و زمان را برای هر الگوریتم اندازه بگیرید.

۶. با استفاده از تابعی MATLAB 'for'، 'size'، و 'zero' یک برنامه MATLAB برای پیدا کردن حاصلضرب دو ماتریس بالا مثلثی A و B و به ترتیب از مرتبه $n \times m$ و $n \times p$ بنویسید. برنامه‌های خود را با استفاده از ماتریسهای زیر بیازمایید.

$$A = \text{triu}(\text{rand}(4, 3))$$

$$B = \text{triu}(\text{rand}(3, 3))$$

۷. برنامه $\text{phousmul}(A, u)$ از MATLAB و از MACOM را با ایجاد یک ماتریس تصادفی A از مرتبه 3×6 و یک بردار تصادفی با ۶ عنصر اجرا کنید. تعداد اعمال متمیز شناور و زمان سپری شده را چاپ کنید.

۸. برنامه phousmul از MATLAB را به $\text{housxy}(A, x, y)$ برای محاسبه حاصلضرب $(I + xy^T)A$ تغییر دهید. برنامه خودتان را با ایجاد یک ماتریس تصادفی 15×15 مانند A و بردارهای x و y با ابعاد مناسب بیازماید. حاصلضرب و تعداد اعمال متمیز شناور و زمان سپری شده را چاپ کنید.

پیشنهادها برای مطالعه بیشتر

یک توصیف واضح و مناسب از این که چگونه می توان بسته های نرم افزاری ریاضی با کیفیت بالا را تدوین کرد در کتاب زیر ارائه شده است.

John Rice, *Matrix Computations and Mathematical Software*, New York: McGraw-Hill, 1981.

این کتاب حاوی فصلی (فصل ۱۲) در مورد پروژه های نرم افزاری است. دانشجویان علاقه مند به مطالعه های تفصیلی نرم افزارهای گوناگون آنها را جالب می یابند. یک مقاله عالی در مرجع Demmel (1984) در مورد قابل اطمینان بودن نرم افزارهای عددی نوشته شده است که واقعاً خواندنی است. کتاب جدید دیگر در این زمینه کتاب زیر است.

T. Coleman and C. Van Loan, *Handbook of Matrix Computations* Philadelphia: SIAM, 1988.

کتابهای نوشته شده توسط رایس و کلمن و ون لوتن^(۱) برای خوانندگان علاقه مند به تدوین نرم افزار برای مسائل ماتریسی ضروری هستند.

کتابهای Forsythe, Malcolm, and Moler (1977) و Hager (1988) حاوی زیرروالهای مفیدی برای محاسبات ماتریسی هستند.

برای مطالعه بیشتر کتابهای زیر را ملاحظه کنید.

Johnston, R, L. *Numerical Methods: A Software Approach*. Toronto: John Wiley, 1982.

Kahaner, D., C. B. Moler, and S. Nash. *Numerical Methods and Software*. Englewood Cliffs, N.J.: Printice-Hall, 1988.

کتاب (1973) Stewart یک منبع ارزشمند برای فراگیری این است که چگونه الگوریتمها را می توان برای اعمال ماتریسی مقدماتی به زمان - کارا و حافظه - اقتصادی تنظیم و تدوین نمود. هر بسته نرم افزاری دارای راهنمای کاربر منحصر به خود می باشد که به طور مفصل تابعها و زیرروالها را توصیف می کند. ما مهمترین آنها را فهرست می کنیم.

۱. راهنمای کاربر *LINPACK*

LINPACK Users' Guide, by J. Dongarra, J. Bunch, C. Moler, and G. W.

Stewart, SIAM, Philadelphia, PA. 1979.

۲. راهنمای کاربر *EISPACK* می تواند از NESC یا IMSL به دست آید. راهنمای

Matrix Eigensystem Routines EISPACK Guide, by B. T. Smith, J. M. Boyle,

J. J. Dongarra, B. S. Garbow, Y. Ikebe, V. C. Klema, and C. B. Moler,

توسط انتشارات Springer-Verlag, Berlin در سال 1976 به صورت ۶ جلد *Lecture Notes در Computer Science* چاپ شده است. توسعه بعدی آن که توسط ناشر Springer-Verlag چاپ شده است توسط افراد زیر تهیه گردیده است.

Garbow, Boyle, Dongarra, and Moler (1977)

۳. راهنمای کاربر *LAPACK* که توسط E. Anderson و دیگران تهیه شده است از SIAM قابل

دسترسی است. آدرس SIAM و شماره تلفن آن عبارتند از :

SIAM (Society for Industrial and Applied Mathematics), 3600 University City Science Center, Philadelphia, PA 19114-2688; (215) 382-9800.

۴. کتابخانه NAG و راهنمای کاربر وابسته به آن را می توان از آدرس زیر به دست آورد.

Numerical Algorithms Group, Inc.

1400 Opus Place, Suite 200

Dowens Grove, IL 60515-5702

Tel: (708) 971-2395

۵. IMSL : کتابخانه نرم افزاری و مستندات IMSL از آدرس زیر قابل دسترسی هستند

IMSL, 2500 Parkwest Tower One, Houston, Texas 77042-3020; (713) 782-6060

۶. MATLAB : نرم افزار و راهنمای کاربر MATLAB از آدرس زیر قابل دسترسی هستند.

The Mathworks, Inc.
Cochituate Place
24 Prime Park Way
Natick, MA 01760-1520
TEL: (508) 653-1415; FAX: (508) 653-2997
e-mail: info@mathworks.com

ویرایش دانشجویی MATLAB توسط ناشر

Prentice-Hall, Englewood Cliffs, NJ, 07632

چاپ شده است. یک دیسک ۵ اینچی برای کامپیوترهای شخصی MS-DOS همراه با کتاب می‌باشد.
۷. کتاب

TEMPLATES for the Solutions of Linear Systems: Building Blocks for Iterative Methods by R. Barrett, M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst.

از SIMA, Philadelphia یا در قالب پست اسکرپت (postscript format) بر روی اینترنت موجود است. برای دریافت فایل پست اسکرپت یکی از روشهای زیر را می‌توانید استفاده کنید.

- anonymous ftp to netlib2.cs.utk.edu
cd linalg
get templates.ps
quit
- send e-mail to netlib@ornl.gov and in the message type:
send templates.ps from linalg

برای اطلاعات بیشتر در مورد دستیابی به نرم‌افزارهای ریاضی الکترونیکی، خواندن مقاله زیر ارزشمند است.

"Distribution of Mathematical software via Electronic Mail" by J. Dongarra and E. Grosse, *Communications of the ACM* 30,5 (1987): 403-7.

بالاخره یک مرور زیبا از الگوریتمهای بلوک شده در مقاله زیر ارائه شده است.

Demmel, Heath, and Van der Vorst: *Parallel Numerical Linear Algebra*. (Acta Numerica, vol. 2, Cambridge, MA: University Press, 1993).

برخی تبدیلات مفید در جبر خطی و کاربردهای آن

هدفهای اصلی این فصل معرفی سه ابزار اساسی ماتریسهای مقدماتی، هاوس هولدر، گیونز و نشان دادن این است که چگونه این ماتریسها می توانند برای رسیدن به تجزیه‌های مهم ماتریسی نظیر LU و شکلهای دیگر آن، QR، و کاهش به یک شکل هستبرگی یک ماتریس مورد استفاده قرار گیرند. کاربردهای تجزیه QR برای ساختن پایه‌های یکا متعامد و تصویرهای متعامد، و به هنگام درآوری تجزیه QR یک ماتریس نیز به طور مختصر ذکر می شوند.

مطالب و ابزار زیر بنایی زیر که قبلاً توسعه یافته‌اند برای فهم راحت مطالب این فصل مورد نیاز خواهند بود: برد و فضای پوچ (بخش ۱-۳-۲)، خواص رتبه (بخش ۱-۳-۳)، متعامد بودن و تصویرها (بخش ۱-۳-۶ و ۱-۳-۷) ماتریسهای خاص (بخش ۱-۴) الگوریتم حذفی گاوس پایه‌ای (الگوریتم ۱-۳-۵) و مفاهیم پایداری و وضعیت و نتایج (بخش ۳-۲ تا ۳-۴).

۵-۱ یک راه کار محاسباتی در جبر خطی عددی

اغلب الگوریتمهای عددی که در این کتاب ارائه می شوند دارای یک ساختار مشترک پایه‌ای هستند که می توانند به صورت زیر توصیف شوند:

۱. مسأله ابتدا به یک مسأله کاهش یافته تبدیل می شود.

۲. سپس مسأله کاهش یافته، با استفاده از ساختار خاصی که توسط مسأله ارائه می شود، حل می گردد.

۳. بالاخره جواب مسأله اصلی از روی جواب مسأله کاهش یافته به دست می آید.

مسأله کاهش یافته معمولاً در بردارنده یک شکل فشرده یا شکلهایی از ماتریس A ، نظیر مثلثی، هسنبرگی (تقریباً مثلثی)، سه قطری، شکل شور حقیقی (شبه سه قطری)، یا دو قطری است. از ساختارهای این شکلهای فشرده است که در حل مسأله کاهش یافته استفاده می شود. برای مثال، همان طور که در فصل ۳ ملاحظه نمودیم، جواب دستگاه خطی $Ax = b$ معمولاً توسط ابتدا مثلثی سازی ماتریس A و سپس حل دستگاه مثلثی معادل به دست می آید. در محاسبه مقادیر ویژه، قبل از به کار بردن تکرارهای QR، ماتریس A به یک شکل هسنبرگی تبدیل می شود. برای محاسبه تجزیه مقدار تکین، ابتدا ماتریس A به یک ماتریس دو قطری تبدیل می شود و سپس مقادیر تکین ماتریس دو قطری محاسبه می شوند. این شکلهای فشرده معمولاً در طی یک سری تبدیلات که معروف به تبدیلات مقدماتی، هاوس هولدر یا گایوز هستند، حاصل می گردند. ما در اینجا تبدیلات را مطالعه خواهیم کرد و نشان خواهیم داد که چگونه آنها می توانند برای رسیدن به شکلهای فشرده مختلف به کار برده شوند.

۲-۵ ماتریسهای مقدماتی و تجزیه LU

در این بخش ما نشان می دهیم که چگونه یک ماتریس را با استفاده از طرح حذفی کلاسیک معروف به طرح حذفی گاوس به مثلثی تبدیل می کنیم.

تعریف ۲-۵-۱ یک ماتریس پایین مثلثی مقدماتی مرتبه n یک ماتریس به شکل زیر است.

$$E = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \dots & \cdot & \cdot & \cdot \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \cdot & \cdot & \cdot & 1 & \ddots & \vdots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \dots & m_{k+1,k} & \ddots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \vdots & \ddots & \ddots & \cdot \\ \cdot & \cdot & \cdot & \dots & m_{n,k} & \dots & \cdot & 1 \end{bmatrix}$$

بنابراین، یک ماتریس مقدماتی یک ماتریس واحد است بجز احتمالاً برای یک تعداد کمی عنصر مخالف صفر که در زیر قطر تنها در یک ستون وجود دارند. اگر عناصر مخالف صفر در ستون k ام به صورت نشان داده شده در فوق قرار گرفته باشد، آنگاه E دارای شکل

$$E = I + m_k e_k^T$$

است که در آن I ماتریس واحد از مرتبه n ، $m_k = (0, 0, \dots, m_{k+1}, 0, \dots, m_n, 0)^T$ و بردار واحد e_k k ام می باشند. توجه کنید که $e_i^T m_k = 0$ ، $i = 1, 2, \dots, k$. از ماتریسهای مقدماتی می توان به صورت مناسب استفاده کرد تا در یک بردار به صورت نشان داده شده در لم زیر ایجاد صفر نمایند.

لم ۱-۲-۵ با مفروض بودن

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad a_1 \neq 0$$

یک ماتریس مقدماتی E وجود دارد به قسمی که Ea مضربی از e_1 باشد.

اثبات: تعریف می کنیم

$$E = \begin{bmatrix} 1 & 0 & 0 & \dots & \dots & 0 \\ -\frac{a_2}{a_1} & 1 & 0 & \dots & \dots & 0 \\ \vdots & 0 & \ddots & & & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ -\frac{a_n}{a_1} & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

پس E یک ماتریس پایین مثلثی مقدماتی است به قسمی که

$$Ea = \begin{bmatrix} a_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

□

$$\widehat{E}_\gamma \begin{bmatrix} a_{\gamma\gamma}^{(1)} \\ a_{\gamma\gamma}^{(1)} \\ \vdots \\ \vdots \\ a_{n\gamma}^{(1)} \end{bmatrix} = \begin{bmatrix} a_{\gamma\gamma}^{(1)} \\ \cdot \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

ضریب‌گرهای $m_{\gamma\gamma}, \dots, m_{n\gamma}, \dots, m_{i\gamma} = -a_{i\gamma}^{(1)} / a_{\gamma\gamma}^{(1)}$ ، $i = \gamma, \dots, n$ را ثبت کنید. تعریف کنید

$$E_\gamma = \begin{bmatrix} 1 & \cdot & \dots & \cdot \\ \cdot & & & \\ \vdots & \widehat{E}_\gamma & & \\ \cdot & & & \end{bmatrix} = \left[\begin{array}{c|c} 1 & \cdot \\ \hline \cdot & \widehat{E}_\gamma \end{array} \right]$$

در نتیجه $A^{(2)} = E_\gamma A^{(1)}$ در ستون دوم و در زیر عنصر (γ, γ) دارای عناصر صفر خواهد بود.

$$A^{(2)} = \begin{bmatrix} a_{11} & a_{1\gamma} & \dots & \dots & a_{1n} \\ \cdot & a_{\gamma\gamma}^{(1)} & \dots & \dots & a_{\gamma n}^{(2)} \\ \cdot & \cdot & a_{\gamma\gamma}^{(2)} & \dots & a_{\gamma n}^{(2)} \\ \vdots & \vdots & \vdots & & \vdots \\ \cdot & \cdot & a_{n\gamma}^{(2)} & \dots & a_{nn}^{(2)} \end{bmatrix}$$

توجه کنید که پیش ضرب $A^{(1)}$ در E_γ صفرهای قبلاً تولید شده در $A^{(1)}$ را تغییر نمی‌دهد. ماتریس $A^{(2)}$ همان ماتریس $A^{(2)}$ الگوریتم ۳-۱-۵ است.

گام k : به طور کلی، در گام k ام، یک ماتریس مقدماتی E_k به قسمی پیدا می‌شود که $A^{(k)} = E_k A^{(k-1)}$ در ستون k ام و در زیر عنصر (k, k) دارای عناصر صفر باشد. E_k در دو گام متوالی محاسبه می‌شود. ابتدا یک ماتریس مقدماتی \widehat{E}_k از مرتبه $n - k + 1$ ساخته می‌شود به طوری که

$$\widehat{E}_k \begin{bmatrix} a_{k,k}^{(k-1)} \\ a_{k+1,k}^{(k-1)} \\ \vdots \\ a_{nk}^{(k-1)} \end{bmatrix} = \begin{bmatrix} a_{kk}^{(k-1)} \\ \cdot \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

و سپس E_k به صورت زیر تعریف می شود

$$E_k = \begin{bmatrix} I_{k-1} & & \\ & \hat{E}_k & \\ & & \cdot \\ & & & \cdot \end{bmatrix}$$

در اینجا I_{k-1} ماتریس حاصل از $(k-1)$ سطر و ستون اول ماتریس واحد $n \times n$ می باشد.

ماتریس $A^{(k)} = E_k A^{(k-1)}$ همان ماتریس $A^{(k)}$ الگوریتم ۳-۱-۵ است.

ضریبگرهای زیر را ثبت کنید:

$$m_{k+1,k}, \dots, m_{nk}; \quad m_{ik} = -\frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \quad i = k-1, \dots, n$$

گام $n-1$: در انتهای گام $n-1$ ، ماتریس $A^{(n-1)}$ یک ماتریس بالا مثلثی بوده، و همان

ماتریس $A^{(n-1)}$ الگوریتم ۳-۱-۵ می باشد.

$$A^{(n-1)} = \begin{bmatrix} a_{11} & a_{12} & \dots & \dots & \dots & a_{1n} \\ \cdot & a_{22}^{(1)} & \dots & \dots & \dots & a_{2n}^{(1)} \\ \cdot & \cdot & a_{33}^{(2)} & \dots & \dots & a_{3n}^{(2)} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & & & \\ \cdot & \cdot & \dots & \dots & \cdot & a_{nn}^{(n-1)} \end{bmatrix}$$

L و U را به دست آورید:

$$\begin{aligned} A^{(n-1)} &= E_{n-1} A^{(n-2)} = E_{n-1} E_{n-2} A^{(n-2)} = \dots \\ &= E_{n-1} E_{n-2} E_{n-3} \dots E_2 E_1 A \end{aligned}$$

قرار دهید

$$U = A^{(n-1)}, \quad L_1 = E_{n-1} E_{n-2} \dots E_2 E_1$$

بنابراین از قبل داریم

$$U = L_1 A$$

چون هر E_k یک ماتریس پایین مثلثی واحد (یک ماتریس پایین مثلثی دارای ۱ در امتداد

قطر) است، در نتیجه L_1 نیز چنین است؛ بنابراین L_1^{-1} وجود دارد. (توجه کنید

که حاصلضرب دو ماتریس مثلثی از یک نوع یک ماتریس مثلثی از همان نوع است).

قرار دهید $L = L_1^{-1}$. آنگاه معادله $U = L_1 A$ به صورت زیر در می آید

$$A = LU$$

این تجزیه A به تجزیه LU معروف است.

تعریف ۵-۲-۳ عناصر $a_{11}, a_{22}^{(1)}, \dots, a_{nn}^{(n-1)}$ عناصر محوری نامیده می شوند و فرایند قبلی برای به دست آوردن تجزیه LU به حذفی گاوس بدون تعویض سطر معروف است. این فرایند معمولاً معروف به حذفی گاوس بدون محورگیری است. فرایند به نام ریاضیدان و ستاره شناس آلمانی کارل - فردریچ گاوس (1777-1855) نامیده شده است.

یک عبارت صریح برای L

چون $E_i^{-1} = I - m_i e_i^T$ و $L = L_1^{-1} = E_1^{-1} E_2^{-1} \dots E_{n-1}^{-1}$ که در آن

$$m_i = (0, 0, \dots, 0, m_{i+1,i}, \dots, m_{n,i})^T$$

داریم:

$$L = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ -m_{21} & 1 & 0 & \dots & 0 \\ -m_{31} & -m_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -m_{n1} & -m_{n2} & \dots & -m_{n,n-1} & 1 \end{bmatrix}$$

بنابراین، L می تواند بدون پیدا کردن یک ماتریس معکوس تشکیل شود.

وجود و منحصر به فرد بودن تجزیه LU

خیلی مهم است که توجه داشته باشیم که برای موجود بودن یک تجزیه LU باید عناصر محوری مخالف صفر باشند. بنابراین تجزیه LU ممکن است حتی برای یک ماتریس خیلی ساده وجود نداشته باشد. اختیار کنید

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

می‌توان نشان داد (تمرین ۳) که اگر A_r ، زیر ماتریس حاصل از r سطر اول و r ستون اول ماتریس A باشد، آنگاه

$$\det(A_r) = a_{11} a_{22}^{(1)} \dots a_{rr}^{(r-1)}$$

بنابراین، اگر $\det(A_r) \neq 0$ ، $r = 1, 2, \dots, n$ ، مخالف صفر باشد، آنگاه تجزیه LU همواره وجود دارد. در حقیقت در این حالت تجزیه LU منحصر به فرد است، زیرا اگر

$$A = L_1 U_1 = L_r U_r$$

آنگاه چون L_1 ، L_r و ماتریس A نامنفرد هستند، از

$$\det(A) = \det(L_1 U_1) = \det(L_1) \cdot \det(U_1)$$

و

$$\det(A) = \det(L_r U_r) = \det(L_r) \cdot \det(U_r)$$

نتیجه می‌شود که U_1 و U_r نیز نامنفرد هستند. بنابراین

$$L_r^{-1} L_1 = U_r U_1^{-1}$$

$L_r^{-1} L_1$ حاصلضرب دو ماتریس پایین مثلثی واحد است، پس یک ماتریس پایین مثلثی واحد است؛ $U_r U_1^{-1}$ حاصلضرب دو ماتریس بالا مثلثی است و در نتیجه بالا مثلثی است. چون یک ماتریس پایین مثلثی واحد فقط هنگامی می‌تواند با یک ماتریس بالا مثلثی برابر باشد که هر دو واحد باشند، از این رو داریم

$$L_1 = L_r, \quad U_1 = U_r$$

قضیه ۵-۲-۱. قضیه LU . فرض کنید A یک ماتریس $n \times n$ با همه کهادهای اصلی پیشرو مخالف صفر باشد. آنگاه A یک تجزیه LU منحصر به فرد دارد:

$$A = LU$$

که در آن L یک ماتریس پایین مثلثی واحد و U یک ماتریس بالا مثلثی است.

تبصره: توجه کنید که در قضیه ۵-۲-۱، اگر عناصر قطری L مشخص نباشند، آنگاه تجزیه منحصر به

فرد نیست.

مثال ۵-۲-۱ یک تجزیه LU برای ماتریس زیر بیابید

$$A = \begin{bmatrix} 2 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 2 & 4 \end{bmatrix}$$

گام ۱: E_1 را محاسبه می‌کنیم. ضریبها عبارتند از: $m_{31} = -\frac{1}{2}$, $m_{21} = -2$.

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix}$$

$$A^{(1)} = E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 1 & \frac{5}{2} \end{bmatrix}$$

گام ۲: E_2 را محاسبه می‌کنیم. ضریب عبارت است: $m_{32} = -1$.

$$\hat{E}_2 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \hat{E}_2 \end{bmatrix}$$

$$A^{(2)} = E_2 A^{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 1 & \frac{5}{2} \end{bmatrix} = \begin{bmatrix} 2 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{5}{2} \end{bmatrix}$$

بنابراین

$$U = \begin{bmatrix} 2 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{5}{2} \end{bmatrix}$$

L را محاسبه می‌کنیم

$$L_1 = E_2 E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -\frac{1}{2} & -1 & 1 \end{bmatrix}$$

$$L = L_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -m_{21} & 1 & 0 \\ -m_{31} & -m_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ \frac{1}{2} & 1 & 1 \end{bmatrix}$$

(توجه کنید که نه L_1 نه معکوس آن لازم نیست که به طور صریح محاسبه شوند).

تشکیل ماتریس L و جزئیات محاسباتی دیگر

۱. هر ماتریس مقدماتی E_k به طور منحصر به فرد توسط $(n - k)$ ضریبگر

$$m_{k+1, k}, \dots, m_{nk}$$

تعیین می‌شود. بنابراین برای ساختن E_k ، کافی است این ضریبگرها ذخیره شوند.

۲. پیش ضرب $A^{(k-1)}$ در ماتریس E_k فقط بر روی سطرها $k + 1$ تا n اثر می‌گذارد. k سطر اول بدون تغییر و همانند $A^{(k-1)}$ باقی می‌ماند و فقط $(n - k)$ سطر آخر تغییر می‌کنند.

بنابراین، اگر $A^{(k)} = E_k A^{(k-1)}$ ، آنگاه

(الف) $a_{ij}^{(k)} = a_{ij}^{(k-1)}$ ، $i = 1, 2, \dots, k$ ، $j = 1, 2, \dots, n$ ؛ $(k$ سطر اول)؛

(ب) $a_{ij}^{(k)} = a_{ij}^{(k-1)} + m_{ik} a_{kj}^{(k-1)}$ ، $i = k + 1, \dots, n$ ؛ $j = k + 1, \dots, n$ ؛ $(n - k)$ سطر

آخر).

(توجه کنید که چگونه عناصر $A^{(k)}$ از عناصر $A^{(k-1)}$ در الگوریتم ۳-۱-۵ به

دست آمده‌اند).

برای مثال، فرض کنید $n = 3$.

$$E_{\gamma}A = A^{(1)} = (a_{ij}^{(1)})$$

$$E_{\gamma} = \begin{bmatrix} 1 & \cdot & \cdot \\ m_{\gamma 1} & 1 & \cdot \\ m_{\gamma 1} & \cdot & 1 \end{bmatrix}$$

$$E_{\gamma}A = \begin{bmatrix} 1 & \cdot & \cdot \\ m_{\gamma 1} & 1 & \cdot \\ m_{\gamma 1} & \cdot & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ \cdot & a_{22}^{(1)} & a_{23}^{(1)} \\ \cdot & a_{32}^{(1)} & a_{33}^{(1)} \end{bmatrix}$$

که در آن

$$a_{22}^{(1)} = a_{22} + m_{\gamma 1}a_{21}$$

$$a_{32}^{(1)} = a_{32} + m_{\gamma 1}a_{31}$$

و غیره.

۳. به محض این که $A^{(k)}$ تشکیل شود، می‌تواند بر روی A نوشته شود.

۴. بردار $(m_{k+1,k}, \dots, m_{nk})$ دارای $(n-k)$ مؤلفه است، و در گام k ام دقیقاً $(n-k)$ صفر

تولید می‌شود؛ بنابراین یک طرح واضح در مورد ذخیره سازی این است که این $(n-k)$ عنصر در مکانهای $(k+1, k)$ ، $(k+2, k)$ ، \dots ، (n, k) از ماتریس A در زیر قطر ذخیره شوند.

۵. عناصر ماتریس بالا مثلثی U نیز می‌توانند در نیمه بالایی A ، شامل قطر، ذخیره شوند. با این طرح ذخیره سازی در انتهای گام $(n-1)$ ام ماتریس $A^{(n-1)}$ به صورت زیر خواهد بود.

$$A \equiv A^{(n-1)} = \begin{bmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ m_{21} & a_{22} & \dots & \dots & a_{2n} \\ m_{31} & m_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ m_{n1} & m_{n2} & \dots & m_{n,n-1} & a_{nn} \end{bmatrix}$$

بنابراین گام k ام نوعی از روش حذفی گاوس برای تجزیه LU از مراحل زیر تشکیل می‌شود.

۱. تشکیل ضریبها و ذخیره سازی آنها در مکانهای مناسب زیر قطر، و

۲. به هنگام درآوردی عناصر سطرهای $(k+1)$ تا n ام و ذخیره آنها در نیمه بالایی A .

براساس بحث قبل، اکنون الگوریتم زیر را ارائه می‌دهیم.

الگوریتم ۵-۲-۱ مثلثی سازی با استفاده از روش حذفی گاوس بدون محورگیری. فرض کنید A

یک ماتریس $n \times n$ باشد. الگوریتم زیر مثلثی سازی A را در صورت وجود انجام می دهد. الگوریتم عناصر U را بر روی عناصر قسمت بالا مثلثی A شامل قطر و ضربگرهای لازم برای محاسبه L را بر روی عناصر زیر قطر A می نویسد.

برای $(n - 1), 2, \dots, k$ انجام دهید.

۱: (ضربگرها را تشکیل دهید) $a_{ik} \equiv m_{ik} = -\frac{a_{ik}}{a_{kk}}$ ($i = k + 1, k + 2, \dots, n$)

۲: (عناصر را به هنگام در آوردن) $a_{ij} \equiv a_{ij} + m_{ik}a_{kj}$ ($i = k + 1, \dots, n; j = k + 1, \dots, n$)

تبصره: الگوریتم ماتریس L را به طور صریح ارائه نمی دهد؛ ولیکن L می تواند از روی ضربگرهای ذخیره شده در هر گام، به همان صورت نشان داده شده در قبل، تشکیل شود (عبارت صریح را برای L ملاحظه کنید).

تذکر **MATCOM**: الگوریتم ۵-۲-۱ در برنامه **LUGSEL** در **MATCOM** پیاده سازی شده است. U و L به طور صریح محاسبه می شوند.

شمارش اعمال ممیز شناور (flops). الگوریتم تقریباً به $\frac{n^3}{3}$ اعمال ممیز شناور (flops) نیاز دارد. این تعداد می تواند به صورت زیر دیده شود: درگام ۱، $(n - 1)$ ضربگر را محاسبه می کنیم و $(n - 1)^2$ عنصر از A را به هنگام در می آوریم. هر ضربگر به یک عمل ممیز شناور (flop) نیاز دارد و به هنگام درآوری هر عنصر نیز به یک عمل ممیز شناور نیاز (flop) نیاز دارد. بنابراین گام ۱ به $[(n - 1)^2 + (n - 1)]$ عمل ممیز شناور (flops) نیاز دارد.

به طور مشابه، گام ۲ به $[(n - 2)^2 + (n - 2)]$ عمل ممیز شناور (flops) نیاز دارد و غیره. به طور کلی، گام k به $[(n - k)^2 + (n - k)]$ عمل ممیز شناور (flops) نیاز دارد. چون $(n - 1)$ گام وجود دارد، داریم

$$\begin{aligned} \text{کل اعمال ممیز شناور (flops)} &= \sum_{k=1}^{n-1} (n-k)^2 + \sum_{k=1}^{n-1} (n-k) \\ &= \frac{n(n-1)(2n-1)}{6} + \frac{n(n-1)}{2} \\ &= \left[\frac{n^3}{3} + O(n^2) \right] \end{aligned}$$

روش حذفی گاوس برای یک ماتریس مستطیلی

فرایند حذفی گاوس که برای یک ماتریس $n \times n$ مانند A توصیف شد، می‌تواند به سادگی برای یک ماتریس $m \times n$ و برای محاسبه تجزیه LU آن در صورت وجود گسترش یابد. فرایند یکسان است، تنها تعداد گام در این حالت برابر است با $k = \min\{m - 1, n\}$. مثال زیر یک مثال تشریحی است. فرض کنید

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$m = 3, \quad n = 2$$

تعداد گامها برابر است با $k = \min\{2, 2\} = 2$.

گام ۱: ضریبها عبارتند از: $m_{31} = -5, m_{21} = -3$.

$$a_{22} \equiv a_{22}^{(1)} = a_{22} + m_{21}a_{12} = -2$$

$$a_{32} \equiv a_{32}^{(1)} = a_{32} + m_{31}a_{12} = -4$$

$$A \equiv A^{(1)} = \begin{bmatrix} 1 & 2 \\ \cdot & -2 \\ \cdot & -4 \end{bmatrix}$$

گام ۲: ضریبها عبارت است از $m_{32} = -2$.

$$A \equiv A^{(2)} = \begin{bmatrix} 1 & 2 \\ \cdot & -2 \\ \cdot & \cdot \end{bmatrix}$$

$$U = \begin{bmatrix} 1 & 2 \\ \cdot & -2 \\ \cdot & \cdot \end{bmatrix}$$

توجه کنید که U در این حالت به جای یک ماتریس بالا مثلثی یک ماتریس بالا ذوزنقه‌ای است:

$$L = \begin{bmatrix} 1 & \cdot & \cdot \\ -m_{21} & 1 & \cdot \\ -m_{31} & -m_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \cdot & \cdot \\ 3 & 1 & \cdot \\ 5 & 2 & 1 \end{bmatrix}$$

تحقیق کنید که

$$LU = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 5 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = A$$

مشکلات روش حذفی گاوس بدون محورگیری

همان گونه که قبلاً ملاحظه نمودیم، روش حذفی گاوس بدون محورگیری با شکست مواجه می شود اگر یکی از عناصر محوری صفر باشد. ولیکن وضع بدتری پیش می آید اگر که هر عنصر محوری نزدیک به صفر باشد: در این حالت روش می تواند تا کامل شدن انجام شود، ولیکن نتایج به دست آمده ممکن است کاملاً غلط باشد.

مثال معروف زیر را از مرجع Forsythe and Moler (1967; 34) ملاحظه کنید: فرض کنید روش حذفی گاوس بدون محورگیری در مورد ماتریس زیر به کار برده شود

$$A = \begin{bmatrix} 0.0001 & 1 \\ 1 & 1 \end{bmatrix}$$

فقط یک گام وجود دارد، ضربگر $m_{21} = -1/10^{-4} = -10^4$

$$U = A^{(1)} = \begin{bmatrix} 0.0001 & 1 \\ . & -10^4 \end{bmatrix}$$

$$L = \begin{bmatrix} 1 & . \\ 10^4 & 1 \end{bmatrix}$$

حاصل ضرب L و U محاسبه شده نتیجه می دهد

$$LU = \begin{bmatrix} 0.0001 & 1 \\ 1 & . \end{bmatrix}$$

که با A تفاوت دارد. چه کسی مقصر است؟

توجه کنید که عنصر محوری $a_{11}^{(1)} = 0.0001$ خیلی نزدیک به صفر (در حساب سه رقمی) است. این عنصر محوری کوچک یک ضربگر بزرگ ارائه داد. ضربگر بزرگ، هنگامی که برای به هنگام درآوری مورد استفاده قرار گرفت، عناصر کوچک را حذف نمود (برای مثال، $(1-10^4)$ مقدار -10^4 را نتیجه داد). خوشبختانه فقط با تعویض سطرها می توان از این عنصر محوری کوچک پرهیز کرد، ماتریس فوق

را با نوشتن سطر اول در سطر دوم و سطر دوم در سطر اول ملاحظه کنید:

$$A' = \begin{bmatrix} 1 & 1 \\ 0/\dots 1 & 1 \end{bmatrix}$$

اکنون روش حذفی گاوس نتیجه می دهد

$$U = A^{(1)} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad L = \begin{bmatrix} 1 & 0 \\ 0/\dots 1 & 1 \end{bmatrix}$$

توجه کنید که در این حالت عنصر محوری برابر $a_{11}^{(1)} = 1$ است. برای حاصلضرب داریم:

$$LU = \begin{bmatrix} 1 & 1 \\ 0/\dots 1 & 1/\dots 1 \end{bmatrix} = A'$$

۵-۲-۲ روش حذفی گاوس با محورگیری جزئی

در مثال قبل، یک تجزیه ماتریس A' پیدا کردیم که یک گونه جایگشت یافته از A است، بدین معنا که سطرها تعویض شده‌اند. هدف اولیه از تجزیه یک ماتریس A به LU استفاده از این تجزیه برای حل یک دستگاه خطی است. به سادگی ملاحظه می‌شود که جواب دستگاه $Ax = b$ با جواب دستگاه $A'x = b'$ یکی است، که در آن b' به همان طریقی به دست آمده است که A' توسط آن محاسبه شده است. بنابراین اگر تعویض سطر می‌تواند به پرهیز از یک عنصر محوری کوچک کمک کند، مسلماً انجام آن مطلوب خواهد بود.

همچنان که این مثال پیشنهاد می‌کند، در روش حذفی گاوس بدون محورگیری شاید بتوان از یک حادثه بد، با مشخص کردن یک «عنصر محوری خوب» (یک عنصر محوری تا حد ممکن بزرگ) در هر گام، قبل از این که فرایند حذفی به کار برده شود، پرهیز کرد. عنصر محوری خوب ممکن است در میان عناصر یک ستون و یا در میان همهٔ عناصر در یک زیر ماتریس جاری باشد. در حالت اول، چون جست‌وجو تنها به‌طور جزئی انجام می‌شود، روش محورگیری جزئی نامیده می‌شود. در حالت دوم، روش محورگیری کامل نامیده می‌شود. مهم است توجه کرد که هدف از محورگیری جلوگیری از رشد زیاد در ماتریسهای کاهش یافته است، که می‌تواند داده‌های اصلی را نابود سازد. یکی از طرق انجام این امر نگه داشتن اندازه ضریب‌ها کوچک‌تر از ۱ است، و این دقیقاً چیزی است که توسط محورگیری انجام می‌شود. ولیکن، ضریب‌های بزرگ لزوماً به معنای ناپایداری نیستند (بحث مربوطه به روش حذفی گاوس بدون محورگیری را برای ماتریسهای معین مثبت در فصل ۶ ملاحظه کنید). ابتدا روش حذفی

گوس با محورگیری جزئی را توصیف می‌کنیم.
فرایند از $n - 1$ گام تشکیل می‌شود.

گام ۱: ستون اول را برای مشخص کردن عنصر با بزرگ‌ترین قدر مطلق آن ستون بررسی کنید. فرض کنید آن عنصر a_{r_1} باشد.

• ماتریس جایگشت P_1 را با تعویض سطر ۱ و سطر r_1 از ماتریس واحد و گذاشتن بقیه سطرها بدون تغییر، تشکیل دهید.

• ماتریس $P_1 A$ را با تعویض سطرها r_1 و ۱ از A تشکیل دهید.

• یک ماتریس پایین مثلثی مقدماتی M_1 به قسمی پیدا کنید که $A^{(1)} = M_1 P_1 A$ دارای عناصر صفر در زیر عنصر $(1, 1)$ بر روی ستون اول باشد. کافی است M_1 طوری ساخته شود که

$$M_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{n1} \end{bmatrix} = \begin{bmatrix} * \\ \circ \\ \vdots \\ \circ \end{bmatrix}$$

توجه کنید که

$$M_1 = \begin{bmatrix} 1 & \circ & \circ & \dots & \dots & \circ \\ m_{21} & 1 & \circ & \dots & \dots & \circ \\ m_{31} & \circ & 1 & \dots & \dots & \circ \\ \vdots & \vdots & \circ & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \circ \\ m_{n1} & \circ & \circ & \dots & \circ & 1 \end{bmatrix}$$

که در آن $m_{21} = -a_{21} / a_{11}$ ، $m_{31} = -a_{31} / a_{11}$ ، $m_{n1} = -a_{n1} / a_{11}$ و

a_{ij} نشان‌دهنده عنصر (i, j) ماتریس جایگشت یافته $P_1 A$ است. ضربگرهای m_{i1} ، $i = 2, \dots, n$ را ذخیره کنید و تعویض سطرها را ثبت نمایید.

$$A^{(1)} = \begin{bmatrix} * & * & \dots & * \\ \circ & * & \dots & * \\ \circ & * & \dots & * \\ \vdots & \vdots & & \vdots \\ \circ & * & \dots & * \end{bmatrix}$$

گام ۲: ستون دوم $A^{(1)}$ و زیر سطر اول را برای مشخص کردن عنصر با بزرگ‌ترین قدر مطلق آن ستون، بررسی کنید. فرض کنید $a_{r_2}^{(1)}$ چنین عنصری باشد. ماتریس جایگشت P_2 را با تعویض سطرهاى دوم و r_2 ماتریس واحد تشکیل دهید و بقیه سطرها را بدون تعویض باقی گذارید. $P_2 A^{(1)}$ را تشکیل دهید.

سپس ماتریس پایین مثلثی مقدماتی M_2 را طوری پیدا کنید که $A^{(2)} = M_2 P_2 A^{(1)}$ دارای عناصر صفر در زیر عنصر $(2, 2)$ ستون دوم باشد. M_2 به صورت زیر ساخته می‌شود. ابتدا، یک ماتریس مقدماتی \widehat{M}_2 از مرتبه $(n - 1)$ به قسمی بسازید که

$$\widehat{M}_2 \begin{bmatrix} a_{22} \\ a_{32} \\ \vdots \\ a_{n2} \end{bmatrix} = \begin{bmatrix} * \\ \cdot \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

سپس تعریف کنید

$$M_2 = \begin{bmatrix} 1 & \cdot & \dots & \cdot \\ \cdot & & & \\ \vdots & & \widehat{M}_2 & \\ \cdot & & & \end{bmatrix}$$

توجه کنید که a_{ij} عنصر (i, j) ماتریس جاری $P_2 A^{(1)}$ را نشان می‌دهد. در انتهای گام ۲ خواهیم داشت

$$A^{(2)} = M_2 P_2 A^{(1)} = \begin{bmatrix} * & * & \dots & \dots & * \\ \cdot & * & \dots & \dots & * \\ \cdot & \cdot & * & \dots & * \\ \vdots & \vdots & \vdots & & \vdots \\ \cdot & \cdot & * & \dots & * \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 1 & \cdot & \cdot & \dots & \dots & \cdot \\ \cdot & 1 & \cdot & \dots & \dots & \cdot \\ \cdot & m_{22} & 1 & \dots & \dots & \cdot \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \cdot & m_{n2} & \cdot & \dots & \cdot & 1 \end{bmatrix}$$

که در آن $m_{ij} = -a_{ij} / a_{jj}$ ، $i = 3, 4, \dots, n$ ، ضربگرهای m_{ij} را ذخیره کنید و تعویض سطرها را ثبت نمایید.

گام k : به طور کلی، در گام k ام، عناصر ستون k ام ماتریس $A^{(k-1)}$ و زیر سطر $(k-1)$ ام را برای مشخص کردن عنصر محوری a_{kk} بررسی کنید. ماتریس جایگشت P_k را تشکیل دهید و یک ماتریس پایین مثلثی مقدماتی M_k طوری پیدا کنید که $A^{(k)} = M_k P_k A^{(k-1)}$ دارای عناصر صفر در زیر عنصر (k, k) بر روی ستون k ام باشد. برای ساختن M_k ابتدا ماتریس \hat{M}_k از مرتبه $(n-k+1)$ به قسمی ساخته می شود که

$$\hat{M}_k \begin{bmatrix} a_{kk} \\ \vdots \\ \vdots \\ a_{nk} \end{bmatrix} = \begin{bmatrix} * \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

و سپس M_k به صورت زیر تعریف می شود

$$M_k = \begin{bmatrix} I_{k-1} & \cdot \\ \cdot & \hat{M}_k \end{bmatrix}$$

که در آن \cdot یک ماتریس صفر است. توجه کنید که عنصر a_{ik} (ماتریس $A^{(k-1)}$ را P_k نشان می دهد.

گام $n-1$: در انتهای گام $(n-1)$ ام، ماتریس $A^{(n-1)}$ یک ماتریس بالا مثلثی خواهد بود. تشکیل U : قرار دهید

$$A^{(n-1)} = U \tag{1-2-5}$$

سپس

$$\begin{aligned} U &= A^{(n-1)} = M_{n-1} P_{n-1} A^{(n-2)} \\ &= M_{n-1} P_{n-1} M_{n-2} P_{n-2} A^{(n-3)} = \dots \\ &= M_{n-1} P_{n-1} M_{n-2} P_{n-2} \dots M_2 P_2 M_1 P_1 A \end{aligned}$$

قرار دهید

$$M_{n-1} P_{n-1} M_{n-2} P_{n-2} \dots M_2 P_2 M_1 P_1 = M \tag{2-2-5}$$

سپس از رابطه قبل تجزیه زیر را برای A داریم:

$$U = MA$$

قضیه ۲-۲-۵ قضیه تجزیه با محورگیری جزئی. با مفروض بودن یک ماتریس نامنفرد $n \times n$ مانند A ، روش حذفی گاوس با محورگیری جزئی یک ماتریس بالا مثلثی U و یک ماتریس پایین مثلثی جایگشت یافته به قسمی ارائه می‌دهد که

$$MA = U$$

ماتریسهای M و U به ترتیب توسط (۲-۲-۵) و (۱-۲-۵) ارائه می‌شوند.

از (۲-۲-۵) به سادگی می‌توان دید که یک ماتریس جایگشت P وجود دارد به طوری که $PA =$

LU . تعریف می‌کنیم

$$P = P_{n-1} \dots P_2 P_1 \quad (۳-۲-۵)$$

$$L = P(M_{n-1} P_{n-1} \dots M_1 P_1)^{-1} \quad (۴-۲-۵)$$

سپس $PA = LU$

نتیجه ۱-۲-۵ قضیه تجزیه LU با محورگیری جزئی. با مفروض بودن یک ماتریس نامنفرد $n \times n$ مانند A ، روش حذفی گاوس با محورگیری جزئی یک تجزیه LU از یک گونه جایگشت یافته از A را نتیجه می‌دهد:

$$PA = LU$$

P یک ماتریس جایگشت ارائه شده توسط (۳-۲-۵)، L یک ماتریس پایین مثلثی واحد ارائه شده توسط (۴-۲-۵) و U یک ماتریس بالا مثلثی ارائه شده توسط (۱-۲-۵) می‌باشند.

مثال ۲-۲-۵ ماتریس A زیر را با استفاده از محورگیری جزئی مثلثی کنید؛ یعنی M و U را به قسمی پیدا کنید که $MA = U$. عنصر محوری توسط \square مشخص شده است.

$$A = \begin{bmatrix} \cdot / \cdot \cdot \cdot 1 & 1 \\ \square & 1 \end{bmatrix}$$

فقط یک گام مورد نیاز است. عنصر محوری ۱ است، و $r_1 = 2$

$$P_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$P_1 A = \begin{bmatrix} 1 & 1 \\ 0 & \dots 1 \end{bmatrix}$$

$$m_{21} = -\frac{-0/\dots 1}{1} = -1.0^{-4}$$

$$M_1 = \begin{bmatrix} 1 & 0 \\ m_{21} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1.0^{-4} & 1 \end{bmatrix}$$

$$M_1 P_1 A = \begin{bmatrix} 1 & 0 \\ -1.0^{-4} & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & \dots 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & \dots 9999 \end{bmatrix} = U$$

$$M = M_1 P_1 = \begin{bmatrix} 1 & 0 \\ -1.0^{-4} & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -1.0^{-4} \end{bmatrix}$$

مثال ۵-۲-۳ ماتریس

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

را با استفاده از محورگیری جزئی مثلثی کنید. A را به صورت $A = MU$ بیان کنید. بعلاوه P و L را

به قسمی پیدا کنید که $PA = LU$

گام ۱: عنصر محوری $a_{11} = 1$ است و $r = 2$

$$P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P_1 A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

$$A^{(1)} = M_{\gamma} P_{\gamma} A = \begin{bmatrix} 1 & 2 & 3 \\ \cdot & \square & 1 \\ \cdot & -1 & -2 \end{bmatrix}$$

گام ۲: عنصر محوری ۱ $a_{22} = 1$ است.

$$P_{\gamma} A^{(1)} = \begin{bmatrix} 1 & 2 & 3 \\ \cdot & 1 & 1 \\ \cdot & -1 & -2 \end{bmatrix}$$

$$P_{\gamma} = I_3$$

(تعویض لازم نیست)

$$\tilde{M}_{\gamma} = \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix}$$

$$M_{\gamma} = \begin{bmatrix} I_1 & \cdot \\ \cdot & \tilde{M}_{\gamma} \end{bmatrix} = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \cdot & 1 \end{bmatrix}$$

$$U = A^{(2)} = M_{\gamma} P_{\gamma} A^{(1)} = \begin{bmatrix} 1 & 2 & 3 \\ \cdot & 1 & 1 \\ \cdot & \cdot & -1 \end{bmatrix}$$

$$M = M_{\gamma} P_{\gamma} M_{\gamma} P_{\gamma} = \begin{bmatrix} \cdot & 1 & \cdot \\ 1 & \cdot & \cdot \\ 1 & -1 & 1 \end{bmatrix}$$

به آسانی تحقیق می‌شود که $A = MU$.

L و P را تشکیل می‌دهیم:

$$P = P_{\gamma} P_1 = \begin{bmatrix} \cdot & 1 & \cdot \\ 1 & \cdot & \cdot \\ \cdot & \cdot & 1 \end{bmatrix}$$

$$L = P(M_{\gamma} P_{\gamma} M_{\gamma} P_{\gamma})^{-1} = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ 1 & -1 & 1 \end{bmatrix}$$

تحقیق رابطه $PA = LU$ آسان است.

تشکیل ماتریس M و جزئیات محاسباتی دیگر

۱. هر ماتریس جایگشت P_k می تواند با ثبت اندیس r_k تشکیل شود، زیرا P_k ماتریس واحد جایگشت یافته ای است که در آن سطرهای k و r_k تعویض شده اند. ولیکن نه ماتریس P_k نه حاصلضرب $P_k A^{(k-1)}$ لازم است به طور صریح تشکیل شوند. این بدین علت است که ماتریس $P_k A^{(k-1)}$ دقیقاً گونه جایگشت یافته $A^{(k-1)}$ است که در آن سطرهای r_k و k تعویض شده اند.
۲. هر ماتریس مقدماتی M_k می تواند دقیقاً با ذخیره کردن $(n - k)$ ضریبگر تشکیل شود. ماتریسهای $M_k P_k A^{(k-1)} = M_k B$ نیز لازم نیستند به طور صریح محاسبه شوند. توجه کنید که عناصر k سطر اول ماتریس $M_k B$ همان عناصر k سطر اول ماتریس B هستند، و عناصر $(n - k)$ سطر باقیمانده به صورت زیر ارائه می شوند:

$$b_{ij} + m_{ik} b_{kj} \quad (i = k+1, \dots, n; \quad j = k+1, \dots, n)$$

۳. ضریبگرها می توانند در جاهای مناسبی از قسمت پایین مثلثی A (زیر قطر) به همان گونه که محاسبه شده اند، ذخیره شوند.
 ۴. ماتریس بالا مثلثی نهایی $U = A^{(n-1)}$ در قسمت بالا مثلثی A ذخیره می شود.
 ۵. اندیسهای محوری r_k در یک آرایه صحیح مجزا با تک اندیس ذخیره می شوند.
 ۶. $A^{(k)}$ ها می توانند به محض تشکیل شدن بر روی A نوشته شوند.
- با در نظر داشتن بحث قبل، اکنون می توانیم الگوریتم عملی زیر را برای تجزیه LU با محورگیری جزئی فرمول بندی کنیم.

الگوریتم ۵-۲-۲ مثلثی سازی با استفاده از روش حذفی گاوس با محورگیری جزئی
 فرض کنید A یک ماتریس $n \times n$ نامنفرد باشد. الگوریتم زیر مثلثی سازی A را با تعویض سطرها، با استفاده از روش حذفی گاوس با محورگیری جزئی انجام می دهد. ماتریس بالا مثلثی U در قسمت بالا مثلثی A ، شامل قطر ذخیره می شود. ضریبگرهای مورد نیاز برای محاسبه ماتریس مثلثی جایگشت یافته M که $MA = U$ در قسمت پایین مثلثی A ذخیره می شوند. اندیسهای جایگشت r_k در یک آرایه مجزا ذخیره می شوند.

برای $k = 1, 2, \dots, n - 1$ انجام دهید.

۱: r_k را به قسمی پیدا کنید که $|a_{r_k k}| = \max_{k \leq i \leq n} |a_{ik}|$ ، اگر $a_{r_k k} = 0$ ، آنگاه متوقف شوید.

در غیر این صورت ادامه دهید

۲: (سطرهای r_k و k را تعویض کنید). $a_{kj} \leftrightarrow a_{r_k j}$ ، $(j = k, k + 1, \dots, n)$.

۳: (ضربگرها را تشکیل دهید) $a_{ik} \equiv m_{ik} = -\frac{a_{ik}}{a_{kk}}$ ($i = k + 1, \dots, n$)

۴: (عناصر را به هنگام در آوردن) $a_{ij} \equiv a_{ij} + m_{ik}a_{kj} = a_{ij} + a_{ik}a_{kj}$ ($i = k + 1, \dots, n; j = k + 1, \dots, n$)

شمارش اعمال ممیز شناور (flops). الگوریتم حدوداً به $n^3/3$ عمل ممیز شناور (flops) و $O(n^2)$ مقایسه نیاز دارد. توجه کنید که جستجو برای عنصر محوری در گام k نیاز به $(n - k)$ مقایسه دارد.

تذکر: الگوریتم ۲-۲-۵ ماتریسهای M و P را به طور صریح ارائه نمی‌دهد. ولیکن این ماتریسها می‌توانند، به صورت بیان شده، به ترتیب از روی ضربگرها و اندیسهای جایگشت به سادگی ساخته شوند. الگوریتم می‌تواند به سادگی برای حالت $m \times n$ ، به صورت نشان داده شده در بخش ۱-۲-۵ تعمیم داده شود.

تبصره: الگوریتم ۲-۲-۵ به سطرهای A در داخلی‌ترین حلقه دسترسی دارد و به این دلیل به الگوریتم حذفی گاوس جهت داده شده سطری (با محورگیری جزئی) معروف است. بعلاوه به الگوریتم kij نیز معروف است؛ توجه کنید که i و j در حلقه‌های داخلی ظاهر می‌شوند. الگوریتم جهت داده شده ستونی می‌تواند به طور مشابه توسعه یابد. یک چنین الگوریتم جهت داده شده ستونی در LINPACK (زیر روال SGEFA در LINPACK) استفاده شده است.

تذکر MATCOM: الگوریتم ۲-۲-۵ در برنامه PARPIV در MATCOM پیاده‌سازی شده است. ماتریسهای M و P به طور صریح محاسبه شده‌اند.

مثال ۴-۲-۵ فرض کنید

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

گام ۱: $k = 1$

۱. عنصر محوری برابر ۷ است، $r_1 = 3$.

۲. سطرهای ۳ و ۱ را تعویض می‌کنیم:

$$A = \begin{bmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 4 \end{bmatrix}$$

۳. ضریبها را تشکیل می‌دهیم:

$$a_{۲۱} = m_{۲۱} = -\frac{۴}{۷} \quad a_{۳۱} = m_{۳۱} = -\frac{۱}{۷}$$

۴. به هنگام در می‌آوریم:

$$A \equiv \begin{pmatrix} ۷ & ۸ & ۹ \\ \cdot & \frac{۳}{۷} & \frac{۶}{۷} \\ \cdot & \frac{۶}{۷} & \frac{۱۹}{۷} \end{pmatrix}$$

گام ۲: ۲ = k

۱. عنصر محوری برابر $\frac{7}{6}$ است، $r_۲ = ۳$.

۲. سطرها ۲ و ۳ را تعویض می‌کنیم:

$$A \equiv \begin{pmatrix} ۷ & ۸ & ۹ \\ \cdot & \frac{۶}{۷} & \frac{۱۹}{۷} \\ \cdot & \frac{۳}{۷} & \frac{۶}{۷} \end{pmatrix}$$

۳. ضریبها را تشکیل می‌دهیم:

$$m_{۳۲} = -\frac{۱}{۲}$$

۴. به هنگام در می‌آوریم:

$$A \equiv \begin{pmatrix} ۷ & ۸ & ۹ \\ \cdot & \frac{۶}{۷} & \frac{۱۹}{۷} \\ \cdot & \cdot & -\frac{۱}{۲} \end{pmatrix}$$

M را تشکیل می‌دهیم:

$$M = \begin{pmatrix} \cdot & \cdot & ۱ \\ ۱ & \cdot & -\frac{۱}{۷} \\ -\frac{۱}{۲} & ۱ & -\frac{۱}{۲} \end{pmatrix}$$

۳-۲-۵ روش حذفی گاوس با محورگیری کامل

در روش حذفی گاوس با محورگیری کامل، در گام k ام، جستجو برای عنصر محوری در میان همه عناصر زیر ماتریس واقع در زیر $(k-1)$ سطر اول انجام می‌گیرد. بنابراین اگر عنصر محوری a_{kk} باشد، برای آوردن این عنصر محوری به مکان (k, k) ، باید تعویض سطرها k و k قبل از تعویض ستونهای k و s انجام شود. این معادل با پیش ضرب ماتریس $A^{(k-1)}$ در یک ماتریس جایگشت P_k به دست آمده از تعویض سطرها k و k و پس ضرب ماتریس $P_k A^{(k-1)}$ در ماتریس جایگشت دیگری مانند Q_k است که از تعویض ستونهای k و s ماتریس واحد I حاصل می‌گردد. سپس روش حذفی گاوس معمولی در مورد ماتریس $P_k A^{(k-1)} Q_k$ اعمال می‌شود، یعنی یک ماتریس پایین مثلثی مقدماتی M_k جستجو می‌شود به قسمی که ماتریس

$$A^{(k)} = M_k P_k A^{(k-1)} Q_k$$

بر روی ستون k ام و در زیر عنصر (k, k) دارای عناصر صفر باشد. البته ماتریس M_k می‌تواند در دو گام ساده‌تر مانند قبل محاسبه شود.

در انتهای گام $(n-1)$ ام، ماتریس $A^{(n-1)}$ یک ماتریس بالا مثلثی است.

قرار می‌دهیم

$$A^{(n-1)} = U$$

سپس

$$\begin{aligned} U &= A^{(n-1)} = M_{n-1} P_{n-1} A^{(n-2)} Q_{n-1} \\ &= M_{n-1} P_{n-1} M_{n-2} P_{n-2} A^{(n-3)} Q_{n-2} Q_{n-1} \\ &= \dots = M_{n-1} P_{n-1} M_{n-2} P_{n-2} \dots M_1 P_1 A Q_1 Q_2 \dots Q_{n-1} \end{aligned}$$

قرار می‌دهیم

$$M_{n-1} P_{n-1} \dots M_1 P_1 = M \quad (۶-۲-۵)$$

$$Q_1 \dots Q_{n-1} = Q \quad (۷-۲-۵)$$

سپس داریم

$$U = MAQ$$

قضیه ۳-۲-۵ قضیه تجزیه با محورگیری کامل. با مفروض بودن یک ماتریس $n \times n$ مانند A ، روش حذفی گاوس با محورگیری کامل یک ماتریس بالا مثلثی U ، یک ماتریس پایین مثلثی

جایگشت یافته M ، و یک ماتریس جایگشت Q به قسمی به دست می‌دهد که

$$MAQ = U$$

که در آن U ، M ، و Q توسط روابط $(5-2-5)$ تا $(7-2-5)$ ارائه می‌شوند.

همانند حالت محورگیری جزئی، از $(4-2-5)$ و $(7-2-5)$ به آسانی ملاحظه می‌شود که تجزیه

$MAQ = U$ می‌تواند به صورت زیر بیان شود.

$$PAQ = LU$$

نتیجه $2-2-5$ قضیه تجزیه LU با محورگیری کامل. روش حذفی گاوس با محورگیری کامل تجزیه

$PAQ = LU$ را نتیجه می‌دهد که در آن P و Q ماتریسهای جایگشت ارائه شده به صورت زیر

هستند

$$P = P_{n-1} \dots P_1$$

$$Q = Q_1 \dots Q_{n-1}$$

و L یک ماتریس پایین مثلثی واحد است که به صورت زیر ارائه می‌شود.

$$L = P(M_{n-1}P_{n-1} \dots M_1P_1)^{-1}$$

مثال $5-2-5$ ماتریس

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

را با استفاده از محورگیری کامل مثلثی کنید.

گام ۱: $k = 1$. عنصر محوری $3 = a_{33}$ است.

$$P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & \cdot & \cdot \\ \cdot & \cdot & 1 \end{bmatrix}$$

$$Q_1 = \begin{bmatrix} \cdot & \cdot & 1 \\ \cdot & 1 & \cdot \\ 1 & \cdot & \cdot \end{bmatrix}$$

$$P_1 A Q_1 = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 1 & \cdot \\ 1 & 1 & 1 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 1 & \cdot & \cdot \\ -\frac{1}{3} & 1 & \cdot \\ -\frac{1}{3} & \cdot & 1 \end{bmatrix}$$

$$A^{(1)} = M_1 P_1 A Q_1 = \begin{bmatrix} 3 & 2 & 1 \\ \cdot & \frac{1}{3} & -\frac{1}{3} \\ \cdot & \frac{1}{3} & \boxed{\frac{2}{3}} \end{bmatrix}$$

گام ۲: $k = 2$ عنصر محوری $a_{33}^{(1)} = \frac{2}{3}$ است.

$$P_2 = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & \cdot & 1 \\ \cdot & 1 & \cdot \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & \cdot & 1 \\ \cdot & 1 & \cdot \end{bmatrix}$$

$$P_2 A^{(1)} Q_2 = \begin{bmatrix} 3 & 1 & 2 \\ \cdot & \frac{2}{3} & \frac{1}{3} \\ \cdot & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

$$\hat{M}_2 = \begin{bmatrix} 1 & \cdot \\ \frac{1}{2} & 1 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \frac{1}{2} & 1 \end{bmatrix}$$

$$U = A^{(2)} = M_2 P_2 A^{(1)} Q_2 = M_2 P_2 (M P_1 A Q_1) Q_2 = \begin{bmatrix} 3 & 1 & 2 \\ \cdot & \frac{2}{3} & \frac{1}{3} \\ \cdot & \cdot & \frac{1}{3} \end{bmatrix}$$

(با استفاده از نتیجه ۵-۲-۲، ماتریسهای P, Q, L و U را برای خودتان طوری پیدا کنید که $(PAQ = LU$).

تشکیل ماتریس M و سایر جزئیات محاسباتی

تذکراهایی مشابه با تذکراهای حالت محورگیری جزئی می توان بیان کرد. ماتریسهای P_k, Q_k ، $M_k, P_k A^{(k-1)} Q_k$ ، و $M_k P_k A^{(k-1)} Q_k$ نباید به طور صریح تشکیل شوند و موجبات استفاده غیر ضروری از حافظه را فراهم سازند. فقط کافی است اندیسها و ضریبها ذخیره شوند.

با در نظر گرفتن بحث مربوط به تشکیل ماتریسهای M_k و ماتریسهای جایگشت P_k ، اکنون یک الگوریتم عملی از حذفی گاوس با محورگیری کامل ارائه می دهیم که تشکیل صریح ماتریسهای $P_k, Q_k, M_k, P_k A Q_k$ را نشان نمی دهد.

الگوریتم ۵-۲-۳ مثلثی سازی با استفاده از روش حذفی گاوس با محورگیری کامل. با مفروض بودن یک ماتریس $n \times n$ ، الگوریتم زیر مثلثی سازی A را با سطرها و ستونهای جایگشت یافته، با استفاده از روش حذفی گاوس با محورگیری کامل انجام می دهد. الگوریتم ماتریس U را بر روی A می نویسد. U در قسمت بالا مثلثی A (شامل قطر) ذخیره می شود و ضریبهای m_{ik} در قسمت پایین مثلثی A ذخیره می شوند. اندیسهای جایگشت r_k و s_k به طور مجزا ذخیره می شوند. برای $k = 1, 2, \dots, n - 1$ انجام دهید.

۱: r_k و s_k را به قسمی پیدا کنید $|a_{ij}| = \max\{|a_{ij}| : i, j \geq k\}$ ، $|a_{r_k s_k}| = \max\{|a_{ij}| : i, j \geq k\}$ ، r_k و s_k را ذخیره کنید. اگر $a_{r_k s_k} = 0$ ، آنگاه متوقف شوید، در غیر این صورت ادامه دهید.

۲: (سطرهای r_k و k را تعویض کنید) $a_{kj} \leftrightarrow a_{r_k j}$ ($j = k, k + 1, \dots, n$)

۳: (ستونهای s_k و k را تعویض کنید) $a_{ik} \leftrightarrow a_{i, s_k}$ ($i = 1, 2, \dots, n$)

۴: (ضریبها را تشکیل دهید) $a_{ik} \equiv m_{ik} = -\frac{a_{ik}}{a_{kk}}$ ($i = k + 1, \dots, n$)

۵: (عناصر را به هنگام درآوردن) $a_{ij} \equiv a_{ij} + m_{ik} a_{kj} = a_{ij} + a_{ik} a_{kj}$ ($i = k + 1, \dots, n; j = k + 1, \dots, n$)

تذکر: الگوریتم ۵-۲-۳ ماتریسهای P, M, Q و L را به طور صریح ارائه نمی دهد؛ این ماتریسها باید به ترتیب از ضریبهای m_{ik} و اندیسهای جایگشت r_k و s_k ، به صورت توصیف شده در فوق تشکیل شوند. همانند قبل، الگوریتم می تواند به حالت $m \times n$ توسعه یابد.

شمارش اعمال متمیز شناور (flops). الگوریتم به $n^3/3$ عمل متمیز شناور (flops) و $n^3/3$ مقایسه نیاز دارد.

تذکر MATCOM : الگوریتم ۵-۲-۳ در برنامه COMPIV در MATCOM پیاده سازی شده است. ماتریسهای M, U, Q به طور صریح محاسبه می شوند.

مثال ۵-۲-۶ فرض کنید

$$A = \begin{bmatrix} 1 & 2 \\ 3 & \boxed{4} \end{bmatrix}$$

فقط یک گام مورد نیاز است.

$k = 1$: عنصر محوری ۴ است.

$$r_1 = 2$$

$$s_1 = 2$$

ابتدا سطرهای دوم و اول تعویض می شوند؛ سپس ستونهای دوم و اول عوض می شوند، عنصر محوری ۴ در مکان (۱, ۱) قرار می گیرد:

$$A \equiv \begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix}$$

(بعد از تعویض سطرهای اول و دوم)

$$A \equiv \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix}$$

(بعد از تعویض ستونهای اول و دوم). ضریب عبارت است از $-\frac{1}{2} = -\frac{1}{2} = -\frac{1}{2}$ $a_{21} \equiv m_{21} = -a_{21}/a_{11} = -\frac{1}{2}$

$$A \equiv \begin{bmatrix} 4 & 3 \\ 0 & -\frac{1}{2} \end{bmatrix}$$

(بعد از به هنگام در آوردن A)

$$P_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Q = Q_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$M = M_1 P_1 = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -\frac{1}{2} \end{bmatrix}$$

۳-۵ پایداری روش حذفی گاوس

پایداری الگوریتمهای حذفی گاوس با اندازه گیری رشد عناصر در ماتریس کاهش یافته $A^{(k)}$ بهتر درک می شود. (توجه کنید که اگرچه محورگیری ضربگرها را توسط واحد کراندار نگه می دارد، ولیکن عناصر ماتریسهای کاهش یافته باز هم می توانند به طور دلخواه رشد کنند). ما تعریف عامل رشد را که در فصل ۳ ارائه شد، در این متن برای خواننده مجدداً یادآور می شویم.

تعریف ۱-۳-۵ عامل رشد ρ برابر نسبت بزرگترین عنصر (از لحاظ اندازه) از $A^{(1)}$ ، A ، ... ، $A^{(n-1)}$ به بزرگترین عنصر (از لحاظ اندازه) از A است:

$$\rho = \frac{\max(\alpha, \alpha_1, \alpha_2, \dots, \alpha_{n-1})}{\alpha}$$

$$\alpha_k = \max_{ij} |a_{ij}^{(k)}| \quad \text{و} \quad \alpha = \max_{ij} |a_{ij}|$$

که در آن

مثال ۱-۳-۵ فرض کنید

$$A = \begin{bmatrix} 0.0001 & 1 \\ 1 & 1 \end{bmatrix}$$

(الف) روش حذفی گاوس بدون محورگیری ارائه می دهد

$$A^{(1)} = U \cong \begin{bmatrix} 0.0001 & 1 \\ . & -1.04 \end{bmatrix}$$

$$\max |a_{ij}^{(1)}| = 1.04$$

$$\max |a_{ij}| = 1$$

$$\rho = \text{عامل رشد} = 1.04$$

(ب) روش حذفی گاوس با محورگیری جزئی نتیجه می دهد

$$A^{(1)} U = U \cong \begin{bmatrix} 1 & 1 \\ . & 1 \end{bmatrix}$$

$$\max |a_{ij}^{(1)}| = 1$$

$$\max |a_{ij}| = 1$$

$$\rho = \text{عامل رشد} = 1$$

طبیعتاً این سؤال مطرح می‌شود: تا چه حد عامل رشد ρ می‌تواند برای یک ماتریس دلخواه بزرگ باشد؟ به این سؤال در زیر پاسخ می‌دهیم.

عامل رشد حذفی گاوس برای محورگیری کامل

برای روش حذفی گاوس با محورگیری کامل داریم:

$$\rho \leq \{n \cdot 2^{1/2} \cdot 3^{1/2} \cdot 4^{1/2} \dots n^{1/(n-1)}\}^{1/2}$$

این یک تابع رشد کند از n است. بعلاوه در عمل این کران هرگز به دست نمی‌آید. در واقع یک حدس ثابت نشده از ویلکینسون ((Wilkinson (1965,213) وجود داشت که برای محورگیری کامل عامل رشد برای ماتریسهای حقیقی $n \times n$ توسط n کراندار می‌شد. متأسفانه، اخیراً این ادعا توسط گولد (Gould (1991)) با ارائه یک ماتریس 13×13 که برای آن روش حذفی گاوس با محورگیری کامل عامل رشد $\rho = 13/0.205$ را ارائه داده تکذیب شد. علی‌رغم نتیجه اخیر گولد، روش حذفی گاوس با محورگیری کامل یک الگوریتم پایدار است.

عامل رشد روش حذفی گاوس برای محورگیری جزئی

برای روش حذفی گاوس با محورگیری جزئی داریم $\rho \leq 2^{n-1}$ ، یعنی،

ρ می‌تواند به بزرگی 2^{n-1} باشد.

متأسفانه، ماتریسهایی که برای آنها این کران به دست می‌آید، می‌توانند ساخته شوند. مثال زیر را ملاحظه کنید:

$$A = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 1 \\ -1 & 1 & 0 & \dots & 0 & 1 \\ \vdots & \ddots & \ddots & & \vdots & \vdots \\ \vdots & & & \ddots & \vdots & \vdots \\ \vdots & & & & \ddots & \vdots \\ -1 & \dots & \dots & \dots & -1 & 1 \end{bmatrix}$$

یعنی:

$$a_{ij} = \begin{cases} 1 & \text{به ازای } n = j \\ -1 & \text{به ازای } i < j \\ 0 & \text{در غیر این صورت} \end{cases}$$

ویلکنسون (Wilkinson (1965, 212)) نشان داد که عامل رشد ρ برای این ماتریس با محورگیری جزئی برابر 2^{n-1} است. برای دیدن آن حالت خاص $n = 4$ را در نظر بگیرید.

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$$

$$A^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & -1 & 1 & 2 \\ 0 & -1 & -1 & 2 \end{bmatrix}$$

$$A^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & -1 & 4 \end{bmatrix}$$

$$A^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 8 \end{bmatrix}$$

بنابراین عامل رشد برابر است با

$$\rho = \frac{\Lambda}{1} = 2^3 = 2^4 - 1$$

تبصره: توجه کنید که این تنها ماتریسی نیست که برای آن $\rho = 2^{n-1}$ می باشد. هایام (Higham (1989)) یک مجموعه ماتریس شناسایی کرده که برای آنها $\rho = 2^{n-1}$ می باشد. ماتریس

$$B = \begin{bmatrix} 0./7248 & 0./7510 & 0./5241 & 0./7510 \\ 0./7317 & 0./1889 & 0./0227 & -0./7510 \\ 0./7298 & -0./3756 & 0./1150 & 0./7511 \\ -0./6993 & -0./7444 & 0./6647 & -0./7500 \end{bmatrix}$$

یک چنین ماتریسی است. اخیراً فوستر (Foster (1994)) و رایب (Wright (1993)) مثالهایی ارائه داده‌اند که برای آنها روش حذفی گاوس با محورگیری جزئی می‌تواند عاملهای رشد بزرگی ارائه دهد. مثالهایی از نوع قبل نادر هستند. در حقیقت برای بسیاری از مثالهای عددی عناصر ماتریسهای $A^{(k)}$ اغلب به کاهش اندازه ادامه می‌دهند. بنابراین روش حذفی گاوس با محورگیری جزئی از لحاظ نظری بدون قید و شرط پایدار نیست، ولیکن در عمل می‌تواند یک الگوریتم پایدار در نظر گرفته شود.

عامل رشد و پایداری روش حذفی گاوس بدون محورگیری

برای روش حذفی گاوس بدون محورگیری، ρ می‌تواند به طور دلخواه بزرگ باشد، بجز برای تعدادی حالت خاص، نظیر ماتریسهای معین مثبت که بعداً ملاحظه خواهیم کرد. بنابراین روش حذفی گاوس بدون محورگیری، در حالت کلی یک الگوریتم کاملاً ناپایدار است.

۴-۵ تبدیلات هاوس هولدر و کاربردها برای تجزیه QR و کاهش هسبرگی

تعریف ۴-۵-۱ یک ماتریس به شکل

$$H = I - \frac{\gamma uu^T}{u^T u}$$

که در آن u یک بردار مخالف صفر است، به افتخار متخصص آنالیز عددی مشهور آمریکایی آلستون هاوس هولدر (Alston Householder (1904-1993))، یک ماتریس هاوس هولدر نامیده می‌شود. یک ماتریس هاوس هولدر همچنین یک بازتابندهٔ مقدماتی یا یک تبدیل هاوس هولدر نیز نامیده می‌شود. شکل ۴-۵-۱ یک تعبیر هندسی از یک تبدیل هاوس هولدر را ارائه می‌دهد.

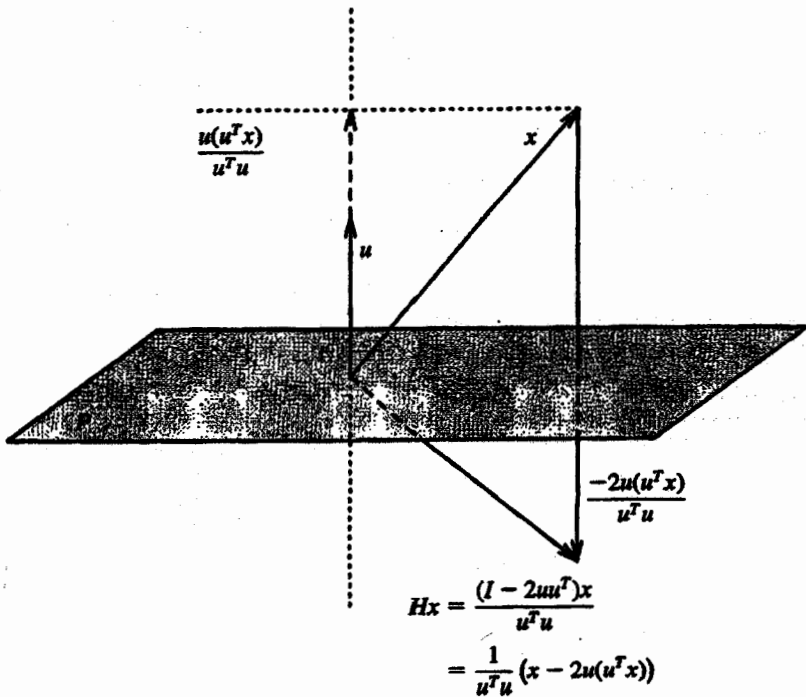
با این تعبیر هندسی، نتایج زیر روشن می‌گردند.

- $\|Hx\|_2 = \|x\|_2$ به ازای هر بردار $x \in \mathbb{R}^n$. یک بازتاب طول بردار را عوض نمی‌کند.
- H یک ماتریس متعامد است. $\|Hx\|_2 = \|x\|_2$ به ازای هر x ایجاب می‌کند که H متعامد باشد.
- $H^2 = I$. Hx بردار را به طرف دیگر P منعکس می‌کند، اما $H^2x = H(Hx) = x$ آن را به x باز می‌گرداند.
- $Hy = y$ به ازای هر بردار $y \in P$. بردارهای P نمی‌توانند به بیرون منعکس شوند.

• H یک مقدار ویژه -1 و $(n-1)$ تا مقدار ویژه 1 دارد. $\{v \in \mathbb{R}^n : v^T u = 0\}$ دارای $(n-1)$ بردار مستقل خطی y_1, \dots, y_{n-1} است و $Hy_i = y_i$ ، $i = 1, \dots, n-1$. بنابراین یک مقدار ویژه $(n-1)$ تایی است. همچنین H ، u را به $-u$ منعکس می‌کند؛ یعنی $Hu = -u$. بنابراین -1 یک مقدار ویژه H است که باید یک مقدار ویژه ساده H باشد زیرا H می‌تواند فقط n مقدار ویژه داشته باشد.

$$\det(H) = (-1) \cdot 1 \dots 1 = -1 \quad \bullet$$

با مفروض بودن $x, y \in \mathbb{R}^n$ با $\|x\|_2 = \|y\|_2$ ؛ اگر بردار u را برابر یک بردار واحد موازی با $x-y$ انتخاب کنیم، آنگاه $H = I - 2uu^T$ بردار x را به y منعکس می‌نماید. اهمیت ماتریسهای هاوس هولدر در این حقیقت نهفته است که آنها می‌توانند برای ایجاد صفر در یک بردار نیز استفاده شوند.



شکل ۱-۵

لم ۱-۴-۵ با مفروض بودن یک بردار مخالف صفر $e_1 \neq x$ یک ماتریس هاوس هولدر H وجود دارد به قسمی که Hx مضربی از e_1 باشد.

اثبات: تعریف می‌کنیم

$$H = I - \frac{\gamma uu^T}{u^T u}$$

با $e_1 \neq x + \text{sign}(x_1) \|x\|$ ، سپس به آسانی ملاحظه می‌شود که Hx مضربی از e_1 است. □

تذکرها:

۱. اگر x_1 صفر باشد علامت آن می‌تواند، + یا - انتخاب شود. از هر امکان سرریز یا پاریز در

محاسبه $\|x\|$ می‌توان با مقیاس کردن بردار x پرهیز کرد. بنابراین بردار u باید از

بردار $\{x/\max_i\{|x_i|\}\}$ به جای خود بردار x تعیین شود.

۲. چون $(x_1 + \text{sign}(x_1) \|x\|, x_2, \dots, x_n)^T = u = x + \text{sign}(x_1) \|x\| e_1$ ، بردار u می‌تواند

بر روی x ذخیره شود.

۳. همچنین $Hx = (\text{sign}(x_1) \|x\|, 0, \dots, 0)^T$.

الگوریتم ۱-۴-۵ ایجاد صفر در یک بردار با یک ماتریس هاوس هولدر. با مفروض بودن یک بردار

مخالف صفر x ، الگوریتم زیر یک بردار u و یک اسکالر σ به قسمی محاسبه می‌کند که

$$Hx = (I - \frac{\gamma uu^T}{u^T u})x = (\sigma, 0, \dots, 0)^T$$

بر روی x ذخیره شود.

Step1: $m = \max_i\{|x_i|\}, i = 1, 2, \dots, n$

Step2: $x_i \equiv u_i = \frac{x_i}{m}, i = 1, 2, \dots, n$

Step3: $\sigma = \text{sign}(u_1) \sqrt{u_1^2 + u_2^2 + \dots + u_n^2}$

Step4: $x_1 \equiv u_1 = u_1 + \sigma$

step5: $\sigma = -m \cdot \sigma$

مثال ۱-۴-۵ با مفروض بودن

$$x = \begin{bmatrix} 0 \\ 4 \\ 1 \end{bmatrix}$$

u را به قسمی پیدا کنید که $Hx = (\sigma, 0, 0)^T$ ، که در آن $H = I - \frac{\gamma uu^T}{u^T u}$

$$m = 4, \quad u_1 = 1/0.308$$

$$u_2 = 1, \quad u_3 = 0/25$$

بنابراین $u = (1/0.308, 1, 0/25)^T$ ، $\sigma = -4 \times 1/0.308 = -4/1232$ ، $Hx = (-4/1232, 0, 0)^T$ به سادگی تحقیق می شود که

$$Hx = (-4/1232, 0, 0)^T$$

شمارش اعمال ممیز شناور (*flops*) و خواص گرد کردن. ایجاد صفر در یک بردار توسط یک ماتریس هاوس هولدر یک رویه ساده و پایدار عددی است. فقط به $2(n+1)$ عمل ممیز شناور (*flops*) و یک ریشه دوم نیاز دارد تا در مکانهای ۲ تا n یک بردار صفر ایجاد نماید. می توان نشان داد (Wilkinson (1965, 152-162)) که اگر \hat{H} ماتریس هاوس هولدر محاسبه شده باشد آنگاه

$$\|H - \hat{H}\| \leq 10\mu$$

بعلاوه اگر حاصلضرب داخلی با دقت مضاعف انجام شود، آنگاه

$$\text{fl}(\hat{H}x) = H(x + e)$$

که در آن

$$\|e\| \leq c\mu \|x\|_2$$

c یک ثابت از مرتبه واحد و μ دقت ماشین است.

تذکر **MATCOM**: الگوریتم ۴-۵-۱ در برنامه **HOUSZERO** در **MATCOM** پیاده سازی شده است.

اکنون نشان خواهیم داد که چگونه حاصلضرب ماتریس - بردار و ماتریس - ماتریس با یک ماتریس هاوس هولدر می توانند فقط با استفاده از بردار تعریف کننده ماتریس هاوس هولدر تشکیل شوند (الگوریتم ۴-۲-۱ را ملاحظه کنید). خود ماتریس هاوس هولدر در عمل نباید تشکیل شود.

محاسبه حاصلضرب ماتریس - بردار با یک ماتریس هاوس هولدر

فرض کنید x یک بردار n بعدی و H یک ماتریس هاوس هولدر $n \times n$ و تعریف شده توسط بردار u باشد. آنگاه

$$Hx = \left(I - \frac{2uu^T}{u^T u} \right) x = x - \beta u (u^T x)$$

$$\beta = \frac{2}{u^T u} \text{ که در آن}$$

الگوریتم ۲-۴-۵ حاصلضرب ماتریس - بردار با یک ماتریس هاوس هولدر. با مفروض بودن یک بردار n بعدی u که ماتریس هاوس هولدر $H = I - \frac{2uu^T}{u^T u}$ را تعریف می‌کند، و یک بردار n بعدی $x = (x_1, x_2, \dots, x_n)^T$ ، الگوریتم زیر حاصلضرب ماتریس - بردار Hx را محاسبه می‌نماید. الگوریتم بر روی x بردار Hx را می‌نویسد.

Step1: Compute $\beta = \frac{2}{u^T u}$

Step2: Compute the sum $s = \sum_{i=1}^n u_i x_i$

Step3: Update $\beta \equiv \beta s$

Step4: For $i = 1, \dots, n$ do

$$x_i \equiv x_i - \beta u_i$$

تبصره: توجه کنید که الگوریتم ۲-۴-۵ یک حالت خاص از الگوریتم ۱-۲-۴ است. در اینجا ما آن را جهت راحتی دوباره تکرار کرده‌ایم.

شمارش اعمال ممیز شناور (flops). الگوریتم فقط به $O(n)$ عمل ممیز شناور (flops) نیاز دارد، در صورتی که برای ضرب معمولی ماتریس - بردار به $O(n^2)$ عمل نیاز داریم.

محاسبه حاصلضرب ماتریس - ماتریس با پیش ضرب در یک ماتریس هاوس هولدر الگوریتم ۱-۲-۴ چگونگی محاسبه حاصلضرب Hx را بدون تشکیل صریح ماتریس هاوس هولدر H توصیف می‌کند. در اینجا توجه می‌کنیم که این حاصلضرب می‌تواند همچنین با الگوریتم ۲-۴-۵ محاسبه شود. فرض کنید $A = (a_1, a_2, \dots, a_n)$ پس $HA = H(a_1, a_2, \dots, a_n) = (Ha_1, Ha_2, \dots, Ha_n)$ اکنون هر ستون Ha_i ، $i = 1, 2, \dots, n$ ، می‌تواند از روی u با استفاده از الگوریتم ۲-۴-۵ محاسبه شود.

محاسبه حاصلضرب ماتریس - ماتریس با پس ضرب در یک ماتریس هاوس هولدر

فرض کنید

$$A = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix}$$

پس

$$AH = \begin{bmatrix} a_1^T H \\ a_2^T H \\ \vdots \\ a_n^T H \end{bmatrix}$$

چون $(a_i^T H)^H = H^T a_i = Ha_i$ ، $i = 1, 2, \dots, n$ ، هر ستون HA می‌تواند توسط حاصلضرب ماتریس - بردار با ماتریس هاوس هولدر H به صورت نشان داده شده در الگوریتم ۱-۲-۴ یا الگوریتم ۲-۴-۵ محاسبه شود.

تذکر: در اینجا تاکید می‌کنیم که در عمل حاصلضرب ماتریس - ماتریس یا ماتریس - بردار با یک ماتریس هاوس هولدر باید به صورت نشان داده شده در فوق، بدون تشکیل صریح ماتریس هاوس هولدر H ، محاسبه شود.

۱-۴-۵ ماتریسهای هاوس هولدر و تجزیه QR

قضیه ۱-۴-۵ قضیه تجزیه QR هاوس هولدر. با مفروض بودن یک ماتریس $n \times n$ مانند A ، یک ماتریس متعامد Q و یک ماتریس بالا مثلثی R وجود دارند به قسمی که

$$A = QR$$

ماتریس Q می‌تواند به صورت $Q = H_1 H_2 \dots H_{n-1}$ نوشته شود، که در آن H_i یک ماتریس هاوس هولدر است.

یک تجزیه از A به صورت QR یک تجزیه QR از A نامیده می‌شود.

همچنان که بعد ملاحظه خواهیم کرد، تجزیه QR نقش خیلی مهمی در حل عددی دستگاههای خطی، مسائل کمترین توانهای دوم، و محاسبه مقادیر ویژه و مقادیر تکین دارد.

اکنون نشان می‌دهیم که چگونه تجزیه QR ماتریس A می‌تواند با استفاده از ماتریسهای هاوس هولدر به دست آید و این یک اثبات ساختنی از قضیه ۱-۴-۵ را فراهم خواهد ساخت.

همانند فرایند تجزیه LU ، این کار می‌تواند در $(n - 1)$ گام انجام شود؛ ولیکن بر خلاف فرایند حذفی گاوس، فرایند هاوس هولدر می‌تواند همواره تا کامل شدن ادامه یابد.

گام ۱: یک ماتریس H_1 به قسمی بسازید که $H_1 A$ در زیر عنصر $(1, 1)$ در ستون اول دارای عناصر صفر باشد:

$$H_1 A = \begin{pmatrix} * & * & \dots & * \\ \cdot & * & \dots & * \\ \cdot & * & \dots & * \\ \vdots & \vdots & & \vdots \\ \cdot & * & \dots & * \end{pmatrix}$$

توجه کنید که کافی است $H_1 = I - \gamma u_n u_n^T / (u_n^T u_n)$ طوری ساخته شود که

$$H_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} = \begin{pmatrix} * \\ \cdot \\ \vdots \\ \cdot \end{pmatrix}$$

سپس $H_1 A$ دارای شکل فوق خواهد بود.

برای استفاده در گام بعد بر روی A ماتریس $H_1 A = A^{(1)}$ را بنویسید. چون $A^{(1)}$ بر روی A نوشته می‌شود، $A^{(1)}$ می‌تواند به صورت زیر نوشته شود.

$$A \equiv A^{(1)} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \cdot & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ \cdot & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

گام ۲: یک ماتریس هاوس هولدر H_2 بسازید به قسمی که $H_2 A^{(1)}$ در زیر عنصر $(2, 2)$ در ستون دوم دارای صفر باشد و صفرهای قبلاً ایجاد شده در ستون اول $A^{(1)}$ در گام ۱ از بین نروند:

$$A^{(2)} \equiv H_2 A^{(1)} = \begin{pmatrix} * & * & * & \dots & * \\ \cdot & * & * & \dots & * \\ \cdot & \cdot & * & \dots & * \\ \vdots & \vdots & \vdots & & \vdots \\ \cdot & \cdot & * & \dots & * \end{pmatrix}$$

H_2 می‌تواند به صورت زیر ساخته شود. ابتدا ماتریس هاوس

هولدر $\hat{H}_\gamma = I_{n-1} - \gamma u_{n-1} u_{n-1}^T / (u_{n-1}^T u_{n-1})$ از مرتبه $n - 1$ به قسمی بسازید که

$$\hat{H}_\gamma \begin{bmatrix} a_{\gamma\gamma} \\ a_{\gamma\gamma} \\ \vdots \\ \vdots \\ a_{n\gamma} \end{bmatrix} = \begin{bmatrix} * \\ \cdot \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

و سپس تعریف کنید.

$$H_\gamma = \begin{bmatrix} 1 & \cdot & \dots & \cdot \\ \cdot & & & \\ \vdots & & \hat{H}_\gamma & \\ \cdot & & & \end{bmatrix}$$

و آنگاه $A^{(\gamma)} = H_\gamma A^{(1)}$ دارای شکلی که قبلاً ارائه شده خواهد بود. بر روی A ، $A^{(\gamma)}$ را بنویسید. تذکر: چون H_γ هم در زیر قطر بر روی ستون اول دارای صفر است، پیش ضرب $A^{(1)}$ در H_γ صفهای قبلاً ایجاد شده در گام ۱ را حفظ می‌کند.

گام k : به طور کلی، در گام k ام، ابتدا یک ماتریس هاوس هولدر

$$\hat{H}_k = I_{n-k+1} - \frac{\gamma u_{n-k+1} u_{n-k+1}^T}{(u_{n-k+1}^T u_{n-k+1})}$$

از مرتبه $n - k + 1$ بسازید به قسمی که

$$\hat{H}_k \begin{bmatrix} a_{kk} \\ \vdots \\ \vdots \\ a_{nk} \end{bmatrix} = \begin{bmatrix} * \\ \cdot \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

و سپس با تعریف

$$H_k = \begin{bmatrix} I_{k-1} & \cdot \\ \cdot & \hat{H}_k \end{bmatrix}$$

$A^{(k)} = H_k A^{(k-1)}$ را محاسبه کنید.

$A^{(k)}$ را بر روی A بنویسید. ماتریس $A^{(k)}$ بر روی ستون k ام در زیر عنصر (k, k) صفر خواهد داشت، و صفهای قبلاً ایجاد شده در ستونهای قبلی از بین نخواهند رفت. در انتهای گام $(n - 1)$ ام ماتریس $A^{(n-1)}$ حاصل یک ماتریس بالا مثلثی R خواهد بود.

چون

$$A^{(k)} = H_k A^{(k-1)}, \quad k = n-1, \dots, 2$$

داریم

$$R = A^{(n-1)} = H_{n-1} A^{(n-2)} = H_{n-1} H_{n-2} A^{(n-3)} \quad (1-4-5)$$

$$= \dots = H_{n-1} H_{n-2} \dots H_2 H_1 A$$

قرار دهید

$$Q^T = H_{n-1} H_{n-2} \dots H_2 H_1 \quad (2-4-5)$$

چون هر ماتریس هاوس هولدر H_k متعامد است، در نتیجه Q^T نیز چنین است. بنابراین

از (۱-۴-۵) داریم.

$$R = Q^T A \quad \text{یا} \quad A = QR \quad (3-4-5)$$

(توجه کنید که $Q^T = H_1^T H_2^T \dots H_n^T$ نیز متعامد است.)

مثال ۲-۴-۵ فرض کنید

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

گام ۱: ۱

ساختن H_1 :

$$H_1 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} * \\ 0 \\ 0 \end{bmatrix}$$

$$u_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + \sqrt{2} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 1 \\ 1 \end{bmatrix}$$

$$H_1 = I_3 - \frac{u_1 u_1^T}{u_1^T u_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

تشکیل $A^{(1)}$:

$$A^{(1)} = H_2 A = \begin{bmatrix} -\sqrt{2} & \frac{-3\sqrt{2}}{2} & 2\sqrt{2} \\ 0 & \frac{1-\sqrt{2}}{2} & \frac{2-\sqrt{2}}{2} \\ 0 & -\frac{(1+\sqrt{2})}{2} & -\frac{(2+\sqrt{2})}{2} \end{bmatrix}$$

نوشتن $A^{(1)}$ بر روی A

$$A \equiv A^{(1)} \equiv \begin{bmatrix} -1/4142 & -2/1213 & -2/8284 \\ 0 & -0/2071 & 0/2929 \\ 0 & -1/2071 & -1/7071 \end{bmatrix}$$

گام $2: k = 2$.

ساختن \hat{H}_2 :

$$\hat{H}_2 \begin{bmatrix} -0/2071 \\ -1/2071 \end{bmatrix} = \begin{bmatrix} * \\ 0 \end{bmatrix}$$

$$u_2 = \begin{bmatrix} -0/2071 \\ -1/2071 \end{bmatrix} - 1/2247 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/4318 \\ -1/2071 \end{bmatrix}$$

$$\hat{H}_2 = \begin{bmatrix} -0/1691 & -0/9856 \\ -0/9856 & 0/1691 \end{bmatrix}$$

ساختن H_2 :

$$H_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0/1691 & -0/9856 \\ 0 & -0/9856 & 0/1691 \end{bmatrix}$$

تشکیل $A^{(2)}$:

$$A^{(2)} = H_3 A^{(1)} = H_3 H_1 A$$

$$= \begin{bmatrix} -1/4142 & -2/1213 & -2/8284 \\ \cdot & 1/2242 & 1/6330 \\ \cdot & \cdot & -0/5774 \end{bmatrix} = R$$

تشکیل Q :

$$Q = H_3 H_1 = \begin{bmatrix} \cdot & 0/8165 & 0/5774 \\ -0/7071 & 0/4082 & -0/5774 \\ -0/7071 & -0/4082 & 0/5774 \end{bmatrix}$$

تشکیل ماتریس Q و سایر جزئیات محاسباتی

۱. هر ماتریس هاوس هولدر H_k به طور منحصر به فرد توسط بردار u_{n-k+1} که \hat{H}_k را تعریف می‌کند، تعیین می‌شود. بنابراین کافی است بردار u_{n-k+1} ذخیره شود. اگر ماتریس Q به طور صریح لازم باشد، می‌تواند از ماتریسهای هاوس هولدر H_1 تا H_{n-1} محاسبه شود.
۲. بردار $u_{n-k+1} = (u_{nk}, \dots, u_{kk})^T$ دارای $(n - k + 1)$ مؤلفه است، در حالی که فقط $(n - k)$ صفر در گام k ام تولید می‌شود. بنابراین یک طرح ممکن برای ذخیره سازی این است که مؤلفه‌های دوم تا $(n - k + 1)$ ام در مکانهای $(k + 1, k)$ ، \dots ، (n, k) از A قرار گیرند. اولین مؤلفه باید به‌طور مجزا در یک آرایه یک بعدی ذخیره شود.
۳. ماتریس $A^{(k)} = H_k A^{(k-1)}$ می‌تواند از روی بردار u_{n-k+1} که H_k را تعریف می‌کند، بدون تشکیل ماتریس H_k محاسبه شود. برای توضیح آن ماتریس

$$A^{(2)} = H_3 A^{(1)}$$

$$= \begin{bmatrix} 1 & \cdot & \dots & \cdot \\ \cdot & & & \\ \vdots & I - \gamma \frac{u_{n-1} u_{n-1}^T}{u_{n-1}^T u_{n-1}} & & \\ \cdot & & & \end{bmatrix} \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & a_{1n}^{(1)} \\ \cdot & & & \\ \vdots & A_1^{(1)} & & \\ \cdot & & & \end{bmatrix}$$

را ملاحظه کنید.

همان طور که مشاهده می‌کنیم اولین سطر $A^{(2)}$ همان اولین سطر $A^{(1)}$ است، عناصر ستون اول $A^{(2)}$ و زیر عنصر قطری همگی برابر صفرند، و سایر عناصر $A^{(2)}$ توسط حاصلضرب ماتریسی

$$\begin{bmatrix} I - \gamma \frac{u_{n-1} u_{n-1}^T}{u_{n-1}^T u_{n-1}} \end{bmatrix} A_1^{(1)}$$

ارائه می‌شوند که می‌توانند فقط توسط ذخیره سازی بردار u_{n-1} به صورت نشان داده شده در الگوریتم ۴-۲-۱، محاسبه شوند (بخش ۵-۴ نیز ملاحظه شود). حالت کلی به طور مشابه می‌تواند

انجام شود. به طور کلی برای محاسبه $A^{(k)}$ از $A^{(k-1)}$ به طریق زیر می‌توان عمل کرد:

(الف) $(k-1)$ سطر اول را بدون تغییر باقی گذارید.

(ب) $(n-k+1)$ سطر بعدی را توسط تشکیل حاصلضرب

$$\begin{bmatrix} I - \gamma \frac{u_{n-k+1} u_{n-k+1}^T}{u_{n-k+1}^T u_{n-k+1}} \end{bmatrix} A_{k-1}^{(k-1)}$$

که در آن $A_{k-1}^{(k-1)}$ زیر ماتریس حاصل از $A^{(k-1)}$ با حذف $(k-1)$ سطر اول و $(k-1)$ ستون

اول می‌باشد، به هنگام در آوری. برای این منظور از الگوریتم ۴-۲-۱ که نشان

می‌دهد حاصلضرب می‌تواند فقط با ذخیره سازی بردار u_{n-k+1} محاسبه شود، استفاده کنید

۴. عناصر مخالف صفر $A^{(k)}$ می‌توانند در مکانهای متناظر از A به محض تشکیل ذخیره شوند.

۵. با طرح ذخیره سازی فوق، در انتهای گام $(n-1)$ ام داریم:

$$A \equiv A^{(n-1)} = \begin{bmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ u_{21} & a_{22} & \dots & \dots & a_{2n} \\ u_{31} & u_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ u_{n1} & u_{n2} & \dots & u_{n,n-1} & a_{nn} \end{bmatrix}$$

آرایه یک بعدی v حاوی مؤلفه‌های اول u_{n-k+1} به صورت زیر ارائه می‌شود.

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-2} \end{bmatrix} \equiv \begin{bmatrix} u_{11} \\ u_{22} \\ \vdots \\ u_{n-1,n-1} \end{bmatrix}$$

بنابراین، همه اطلاعات لازم برای محاسبه Q در قسمت پایین مثلثی ماتریس A و در زیر قطر و در آرایه تک بعدی v که حاوی مؤلفه‌های اول بردارهای تعریف کننده ماتریسهای H_1 تا H_{n-1} است، قرار دارند. ماتریس Q اگر لازم باشد، می‌تواند از روی بردارهای تعریف کننده ماتریسهای H_1, H_2, \dots, H_{n-1} محاسبه شود.

الگوریتم ۳-۴-۵ تجزیه QR هاوس هولدر. با مفروض بودن یک ماتریس $n \times n$ مانند A ، الگوریتم زیر بردارهای $(u_{kk}, \dots, u_{nk})^T = u_{n-k+1}$ ، $k = 1, \dots, n-1$ ، تعریف کننده‌های ماتریسهای هاوس هولدر H_1 تا H_{n-1} ، و ماتریس بالا مثلثی R را طوری به وجود می‌آورد که $A = QR$ با $Q = H_1 H_2 \dots H_{n-1}$. مؤلفه‌های $u_{k+1, k}$ تا u_{nk} در مکانهای $(k+1, k)$ تا (n, k) از A ذخیره می‌شوند. مؤلفه‌های اول u_{kk} در یک آرایه تک بعدی $v = (v_1, v_2, \dots, v_n)^T$ ذخیره می‌شوند. عناصر مخالف صفر $A^{(k)} = H_k A^{(k-1)}$ ، $k = 1, \dots, n-1$ ، بر روی عناصر نظیرشان از A نوشته می‌شوند؛ یعنی R بر روی قسمت بالا مثلثی A نوشته می‌شود.

برای $k = 1, 2, \dots, n-1$ انجام دهید

۱. بردار $u_{n-k+1} = (u_{k, n-k+1}, \dots, u_{n, n-k+1})^T$ تعریف کننده ماتریس \hat{H}_k و اسکالر σ را به قسمی پیدا کنید که:

$$\hat{H}_k \begin{bmatrix} a_{kk} \\ \vdots \\ a_{nk} \end{bmatrix} = \begin{bmatrix} \sigma \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

(از الگوریتم ۱-۴-۵ استفاده کنید).

۲. σ را بر روی a_{kk} بنویسید:

$$a_{kk} \equiv \sigma$$

۳. بردار u_{n-k+1} را به صورت زیر ذخیره کنید:

$$a_{ik} \equiv u_{ik}, \quad i = k+1, \dots, n$$

$$v_k \equiv u_{kk}$$

۴. محاسبه کنید

$$\beta = \frac{v_k}{u_{n-k+1}^T u_{n-k+1}}$$

۵. عناصر زیر ماتریس A را که شامل سطرهای k تا n و ستونهای $k+1$ تا n می‌باشند، به

هنگام در آورید. برای $k = 1, \dots, n$ z انجام دهید

$$1. s = \beta \sum_{i=k}^n u_{ik} a_{ij}$$

$$2. a_{ij} \equiv a_{ij} - s u_{ik} \quad (i = k, k+1, \dots, n)$$

مثال ۳-۴-۵ تجزیه QR ماتریس زیر را با استفاده از روش هاوس هولدر پیدا کنید (الگوریتم ۳-۴-۵).

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ 3 & 8 & 9 \end{bmatrix}$$

$k = 1$: با استفاده از الگوریتم ۳-۴-۵، بردار $u_1 = (u_{11}, u_{21}, u_{31})^T$ را طوری پیدا می‌کنیم که

$$\hat{H}_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & - & \\ & & - \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \sigma \\ 0 \\ 0 \end{bmatrix}$$

$$u_1 = (1/58.06, 0/66.67, 1)^T$$

$$\sigma = -3/74.17$$

اکنون عناصر A را به هنگام درمی‌آوریم:

$$A \equiv H_1 A = \begin{bmatrix} -3/74.17 & -9/62.14 & -11/75.95 \\ 0 & 0.982 & 0.7745 \\ 0 & 0.6473 & -0.3382 \end{bmatrix}$$

توجه کنید که فقط لازم است عناصر $a_{13}, a_{12}, a_{23}, a_{22}, a_{21}, a_{11}$ ؛ و a_{33}, a_{32}, a_{31} را به هنگام درآوریم، زیرا $\sigma = -3/74.17, a_{11} \equiv 0, a_{21} \equiv 0, a_{31} \equiv 0$. بعلاوه توجه کنید که $\hat{H}_1 = H_1$.

$k = 2$: با استفاده از الگوریتم ۳-۴-۵، بردار $u_2 = (u_{22}, u_{32})^T$ را به قسمی پیدا می‌کنیم که

$$\hat{H}_2 \begin{bmatrix} 0.982 \\ 0.6473 \end{bmatrix} = \begin{bmatrix} \sigma \\ 0 \end{bmatrix}$$

$$u_2 = (1/163.1, 1)^T, \sigma = -0/65.47$$

اکنون فقط عناصر a_{33} و a_{32} را به هنگام درمی‌آوریم. توجه کنید که $a_{33} = \sigma = -0/65.47$ و

$a_{۳۳} = ۰$ ؛ بقیه عناصر بدون تغییر باقی می‌مانند.

$$a_{۲۳} = ۰/۲۱۸۲, \quad a_{۳۳} = -۰/۸۱۶۵$$

بنابراین

$$A \equiv H_2 H_1, \quad A = R = \begin{bmatrix} -۳/۷۴۱۷ & -۹/۶۲۱۴ & -۱۱/۷۵۹۵ \\ \cdot & -۰/۶۵۴۷ & ۰/۲۱۸۲ \\ \cdot & \cdot & -۰/۸۱۶۵ \end{bmatrix}$$

تشکیل Q : Q می‌تواند از روی بردارهای u_2 و u_3 تشکیل شود.

$$Q^T = H_2 H_1 = \begin{bmatrix} ۱ & \cdot & \cdot \\ \cdot & I - \frac{\gamma u_2 u_2^T}{u_2^T u_2} & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \gamma u_2 u_2^T \\ I - \frac{\gamma u_2 u_2^T}{u_2^T u_2} \\ \cdot \end{bmatrix}$$

$$= \begin{bmatrix} -۰/۲۶۷۳ & ۰/۸۷۲۶ & ۰/۴۰۸۲ \\ -۰/۵۳۴۵ & ۰/۲۱۸۲ & -۰/۸۱۶۵ \\ -۰/۸۰۸۱ & -۰/۴۳۶۴ & ۰/۴۰۸۲ \end{bmatrix}$$

(اکنون تحقیق کنید که $Q^T A = R$.)

شمارش اعمال ممیز شناور. الگوریتم تقریباً به $\frac{2}{3}n^3$ عمل ممیز شناور (flops) فقط برای محاسبه نیاز دارد. این امر می‌تواند به صورت زیر ملاحظه شود.

ساختن \hat{H}_k (و در نتیجه H_k) حدوداً به $2(n-k)$ عمل ممیز شناور (flops) نیاز دارد، در صورتی که $A^{(k)} = H_k A$ (با در نظر گرفتن مزیت ساختار خاص H_k) تقریباً به $2(n-k)^2$ عمل ممیز شناور (flops) نیاز دارد. بنابراین

$$\begin{aligned} \text{تعداد کل اعمال ممیز شناور} &= 2 \sum_{k=1}^{n-1} [(n-k)^2 + (n-k)] \\ &= 2[(n-1)^2 + (n-2)^2 + \dots + 1^2] + 2[(n-1) \\ &\quad + (n-2) + \dots + 1] \\ &= 2 \frac{n(n-1)(2n-1)}{6} + 2 \frac{n(n-1)}{2} \\ &\approx \frac{2n^3}{3} \end{aligned}$$

تذکر: در شمارش قبلی ساختن صریح Q به حساب آورده نشده است. Q فقط به شکل حاصلضرب موجود است. باید توجه داشت که در بیشتر کاربردهای عملی، کافی است Q را به شکل حاصلضرب در اختیار داشته باشیم؛ در بسیاری از کاربردها Q ابدأ لازم نیست. اگر Q به طور صریح لازم باشد، $\sum_{p=1}^n n^3$ عمل ممیز شناور (flops) دیگر لازم خواهد بود (تمرین ۲۲).

خاصیت گرد کردن. در صورت وجود خطاهای گرد کردن، الگوریتم تجزیه QR یک ماتریس کمی اختلال یافته را محاسبه می‌کند. بویژه، می‌توان نشان داد (Wilkinson (1965, 236)) که اگر ماتریس محاسبه شده R توسط \tilde{R} نمایش داده شود، آنگاه یک ماتریس متعامد \tilde{Q} وجود دارد به قسمی که

$$A + E = \tilde{Q}\tilde{R}$$

ماتریس خطای E در رابطه زیر صدق می‌کند

$$\|E\|_F \leq \phi(n)\mu \|A\|_F$$

که در آن $\phi(n)$ یک تابع رشد آهسته از n و μ دقت ماشین است. اگر حاصلضربهای داخلی در دقت مضاعف انجام شوند، آنگاه می‌توان نشان داد (Golub and Wilkinson (1966)) که $\phi(n) = 12/5 n$ بنابراین الگوریتم پایدار است.

تذکر MATCOM: الگوریتم ۳-۴-۵ در برنامه HOURSQR در MATCOM پیاده‌سازی شده است. Q و R به طور صریح محاسبه می‌شوند.

۲-۴-۵ تجزیه QR هاوس هولدر یک ماتریس نامربعی

در بسیاری از کاربردها (نظیر مسائل کمترین توانهای دوم)، تجزیه QR یک ماتریس $m \times n$ مانند A مورد نیاز است. روش هاوس هولدر قبلی می‌تواند برای به دست آوردن تجزیه QR یک چنین ماتریس A بی نیز به کار رود. فرایند از $s = \min\{n, m - 1\}$ گام تشکیل می‌شود؛ ماتریسهای هاوس هولدر H_1, H_2, \dots, H_s به طور متوالی ساخته می‌شوند؛ در نتیجه

$$H_s H_{s-1} \dots H_2 H_1 A = Q^T A = \begin{cases} \begin{bmatrix} R_1 \\ \cdot \\ \cdot \end{bmatrix}, & \text{اگر } m > n \\ (R_1, S), & \text{اگر } m \leq n \end{cases}$$

شمارش اعمال ممیز شناور (flops) و خاصیت گرد کردن. روش هاوس هولدر در این حالت به اعمال زیر نیاز دارد

$$۱. \quad n^2(m - \frac{n}{3}) \text{ عمل ممیز شناور (flops) اگر } m \geq n$$

$$۲. \quad m^2(n - \frac{m}{3}) \text{ عمل ممیز شناور (flops) اگر } m \leq n$$

خاصیت گرد کردن همانند حالت قبل است. تجزیه QR یک ماتریس مستطیلی با استفاده از تبدیلهای هاوس هولدر پایدار است.

تذکر MATCOM : در برنامه HOUSQRN در MATCOM این روش را برای تجزیه QR هاوس هولدر یک ماتریس نامربعی پیاده‌سازی شده است.

مثال ۴-۴-۵ فرض کنید

$$A = \begin{bmatrix} 1 & 1 \\ \cdot/\cdot\cdot\cdot 1 & \cdot \\ \cdot & \cdot/\cdot\cdot\cdot 1 \end{bmatrix}$$

$$s = \min\{2, 2\} = 2$$

گام ۱: تشکیل H_1 .

$$u_2 = \begin{bmatrix} 1 \\ \cdot/\cdot\cdot\cdot 1 \\ \cdot \end{bmatrix} + \sqrt{1 + (\cdot/\cdot\cdot\cdot 1)^2} \begin{bmatrix} 1 \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} 2 \\ \cdot/\cdot\cdot\cdot 1 \\ \cdot \end{bmatrix}$$

$$H_1 = I - \frac{u_2 u_2^T}{u_2^T u_2} = \begin{bmatrix} -1 & -\cdot/\cdot\cdot\cdot 1 & \cdot \\ -\cdot/\cdot\cdot\cdot 1 & 1 & \cdot \\ \cdot & \cdot & 1 \end{bmatrix}$$

$$A^{(1)} = H_1 A = \begin{bmatrix} -1 & -1 \\ \cdot & -\cdot/\cdot\cdot\cdot 1 \\ \cdot & \cdot/\cdot\cdot\cdot 1 \end{bmatrix}$$

(در عمل $H_1 A$ باید به طور ضمنی از روی u_2 تشکیل شود)

گام ۲: تشکیل H_2 .

$$u_1 = \begin{bmatrix} -\cdot/\cdot\cdot\cdot 1 \\ \cdot/\cdot\cdot\cdot 1 \end{bmatrix} - \sqrt{(-\cdot/\cdot\cdot\cdot 1)^2 + (\cdot/\cdot\cdot\cdot 1)^2} \begin{bmatrix} 1 \\ \cdot \end{bmatrix}$$

$$= 1.0^{-2} \begin{bmatrix} -2/\sqrt{141} \\ \cdot/\cdot\cdot\cdot 1 \end{bmatrix}$$

$$\hat{H}_\gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \gamma \frac{u_1 u_1^T}{u_1^T u_1} = \begin{bmatrix} -\gamma/\gamma_0 \gamma_1 & \gamma/\gamma_0 \gamma_1 \\ \gamma/\gamma_0 \gamma_1 & \gamma/\gamma_0 \gamma_1 \end{bmatrix}$$

$$H_\gamma = \begin{bmatrix} \cdot & -\gamma/\gamma_0 \gamma_1 & \gamma/\gamma_0 \gamma_1 \\ \cdot & \gamma/\gamma_0 \gamma_1 & \gamma/\gamma_0 \gamma_1 \end{bmatrix}$$

تشکیل R :

$$H_\gamma A^{(1)} = \begin{bmatrix} -1 & -1 \\ \cdot & \cdot/\dots 1 \\ \cdot & \cdot \end{bmatrix} = \begin{bmatrix} R_1 \\ \cdot \end{bmatrix}$$

(در عمل $H_\gamma A^{(1)}$ باید به طور ضمنی از روی u_1 تشکیل شود).

تشکیل Q و R_1 :

$$Q = H_1 H_\gamma = \begin{bmatrix} -1 & \cdot/\dots 1 & -\gamma/\dots 1 \\ -\gamma/\dots 1 & -\gamma/\gamma_0 \gamma_1 & \gamma/\gamma_0 \gamma_1 \\ \cdot & \gamma/\gamma_0 \gamma_1 & \gamma/\gamma_0 \gamma_1 \end{bmatrix}$$

$$R_1 = \begin{bmatrix} -1 & -1 \\ \cdot & \cdot/\dots 1 \end{bmatrix}$$

۳-۴-۵ ماتریسهای هاوس هولدر و کاهش به شکل هسنبرگی

قضیه ۴-۵-۲ قضیه کاهش هسنبرگی. یک ماتریس $n \times n$ دلخواه می تواند همواره با تبدیلات تشابهی متعامد زیر به یک ماتریس بالا هسنبرگی تبدیل شود.

$$PAPT = H_u$$

همان گونه که قبلاً متذکر شدیم، کاهش به یک شکل هسنبرگی در محاسبات مقادیر ویژه خیلی اهمیت دارد. به طور معمول قبل از فرایند محاسبات مقادیر ویژه (معروف به تکرارهای QR) ماتریس A به یک ماتریس بالا هسنبرگی تبدیل می شود. در کاربردهای بسیار دیگری نظیر نظریه کنترل و پردازش سیگنال نیز شکلهای هسنبرگی ابزار مفیدی هستند.

ایده تجزیه متعامد با استفاده از ماتریسهای هاوس هولدر توصیف شده در بخش قبل می تواند به سادگی برای به دست آوردن P و H_u توسعه یابد.

ماتریس P به صورت حاصلضرب $(n-2)$ ماتریس هاوس هولدر P_1 تا P_{n-2} ساخته می شود. P_1 برای به وجود آوردن صفر در ستون اول A و در زیر عنصر $(2, 1)$ ساخته می شود. P_2 برای

به وجود آوردن صفر در زیر عنصر $(۲, ۲)$ ستون دوم ماتریس $P_۱AP_۱^T$ تعیین می‌شود، و غیره. فرایند از $(n - ۲)$ گام تشکیل می‌شود. (توجه کنید که یک ماتریس هسنبرگی $n \times n$ حاوی حداقل $(n - ۲)(n - ۱)/۲$ صفر است.)

گام ۱: یک ماتریس هاوس هولدر $\hat{P}_۱$ از مرتبه $n - ۱$ طوری پیدا کنید که

$$\hat{P}_۱ \begin{bmatrix} a_{۲۱} \\ a_{۳۱} \\ \vdots \\ a_{n۱} \end{bmatrix} = \begin{bmatrix} * \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

تعریف کنید

$$P_۱ = \begin{bmatrix} I_۱ & \cdot \\ \cdot & \hat{P}_۱ \end{bmatrix}$$

و محاسبه کنید

$$A^{(۱)} = P_۱AP_۱^T$$

بر روی A ماتریس $A^{(۱)}$ را بنویسید. پس

$$A \equiv A^{(۱)} = \begin{bmatrix} * & * & \dots & * \\ * & * & \dots & * \\ \cdot & * & \dots & * \\ \vdots & \vdots & \dots & \vdots \\ \cdot & * & \dots & * \end{bmatrix}$$

گام ۲: یک ماتریس هاوس هولدر $\hat{P}_۲$ از مرتبه $(n - ۲)$ طوری پیدا کنید که

$$\hat{P}_۲ \begin{bmatrix} a_{۳۲} \\ \vdots \\ a_{n۲} \end{bmatrix} = \begin{bmatrix} * \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

تعریف کنید

$$P_۲ = \begin{bmatrix} I_۲ & \cdot \\ \cdot & \hat{P}_۲ \end{bmatrix}$$

و $A^{(۲)} = P_۲A^{(۱)}P_۲^T$ را حساب کنید.

بر روی A ماتریس $A^{(2)}$ را بنویسید. پس

$$A \equiv A^{(2)} = \begin{bmatrix} * & * & * & \dots & * \\ * & * & * & \dots & * \\ \cdot & * & * & \dots & * \\ \cdot & \cdot & * & \dots & * \\ \vdots & \vdots & \vdots & & \vdots \\ \cdot & \cdot & * & \dots & * \end{bmatrix}$$

اکنون گام کلی k می تواند به سادگی نوشته شود.

در انتهای $(n - 2)$ گام، ماتریس $A^{(n-2)}$ یک ماتریس بالا هسنبرگی H_u است. همان طور که در زیر ملاحظه می کنید H_u به طور متعامد مشابه با A است:

$$\begin{aligned} H_u &= A^{(n-2)} = P_{n-2} A^{(n-2)} P_{n-2}^T \\ &= P_{n-2} (P_{n-2} A^{(n-2)} P_{n-2}^T) P_{n-2}^T \\ &\vdots \\ &= (P_{n-2} P_{n-3} \dots P_1) A (P_1^T P_1^T \dots P_{n-2}^T P_{n-2}^T) \end{aligned} \quad (4-4-5)$$

قرار دهید

$$P = P_{n-2} P_{n-3} \dots P_1 \quad (5-4-5)$$

سپس داریم $H_u = P A P^T$. چون هر ماتریس هاوس هولدر P_i متعامد است، ماتریس P که حاصلضرب $(n - 2)$ ماتریس هاوس هولدر می باشد نیز متعامد است.

تذکره: مهم است توجه داشته باشیم به این که چون P_k دارای شکل

$$\begin{bmatrix} I_k & \cdot \\ \cdot & \hat{P}_k \end{bmatrix}$$

است، پس ضرب $P_k A$ در P_k^T صفرهای قبلاً به وجود آمده در $P_k A$ را از بین نمی برد. برای مثال فرض کنید $n = 4$ و $k = 1$. پس

$$P_1 = \begin{bmatrix} 1 & \circ & \circ & \circ \\ \circ & * & * & * \\ \circ & * & * & * \\ \circ & * & * & * \end{bmatrix}$$

$$P_1 A = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ \circ & * & * & * \\ \circ & * & * & * \end{bmatrix}$$

$$P_1 A P_1^T = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ \circ & * & * & * \\ \circ & * & * & * \end{bmatrix} \begin{bmatrix} 1 & \circ & \circ & \circ \\ \circ & * & * & * \\ \circ & * & * & * \\ \circ & * & * & * \end{bmatrix} = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ \circ & * & * & * \\ \circ & * & * & * \end{bmatrix}$$

مثال ۵-۴-۵ فرض کنید

$$A = \begin{bmatrix} \circ & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

چون $n = 3$ ، ما فقط باید یک گام انجام دهیم. \hat{P}_1 را تشکیل می‌دهیم:

$$\hat{P}_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} * \\ \circ \end{bmatrix}$$

$$u_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \sqrt{2} e_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \sqrt{2} \begin{bmatrix} 1 \\ \circ \end{bmatrix} = \begin{bmatrix} 1 + \sqrt{2} \\ 1 \end{bmatrix}$$

$$\hat{P}_1 = I_2 - 2 \frac{u_2 u_2^T}{u_2^T u_2} = \begin{bmatrix} 1 & \circ \\ \circ & 1 \end{bmatrix} - \frac{2}{2+2\sqrt{2}} \begin{bmatrix} 5/2+2\sqrt{2} & 2/2+2\sqrt{2} \\ 2/2+2\sqrt{2} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -\circ/\sqrt{2} & -\circ/\sqrt{2} \\ -\circ/\sqrt{2} & \circ/\sqrt{2} \end{bmatrix}$$

P_1 نیز به شکل زیر خواهد بود:

$$P_1 = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & & \\ \cdot & \hat{P}_1 & \cdot \end{bmatrix} = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & -\cdot/\sqrt{0.71} & -\cdot/\sqrt{0.71} \\ \cdot & -\cdot/\sqrt{0.71} & \cdot/\sqrt{0.71} \end{bmatrix}$$

$$A \equiv A^{(1)} = P_1 A P_1^T = \begin{bmatrix} \cdot & -2/1213 & \cdot/\sqrt{0.71} \\ -1/4142 & 3/5000 & -\cdot/5000 \\ \cdot & 1/5000 & -\cdot/5000 \end{bmatrix} = H_u$$

همه محاسبات با استفاده از چهار رقم حسابی انجام شده‌اند.

تشکیل ماتریس P و سایر جزئیات محاسباتی

- هر ماتریس هاوس هولدر P_k به طور منحصر به فرد توسط u_{n-k} که ماتریس \hat{P}_k را تعریف می‌کند، تعیین می‌شود. بنابراین کافی است بردار u_{n-k} را ذخیره کنیم. اگر ماتریس P به طور صریح موردنیاز باشد، می‌توان آن را از روی ماتریسهای هاوس هولدر P_1 تا P_{n-2} حساب کرد.
- بردار $u_{n-k} = (u_{k+1,k}, \dots, u_{n,k})^T$ دارای $n-k$ مؤلفه است، در حالی که تعداد صفرهای تولید شده در گام k ام برابر $(n-k+1)$ می‌باشد. مؤلفه‌های دوم تا $(n-k)$ ام از u_{n-k} می‌توانند در مکانهای $(k+2, k)$ تا (n, k) از A ذخیره شوند، و اولین مؤلفه به طور مجزا ذخیره شود.
- ماتریس $A^{(k)} = P_k A^{(k-1)} P_k^T$ می‌تواند بدون تشکیل صریح P_k محاسبه شود. برای دیدن آن، ماتریس زیر را ملاحظه کنید.

$$A^{(k)} = P_k A^{(k-1)} P_k^T = \begin{bmatrix} I_k & \cdot \\ \cdot & \hat{P}_k \end{bmatrix} \begin{bmatrix} B & C \\ \cdot & b & D \end{bmatrix} \begin{bmatrix} I_k & \cdot \\ \cdot & \hat{P}_k^T \end{bmatrix}$$

در اینجا B یک ماتریس هسنبرگی $k \times k$ ، \hat{P}_k ماتریس هاوس هولدر از مرتبه $(n-k)$ ، و b یک بردار است. بردار $\hat{P}_k b$ ضربی از e_1 است. ماتریس $\hat{P}_k D (\hat{P}_k)^T$ می‌تواند از روی بردار تعریف‌کننده \hat{P}_k ، به صورت تعریف شده در قبل (بخش ۴-۵) محاسبه شود.

- عناصر غیر صفر $A^{(k)}$ می‌توانند بر روی A به محض تشکیل ذخیره شوند. بنابراین در انتهای گام $(n-2)$ ام داریم:

$$A \equiv A^{(n-2)} = \begin{bmatrix} a_{11} & a_{12} & \dots & \dots & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \dots & \dots & a_{2n} \\ u_{31} & a_{32} & a_{33} & \dots & \dots & a_{3n} \\ \vdots & u_{42} & \ddots & \ddots & \dots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \dots \\ u_{n1} & u_{n2} & \dots & u_{n,n-2} & a_{n,n-1} & a_{nn} \end{bmatrix}$$

آرایه یک بعدی v حاوی مؤلفه‌های اول بردارهای u_{n-k} به صورت زیر ارائه می‌شود

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_{n-2} \end{bmatrix} \equiv \begin{bmatrix} u_{21} \\ u_{32} \\ u_{43} \\ \vdots \\ u_{n-1, n-2} \end{bmatrix}$$

بنابراین همه اطلاعات لازم برای محاسبه P در قسمت پایین مثلثی A در زیر قطر و در آرایه تک بعدی $v = (v_1, v_2, \dots, v_{n-2})^T$ که حاوی عناصر اول بردارهای تعریف کننده P_1 تا P_{n-2} می‌باشد، قرار دارند.

الگوریتم ۴-۴-۵ کاهش هسنبرگی هاوس هولدر. با مفروض بودن یک ماتریس $n \times n$ مانند A ، الگوریتم زیر بردارهای $u_{n-k} = (u_{k+1, k}, \dots, u_{nk})^T$ ، $k = 1, \dots, n-1$ ، را که ماتریسهای هاوس هولدر \hat{P}_k ، $k = 1, 2, \dots, n-2$ (و در نتیجه P_k) را تعریف می‌کنند، به قسمی تعیین می‌کند که با $P = P_{n-2}P_{n-3} \dots P_2P_1$ ماتریس $H_u = PAP^T$ یک ماتریس بالا هسنبرگی باشد. مؤلفه‌های $u_{k+2, k}$ تا u_{nk} از بردار u_{n-k} در مکانهای متناظر $(k+2, k)$ تا (n, k) از A ذخیره می‌شوند. مؤلفه‌های اول $u_{k+1, k}$ در آرایه مجزا $v = (v_1, v_2, \dots, v_{n-2})^T$ ذخیره می‌شوند. عناصر مخالف صفر $A^{(k)} = P_k A P_k^T$ بر روی عناصر A نوشته می‌شوند. برای $k = 1, 2, \dots, n-2$ مراحل زیر را انجام دهید.

۱: بردار $u_{n-k} = (u_{k+1, k}, \dots, u_{nk})^T$ که ماتریس هاوس هولدر

$$\hat{P}_k = I_{n-k} - \gamma \frac{u_{n-k} u_{n-k}^T}{u_{n-k}^T u_{n-k}}$$

مرتبه $(n-k)$ را تعریف می‌کند و یک اسکالر σ را طوری تعیین کنید که

$$\hat{P}_k = \begin{bmatrix} a_{k+1,k} \\ \vdots \\ \vdots \\ a_{nk} \end{bmatrix} = \begin{bmatrix} \sigma \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

۲: از الگوریتم ۵-۴-۱ استفاده کنید.

۳: σ را بر روی $a_{k+1,k}$ بنویسید

$$a_{k+1,k} \equiv \sigma$$

۴: بردار $u_{n-k} = (u_{k+1,k}, \dots, u_{n,k})^T$ را به صورت زیر ذخیره کنید

$$v_k \equiv u_{k+1,k}$$

$$a_{k+i,k} \equiv u_{k+i,k}, \quad i = 2, \dots, n-k$$

۵: محاسبه کنید

$$\beta = \frac{\gamma}{u_{n-k}^T u_{n-k}}$$

۵: عناصر A واقع در ستونهای $k+1$ تا n و سطرهای $k+1$ تا n را با پیش ضرب به هنگام درآوردن، و سپس عناصر واقع در ستونهای $k+1$ تا n و سطرهای 1 تا n را با پس ضرب به هنگام درآوردن.

پیش ضرب:

For $j = k+1, \dots, n$ do

$$1. \quad s = \beta \sum_{i=k+1}^n u_{ik} a_{ij}$$

$$2. \quad a_{ij} \equiv a_{ij} - s u_{ik} (i = k+1, \dots, n)$$

پس ضرب:

For $i = k+1, \dots, n$ do

$$1. \quad s = \beta \sum_{j=k+1}^n a_{ij} u_{jk}$$

$$2. \quad a_{ij} \equiv a_{ij} - s u_{jk} (j = k+1, \dots, n)$$

مثال ۵-۴-۶ ماتریس

$$A = \begin{bmatrix} 1 & 2 & 5 \\ 3 & 7 & 9 \\ 2 & 5 & 3 \end{bmatrix}$$

را با استفاده از الگوریتم ۵-۴-۴ به یک ماتریس بالا هسنبرگی H_u تبدیل کنید و P را بیابید.

فقط یک گام باید انجام شود. ابتدا بردار $u_p = (u_{p1}, u_{p2})^T$ را به قسمی پیدا می‌کنیم

که

$$\hat{P}_1 \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \left[I - \frac{2u_p u_p^T}{u_p^T u_p} \right] \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} \sigma \end{bmatrix}$$

$$u_p = (2/20.19, 0/6667)^T$$

$$\sigma = -3/60.56$$

اکنون عناصر A را به هنگام درمی‌آوریم، توجه کنید که، $a_{21} = \sigma = -3/60.56$ و $a_{31} = 0$ و

a_{11} بدون تغییر باقی می‌ماند.

$$a_{12} \equiv -4/4376,$$

$$a_{13} \equiv 3/0.5096,$$

$$a_{22} \equiv 12/2308,$$

$$a_{23} \equiv -2/8462,$$

$$a_{32} \equiv -1/1528,$$

$$a_{33} \equiv -2/2308.$$

بنابراین

$$A \equiv H_u = P_1 A P_1^T = \begin{bmatrix} 1/0.00 & -4/4376 & 3/0.509 \\ -3/60.56 & 12/2308 & -2/8462 \\ 0 & 1/1528 & -2/2308 \end{bmatrix}$$

P_1 را تشکیل می‌دهیم:

$$\hat{P}_1 = I - \frac{2u_p u_p^T}{u_p^T u_p} = \begin{bmatrix} -0/8321 & -0/5547 \\ -0/5547 & 0/8321 \end{bmatrix}$$

بنابراین

$$P_1 = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & -./۸۳۲۱ & -./۵۵۴۷ \\ \cdot & -./۵۵۴۷ & ./۸۳۲۱ \end{bmatrix}$$

شمارش اعمال ممیز شناور (flops). الگوریتم به $\frac{5}{3}n^3$ عمل ممیز شناور (flops) برای محاسبه H_u نیاز دارد. این تعداد شامل محاسبه صریح P نمی‌شود. P می‌تواند به شکل حاصلضرب ذخیره شود. اگر P به طور صریح محاسبه شود، $\frac{4}{3}n^3$ عمل ممیز شناور (flops) دیگر مورد نیاز خواهد بود. ولیکن هنگامی که n بزرگ است حافظه لازم برای تشکیل P مشکل‌زا است.

خاصیت گرد کردن. الگوریتم پایدار است. می‌توان نشان داد ((Wilkinson (1965, 351) که H_u محاسبه شده به طور متعامد مشابه با یک ماتریس نزدیک $A + E$ است که در آن

$$\|E\|_F \leq cn^2\mu \|A\|_F$$

در اینجا c یک ثابت از مرتبه واحد است.

اگر امکان ذخیره‌سازی حاصلضربهای داخلی در دقت مضاعف وجود داشته باشد، آنگاه در کران قبلی به جای n^2 می‌توان n را قرار داد، در نتیجه در این حالت داریم:

$$\|E\|_F \leq cn\mu \|A\|_F$$

که خیلی مطلوب است.

تذکر **MATCOM** : الگوریتم ۴-۴-۵ در برنامه HOUSHESS در MATCOM پیاده‌سازی شده است. ماتریس P به طور صریح محاسبه شده است.

کاهش سه قطری

اگر ماتریس A متقارن باشد، آنگاه از

$$PAPT^T = H_u$$

بلافاصله نتیجه می‌شود که ماتریس بالا هسنبرگی H_u نیز متقارن است و در نتیجه سه قطری است. بنابراین اگر الگوریتم در مورد یک ماتریس متقارن A به کار برده شود، ماتریس حاصل H_u یک ماتریس سه قطری T خواهد بود. بعلاوه به وضوح می‌توان از مزیت متقارن بودن A برای تغییر الگوریتم استفاده

کرد. برای مثال یک صرفه‌جویی قابل ملاحظه می‌تواند در ذخیره‌سازی با استفاده از متقارن بودن هر ماتریس $A^{(k)}$ ، انجام پذیرد.

در مقایسه با $\frac{5}{3}n^3$ عمل ممیز شناور (flops) لازم برای محاسبه H_u الگوریتم متقارن فقط به $\frac{1}{3}n^3$ عمل ممیز شناور (flops) برای محاسبه T نیاز دارد. خاصیت گرد کردن اساساً همانند الگوریتم نامتقارن است. الگوریتم پایدار است.
مثال ۷-۴-۵ فرض کنید

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

چون $n = 3$ ، فقط یک گام باید انجام شود. \hat{P}_1 را تشکیل می‌دهیم:

$$\hat{P}_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} * \\ * \\ * \end{bmatrix}$$

$$u_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \sqrt{2}e_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \sqrt{2} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 + \sqrt{2} \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{aligned} \hat{P}_1 &= I_3 - \frac{2u_1u_1^T}{u_1^Tu_1} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 0.7929 \begin{bmatrix} 5/1284 & 2/4142 \\ 2/4142 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -0.7071 & -0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \end{aligned}$$

بنابراین

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \hat{P}_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.7071 & -0.7071 \\ 0 & -0.7071 & 0.7071 \end{bmatrix}$$

و در نتیجه

$$H_u = P_1 A P_1^T = \begin{bmatrix} 0 & -1/4142 & 0 \\ -1/4142 & 2/5000 & 0.5000 \\ 0 & 0.5000 & 0.5000 \end{bmatrix}$$

(توجه کنید که H_u سه قطری متقارن است)

بنابراین وقتی یک بردار n بعدی

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

توسط یک ماتریس دوران گیونز $J(i, j, \theta)$ پیش ضرب می‌شود، فقط مؤلفه‌های i ام و j ام x تحت تأثیر قرار می‌گیرند؛ سایر مؤلفه‌ها بدون تغییر باقی می‌مانند.

توجه کنید که چون $1 = c^2 + s^2$ ، $J(i, j, \theta)^T = I$ ، بنابراین دوران $J(i, j, \theta)$ متعامد است.

اگر $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ یک بردار دو بعدی باشد، آنگاه با یک تحقیق ساده ملاحظه می‌شود که با

$$c = \frac{x_1}{\sqrt{x_1^2 + x_2^2}}, \quad s = \frac{x_2}{\sqrt{x_1^2 + x_2^2}}$$

دوران گیونز $J(1, 2, \theta) = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}$ به قسمی است که $J(1, 2, \theta)x = \begin{bmatrix} * \\ \cdot \end{bmatrix}$.

فرمول قبلی برای محاسبه c و s ممکن است موجب پاریزی یا سرریز گردد. ولیکن آرایش مجدد فرمول به صورت ساده زیر از این امکان جلوگیری به عمل می‌آورد.

الگوریتم ۵-۵-۱ محاسبه پارامترهای گیونز. الگوریتم زیر پارامترهای c و s ماتریس گیونز $J = (1, 2, \theta)$ را به قسمی تعیین می‌کند که با مفروض بودن یک بردار مخالف صفر دو بعدی x ، بردار $J(1, 2, \theta)x$ دارای یک صفر در مکان دوم باشد.

$$1. \text{ اگر } |x_2| \geq |x_1|$$

$$t = \frac{x_1}{x_2}, \quad s = \frac{1}{\sqrt{1+t^2}}, \quad c = st$$

$$2. \text{ اگر } |x_2| < |x_1|$$

$$t = \frac{x_2}{x_1}, \quad c = \frac{1}{\sqrt{1+t^2}}, \quad s = ct$$

شمارش اعمال ممیز شناور (flops). الگوریتم حدوداً به ۴ عمل ممیز شناور (flops) و یک ریشه دوم نیاز دارد.

تذکر MATCOM : الگوریتم ۱-۵-۵ در برنامه GIVZERO در MATCOM پیاده‌سازی شده است.

مثال ۱-۵-۵ فرض کنید

$$x = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

چون $|x_1| > |x_2|$ ، از $t = \frac{1}{4}$ ، $c = \sqrt{1 + \frac{1}{4}} = 2/\sqrt{5}$ ، و $s = 1/\sqrt{5}$ استفاده می‌کنیم.

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} x = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \sqrt{5} \\ 2 \\ 0 \end{pmatrix}$$

اکنون نشان می‌دهیم که چگونه پیش ضرب ماتریس A در یک ماتریس گیونز می‌تواند بدون محاسبه صریح ماتریس گیونز انجام شود. ملاحظات مشابه برای پس ضرب نیز وجود دارد.

الگوریتم ۲-۵-۵ محاسبه پیش ضرب در یک ماتریس گیونز. با مفروض بودن یک ماتریس $m \times n$ مانند A ، پارامترهای گیونز c و s ، و اندیسهای i و j ($1 \leq i \leq j \leq m$)، الگوریتم زیر بر روی ماتریس A ($J(i, j, \theta)$) را می‌نویسد.

For $k = 1, \dots, n$ do

$$a \equiv a_{ik}$$

$$b \equiv a_{jk}$$

$$a_{ik} \equiv ac + bs$$

$$a_{jk} \equiv -as + bc$$

شمارش اعمال ممیز شناور. الگوریتم حدوداً به $4n$ عمل ممیز شناور (flops) نیاز دارد.

تذکر MATCOM : الگوریتم ۲-۵-۵ در برنامه PGIVMUL در MATCOM پیاده‌سازی شده است.

صفر کردن عناصر خاص در یک بردار

دورانهای گیونز بویژه در ایجاد صفر در یک مکان خاص در یک بردار مفید است. بنابراین اگر

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_k \\ \vdots \\ x_n \end{pmatrix}$$

و اگر ما مایل باشیم که فقط x_k را صفر کنیم، ما می‌توانیم دوران $J(i, k, \theta)$ ، $(i < k)$ ، را به قسمی بسازیم که $J(i, k, \theta)x$ دارای صفر در مکان k ام باشد.

برای ساختن $J(i, k, \theta)$ ، ابتدا یک دوران گیونز $\begin{pmatrix} c & s \\ -s & c \end{pmatrix}$ به قسمی می‌سازیم که

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} x_i \\ x_k \end{pmatrix} = \begin{pmatrix} * \\ 0 \end{pmatrix}$$

سپس ماتریس $J(i, k, \theta)$ را با قرار دادن c در مکانهای (i, i) و (k, k) ، و قرار دادن s و $-s$ به ترتیب در مکانهای (i, k) و (k, i) و پر کردن بقیه ماتریس با عناصر ماتریس واحد، تشکیل می‌دهیم.

مثال ۲-۵-۵ فرض کنید

$$x = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}$$

فرض کنید می‌خواهیم یک صفر در مکان سوم ایجاد کنیم، یعنی $k = 3$ ، انتخاب می‌کنیم $i = 2$.
گام ۱: یک دوران 2×2 به قسمی می‌سازیم که

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} -1 \\ 3 \end{pmatrix} = \begin{pmatrix} * \\ 0 \end{pmatrix}, \quad c = \frac{-1}{\sqrt{10}}, \quad s = \frac{3}{\sqrt{10}}$$

گام ۲: تشکیل می‌دهیم

$$J(2, 3, \theta) = \begin{pmatrix} 1 & \cdot & \cdot \\ \cdot & \frac{-1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \cdot & \frac{-3}{\sqrt{10}} & \frac{-1}{\sqrt{10}} \end{pmatrix}$$

سپس

$$J(2, 3, \theta)x = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & \frac{-1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \cdot & \frac{-3}{\sqrt{10}} & \frac{-1}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ \cdot \\ \cdot \end{bmatrix}$$

ایجاد صفر در یک بردار بجز احتمالاً در اولین امکان

با مفروض بودن یک بردار n بعدی x ، اگر مایل باشیم همه عناصر x را بجز اولین عنصر، صفر سازیم، می‌توانیم $J(1, 2, \theta)x$ ، $J(1, 3, \theta)x^{(1)}$ ، $x^{(2)} = J(1, 4, \theta)x^{(2)}$ ، $x^{(3)}$ و غیره را بسازیم. در نتیجه با

$$P = J(1, n, \theta) \dots J(1, 3, \theta)J(1, 2, \theta)$$

Px را خواهیم داشت که ضربی از e_i است. چون هر دوران متعامد است، پس P نیز چنین است.

مثال ۳-۵-۵ فرض کنید

$$x = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

$$J(1, 2, \theta) = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & \cdot \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \cdot \\ \cdot & \cdot & 1 \end{bmatrix}$$

$$x^{(1)} = J(1, 2, \theta)x = \begin{bmatrix} \sqrt{2} \\ \cdot \\ 2 \end{bmatrix}$$

$$J(1, 3, \theta) = \begin{bmatrix} \sqrt{\frac{2}{6}} & \cdot & \frac{2}{\sqrt{6}} \\ \cdot & 1 & \cdot \\ \frac{-2}{\sqrt{6}} & \cdot & \sqrt{\frac{2}{6}} \end{bmatrix}$$

سپس

$$J(1, 3, \theta)x^{(1)} = \begin{bmatrix} \sqrt{6} \\ \cdot \\ \cdot \end{bmatrix}$$

$$P = J(1, 3, \theta)J(1, 2, \theta) = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{\sqrt{2}}{\sqrt{6}} & \frac{\sqrt{2}}{\sqrt{6}} & \frac{\sqrt{2}}{\sqrt{6}} \end{bmatrix}$$

$$Px = \begin{bmatrix} \sqrt{6} \\ \cdot \\ \cdot \end{bmatrix}$$

شمارش اعمال ممیز شناور و خاصیت گرد کردن. ایجاد صفر در یک بردار با استفاده از دورانهای گیونز حدوداً دو برابر گران‌تر از استفاده از ماتریسهای هاوس هولدر است. به طور دقیق‌تر فرایند فقط به $\frac{1}{4}$ برابر اعمال ممیز شناور هاوس هولدر نیاز دارد. فرایند به همان پایداری روش هاوس هولدر است.

ایجاد صفرها در مکانهای خاصی از یک ماتریس

ایده ایجاد صفر در مکانهای خاصی از یک بردار به طور بدیهی می‌تواند به ایجاد صفر در مکانهای خاصی از یک ماتریس نیز توسعه یابد. بنابراین اگر می‌خواهیم یک صفر در مکان (j, i) $[j > i]$ یک ماتریس A ایجاد کنیم، یک راه برای این کار ساختن دوران $J(i, j, \theta)$ بوده که فقط بر سطرهای i ام و j ام اثر می‌گذارد و به قسمی است که $J(i, j, \theta)A$ دارای یک صفر در مکان (j, i) خواهد بود. پس فرایند به صورت زیر است.

الگوریتم ۵-۵-۳ ایجاد صفر در یک مکان خاص از یک ماتریس با استفاده از دورانهای گیونز. با مفروض بودن یک ماتریس $n \times n$ مانند A ، الگوریتم زیر بر روی A ماتریس $J(i, j, \theta)A$ را به قسمی می‌نویسد؛ که این ماتریس آخر در مکان (j, i) دارای یک صفر باشد.

گام ۱: مقادیر $c = \cos(\theta)$ ، $s = \sin(\theta)$ را به قسمی پیدا کنید که

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} a_{ii} \\ a_{ji} \end{pmatrix} = \begin{pmatrix} * \\ 0 \end{pmatrix}$$

گام ۲: ماتریس $J(i, j, \theta)A$ را بر روی ماتریس A با تشکیل حاصلضرب ضمنی، به صورت نشان داده شده در الگوریتم ۵-۲-۵، بنویسید.

تبصره: توجه کنید که روشهای دیگری برای انجام این عمل نیز وجود دارند. برای مثال، می‌توانیم $J(k, j, \theta)$ اثرکننده بر سطرهای k ام و j ام را به قسمی تشکیل دهیم که $J(k, j, \theta)A$ دارای یک صفر در مکان (j, k) باشد.

تذکر MATCOM: الگوریتم ۵-۳-۵ در برنامه GIVZERO در MATCOM پیاده‌سازی شده است. مثال ۵-۴-۵ فرض کنید

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 3 & 4 \\ 4 & 5 & 6 \end{pmatrix}$$

یک صفر در مکان $(2, 1)$ ماتریس با استفاده از $J(1, 2, \theta)$ ایجاد کنید.

گام ۱: c و s را به قسمی پیدا می‌کنیم که

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} * \\ 0 \end{pmatrix}$$

$$c = \frac{1}{\sqrt{10}}, \quad s = \frac{3}{\sqrt{10}}$$

گام ۲: $J(1, 2, \theta)$ را تشکیل می‌دهیم

$$J(1, 2, \theta) = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} & 0 \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$J(1, 2, \theta)A = \begin{pmatrix} \frac{10}{\sqrt{10}} & \frac{11}{\sqrt{10}} & \frac{15}{\sqrt{10}} \\ 0 & -\frac{3}{\sqrt{10}} & -\frac{5}{\sqrt{10}} \\ 4 & 5 & 6 \end{pmatrix}$$

مثال ۵-۵-۵ فرض کنید

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 4 & 5 & 6 \end{bmatrix}$$

یک صفر در مکان (۳, ۱) با استفاده از $J(2, 3, \theta)$ ایجاد کنید.گام ۱: c و s را به قسمی پیدا می‌کنیم که

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} * \\ 0 \end{bmatrix}$$

$$c = \frac{2}{\sqrt{20}}, \quad s = \frac{4}{\sqrt{20}}$$

گام ۲: $J(2, 3, \theta)$ را تشکیل می‌دهیم

$$J(2, 3, \theta) = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & \frac{2}{\sqrt{20}} & \frac{4}{\sqrt{20}} \\ \cdot & \frac{-4}{\sqrt{20}} & \frac{2}{\sqrt{20}} \end{bmatrix}$$

$$A^{(1)} = J(2, 3, \theta)A = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & \frac{2}{\sqrt{20}} & \frac{4}{\sqrt{20}} \\ \cdot & \frac{-4}{\sqrt{20}} & \frac{2}{\sqrt{20}} \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 4 & 5 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2 & 3 \\ \sqrt{20} & \frac{26}{\sqrt{20}} & \frac{32}{\sqrt{20}} \\ \cdot & \frac{-2}{\sqrt{20}} & \frac{-4}{\sqrt{20}} \end{bmatrix}$$

۵-۵-۱ دورانهای گیونز و تجزیه QR

از بحث قبل واضح است که نظیر ماتریسهای هاوس هولدر، دورانها نیز می‌توانند برای پیدا کردن تجزیه

QR یک ماتریس استفاده شوند. ولیکن روش گیونز تقریباً دو برابر گران‌تر از روش هاوس هولدر است. با در نظر گرفتن این مطلب، به نظر می‌رسد، تجزیه QR با استفاده از دورانهای گیونز بویژه در تکرارهای QR برای محاسبات مقدار ویژه و در حل دستگاههای خطی با ماتریسهای ساخت یافته، نظیر توپلیتز مفید هستند. همچنین در محاسبات موازی بسیاری از مسائل جبرخطی مهم دورانهای گیونز ابزار مهمی به حساب می‌آیند.

در اینجا الگوریتم را برای تجزیه QR یک ماتریس $m \times n$ مانند A و با استفاده از روش گیونز ارائه می‌دهیم.

ایده اصلی دقیقاً مانند ایده اصلی هاوس هولدر است: ماتریسهای متعامد Q_1, Q_2, \dots, Q_k را با استفاده از دورانهای گیونز به قسمی محاسبه کنید که $A^{(1)} = Q_1 A$ دارای صفر در زیر عنصر $(1, 1)$ در ستون اول باشد، $A^{(2)} = Q_2 A^{(1)}$ دارای صفر در زیر عنصر $(2, 2)$ در ستون دوم باشد، و غیره. هر Q_i به صورت حاصلضرب دورانهای گیونز تولید می‌شود. یک راه برای تشکیل $\{Q_i\}$ عبارت است از:

$$Q_1 = J(1, m, \theta)J(1, m-1, \theta) \dots (J(1, 2, \theta))$$

$$Q_2 = J(2, m, \theta)J(2, m-1, \theta) \dots (J(2, 3, \theta))$$

و غیره.

فرض کنید $s = \min(n, m-1)$. آنگاه

$$R = A^{(s)} = Q_s A^{(s-1)} = Q_s Q_{s-1} A^{(s-2)} = \dots$$

$$= Q_s Q_{s-1} \dots Q_2 Q_1 A = Q^T A$$

اکنون داریم: $A = QR$ با $Q^T = Q_s Q_{s-1} \dots Q_2 Q_1$.

قضیه ۵-۵-۱ قضیه تجزیه QR گیونز. با مفروض بودن یک ماتریس $m \times n$ مانند A به تعداد

$s = \min(n, m-1)$ ماتریس متعامد Q_1, Q_2, \dots, Q_s تعریف شده توسط

$Q_i = J(i, m, \theta)J(i, m-1, \theta) \dots J(i, i+1, \theta)$ وجود دارد به قسمی که اگر

$$Q = Q_1^T Q_2^T \dots Q_s^T$$

آنگاه

$$A = QR$$

که در آن R یک ماتریس با همان مرتبه A و با عناصر صفر در زیر قطر می باشد.

تشکیل ماتریس Q و جزئیات محاسباتی دیگر

در عمل نیازی به تشکیل ماتریسهای $n \times n$ ، $J(k, l, \theta)$ و $J(k, l, \theta)A$ به طور صریح نیست. توجه کنید که $J(k, l, \theta)$ می تواند فقط با معلوم بودن c و s تعیین شود و $J(k, l, \theta)A$ به جای سطرهای k ام و l ام A ترکیب خطی آنها را قرار می دهد. بویژه سطر k ام ماتریس $J(k, l, \theta)A$ مساوی c برابر سطر k ام A به اضافه s برابر سطر l ام می باشد. به طور مشابه، سطر l ام ماتریس $J(k, l, \theta)A$ مساوی $-s$ برابر سطر k ام A به اضافه c برابر سطر l ام می باشد. اگر ماتریس متعامد Q به طور صریح لازم باشد، آنگاه می توان آن را از حاصلضرب

$$Q = Q_1^T Q_2^T \dots Q_s^T$$

به دست آورد که در آن Q_i حاصلضرب $m - i$ دوران گیونز زیر است.

$$Q_i = J(i, m, \theta)J(i, m - 1, \theta) \dots J(i, i + 1, \theta)$$

الگوریتم ۵-۵-۴ تجزیه QR با استفاده از دورانهای گیونز

الگوریتم زیر دورانهای گیونز تعریف کننده یک ماتریس دوران Q و یک ماتریس R را همانند قضیه ۵-۵-۱ محاسبه می کند. الگوریتم R را بر روی A می نویسد.

برای $k = 1, 2, \dots, \min\{n, m - 1\}$ انجام دهید.

برای $l = k + 1, \dots, m$ انجام دهید.

۱. c و s را به قسمی پیدا کنید که

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} a_{kk} \\ a_{lk} \end{bmatrix} = \begin{bmatrix} * \\ 0 \end{bmatrix}$$

۲. اندیسهای k و l و اعداد c و s را ذخیره کنید.

۳. بر روی A ماتریس $J(k, l, \theta)A$ را (با استفاده از الگوریتم ۵-۵-۲) بنویسید.

شمارش اعمال ممیز شناور. الگوریتم به $2n^2(m - n/2)$ عمل ممیز شناور (flops) نیاز دارد. در

این شمارش محاسبه Q منظور نشده است. بنابراین الگوریتم تقریباً دو برابر گرانتر از الگوریتم هاوس

هولدر برای تجزیه QR است. در اینجا فرض بر این است که $m \geq n$.

خاصیت گرد کردن. الگوریتم کاملاً پایدار است. می توان نشان داد که برای $m = n$ ماتریسهای محاسبه شده \hat{Q} و \hat{R} در رابطه

$$\hat{R} = \hat{Q}^T (A + E)$$

صدق می کنند که در آن

$$\|E\|_F \leq 6\mu n^{3/2} (1 + 6\mu)^{2n-4} \|A\|_F$$

(Wilkinson (1965, 243)).

تذکر MATCOM : الگوریتم ۵-۵-۴ در برنامه GIVQR در MATCOM پیاده سازی شده است.

۲-۵-۵ منحصر به فرد بودن در تجزیه QR

ملاحظه کرده ایم که تجزیه QR یک ماتریس A همیشه وجود دارد و این تجزیه می تواند به طریقه های مختلف به دست آید. بنابراین باید این سؤال مطرح باشد که آیا این تجزیه منحصر به فرد است. در زیر ملاحظه خواهیم کرد که اگر A نامنفرد باشد، آنگاه تجزیه ذاتاً منحصر به فرد است. بعلاوه اگر عناصر قطری ماتریسهای بالا مثلثی تجزیه های به دست آمده به دو روش مختلف مثبت باشند، آنگاه این تجزیه های QR از A دقیقاً یکی هستند. برای دیدن این مطلب، فرض کنید

$$A = Q_1 R_1 = Q_2 R_2 \quad (1-5-5)$$

چون A نامنفرد است، آنگاه R_1 و R_2 نیز نامنفرد هستند. توجه کنید که

$$\det(A) = \det(Q_1 R_1) = \det(Q_1) \det(R_1)$$

و

$$\det(A) = \det(Q_2 R_2) = \det(Q_2) \det(R_2)$$

چون Q_1 و Q_2 متعامدند درمیان آنها برابر ± 1 هستند. بنابراین $\det R_1$ و $\det R_2$ مخالف صفر هستند. از (۵-۵-۱) داریم:

$$Q_2^T Q_1 = R_2 R_1^{-1} = V$$

بنابراین $V^T V = Q_1^T Q_2 Q_2^T Q_1 = I$. بعلاوه V یک ماتریس بالا مثلثی است (زیرا R_1^{-1} و R_2 هر دو بالا مثلثی هستند)، بنابراین با برابر قرار دادن دو طرف، ملاحظه می کنیم که V باید یک ماتریس قطری با

عناصر قطری ± 1 باشد. بنابراین

$$R_2 R_1^{-1} = V = \text{diag}(d_1, d_2, \dots, d_n), \quad d_i = \pm 1, i = 1, 2, \dots, n$$

این بدان معنا است که R_1 و R_2 بجز در علامتهای سطرهایشان یکی هستند. به طور مشابه، از

$$Q_2^T Q_1 = V$$

ملاحظه می‌کنیم که Q_1 و Q_2 نیز بجز در علامتهای ستونهایشان یکی هستند. اگر عناصر قطری R_1 و R_2 مثبت باشند، آنگاه V ماتریس واحد است. بنابراین

$$R_2 = R_1$$

و

$$Q_2 = Q_1$$

این نتیجه می‌تواند به سادگی برای حالتی که در آن A یک ماتریس $m \times n$ بوده و دارای ستونهای مستقل خطی باشد، تعمیم یابد.

قضیه ۵-۵-۲ قضیه منحصربه‌فرد بودن QR . فرض کنید A یک ماتریس $m \times n$ بوده و دارای ستونهای مستقل خطی باشد. آنگاه یک ماتریس منحصربه‌فرد $m \times n$ مانند Q با ستونهای متعامد و یک ماتریس بالا مثلثی منحصربه‌فرد $n \times n$ مانند R با عناصر قطری مثبت وجود دارد به طوری که

$$A = QR$$

مثال ۵-۵-۶ ماتریس R تجزیه QR ماتریس

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

را با استفاده از دورانه‌های گیونز پیدا می‌کنیم و منحصربه‌فرد بودن این ماتریس را تحقیق می‌کنیم.

گام ۱: اعداد c و s را به قسمی پیدا می‌کنیم که

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} = \begin{bmatrix} * \\ \cdot \end{bmatrix}$$

$$a_{11} = 0, \quad a_{21} = 1$$

$$c = 0, \quad s = 1$$

$$J(1, \gamma, \theta) = \begin{bmatrix} \cdot & 1 & \cdot \\ -1 & \cdot & \cdot \\ \cdot & \cdot & 1 \end{bmatrix}$$

$$\begin{aligned} A \equiv J(1, \gamma, \theta)A &= \begin{bmatrix} \cdot & 1 & \cdot \\ -1 & \cdot & \cdot \\ \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} \cdot & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 2 & 3 \\ \cdot & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \end{aligned}$$

اعداد c و s را به قسمی پیدا می‌کنیم که

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} = \begin{bmatrix} * \\ \cdot \end{bmatrix}$$

$$a_{11} = 1, \quad c = \frac{1}{\sqrt{2}}, \quad s = \frac{1}{\sqrt{2}}$$

$$J(1, \gamma, \theta) = \begin{bmatrix} \frac{1}{\sqrt{2}} & \cdot & \frac{1}{\sqrt{2}} \\ \cdot & 1 & \cdot \\ \frac{-1}{\sqrt{2}} & \cdot & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$A \equiv J(1, \gamma, \theta)A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \cdot & \frac{1}{\sqrt{2}} \\ \cdot & 1 & \cdot \\ \frac{-1}{\sqrt{2}} & \cdot & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ \cdot & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{2} & \frac{3}{\sqrt{2}} & 2\sqrt{2} \\ \cdot & -1 & -1 \\ \cdot & \frac{-1}{\sqrt{2}} & -\sqrt{2} \end{bmatrix}$$

گام ۲: اعداد c و s را به قسمی پیدا می‌کنیم که

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} a_{22} \\ a_{32} \end{bmatrix} = \begin{bmatrix} * \\ \cdot \end{bmatrix}$$

$$a_{22} = -1, \quad a_{32} = -\frac{1}{\sqrt{2}}, \quad c = -\frac{\sqrt{2}}{\sqrt{3}}, \quad s = -\frac{1}{\sqrt{3}}$$

$$J(\gamma, \nu, \theta) A = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & \frac{-\sqrt{2}}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \\ \cdot & \frac{1}{\sqrt{3}} & -\frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{3}{\sqrt{2}} & 2\sqrt{2} \\ \cdot & -1 & -1 \\ \cdot & -\frac{1}{\sqrt{2}} & -\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{2} & \frac{3}{\sqrt{2}} & 2\sqrt{2} \\ \cdot & \frac{\sqrt{3}}{\sqrt{2}} & \frac{2\sqrt{2}}{\sqrt{3}} \\ \cdot & \cdot & \frac{1}{\sqrt{3}} \end{bmatrix}$$

$$= \begin{bmatrix} 1/4142 & 2/1213 & 2/8284 \\ \cdot & 1/2247 & 1/6330 \\ \cdot & \cdot & \cdot/5774 \end{bmatrix} = R$$

(با استفاده از محاسبات چهار رقمی)

تبصره: توجه کنید که ماتریس بالا مثلثی به دست آمده در اینجا ذاتاً همان ماتریس بالا مثلثی ارائه شده

توسط روش هاوس هولدر است (مثال ۵-۴-۲) و تنها در علامتهای سطرهای اول و سوم با آن تفاوت دارد.

۳-۵-۵ دورانهای گیونز و کاهش به شکل هسنبرگی

ماتریسهای گیونز همچنین می توانند برای تبدیل یک ماتریس $n \times n$ مانند A به یک ماتریس بالا هسنبرگی H_u توسط تبدیلهای متشابه متعامد، مورد استفاده قرار گیرند: $PAP^T = H_u$. ولیکن برای انجام آن، دورانهای گیونز باید به طریق خاصی ساخته شوند. برای مثال، در گام اول دورانهای گیونز $J(\gamma, 3, \theta)$ ، $J(\gamma, 4, \theta)$ ، ...، $J(\gamma, n, \theta)$ به طور متوالی به قسمی محاسبه می شوند که با $P_1 = J(\gamma, n, \theta) \dots J(\gamma, 4, \theta)J(\gamma, 3, \theta)$ داشته باشیم

$$P_1 A P_1^T = A^{(1)} = \begin{bmatrix} * & * & \dots & * \\ * & * & \dots & * \\ \cdot & * & \dots & * \\ \vdots & \vdots & & \vdots \\ \cdot & * & \dots & * \end{bmatrix}$$

در گام دوم، دورانهای $J(\gamma, 4, \theta)$ ، $J(\gamma, 5, \theta)$ ، ...، $J(\gamma, n, \theta)$ به طور متوالی به قسمی محاسبه می شوند که با $P_2 = J(\gamma, n, \theta)J(\gamma, n-1, \theta) \dots J(\gamma, 4, \theta)$ داشته باشیم

$$P_2 A^{(1)} P_2^T = A^{(2)} = \begin{bmatrix} * & * & * & \dots & * \\ * & * & * & \dots & * \\ \cdot & * & * & \dots & * \\ \cdot & \cdot & * & \dots & * \\ \vdots & \vdots & \vdots & & \vdots \\ \cdot & \cdot & * & \dots & * \end{bmatrix}$$

و غیره. در انتهای گام $(n-2)$ ام، ماتریس $A^{(n-2)}$ ماتریس بالا هسنبرگی H_u است. ماتریس تبدیل P به صورت

$$P = P_{n-2} P_{n-3} \dots P_2 P_1$$

است که در آن $P_i = J(i+1, n, \theta)J(i+1, n-1, \theta) \dots J(i+1, i+2, \theta)$.

تشکیل ماتریس P و جزئیات محاسباتی دیگر

تشکیل صریح $J(p+1, q, \theta)$ و $J(p+1, q, \theta)A$ لازم نیست، زیرا این ماتریسها به طور کامل توسط p, q, c و s تعیین می‌شوند (بخش واقع در انتهای الگوریتم تجزیه QR با استفاده از دورانهای گیونز را ملاحظه کنید). اگر P به طور صریح لازم باشد، می‌توان آن را با استفاده از رابطه

$$P = P_{n-2} \dots P_3 P_1$$

تشکیل داد که در آن $J(i+1, i+2, \theta) \dots J(i+1, n-1, \theta) J(i+1, n, \theta) P_i$.

الگوریتم ۵-۵-۵ کاهش هسنبرگی گیونز. با مفروض بودن یک ماتریس $n \times n$ مانند A ، الگوریتم زیر بر روی A ماتریس $H_{ii} = PAP^T$ را می‌نویسد، که در آن H_{ii} یک ماتریس بالا هسنبرگی است.

برای $p = 1, 2, \dots, n-2$ انجام دهید.

برای $q = p+2, \dots, n$ انجام دهید.

۱: $s = \sin(\theta)$ ، $c = \cos(\theta)$ را به قسمی پیدا کنید که

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} a_{p+1,p} \\ a_{q,p} \end{bmatrix} = \begin{bmatrix} * \\ 0 \end{bmatrix}$$

۲. اندیسه‌های p و q و اعداد c و s را ذخیره کنید.

۳. بر روی A ، ماتریس $J(p+1, q, \theta)AJ(p+1, q, \theta)^T$ را بنویسید.

شمارش اعمال ممیز شناور. الگوریتم حدوداً به $\frac{1}{3}n^3$ عمل ممیز شناور (flops) نیاز دارد. در صورتی که روش هاوس هولدر به $\frac{5}{3}n^3$ عمل ممیز شناور (flops) نیاز دارد. بنابراین کاهش گیونز به شکل هسنبرگی حدوداً دو برابر گران‌تر از کاهش هاوس هولدر است. اگر ماتریس A متقارن باشد، آنگاه الگوریتم حدوداً به $\frac{4}{3}n^3$ عمل ممیز شناور (flops) برای تبدیل ماتریس A به یک ماتریس سه قطری T نیاز دارد؛ باز هم این دو برابر اعمال ممیز شناور (flops) مورد نیاز در روش هاوس هولدر برای انجام همان کار می‌باشد.

خاصیت گرد کردن. خاصیت گرد کردن ذاتاً نظیر روش هاوس هولدر است. این روش از لحاظ عددی پایدار است.

تذکر MATCOM: الگوریتم ۵-۵-۵ در برنامه GIVHESS در MATCOM پیاده‌سازی شده است. ماتریس P به طور صریح محاسبه شده است.

مثال ۷-۵-۵

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

اعداد c و s را به قسمی پیدا می‌کنیم که

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{31} \end{bmatrix} = \begin{bmatrix} * \\ 0 \end{bmatrix}$$

$$a_{21} = a_{31} = 1, \quad c = \frac{1}{\sqrt{2}}, \quad s = \frac{1}{\sqrt{2}}$$

$$J(\gamma, \nu, \theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},$$

$$A \equiv J(\gamma, \nu, \theta)AJ(\gamma, \nu, \theta)^T = \begin{bmatrix} 0 & 2/1213 & 0/1711 \\ 1/4142 & 3/5000 & -/5000 \\ 0 & -1/5000 & -/5000 \end{bmatrix}$$

= بالا هسنبرگی

ملاحظه: توجه کنید که ماتریس بالا هسنبرگی به دست آمده در اینجا ذاتاً همان ماتریس به دست آمده توسط روش هاوس هولدر است (مثال ۵-۴-۵)؛ فقط عناصر هم قطرها در علامت تفاوت دارند.

۴-۵-۵ منحصر به فرد بودن در کاهش هسنبرگی

مثال قبل و ملاحظه ذکر شده در آن سؤال منحصر به فرد بودن را دوباره در کاهش هسنبرگی مطرح می‌سازد. در پایان یک گونه ساده شده از آنچه را که معروف به قضیه Q ضمنی است، بیان می‌کنیم. برای بیان قضیه و اثبات کامل آن مرجع (Golub and Van Loan (1989, 367-368) را ملاحظه کنید.

قضیه ۳-۵-۵ قضیه Q ضمنی. فرض کنید P و Q ماتریسهای متعامد باشند به قسمی که $Q^T A Q = H_p$ و $P^T A P = H_1$ فرض کنید که Q و P ستونهای اول یکسان داشته باشند. آنگاه H_1 و H_p ذاتاً یکی هستند به این مفهوم که $H_p = D^{-1} H_1 D$ که در آن

$$D = \text{diag}(\pm 1, \dots, \pm 1)$$

مثال ۵-۵-۸ ماتریس

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

را یک بار دیگر ملاحظه کنید. روش هاوس هولدر (مثال ۵-۴-۵) نتیجه زیر را ارائه داد.

$$H_1 = P_1 A P_1^T = \begin{bmatrix} \cdot & -2/1213 & \cdot/7171 \\ -1/4142 & 3/5000 & -\cdot/5000 \\ \cdot & 1/5000 & -\cdot/5000 \end{bmatrix}$$

روش گیونز (مثال ۷-۵-۵) نتیجه زیر را ارائه می دهد

$$H_2 = J(2, 3, \theta) A J(2, 3, \theta)^T = \begin{bmatrix} \cdot & 2/1213 & \cdot/7171 \\ 1/4142 & 3/5000 & \cdot/5000 \\ \cdot & -1/5000 & -\cdot/5000 \end{bmatrix}$$

با نمادگذاری قضیه ۳-۵-۵ داریم:

$$P = P_1^T, \quad Q^T = J(2, 3, \theta)$$

P و Q هر دو ستون اول یکسان دارند، یعنی ستون اول ماتریس واحد. ملاحظه می کنیم که

$$H_2 = D^{-1} H_1 D$$

و

$$D = \text{diag}(1, -1, 1)$$

۵-۶ پایه های یکا متعامد و تصویرهای متعامد

فرض کنید که A یک ماتریس $m \times n$ باشد که در آن $m \geq n$. تجزیه QR ماتریس A را در نظر بگیرید:

$$Q^T A = \begin{bmatrix} R_1 \\ \cdot \end{bmatrix}$$

فرض کنید که A دارای رتبه کامل n باشد. ماتریس Q را به صورت $Q = (Q_1, Q_2)$ افزایش می کنیم که در آن Q_1 دارای n ستون است. سپس ستونهای Q_1 یک پایه یکا متعامد برای $R(A)$ تشکیل می دهند.

به طور مشابه، ستونهای Q_2 یک پایه یکا متعامد برای متمم متعامد $R(A)$ تشکیل می دهند. بنابراین ماتریس

$$P_A = Q_1 Q_1^T$$

یک تصویر متعامد بر روی $R(A)$ و ماتریس $P_A^\perp = Q_2 Q_2^T$ تصویر بر روی متمم متعامد $R(A)$ است. از آن جا که متمم متعامد $R(A)$ توسط $R(A)^\perp = N(A^T)$ نمایش داده می شود. P_A^\perp را با P_N نمایش می دهیم.

مثال ۵-۶-۱ فرض کنید

$$A = \begin{bmatrix} 1 & 1 \\ \cdot/\cdot\cdot\cdot 1 & \cdot \\ \cdot & \cdot/\cdot\cdot\cdot 1 \end{bmatrix}$$

با استفاده از نتایج مثال ۵-۴-۴ داریم

$$Q = \begin{bmatrix} -1 & \cdot/\cdot\cdot\cdot 1 & -\cdot/\cdot\cdot\cdot 1 \\ -\cdot/\cdot\cdot\cdot 1 & -\cdot/\cdot\cdot\cdot 1 & \cdot/\cdot\cdot\cdot 1 \\ \cdot & \cdot/\cdot\cdot\cdot 1 & \cdot/\cdot\cdot\cdot 1 \end{bmatrix}$$

$$Q_1 = \begin{bmatrix} -1 & \cdot/\cdot\cdot\cdot 1 \\ -\cdot/\cdot\cdot\cdot 1 & -\cdot/\cdot\cdot\cdot 1 \\ \cdot & \cdot/\cdot\cdot\cdot 1 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} -\cdot/\cdot\cdot\cdot 1 \\ \cdot/\cdot\cdot\cdot 1 \\ \cdot/\cdot\cdot\cdot 1 \end{bmatrix}$$

$$P_A = Q_1 Q_1^T = \begin{bmatrix} 1/\cdot\cdot\cdot & \cdot/\cdot\cdot\cdot & \cdot/\cdot\cdot\cdot \\ \cdot/\cdot\cdot\cdot & \cdot/\cdot\cdot\cdot & -\cdot/\cdot\cdot\cdot \\ \cdot/\cdot\cdot\cdot & -\cdot/\cdot\cdot\cdot & \cdot/\cdot\cdot\cdot \end{bmatrix}$$

$$P_N = P_A^\perp = Q_2 Q_2^T = \begin{bmatrix} -\cdot/\cdot\cdot\cdot 1 \\ \cdot/\cdot\cdot\cdot 1 \\ \cdot/\cdot\cdot\cdot 1 \end{bmatrix} \begin{bmatrix} -\cdot/\cdot\cdot\cdot 1 \\ \cdot/\cdot\cdot\cdot 1 \\ \cdot/\cdot\cdot\cdot 1 \end{bmatrix}^T$$

$$= \begin{bmatrix} \cdot/\cdot\cdot\cdot & -\cdot/\cdot\cdot\cdot & -\cdot/\cdot\cdot\cdot \\ -\cdot/\cdot\cdot\cdot & \cdot/\cdot\cdot\cdot & \cdot/\cdot\cdot\cdot \\ -\cdot/\cdot\cdot\cdot & \cdot/\cdot\cdot\cdot & \cdot/\cdot\cdot\cdot \end{bmatrix}$$

تصویر یک بردار

با مفروض بودن یک بردار m بعدی مانند b ، بردار b_R ، تصویر b بر روی $R(A)$ توسط رابطه زیر ارائه می‌شود

$$b_R = P_A b$$

به طور مشابه، b_{\perp} ، تصویر b بر روی متمم متعامد $R(A)$ توسط رابطه زیر ارائه می‌شود.

$$b_{\perp} = P_A^{\perp} b = P_N b$$

که در آن P_A^{\perp} را با P_N نمایش داده‌ایم.

توجه کنید که $b = b_R + b_{\perp}$. دوباره، چون مکمل متعامد $R(A)$ ، $N(A^T)$ (یعنی فضای پرچ A^T) است، برای راحتی نمادگذاری b_{\perp} را با b_N نمایش می‌دهیم.

تذکره: چون $b_N = b - b_R$ ، به نظر می‌رسد که محاسبه b_N دقیقاً با کم کردن b_R از b به محض این که b_R محاسبه شد، انجام شود. این راه توصیه نمی‌شود، زیرا در محاسبه b_N از $b_N = b - b_R$ می‌تواند حذف صورت پذیرد اگر $b_R \approx b$ باشد.

مثال ۵-۶-۲ فرض کنید

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

سپس

$$Q = \begin{bmatrix} -0.7071 & -0.5774 & -0.4082 \\ 0 & -0.5774 & 0.8165 \\ -0.7071 & 0.5774 & 0.4082 \end{bmatrix}$$

$$Q_1 = \begin{bmatrix} -0.7071 & -0.5774 \\ 0 & -0.5774 \\ -0.7071 & 0.5774 \end{bmatrix}$$

$$P_A = Q_1 Q_1^T = \begin{bmatrix} 0.8333 & 0.3333 & 0.1667 \\ 0.3333 & 0.3333 & -0.3333 \\ 0.1667 & -0.3333 & 0.8333 \end{bmatrix}$$

$$b_R = P_A b = \begin{bmatrix} ۱/۳۳۳۳ \\ ۰/۳۳۳۳ \\ ۰/۶۶۶۷ \end{bmatrix}$$

$$P_N = \begin{bmatrix} ۰/۱۶۶۷ & -۰/۳۳۳۳ & -۰/۱۶۶۷ \\ -۰/۳۳۳۳ & ۰/۶۶۶۷ & ۰/۳۳۳۳ \\ -۰/۱۶۶۷ & ۰/۳۳۳۳ & ۰/۱۶۶۷ \end{bmatrix}$$

$$b_N = P_N b = \begin{bmatrix} -۰/۳۳۳۳ \\ ۰/۶۶۶۷ \\ ۰/۳۳۳۳ \end{bmatrix}$$

مثال ۳-۶-۵ فرض کنید

$$A = \begin{bmatrix} ۱ & ۱ \\ ۰/۰۰۰۱ & ۰ \\ ۰ & ۰/۰۰۰۱ \end{bmatrix}, \quad b = \begin{bmatrix} ۲ \\ ۰/۰۰۰۱ \\ ۰/۰۰۰۱ \end{bmatrix}$$

سپس

$$Q = \begin{bmatrix} -۱ & -۰/۰۰۰۱ & -۰/۰۰۰۱ \\ -۰/۰۰۰۱ & -۰/۷۰۷۱ & ۰/۷۰۷۱ \\ ۰ & ۰/۷۰۷۱ & ۰/۷۰۷۱ \end{bmatrix}$$

$$P_A = Q_1 Q_1^T = \begin{bmatrix} ۱/۰۰۰ & ۰/۰۰۰۰ & ۰/۰۰۰۰ \\ ۰/۰۰۰ & ۰/۵۰۰۰ & -۰/۵۰۰۰ \\ ۰/۰۰۰۰ & -۰/۵۰۰۰ & ۰/۵۰۰۰ \end{bmatrix}$$

$$b_R = P_A * b = \begin{bmatrix} ۲ \\ ۰/۰۰۰۱ \\ ۰/۰۰۰۱ \end{bmatrix}$$

$$b_N = \begin{bmatrix} ۰ \\ ۰ \\ ۰ \end{bmatrix}$$

تصویر یک ماتریس بر روی برد یک ماتریس دیگر

فرض کنید $B = (b_1, \dots, b_n)$ یک ماتریس $m \times n$ باشد. سپس می‌توانیم در مورد تصویر کردن هر

ستون از B بر روی $R(A)$ و بر روی متمم متعامد آن فکر کنیم. بنابراین ماتریس

$$B_R = P_A(b_1, \dots, b_n) = P_A B$$

تصویر B بر روی $R(A)$ است. به طور مشابه ماتریس

$$B_N = P_N B$$

تصویر B بر روی متمم متعامد $R(A)$ است.

مثال ۵-۶-۷ فرض کنید

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 4 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

آنگاه $A = QR$ نتیجه می‌دهد

$$Q = \begin{bmatrix} -0.2182 & -0.8165 & -0.5345 \\ -0.4364 & -0.4082 & 0.8018 \\ -0.8729 & 0.4082 & -0.2673 \end{bmatrix}$$

پایه یکا متعامد برای $R(A)$ برابر است با

$$Q_1 = \begin{bmatrix} -0.2182 & -0.8165 \\ -0.4364 & -0.4082 \\ -0.8729 & 0.4082 \end{bmatrix}$$

$$P_A = Q_1 Q_1^T = \begin{bmatrix} 0.7143 & 0.4286 & -0.1429 \\ 0.4286 & 0.3571 & 0.2143 \\ -0.1429 & 0.2143 & 0.9286 \end{bmatrix}$$

تصویر متعامد B بر روی $R(A)$ عبارت است از:

$$B_R = P_A B = \begin{bmatrix} 1/1429 & 2/1429 & 3/1429 \\ 1/7857 & 2/7857 & 3/7857 \\ 3/0.714 & 4/0.714 & 5/0.714 \end{bmatrix}$$

۵-۷ تجزیه QR با محورگیری ستونی

اگر یک ماتریس $m \times n$ ($m \geq n$) مانند A دارای رتبه $r < n$ باشد، تجزیه QR ممکن است برای تولید یک پایه یکا متعامد از $R(A)$ به کار گرفته نشود.

برای دیدن این مطلب، مثال ساده 2×2 زیر را ملاحظه کنید (Björck, (1994b):

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \begin{bmatrix} 0 & s \\ 1 & c \end{bmatrix} = QR$$

اگر c و s به قسمی ساخته شوند که $1 = c^2 + s^2$ ، $1 < 2$ ، $\text{rank}(A) = 1$ ، ستونهای Q نه یک پایه یکا متعامد از $R(A)$ و نه یک پایه یکا متعامد از متمم آن را تشکیل نمی دهند.

ولیکن خوشبختانه فرایند تجزیه QR (برای مثال روش هاوس هولدر) می تواند تغییر یابد و یک پایه یکا متعامد نتیجه دهد. در اینجا ایده این است که یک ماتریس جایگشت P به قسمی ساخته شود که داشته باشیم

$$AP = QR$$

که در آن

$$R = \begin{bmatrix} R_{11} & R_{12} \\ \cdot & \cdot \end{bmatrix}$$

در اینجا R_{11} یک ماتریس بالا مثلثی $r \times r$ است که r رتبه A می باشد، و Q متعامد است. سپس r ستون اول Q یک پایه یکا متعامد از $R(A)$ تشکیل خواهد داد. قضیه زیر وجود چنین تجزیه ای را تضمین می کند.

قضیه ۵-۷-۱ قضیه محورگیری ستونی QR. فرض کنید A یک ماتریس $m \times n$ با $\text{rank}(A) = r$ باشد. آنگاه یک ماتریس جایگشت $n \times n$ مانند P و یک ماتریس متعامد $m \times m$ مانند Q وجود دارد به قسمی که

$$Q^T AP = \begin{bmatrix} R_{11} & R_{12} \\ \cdot & \cdot \end{bmatrix}$$

که در آن R_{11} یک ماتریس بالا مثلثی $r \times r$ با عناصر قطری مخالف صفر است.

اثبات: چون $\text{rank}(A) = r$ ، یک ماتریس جایگشت P وجود دارد به قسمی که

$$AP = (A_1, A_r)$$

و A_1 یک ماتریس $m \times r$ با ستونهای مستقل خطی است. اکنون تجزیه QR ماتریس A_1 را در نظر بگیرید.

$$Q^T A_1 = \begin{bmatrix} R_{11} \\ \cdot \\ \cdot \end{bmatrix}$$

که در آن R_{11} یک ماتریس بالا مثلثی $r \times r$ با عناصر قطری مخالف صفر است. پس

$$Q^T A P = (Q^T A_1, Q^T A_2) = \begin{bmatrix} R_{11} & R_{12} \\ \cdot & R_{22} \end{bmatrix}$$

چون $\text{rank}(Q^T A P) = \text{rank}(A) = r$ و $\text{rank}(R_{11}) = r$ باید داشته باشیم $R_{22} = 0$.

ایجاد ماتریس جایگشت P

راههای مختلفی برای ایجاد این ماتریس جایگشت P وجود دارد. در اینجا یکی از این راهها را ذکر می‌کنیم که به تجزیه QR با محورگیری ستونی معروف است.

ماتریس جایگشت P از حاصلضرب r ماتریس جایگشت P_1 تا P_r تشکیل می‌شود. این ماتریسهای جایگشت یکی یکی بر A قبل از تشکیل ماتریسهای هاوس هولدر برای ایجاد صفرها در ستونهای مناسب، اعمال می‌شوند. به طور مشخص تر کار زیر انجام می‌شود:

گام ۱: ستونی از A را پیدا کنید که دارای ماکزیمم نرم باشد. ستونهای A را به نحوی تعویض نمایید که ستون با ماکزیمم نرم در ستون اول قرار گیرد.

این معادل با تشکیل یک ماتریس جایگشت P_1 است به قسمی که ماتریس AP_1 دارای ستون اول با ماکزیمم نرم باشد. یک ماتریس هاوس هولدر H_1 به قسمی تشکیل دهید که

$$A^{(1)} = H_1 AP_1$$

دارای صفر در ستون اول و در زیر عنصر $(1,1)$ باشد.

گام ۲: ستون با نرم ماکزیمم زیر ماتریس $\tilde{A}^{(1)}$ حاصل از حذف سطر اول و ستون اول ماتریس A را پیدا کنید. ستونهای $A^{(1)}$ را به قسمی تعویض کنید که ستون با ماکزیمم نرم $\tilde{A}^{(1)}$ در ستون اول $\tilde{A}^{(1)}$ قرار گیرد. این کار معادل است با ساختن یک ماتریس جایگشت \tilde{P}_2 است به قسمی که اولین ستون $\tilde{P}_2 A^{(1)}$ دارای ماکزیمم نرم باشد، و سپس پس ضرب کردن ماتریس $A^{(1)}$ در یک ماتریس جایگشت P_2 که به صورت زیر تعریف می‌شود.

$$P_2 = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & & \hat{P}_2 & \\ \vdots & & & \end{bmatrix}$$

اکنون یک ماتریس هاوس هولدر بسازید به قسمی که

$$A^{(2)} = H_2 A^{(1)} P_2 = H_2 H_1 A P_1 P_2$$

در ستون دوم در زیر عنصر $(2, 2)$ دارای صفر باشد. مانند قبل H_2 می تواند در دو گام مانند

بخش ۱-۴-۵ ساخته شود:

- اکنون گام k ام می تواند به سادگی نوشته شود.
 - فرایند ادامه می یابد تا عناصر زیر قطر ماتریس جاری همگی صفر شوند.
- فرض کنید r گام مورد نیاز باشد. آنگاه در انتهای گام r داریم:

$$\begin{aligned} A &\equiv A^{(r)} = H_r \dots H_1 A P_1 \dots P_r \\ &= Q^T A P = R = \begin{bmatrix} R_{11} & R_{12} \\ \cdot & \cdot \end{bmatrix} \end{aligned}$$

شمارش اعمال ممیز شناور و ملاحظات ذخیره سازی. روش قبلی به $2mnr - r^2(m+n) + 2r^3/3$ عمل ممیز شناور نیاز دارد. ماتریس Q ، همانند تجزیه هاوس هولدر، به شکل عاملها ذخیره می شود. ماتریس Q می تواند به شکل عاملها در قسمت زیر قطر A ذخیره شود، و R می تواند بر روی A نوشته شود.

مثال ۱-۷-۵ فرض کنید

$$A = \begin{bmatrix} 0 & 0 \\ \frac{1}{2} & 1 \\ \frac{1}{2} & 1 \end{bmatrix}_{2 \times 2} = (a_1, a_2)$$

a_2 دارای بزرگترین نرم است. پس

$$P_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$AP_1 = \begin{bmatrix} 0 & 0 \\ 1 & \frac{1}{\sqrt{2}} \\ 1 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$H_1 = \begin{bmatrix} 0 & \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{2} & \frac{-1}{2} \\ \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & \frac{1}{2} \end{bmatrix}$$

$$A^{(1)} = H_1 AP_1 = \begin{bmatrix} 0 & \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{2} & \frac{-1}{2} \\ \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & \frac{1}{\sqrt{2}} \\ 1 & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\sqrt{2} & \frac{-1}{\sqrt{2}} \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} = R$$

$$R = \begin{bmatrix} R_1 & R_2 \\ \cdot & \cdot \end{bmatrix}$$

بنابراین برای این مثال داریم:

$$Q = H_1^T = H_1,$$

$$P = P_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

ماتریس A دارای رتبه ۱ است، زیرا $R_1 = -\sqrt{2}$ یک ماتریس 1×1 است. بردار ستونی

$$\begin{bmatrix} 0 \\ -1 \\ \sqrt{2} \\ -1 \\ \sqrt{2} \end{bmatrix}$$

یک پایهٔ یکا متعامد $R(A)$ را تشکیل می‌دهد.

تجزیهٔ متعامد کامل

به راحتی ملاحظه می‌شود که زیر ماتریس (R_{11}, R_{12}) می‌تواند باز هم با استفاده از تبدیلهای متعامد کاهش یابد و نتیجه به صورت زیر باشد.

$$\begin{bmatrix} T & 0 \\ 0 & 0 \end{bmatrix}$$

قضیهٔ ۵-۷-۲ قضیهٔ متعامد سازی کامل. با مفروض بودن $A_{m \times n}$ با $\text{rank}(A) = r$ ، ماتریسهای متعامد $Q_{m \times m}$ و $W_{n \times n}$ وجود دارند به قسمی که

$$Q^T A W = \begin{bmatrix} T & 0 \\ 0 & 0 \end{bmatrix}$$

که در آن T یک ماتریس بالا مثلثی $r \times r$ با عناصر قطری مثبت است.

اثبات: اثبات به عنوان یک تمرین واگذار می‌شود.

این تجزیه از A تجزیهٔ متعامد کامل نامیده می‌شود.

QR آشکارساز رتبه

فرایند قبل، معروف به تجزیه QR با محورگیری ستونی، توسط گولوب ((Golub (1965) پیشنهاد شده است. این تجزیه به تجزیه QR آشکارساز رتبه معروف است، زیرا در حساب دقیق رتبهٔ ماتریس A را آشکار می‌سازد، که مرتبه ماتریس بالا مثلثی نامفرد R_{11} است. ولیکن به هنگام وجود خطاهای گردکردن، در واقع خواهیم داشت:

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}$$

و اگر R_{22} از نظر اندازه‌ای کوچک باشد (برای مثال $\|R_{22}\|$ از رتبه $O(\mu)$ باشد، که در آن μ دقت

ماشین است)، آنگاه کاهش تمام خواهد شد. بنابراین از بحث قبل ملاحظه می‌کنیم که، با مفروض بودن یک ماتریس $m \times n$ مانند A ($m \geq n$)، اگر یک ماتریس جایگشت P وجود داشته باشد به قسمی که

$$Q^T A P = R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}$$

و در آن R_{11} از مرتبه $r \times r$ و R_{22} از نظر اندازه‌ای کوچک باشد، آنگاه خواهیم گفت که A دارای رتبه عددی r است. (برای مطالب بیشتر در مورد رتبه عددی بخش ۱۰-۵-۵ را ملاحظه کنید).

متأسفانه، عکس آن درست نیست. یک مثال نقص مشهور که کاهان (Kahan (1966)) ارائه داده است، نشان می‌دهد که یک ماتریس می‌تواند تقریباً رتبه ناقص داشته باشد بدون آن که به هیچ وجه دارای $\|R_{22}\|$ کوچک باشد.

ماتریس

$$A = \text{diag} (1, s, \dots, s^{n-1}) \begin{bmatrix} 1 & -c & -c & \dots & -c \\ & 1 & -c & \dots & -c \\ & & 1 & & \vdots \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix}$$

را با $1 = c^2 + s^2$ و $c, s > 0$ در نظر بگیرید. برای $n = 100$ ، $c = 0.2$ می‌توان نشان داد که A تقریباً منفرد است (کوچکترین مقدار تکیین برابر $O(10^{-8})$ است). از طرف دیگر $r_{nn} = s^{n-1} = 0.133$ ، که کوچک نیست، بنابراین R نمی‌تواند تقریباً منفرد باشد.

این سؤال که آیا در هر مرحله‌ای R_{pp} برای هر ماتریسی حقیقتاً کوچک می‌شود، توسط چان (Chan (1987)) و پس از آن توسط هانک و پان (Hang and Pan (1992)) بررسی شده است.

۸-۵ اصلاح یک تجزیه QR

فرض کنید که تجزیه QR یک ماتریس $m \times k$ مانند $A = (a_1, \dots, a_k)$ ، $m \geq k$ ، به دست آمده باشد. اکنون یک بردار a_{k+1} به انتها اضافه می‌کنیم تا یک ماتریس جدید

$$A' = (a_1, \dots, a_k, a_{k+1})$$

به دست آید. طبیعی است که تعجب کنیم که چگونه تجزیه QR ماتریس A' می‌تواند از تجزیه QR

مفروض A بدون پیدا کردن مستقیم آن، حاصل گردد.

این مسأله به هنگام درآوری تجزیه QR نامیده می شود. سطح پایین درآوری متشابهاً تعریف می شود، به هنگام درآوری و به سطح پایین تر درآوری تجزیه QR در کاربردهای عملی گوناگون نظیر پردازش سیگنال و تصویر مطرح می شوند.

یک الگوریتم ساده با استفاده از ماتریسهای هاوس هولدر برای به هنگام درآوری مسأله ارائه می دهیم (Lawson and Hanson, (1974, 174-179)).

الگوریتم ۱-۸-۵ به هنگام درآوری تجزیه QR با استفاده از ماتریسهای هاوس هولدر. الگوریتم زیر تجزیه QR ماتریس $A' = (a_1, \dots, a_k, a_{k+1})$ را با مفروض بودن تجزیه QR هاوس هولدر ماتریس $A = (a_1, \dots, a_k)$ محاسبه می کند.

گام ۱: $b_{k+1} = H_k \dots H_1 a_{k+1}$ را محاسبه کنید که در آن H_1 تا H_k آن ماتریسهای هاوس هولدری هستند که

$$Q^T A = H_k \dots H_1 A = \begin{bmatrix} R_1 \\ \cdot \\ \cdot \end{bmatrix} = R$$

گام ۲: یک ماتریس هاوس هولدر H_{k+1} به قسمی پیدا کنید که $H_{k+1} b_{k+1} = r_{k+1}$ در عناصر $m, \dots, k+2$ صفر باشد.

گام ۳: $R' = \begin{bmatrix} R_1 \\ \cdot \\ \cdot \\ r_{k+1} \end{bmatrix}$ را تشکیل دهید.

گام ۴: $Q' = H_{k+1} \dots H_1$ را تشکیل دهید.

قضیه ۱-۸-۵ R' و Q' تعریف شده در فوق به قسمی هستند که $(Q')^T A' = R'$.

مثال ۱-۸-۵ فرض کنید

$$A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

آنگاه

$$H_1 = Q_1^T = \begin{bmatrix} -./۲۶۷۳ & -./۵۳۴۵ & -./۸۰۱۸ \\ -./۵۳۴۵ & ./۷۷۴۵ & -./۳۳۸۲ \\ -./۸۰۱۸ & -./۳۳۸۲ & ./۴۹۲۷ \end{bmatrix}$$

$$R = \begin{bmatrix} -۳/۷۴۱۷ \\ \cdot \\ \cdot \end{bmatrix}$$

$$A' = \begin{bmatrix} ۱ & ۱ \\ ۲ & ۴ \\ ۳ & ۵ \end{bmatrix}$$

گام ۱:

$$H_1 = \begin{bmatrix} -./۲۶۷۳ & -./۵۳۴۵ & -./۸۰۱۸ \\ -./۵۳۴۵ & ./۷۷۴۵ & -./۳۳۸۲ \\ -./۸۰۱۸ & -./۳۳۸۲ & ./۴۹۲۷ \end{bmatrix}$$

$$R = \begin{bmatrix} -۳/۷۴۱۷ \\ \cdot \\ \cdot \end{bmatrix}$$

$$b_r = H_1 a_r = \begin{bmatrix} -۶/۴۱۴۳ \\ ./۸۲۲۷ \\ ./۳۰۹۱ \end{bmatrix}$$

گام ۲:

$$H_r = \begin{bmatrix} ۱ & \cdot & \cdot \\ \cdot & -./۹۴۲۶ & -./۳۳۳۹ \\ \cdot & -./۳۳۳۹ & ./۹۴۲۶ \end{bmatrix}, \quad r_r = \begin{bmatrix} -۶/۴۱۴۳ \\ -./۹۲۵۸ \\ \cdot \end{bmatrix}$$

گام ۳:

$$R' = (R, r_r) = \begin{bmatrix} -۳/۷۴۱۷ & -۶/۴۱۴۳ \\ \cdot & -./۹۲۵۸ \\ \cdot & \cdot \end{bmatrix}$$

$$Q' = H_r H_1 = \begin{bmatrix} -./۲۶۷۳ & -./۵۳۴۵ & -./۸۰۱۸ \\ ./۷۷۱۵ & -./۶۱۷۲ & ./۱۵۴۳ \\ -./۵۷۷۳ & -./۵۷۷۴ & ./۵۷۷۴ \end{bmatrix}$$

$$(Q')^T R' = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 5 \end{bmatrix} = A' \quad \text{تحقیق:}$$

۹-۵ خلاصه و جدول مقایسه‌ها

۱. سه ماتریس مهم عبارتند از: ماتریسهای مقدماتی، هاوس هولدر، و گیونز.

(الف) یک ماتریس پایین مثلثی مقدماتی یک ماتریس پایین مثلثی به شکل $E = I + m_k e_k^T$

است که در آن $m_k = (0, 0, \dots, 0, m_{k+1,k}, \dots, m_{n,k})^T$. اگر E به صورت ارائه شده در

$$E^{-1} = I - m_k e_k^T$$

اینجا باشد، آنگاه

(ب) یک ماتریس هاوس هولدر یک ماتریس $H = I - \gamma u u^T / u^T u$ ، $n \times n$ است، که در آن u

یک بردار n بعدی است. یک ماتریس هاوس هولدر متقارن و متعامد است.

(ج) یک ماتریس گیونز $J(i, j, c, s)$ یک ماتریس واحد است بجز در عناصر (i, j) ، (j, i) ،

(j, j) ، و (i, i) که به ترتیب عبارتند از c ، s ، $-s$ و c و بعلاوه $c^2 + s^2 = 1$. یک

ماتریس گیونز یک ماتریس متعامد است.

۲. دو تجزیه ماتریسی مهم، LU و QR هستند.

(الف) یک تجزیه $A = LU$ به شکل $A = LU$ که در آن L پایین مثلثی واحد و U بالا مثلثی است،

یک تجزیه LU ماتریس A نامیده می‌شود. یک تجزیه LU ماتریس A همیشه وجود

ندارد. اگر کدهای اصلی پیشرو A همگی مخالف صفر باشند، آنگاه تجزیه LU ماتریس A

وجود دارد و منحصر به فرد است (قضیه ۵-۲-۱).

تجزیه LU یک ماتریس A در صورت وجود، با استفاده از ماتریسهای پایین مثلثی

مقدماتی حاصل می‌گردد. فرایند حذفی گاوس بدون محورگیری نامیده می‌شود.

فرایند کارا است، فقط نیاز به $n^3/3$ عمل ممیز شناور (flops) دارد، اما پایدار نیست.

استفاده آن در عمل توصیه نمی‌شود مگر A یک ماتریس معین مثبت متقارن یا غالب قطری

ستونی باشد. برای تجزیه A به LU به یک طریق پایدار تعویضهای سطری (حذفی گاوس

با محورگیری جزئی) یا تعویضهای هم سطری و هم ستونی (حذفی گاوس با محورگیری

کامل) برای مشخص کردن یک عنصر محوری مناسب لازم خواهد بود. حذفی گاوس با

محورگیری جزئی و کامل به ترتیب تجزیه‌های $MA = U$ و $MAQ = U$ را نتیجه

می‌دهند.

(ب) هر ماتریس A همواره می‌تواند به شکل $A = QR$ نوشته شود، که در آن Q متعامد و R

بالا مثلثی است. این تجزیه QR ماتریس A نامیده می‌شود. تجزیه QR یک ماتریس ذاتاً منحصراً به فرد است اگر A نامنفرد باشد (قضیه ۵-۵-۲). تجزیه QR می‌تواند با استفاده از ماتریسهای هاوس هولدر یا گیونز حاصل گردد. هر دو روش دارای پایداری عددی تضمین شده هستند. روش هاوس هولدر کاراتر از روش گیونز است ($2n^3/3$ عمل ممیز شناور در مقابل $4n^3/3$ عمل ممیز شناور (به طور تقریبی))، اما ماتریسهای گیونز در محاسبات ماتریسی موازی و محاسبات با ماتریسهای ساخت یافته ابزار مفیدی به حساب می‌آیند. برای یک نوع متفاوت از تجزیه QR فرایندهای گرام اشمیت و گرام اشمیت اصلاح شده در فصل ۷ توصیف خواهند شد.

۳. کاهش هسنبرگی: شکل هسنبرگی یک ماتریس یک شکل فشرده خیلی مفید است. استفاده آن را در تمام کتاب ملاحظه خواهیم کرد. یک ماتریس دلخواه A همواره می‌تواند با تبدیلات تشابهی متعامد به یک ماتریس بالا هسنبرگی تبدیل شود: با مفروض بودن یک ماتریس $n \times n$ مانند A ، همواره یک ماتریس متعامد مانند P وجود دارد به قسمی که $PAP^T = H_u$ یک ماتریس بالا هسنبرگی باشد (قضیه ۵-۴-۲).

این کاهش می‌تواند با استفاده از ماتریسهای مقدماتی، هاوس هولدر یا گیونز حاصل گردند. ما روشهای مبتنی بر ماتریسهای هاوس هولدر و گیونز را توصیف کرده‌ایم (الگوریتمهای ۵-۴-۴ و ۵-۵-۵). هر دو روش دارای پایداری تضمین شده هستند، اما دوباره، روش هاوس هولدر کاراتر از روش گیونز است.

در مورد منحصراً به فرد بودن در کاهش هسنبرگی صورت قضیه Q ضمنی (قضیه ۵-۵-۳) را ملاحظه کنید. این قضیه اساساً بیان می‌کند که اگر یک ماتریس با تبدیلات تشابهی به دو ماتریس بالا هسنبرگی کاهش ناپذیر با استفاده از دو ماتریس تبدیل P و Q تبدیل شود، آنگاه H_1 و H_2 ذاتاً یکی هستند به شرط آن که P و Q دارای ستون اول یکسان باشند.

۴. پایه‌های متعامد و تصویرهای متعامد: اگر $Q^TA = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ تجزیه QR یک ماتریس

$m \times m$ ($m \geq n$) باشد، آنگاه ستونهای Q_1 یک پایه یکا متعامد برای $R(A)$ و ستونهای Q_2 یک پایه یکا متعامد برای متمم متعامد $R(A)$ تشکیل می‌دهند، در اینجا $Q = (Q_1, Q_2) = Q_1$ دارای n ستون است. فرض می‌شود A دارای رتبه کامل است.

اگر فرض کنیم که $B = (b_1, \dots, b_n)$ یک ماتریس $m \times n$ باشد، آنگاه ماتریس

$R(A)$ بر روی B متعامد $P_A = Q_1 Q_1^T$ که در آن $B_R = P_A(b_1, \dots, b_n) = P_A B$ است. به طور مشابه ماتریس $B_N = P_N B$ تصویر B بر روی متمم متعامد $R(A)$ است، که در آن $P_N = P_A^\perp = Q_2 Q_2^T$

۵. تجزیه QR با محورگیری ستونی، QR آشکار ساز رتبه، و اصلاح نمودن یک تجزیه QR ارائه شدند. اگر A رتبه - ناقص باشد، آنگاه تجزیه QR معمولی ممکن است یک پایه یکا متعامد از $R(A)$ تولید نکند. در چنین حالتی لازم است فرایند تجزیه QR اصلاح گردد. یک اصلاح که تجزیه QR با محورگیری ستونی نامیده می شود، در این فصل مورد بحث قرار گرفت. یک چنین تجزیه ای همواره وجود دارد (قضیه ۵-۷-۱). یک فرایند برای رسیدن به یک چنین تجزیه ای با استفاده از ماتریسهای هاوس هولدر که توسط گولوب ارائه شده است، در بخش ۵-۷ به طور مختصر توصیف شد.

تجزیه QR با محورگیری ستونی در حساب دقیق رتبه A را آشکار می سازد. به این علت تجزیه QR آشکار ساز رتبه نامیده می شود. ولیکن به هنگام وجود خطاهای گرد کردن یک چنین روش تعیین رتبه پیچیده بوده و قابل اطمینان نمی باشد. بالاخره یک الگوریتم ساده برای اصلاح تجزیه QR ارائه دادیم (به هنگام درآوری QR).

۶. در جدول ۵-۱-۵ خاصیت‌های کارایی و پایداری برخی از این محاسبات مهم خلاصه شده اند. فرض می کنیم که A یک ماتریس $m \times n$ ($m \geq n$) است.

۷. حذفی گاوس بدون محورگیری ناپایدار است؛ حذفی گاوس با محورگیری جزئی در عمل پایدار است؛ حذفی گاوس با محورگیری کامل پایدار است. از جدول ۵-۱ ملاحظه می کنیم که فرایندهای تجزیه QR و کاهش به یک ماتریس هسنبرگی با استفاده از تبدیلات هاوس هولدر کارا تر هستند. روشها از لحاظ عددی پایدار هستند. ولیکن همانطور که قبلاً متذکر شدیم، تبدیلات گیونز برای محاسبات ماتریسی با ماتریسهای ساخت یافته مفید هستند، و آنها برای الگوریتمهای ماتریسی موازی ابزار مهمی به حساب می آیند. بعلاوه ذکر این نکته مهم است که روش حذفی گاوس می تواند برای تبدیل یک ماتریس دلخواه به یک ماتریس بالا هسنبرگی به وسیله تبدیلات تشابهی مورد استفاده قرار گیرد، برای جزئیات، مرجع (Wilkinson 1965, 353-5) را ملاحظه کنید.

۱-۵ جدول مقایسه‌ها

پایداری	تعداد اعمال ممیز شناور (flops) (تقریبی)	روش	مسئله
ناپایدار	$\frac{mn^2}{2} - \frac{n^2}{6}$	حذفی گاوس بدون تعویض سطری	تجزیه
در عمل پایدار	$\frac{mn^2}{2} - \frac{n^2}{6}$ ($O(n^2)$ مقایسه +)	حذفی گاوس با محورگیری جزئی	تجزیه $MA = U$
پایدار	$\frac{mn^2}{2} - \frac{n^2}{6}$ ($O(n^3)$ مقایسه +)	حذفی گاوس با محورگیری کامل	تجزیه $MAQ = U$
پایدار	$n^2(m - \frac{n}{3})$	هاوس هولدر	تجزیه
پایدار	$2n^2(m - \frac{n}{3})$	گیونز	تجزیه
پایدار	$\frac{5n^3}{3}$	هاوس هولدر	کاهش هسبرگی یک $n \times n$ ماتریس
پایدار	$\frac{10n^3}{3}$	گیونز	کاهش هسبرگی یک $n \times n$ ماتریس
پایدار	$2mnr$ $-r^2(m+n) + \frac{2nr^3}{3}$ $r = \text{rank}(A)$	هاوس هولدر	تجزیه با محورگیری ستونی

تمرینها

تذکر: در صورت لزوم از MATLAB استفاده کنید.

۱. (الف) نشان دهید که یک ماتریس پایین مثلثی مقدماتی دارای شکل زیر است

$$E = I + m_k e_k^T$$

که در آن $m_k = (0, 0, \dots, 0, m_{k+1,k}, \dots, m_{n,k})^T$

(ب) نشان دهید که معکوس ماتریس E ذکر شده در (الف) به صورت زیر ارائه می‌شود.

$$E^{-1} = I - m_k e_k^T$$

۲. (الف) با مفروض بودن

$$a = \begin{bmatrix} 0/\dots\dots\dots 1 \\ 1 \end{bmatrix}$$

و استفاده از حساب پنج رقمی یک ماتریس مقدماتی E به قسمی پیدا کنید که Ea مضربی از e_1 باشد.

(ب) با استفاده از محاسبات انجام شده در (الف) تجزیه LU ماتریس زیر را پیدا کنید.

$$A = \begin{bmatrix} 0/\dots\dots\dots 1 & 1 \\ & 1 & 2 \end{bmatrix}$$

(ج) فرض کنید \hat{L} و \hat{U} محاسبه شده‌های L و U قسمت (الف) باشند. مقدار

$$\frac{\|A - \hat{L}\hat{U}\|_F}{\|A\|_F}$$

را پیدا کنید که در آن $\| \cdot \|_F$ نرم فروبنیوس می‌باشد.

۳. نشان دهید که عناصر محوری $a_{11}, a_{22}^{(1)}, \dots, a_{nn}^{(n-1)}$ مخالف صفر هستند اگر و فقط اگر کدهای اصلی پیشرو مخالف صفر باشند. راهنمایی: نشان دهید که

$$\det(A_r) = a_{11} a_{22}^{(1)} \dots a_{rr}^{(r-1)}$$

۴. فرض کنید A یک ماتریس معین مثبت متقارن باشد. در انتهای گام اول تجزیه LU ماتریس A داریم:

$$A^{(1)} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \cdot & & & \\ \cdot & & & \\ \vdots & & A' & \\ \cdot & & & \end{bmatrix}$$

ثابت کنید که A' نیز متقارن و معین مثبت است. بنابراین نشان دهید که تجزیه LU یک ماتریس معین مثبت متقارن با استفاده از حذفی گاوس بدون محورگیری همواره وجود دارد و منحصر به فرد است.

۵. (الف) تمرین ۴ را برای هنگامی که A یک ماتریس غالب قطری است تکرار کنید؛ یعنی نشان دهید که تجزیه LU یک ماتریس غالب قطری همواره وجود دارد و منحصر به فرد است.

(ب) با استفاده از (الف)، نشان دهید که یک ماتریس غالب قطری نامنفرد است.

۶. فرض کنید تجزیه LU ماتریس A وجود داشته باشد، ثابت کنید که

(الف) A می‌تواند به صورت

$$A = LDU_1$$

نوشته شود، که در آن D یک ماتریس قطری و L و U_1 به ترتیب ماتریسهای پایین مثلثی و

بالا مثلثی واحد هستند.

(ب) اگر A متقارن باشد، آنگاه

$$A = LDL^T$$

(ج) اگر A متقارن و معین مثبت باشد، آنگاه

$$A = HH^T$$

که در آن H یک ماتریس پایین مثلثی با عناصر قطری مثبت است. این تجزیه به تجزیه

چولسکی^(۱) معروف است.

۷. با فرض این که تجزیه LU وجود دارد، یک الگوریتم برای محاسبه مستقیم U سطر به سطر و

L ستون به ستون از معادله

$$A = LU$$

تدوین نمایید. این فرایند معروف به کاهش دولیتل^(۲) است.

۸. یک الگوریتم برای محاسبه تجزیه

$$A = LU$$

تدوین نمایید که در آن U یک ماتریس بالا مثلثی واحد و L یک ماتریس پایین مثلثی هستند.

این فرایند به کاهش کرات^(۳) معروف است. راهنمایی: یک الگوریتم از $A = LU$ به دست

آورید.

۹. کاهشهای دولیتل و کرات را با حذفی گاوس از نظر تعداد اعمال ممیز شناور (flops)، نیازهای

ذخیره‌سازی و امکان ضرب داخلی در دقت مضاعف، مقایسه کنید.

۱۰. یک ماتریس G به شکل

$$G = I - ge_k^T$$

یک ماتریس گاوس - جردن نامیده می شود. نشان دهید که با مفروض بودن یک بردار x با خاصیت $e_k^T x \neq 0$ یک ماتریس گاوس - جردن وجود دارد به قسمی که Gx مضربی از e_k باشد.

(الف) یک الگوریتم برای ساختن متوالی ماتریسهای گاوس - جردن G_1, G_2, \dots, G_n به قسمی تدوین نمایید که $A(G_n G_{n-1} \dots G_1)$ یک ماتریس قطری باشد. این فرایند معروف به کاهش گاوس - جردن است.

(ب) شرایطی را به دست آورید که تحت آنها کاهش گاوس - جردن بتواند تا تکمیل شدن انجام شود.

(ج) تعداد اعمال ممیز شناور (flops) را برای الگوریتم ارائه دهید و آن را با تعداد اعمال ممیز شناور (flops) روش حذفی گاوس، کاهشهای کرات و دولیتل مقایسه نمایید.

۱۱. با مفروض بودن

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

تجزیه های زیر را به دست آورید.

(الف) تجزیه LU ماتریس A با استفاده از حذفی گاوس و کاهش دولیتل.

(ب) تجزیه LU ماتریس A با استفاده از کاهش کرات (توجه کنید که U ماتریس واحد بالامثلثی و L ماتریس پایین مثلثی است).

۱۲. کاهش گاوس - جردن را در مورد ماتریس A تمرین ۱۱ به کار ببرید.

۱۳. (الف) فرض کنید A یک ماتریس $m \times n$ باشد و فرض کنید $r = \min\{m-1, n\}$. یک الگوریتم برای ساختن ماتریسهای مقدماتی E_1, \dots, E_r تدوین نمایید به قسمی که ماتریس

$$E_r E_{r-1} \dots E_1 A$$

یک ماتریس بالا دوزنقه ای U باشد. الگوریتم باید U را بر روی A بنویسید.

(ب) نشان دهید که الگوریتم حدوداً به $r^3/3$ عمل ممیز شناور (flops) نیاز دارد.

(ج) الگوریتم را در مورد ماتریس زیر به کار ببرید.

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 6 & 7 \end{bmatrix}$$

۱۴. یک ماتریس سه قطری A با عناصر غیر قطری مخالف صفر مفروض است، یک مجموعه از شرایط ساده بر روی عناصر A بنویسید که تضمین کند که حذفی گاوس می تواند تا تکمیل شدن انجام شود.

۱۵. فرض کنید که تجزیه LU وجود دارد با استفاده از

$$A = LU$$

نشان دهید که اگر A سه قطری باشد، L و U هر دو دو قطری خواهند بود. یک طرح برای محاسبه L و U در این حالت تدوین نمایید و طرح خودتان را برای پیدا کردن تجزیه LU ماتریس زیر به کار ببرید.

$$A = \begin{bmatrix} 4 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 1 & 4 \end{bmatrix}$$

۱۶. ثابت کنید که ماتریس L در هر کدام از تجزیه های $PA = LU$ و $PAQ = LU$ ، که به ترتیب از روش حذفی گاوس با محورگیری جزئی و کامل به دست آمده اند، پایین مثلثی واحد است.

۱۷. با مفروض بودن

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 2 & 3 & 4 & 5 \end{bmatrix}$$

یک ماتریس جایگشت P ، یک ماتریس پایین مثلثی واحد L ، یک ماتریس بالا مثلثی U پیدا کنید به قسمی که $PA = LU$.

۱۸. برای هر کدام از ماتریسهای زیر پیدا کنید:

(الف) ماتریسهای جایگشت P_1 و P_2 و ماتریسهای مقدماتی M_1 و M_2 به طوری که $MA = M_2 P_1 M_1 P_1 A$ یک ماتریس بالا مثلثی باشد.

(ب) ماتریسهای جایگشت P_1 ، P_2 ، Q_1 ، Q_2 و ماتریسهای مقدماتی M_1 و M_2 به طوری که $MAQ = M_2 (P_2 (M_1 P_1 A Q_1) Q_2)$ یک ماتریس بالا مثلثی باشد.

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix} \quad (i)$$

$$A = \begin{bmatrix} 100 & 99 & 98 \\ 98 & 55 & 11 \\ 0 & 1 & 11 \end{bmatrix} \quad (ii)$$

$$A = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \\ -1 & -1 & 1 \end{bmatrix} \quad (iii)$$

$$A = \begin{bmatrix} 0.0003 & 1/566 & 1/234 \\ 1/5660 & 2/000 & 1/0.18 \\ 1/2340 & 1/0.18 & -3/000 \end{bmatrix} \quad (iv)$$

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix} \quad (v)$$

- (ج) هر تجزیه را به شکل $PAQ = U$ بیان کنید (توجه کنید که برای حذفی گاوس بدون محورگیری و با محورگیری جزئی داریم $Q = I$).
- (د) عامل رشد را برای هر حالت حساب کنید.

بخشهای ۴-۵ الی ۶-۵

۱۹. فرض کنید x یک بردار n بعدی باشد. یک الگوریتم برای محاسبه یک ماتریس هاوس هولدر

$$H = I - \frac{uu^T}{u^T u}$$

ارائه دهید به قسمی که Hx در مکانهای $(r+1)$ تا n ؛ $r < n$ دارای صفر باشد چند عمل ممیز شناور (flops) برای پیاده سازی الگوریتم مورد نیاز است؟ با مفروض بودن $x = (1, 2, 3)^T$ الگوریتم خودتان را برای ساختن H به کار ببرید به قسمی که Hx دارای یک صفر در مکان سوم باشد.

۲۰. فرض کنید H یک ماتریس بالا هسنبرگی کاهش ناپذیر از مرتبه n باشد، یک الگوریتم برای مثلثی سازی H با استفاده از روشهای زیر تدوین نمایید.

(الف) حذفی گاوس

(ب) تبدیلهای هاوس هولدر

(ج) دورانهای گیونز

۲۱. فرض کنید

$$H = \begin{bmatrix} 1.0 & 1 & 1 & 1 \\ 2 & 1.0 & 1 & 1 \\ 0 & 1 & 1.0 & 1 \\ 0 & 0 & 1 & 1.0 \end{bmatrix}$$

 H را با استفاده از روشهای زیر مثلثی سازید.

(الف) روش حذفی گاوس

(ب) روش هاوس هولدر

(ج) روش گیونز

۲۲. فرض کنید، که $\hat{H}_k = I - \gamma uu^T / u^T u$ ، که در آن u یک بردار $(n - k + 1)$ بعدی است.

تعریف کنید

$$H_k = \begin{bmatrix} I_{k-1} & 0 \\ 0 & \hat{H}_k \end{bmatrix}$$

چند عمل ممیز شناور (flops) برای محاسبه $H_k A$ مورد نیاز خواهد بود، که در آن A دلخواه و $n \times n$ است؟ شمارش شما باید ساختار خاص ماتریس \hat{H}_k را نیز منظور نماید.با استفاده از این نتیجه، نشان دهید که روش هاوس هولدر حدوداً به $\frac{2}{3} n^3$ عمل ممیز شناور(flops) برای به دست آوردن R و $\frac{2}{3} n^3$ عمل ممیز شناور (flops) دیگر برای به دست آوردن Q در تجزیه QR ماتریس A نیاز دارد.۲۳. نشان دهید که $n^2(m - \frac{n}{3})$ عمل ممیز شناور (flops) برای محاسبه R در تجزیه QR یکماتریس $m \times n$ مانند A ($m \geq n$) با استفاده از روش هاوس هولدر مورد نیاز است.

۲۴. با مفروض بودن

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \\ 5 & 6 \end{bmatrix}$$

(الف) ماتریسهای H_1 و H_2 را پیدا کنید به طوری که

$$H_2 H_1 A = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

که در آن R_1 یک ماتریس بالا مثلثی 2×2 باشد.

(ب) پایه‌های یکا متعامد را برای $R(A)$ و برای متمم متعامد $R(A)$ پیدا کنید.

(ج) تصویرهای متعامد A را بر روی $R(A)$ و بر روی متمم متعامد آن پیدا کنید؛ یعنی P_A

و P_{A^\perp} را پیدا کنید.

۲۵. با مفروض بودن

$$b = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

و A همانند تمرین ۲۴، b_R و b_N را بیابید

۲۶. با مفروض بودن

$$B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 5 \end{bmatrix}$$

و A همانند تمرین ۱۱، B_R و B_N را پیدا کنید.

۲۷. فرض کنید H یک ماتریس بالا هسنبرگی $n \times n$ باشد و فرض کنید

$$H = QR$$

که در آن Q متعامد و R بالا مثلثی و توسط دورانهای گیونز به دست آمده باشد. ثابت کنید که Q نیز بالا هسنبرگی است.

۲۸. یک الگوریتم برای محاسبه AH تدوین نمایید که در آن A یک ماتریس دلخواه $m \times n$ و H یک ماتریس هاوس هولدر باشند. الگوریتم چند عمل ممیز شناور (flops) نیاز خواهد داشت؟ الگوریتم شما باید از ساختار H بهره گیرد.

۲۹. یک الگوریتم برای محاسبه AJ تدوین نمایید که در آن A از مرتبه $m \times n$ و J یک دوران گیونز است (الگوریتم شما باید از ساختار ماتریس J بهره گیرد) چند عمل ممیز شناور (flops) مورد نیاز خواهد بود.

۳۰. نشان دهید که تعداد اعمال ممیز شناور برای محاسبه R در تجزیه QR یک ماتریس $m \times n$ مانند A ($m \geq n$) با استفاده از دورانهای گیونز حدوداً $\frac{2}{3}n^2(m - \frac{n}{3})$ است.

۳۱. یک الگوریتم برای محاسبه

$$Q = H_1 H_2 \dots H_n$$

ارائه دهید، که در آن H_1 تا H_n ماتریسهای هاوس هولدر در تجزیه QR یک ماتریس $m \times n$ مانند A ، ($m > n$)، هستند. نشان دهید که الگوریتم می‌تواند با تقریباً $\frac{2}{3}(m^2n - mn^2 + n^3)$ عمل ممیز شناور (flops) پیاده‌سازی شود.

۳۲. نشان دهید که ماتریس متعامد

$$Q = Q_1^T Q_2^T \dots Q_{n-1}^T$$

که در آن Q_i حاصلضرب $(m-i)$ دوران گیونز است، همانند بخش ۵-۵-۱، می‌تواند با $\frac{2}{3}(m^2 - n^2)$ عمل ممیز شناور (flops) محاسبه شود.

۳۳. فرض کنید

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

تجزیه QR ماتریس A را با استفاده از روشهای زیر پیدا کنید.

(الف) روش هاوس هولدر

(ب) روش گیونز

نتایج را مقایسه کنید.

۳۴. هر دو روش کاهش هاوس هولدر و گیونز را در مورد ماتریس

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

به کار ببرید تا آن را با تبدیلات تشابهی به ماتریس هسنبرگی کاهش دهند. نتایج را مقایسه کنید.

۳۵. (الف) نشان دهید که $\frac{5}{3}n^3$ عمل ممیز شناور برای محاسبه ماتریس هسنبرگی H_u با استفاده از

روش هاوس هولدر لازم می‌باشد. راهنمایی:

$$\sum_{k=1}^{n-2} 2(n-k)^2 + \sum_{k=1}^{n-2} 2n(n-k) \approx \frac{2}{3}n^3 + n^2 = \frac{5}{3}n^3$$

(ب) نشان دهید که اگر ماتریس تبدیل P به طور صریح لازم باشد، $\frac{2}{3}n^3$ عمل ممیز شناور (flops) دیگر لازم خواهد بود.

(ج) تعداد اعمال ممیز شناور (flops) مربوط به کاهش به شکل هسنبیگی با استفاده از دورانهای گیونز را تعیین کنید.

(د) اگر A متقارن باشد، آنگاه نشان دهید که تعداد اعمال متناظر در قسمت (الف) برابر $\frac{2}{3}n^3$ است.

۳۶. با مفروض بودن یک ماتریس بالا هسنبیگی کاهش ناپذیر H ، نشان دهید که ماتریس X تعریف شده توسط $X = (e_1, He_1, \dots, H^{n-1}e_1)$ نامنفرد است و به قسمی است که $X^{-1}HX$ یک ماتریس همراه و به شکل بالا هسنبیگی است.
 (الف) چه مشکلات عددی برای این محاسبات وجود دارد.
 (ب) ماتریس زیر را به یک شکل همراه تبدیل کنید.

$$H = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 \times 10^{-5} & 4 & 4 & 4 \\ \cdot & 1 \times 10^{-3} & 1 & 2 \\ \cdot & \cdot & 1 & 1 \end{bmatrix}$$

۳۷. (الف) با مفروض بودن جفت (A, b) ، که در آن A یک ماتریس $n \times n$ و b یک بردار ستونی است، یک الگوریتم برای محاسبه یک ماتریس متعامد P تدوین نمایید به قسمی که ماتریس PAP^T برابر ماتریس بالا هسنبیگی H_{ii} باشد و Pb مضربی از بردار e_1 باشد.
 (ب) نشان دهید که H_{ii} کاهش ناپذیر است و b مضرب مخالف صفری از e_1 است اگر و فقط اگر $\text{rank}(b, Ab, \dots, A^{n-1}b) = n$.

(ج) الگوریتم قسمت (الف) را در مورد ماتریس و بردار زیر به کار ببرید.

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

بخشهای ۵-۷ و ۵-۸

۳۸. با مفروض بودن

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

یک ماتریس جایگشت P ، یک ماتریس متعامد Q ، و یک ماتریس بالا مثلثی R با استفاده از روش هاوس هولدر به قسمی پیدا کنید که $AP = QR$. ماتریس جایگشت P باید بر طبق معیار داده شده در کتاب انتخاب شود.

۳۹. اثباتی برای قضیه متعامد سازی کامل ارائه دهید (قضیه ۵-۷-۲).

۴۰. یک الگوریتم برای اصلاح تجزیه QR یک ماتریس A تدوین نمایید که توسط آن یک ستون برداشته شود.

برنامه‌های MATLAB و MATCOM و مسائل

توجه کنید که شما به برنامه‌های `givqr`، `compiv`، `housqr`، `phousmul` و `givhess` از MATLAB نیاز خواهید داشت.

۱. (الف) یک برنامه MATLAB به نام `elm(v)` بنویسید که یک ماتریس پایین مثلثی مقدماتی

E تولید کند به طوری که Ev مضربی از e_1 باشد. در اینجا v یک بردار n بعدی است.

(ب) یک برنامه MATLAB به نام `elmul(A, E)` بنویسید که حاصلضرب EA را محاسبه

کند، که در آن E یک ماتریس پایین مثلثی مقدماتی است. برنامه شما باید قادر باشد مزایای ساختار ماتریس E را در نظر بگیرد.

(ج) با استفاده از `elm` و `elmlu` یک برنامه MATLAB به نام `elmlu` بنویسید که تجزیه LU

یک ماتریس را در صورت وجود پیدا کند:

$$[L, U] = \text{elmlu}(A)$$

با استفاده از برنامه `elmlu` خودتان L و U را برای ماتریسهای زیر پیدا کنید

$$A = \begin{bmatrix} \cdot & \cdots & 1 & & \\ & & & & 1 \\ & & & & & \\ & & & & & & \\ & & & & & & & 1 \end{bmatrix}, \quad A = \begin{bmatrix} & & 1 & & \\ \cdot & \cdots & & & \\ & & & & 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 1.0 & & & & \\ & 1.0 & & & \\ & & 1.0 & & \\ & & & 1.0 & \\ & & & & 1.0 \end{bmatrix}, \quad A = \text{diag}(1, 2, 3)$$

ماتریس هیلبرت 5×5

اکنون برای هر حالت (i) حاصلضرب LU و $\|A - LU\|_F$ (ii) را پیدا کنید و نتایج را بنویسید.

۲. برنامه `elm1u` خودتان را برای منظور کردن محورگیری جزئی اصلاح کنید.

$$[M, U] = \text{elm1upp}(A)$$

برنامه خودتان را با هر یک از ماتریسهای مسأله ۱ آزمایش کنید. نتایج خودتان را با نتایج به دست آمده توسط برنامه MATLAB مقایسه کنید:

$$[L, U] = \text{lu}(A)$$

۳. یک برنامه MATLAB به نام `parpiv` برای محاسبه M و U به قسمی بنویسید که $MA = U$ ، از محورگیری جزئی استفاده کنید.

$$[M, U] = \text{parpiv}(A)$$

(این برنامه باید الگوریتم ۵-۲-۲ را پیاده سازی کند).

$U, M, \|MA - U\|_F$ و $\|MA - U\|_2$ را برای هر یک از ماتریسهای A مسأله ۱ چاپ کنید.

۴. از برنامه `compiv` از MATLAB استفاده کنید و U, Q, M و

$$\|MAQ - U\|_F \text{ و } \|MAQ - U\|_2$$

را برای هر یک از ماتریسهای مسأله ۱ چاپ کنید.

۵. (الف) یک برنامه MATLAB به نام `houzero` که یک ماتریس هاوس هولدر H تولید می کند، به قسمی بنویسید که Ha مضربی از e_1 باشد، که در آن a یک بردار n بعدی است.

(ب) با استفاده از `houzero` و `phousmul` (از MATLAB)، یک برنامه MATLAB به نام

`houqr` بنویسید که تجزیه QR ماتریس A را پیدا کند:

$$[Q, R] = \text{houqr}(A)$$

(این برنامه باید الگوریتم ۵-۴-۳ را پیاده سازی کند).

(ج) با استفاده از برنامه خودتان `houqr` ماتریسهای Q و R را به قسمی پیدا کنید که

$$A = QR$$

اکنون مقادیر

$$\|I - Q^T Q\|_F \quad (i)$$

$$\|A - QR\|_F \quad (ii)$$

را پیدا کنید و نتایج را با نتایج حاصل از استفاده از برنامه MATLAB زیر مقایسه کنید.

$$[Q, R] = qr(A)$$

(د) قسمت (ج) را با برنامه **givqr** از **MATCOM** که تجزیه QR را با استفاده از دورانه‌های گیونز به جای **housqr** انجام می‌دهد، تکرار کنید.

۶. برنامه **givqr(A)** از **MATCOM** را با هر یک از ماتریسهای مسأله ۱ اجرا کنید. سپس با به کار بردن $[Q, R] = qr(A)$ از **MATLAB** دوباره در مورد همان ماتریسها، منحصر به فرد بودن تجزیه QR را برای هر ماتریس A تحقیق کنید.

۷. با به کار بردن **givhess(A)** از **MATCOM** و تابع **hess(A)** از **MATLAB** در مورد هر یک از ماتریسهای مسأله ۱، قضیه QR ضمنی (قضیه ۵-۵-۳) (منحصر به فرد بودن کاهش هسبرگی)) را تحقیق کنید.

۸. با استفاده از نتایج مسأله ۵، یک پایه یکا متعامد برای $R(A)$ ، یک پایه یکا متعامد برای متمم متعامد $R(A)$ ، تصویر متعامد بر روی $R(A)$ ، و تصویر بر روی مکمل متعامد A را برای هر یک از ماتریسهای متعامد مسأله ۱ بیابید.

۹. برای نوشتن یک برنامه **MATLAB** به نام **housqrp** که تجزیه QR را با محورگیری ستونی برای یک ماتریس A محاسبه کند، محورگیری ستونی با نرم ماکزیمم را در **housqr** منظور کنید. برنامه خودتان را با هر یک از ماتریسها مسأله ۱ آزمایش کنید:

$$[Q, R, P] = \text{housqrp}(A)$$

نتایج خودتان را با نتایج حاصل از استفاده از برنامه **MATLAB** زیر مقایسه کنید.

$$[Q, R, P] = qr(A)$$

تذکر: برخی از برنامه‌هایی را که از شما خواسته شده است که بنویسید، نظیر **parpiv**، **houszero**، یا **housqr**، در **MATCOM** یا در ضمیمه B هستند. اما نوشتن برنامه‌ها توسط خودتان ایده خوبی است.

پیشنهادها برای مطالعه بیشتر

موضوعهای ارائه شده در این فصل موضوعهای استاندارد هستند و می توان آنها را در هر کتاب جبرخطی پیدا کرد. کتابهای Golub and Van Loan (1983, 1989) و G. W. Stewart (1973) منابع غنی جهت اطلاعات بیشتر در این زمینه هستند. در این فصل ما مکرراً از آنها استفاده کرده ایم.

بویژه کتاب محاسبات ماتریسی (*Matrix Computations*) حاوی یک الگوریتم مفصل برای تجزیه QR با محورگیری ستونی با استفاده از تبدیلات هاوس هولدر است.

کتاب Lawson and Hanson (1974) حاوی یک بحث عمیق در مورد مثلثی سازی با استفاده از تبدیلات هاوس هولدر و گیونز و تجزیه QR با استفاده از محورگیری ستونی است. (فصلهای ۱۰ و ۱۵). جزئیات تحلیل خطا روشهای هاوس هولدر و گیونز برای تجزیه QR و کاهش به شکل هسنبرگی در کتاب Wilkinson (1965) وجود دارد.

برای تحلیل خطای تجزیه QR با استفاده از تبدیلات گیونز و انواع تبدیلات گیونز کتاب Gentleman (1975) را ملاحظه کنید.

یک بحث زیبا در مورد تصویرهای متعامد در کتاب زیر ارائه شده است.

Philip E. Gill, Walter Murray, and Margaret H. Wright; *Numerical Linear Algebra and Optimization* (Reading, Mass.: Addison Wesley, 1991)

مقاله جدید Chandrasekaran and Ipsen (1994c) حاوی یک دید کامل از تحقیقات فعلی در مورد تجزیه QR آشکارساز - رتبه است.

برای مطالب بیشتر در مورد مسائل به هنگام درآوری و به سطح پایین تر درآوری مراجع زیر را ملاحظه کنید.

Daniel, Gragg, Kaufman, and Stewart (1976), Gill, Murray, and Saunders (1974), Bojanczyk, Brent, Van Dooren, and De Hoog (1987), and Nazareth (1989).

خواننده های کنجکاو ممکن است مایل به دیدن مقاله های اصلی Givens (1958) و Householder (1958) در مورد مثلثی سازی متعامد (یکانی) یک ماتریس باشند. بیوگرافی والنس گیونز، آکستون هاوس هولدر، و جیمز. اچ. ویلکینسون را می توان به ترتیب در مجلات زیر پیدا کرد.

SIAM News Letter, July, 1993,

SIAM News Letter, October, 1993,

SIAM Calendar 1994.

جوابهای عددی دستگاههای خطی

۱-۶ مقدمه

هدفهای اصلی این فصل مطالعه روشهای عددی برای حل دستگاههای خطی هستند. این مطالعه برخی نتایج نظری پایه‌ای در مورد وجود و منحصر به فرد بودن جوابها، بحثهای کاربردهای عملی که منجر به دستگاههای خطی می‌شوند، و تحلیل اختلالهای دستگاههای خطی را نیز شامل می‌شود.

مسئله حل دستگاه خطی

$$Ax = b$$

در دامنه وسیعی از کاربردها مطرح می‌شود. به عنوان یک حقیقت باید گفته شود که جوابهای عددی تقریباً همه مسائل مهندسی عملی و علوم کاربردی به طور معمول نیاز به جواب یک مسئله خطی دارند. (بخش ۳-۶ ملاحظه شود).

در این فصل ما فقط در مورد دستگاههای خطی نامنفرد بحث می‌کنیم. حالتی که در آن ماتریس A مربعی نیست یا دستگاه دارای بیش از یک جواب است در فصل ۷ مورد بررسی قرار می‌گیرد. یک روش که قاعده کرامر نامیده می‌شود، و در درسهای جبر خطی مقدماتی دوره کارشناسی تدریس می‌شود، از نقطه نظر نظری اهمیت دارد.

قاعده کرامر

فرض کنید A یک ماتریس نامنفرد از مرتبه n و b یک بردار n بعدی باشد. جواب دستگاه

$$Ax = b$$

توسط رابطه

$$x_i = \frac{\det(A_i)}{\det(A)}, \quad i = 1, \dots, n$$

داده می شود که در آن A_i ماتریس حاصل از قرار دادن بردار b به جای ستون i ام A می باشد

$$\text{و } x = (x_1, x_2, \dots, x_n)^T$$

تبصره: قاعده کرامر از نظر محاسباتی به هیچ وجه عملی نیست. برای مثال، حل یک دستگاه خطی با ۲۰ معادله و ۲۰ مجهول توسط روش کرامر، با استفاده از تعریف معمول دترمینان، به بیش از یک میلیون سال حتی با کامپیوترهای سریع، زمان نیاز دارد ((Forsythe, Malcom, and Moler (1977, 30)).

برای یک دستگاه $n \times n$ ، در حدود $O(n!)$ عمل ممیز شناور (flops) مورد نیاز خواهد بود.

دو نوع روش معمول برای محاسبات عددی استفاده می شوند.

۱. روشهای مستقیم

۲. روشهای تکراری

روشهای مستقیم از یک تعداد متناهی گام تشکیل می شوند که همگی باید برای هر روش مفروض قبل از به دست آمدن جواب، انجام شوند. از طرف دیگر روشهای تکراری مبتنی بر محاسبه یک دنباله از تقریبات برای جواب x هستند، و یک کاربر می تواند هر گاه که یک جواب با دقت معین به دست آمد، یا یک تعداد معین تکرار کامل شد، متوقف گردد. هم روشهای مستقیم و هم روشهای تکراری در این فصل مورد مطالعه قرار خواهند گرفت.

در بخش ۶-۲ نتایج نظری اساسی را (بدون اثبات) برای وجود منحصر به فرد بودن جواب برای دستگاههای خطی بیان خواهیم کرد. در بخش ۶-۳ در مورد چندین کاربرد مهندسی که منجر به مسائل دستگاههای خطی می شوند، بحث خواهیم کرد، و در بخش ۶-۴ روشهای مستقیم را برای حل دستگاههای خطی توصیف خواهیم کرد.

همان گونه که در فصل ۳ ملاحظه کردیم، حساسیت جواب یک دستگاه خطی که ناشی از اختلالهای داده های مسأله می باشد، توسط عدد وضعیت ماتریس A اندازه گیری می شود. در این فصل به طور مفصل تحلیل حساسیت و اثر عدد وضعیت را بر روی دقت جواب محاسبه شده ارائه خواهیم کرد

(بخشهای ۶-۶ و ۶-۷).

در موارد زیادی، بویژه در مسائل بدوضع، لازم است که جواب محاسبه شده را با هزینه کم با استفاده از تجزیه‌ای که از قبل برای محاسبه جواب اصلی به دست آمده، تصفیه نماییم. فرایند به روش تصفیه تکراری معروف است. این روش را در بخش ۶-۹ بعد از بحثهایمان در مورد اختلال مؤلفه به مؤلفه (بخش ۶-۸) بحث خواهیم کرد. در بخش ۶-۱۰ روشهای تکراری: روشهای ژاکوبی، گاوس سایدل، SOR ، گرادیان مزدوج و $GMRES$ را مورد مطالعه قرار خواهیم داد.

محاسبه دترمینان و معکوس یک ماتریس بندرت در عمل مورد نیاز خواهند بود. بیشتر مسائلی که ظاهراً حاوی معکوس هستند می‌توانند به صورت جوابهای مسائل دستگاههای خطی دوباره فرمول‌بندی شوند. محاسبه معکوس یک ماتریس حدوداً سه برابر گران‌تر از حل یک دستگاه خطی است. ولیکن کاربردهای آماری و مهندسی وجود دارند که محاسبه صریح معکوس را نیاز دارند، زیرا عناصر ماتریس معکوس دارای معنای فیزیکی هستند. بنابراین بحث مختصری در مورد محاسبه معکوس و دترمینان ارائه خواهیم داد.

ابزار و مفاهیم مهم زیر که قبلاً توضیح داده شده‌اند، برای یادگیری و فهم روان این فصل لازم خواهند بود. ماتریسهای خاص (بخش ۱-۴)، مفاهیم و نتایج در مورد نرمهای برداری و ماتریسی (بخش ۱-۷)، همگرایی دنباله ماتریسی و ماتریسهای همگرا (بخش ۱-۷-۳)، تجزیه‌های LU (بخشهای ۱-۲-۵، ۲-۲-۵، ۳-۲-۵)، مفهوم عامل رشد (بخش ۳-۵)، تجزیه QR یک ماتریس (بخشهای ۱-۴-۵ و ۱-۵-۵)، مفاهیم وضعیت و پایداری (بخشهای ۲-۳، ۳-۳ و ۴-۳)، و برخی اطلاعات اساسی در مورد معادلات دیفرانسیل معمولی و جزئی.

۶-۲ نتایج اساسی در مورد وجود، منحصر به فرد بودن، و پایایی جوابها

دستگاه ناهمگن m معادله و n مجهول زیر را در نظر بگیرید.

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

⋮

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m$$

به شکل ماتریسی، دستگاه به صورت

$$Ax = b$$

نوشته می‌شود که در آن

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, \quad b \neq 0.$$

با مفروض بودن یک ماتریس $m \times n$ بعدی A و یک بردار m بعدی b ، اگر یک بردار x وجود داشته باشد که در $Ax = b$ صدق نماید، آنگاه گفته می‌شود که دستگاه سازگار است. در غیر این صورت ناسازگار است. طبیعی است سؤال شود که آیا دستگاه داد شده سازگار است و اگر سازگار است، چند جواب وجود دارد و چه موقع جواب منحصر به فرد است و غیره. ما قضیه زیر را بیان می‌کنیم.

قضیه ۶-۲-۱ قضیه وجود و منحصر به فرد بودن برای یک دستگاه ناهمگن

۱. دستگاه $Ax = b$ سازگار است اگر و فقط اگر $b \in R(A)$ ، به بیان دیگر اگر $\text{rank}(A) = \text{rank}(A, b)$.
۲. اگر دستگاه سازگار باشد و ستونهای A مستقل خطی باشند، آنگاه جواب منحصر به فرد است.
۳. اگر دستگاه سازگار باشد و ستونهای A وابسته خطی باشند، آنگاه دستگاه دارای تعداد نامتناهی جواب است.

دستگاههای همگن

اگر داشته باشیم $b = 0$ ، آنگاه دستگاه $Ax = 0$ یک دستگاه همگن نامیده می‌شود. یک دستگاه همگن همواره جواب دارد؛ یعنی $x = 0$. این جواب بدیهی است.

قضیه ۶-۲-۲ قضیه وجود برای یک دستگاه همگن

۱. دستگاه $Ax = 0$ دارای یک جواب غیربدیهی است اگر و فقط اگر ستونهای A وابسته خطی باشند. اگر $Ax = 0$ دارای یک جواب غیربدیهی باشد، دارای تعداد نامتناهی جواب خواهد بود.
۲. اگر $m = n$ آنگاه $Ax = 0$ دارای یک جواب غیربدیهی خواهد بود اگر و فقط اگر A منفرد باشد.

قضیه ۶-۲-۳ قضیه پایایی جواب. یک جواب دستگاه سازگار

$$Ax = b$$

تحت هیچ کدام از اعمال زیر تغییر نمی‌کند.

۱. دو معادله دلخواه تعویض شوند.
۲. یک معادله در یک عدد مخالف صفر ضرب شود.

۳. مضرب مخالف صفری از یک معادله به معادله دیگر اضافه شود.

دو دستگاه که یکی از دیگری با به کار بردن هر یک از اعمال فوق به دست آید، دستگاههای معادل نامیده می شوند. قضیه ۶-۲-۳ بیان می کند که دو دستگاه معادل دارای مجموعه جوابهای یکسان هستند.

۶-۳ برخی کاربردهای به وجود آورنده مسائل دستگاههای خطی

شاید این اغراق نباشد که مسائل دستگاههای خطی تقریباً در همه کاربردهای عملی پیش می آیند. در اینجا مثالهایی از مهندسی برق، مکانیک، شیمی، و راه و ساختمان ارائه می دهیم. با یک مسأله از یک مدار الکتریکی شروع می کنیم.

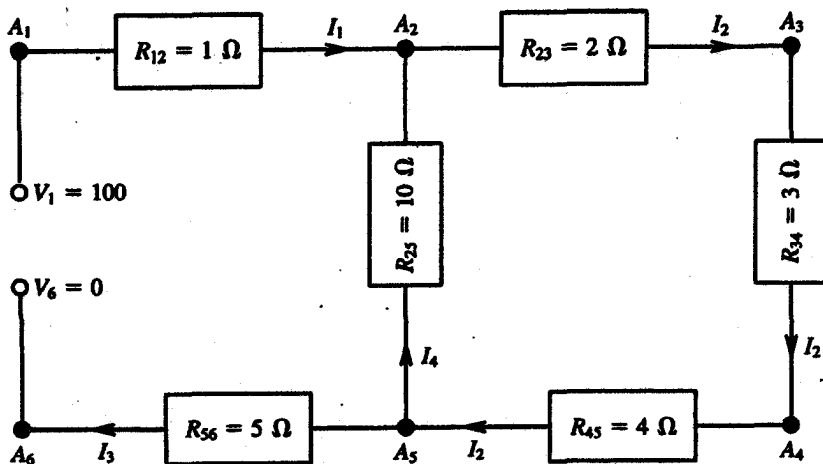
۶-۳-۱ یک مسأله مدار الکتریکی

نمودار یک مدار الکتریکی نشان داده شده در شکل ۶-۱ را ملاحظه کنید. می خواهیم مقدار جریان بین گره های $A_1, A_2, A_3, A_4, A_5, A_6$ را تعیین کنیم. قانون کیرشهوف در مورد جریان بیان می دارد که مجموع جبری همه جریانهای وارد شونده در یک گره باید برابر صفر باشد. با به کار بردن این قانون در گره A_2 داریم:

$$I_1 - I_2 + I_4 = 0 \quad (۱-۳-۶)$$

در گره A_5 :

$$I_2 - I_3 - I_4 = 0 \quad (۲-۳-۶)$$



شکل ۶-۱

در گره A_3 :

$$I_7 - I_2 = 0 \quad (3-3-6)$$

در گره A_4 :

$$I_7 - I_2 = 0 \quad (4-3-6)$$

اکنون آفت ولتاژ را حول هر حلقه بسته از جریان، $A_1 A_2 A_3 A_4 A_5 A_6 A_1$ ، $A_1 A_2 A_3 A_4 A_5 A_6 A_1$ ، $A_1 A_2 A_3 A_4 A_5 A_6 A_1$ در نظر بگیرید. قانون کیرشهوف در مورد ولتاژ بیان می‌کند که خالص آفت ولتاژ حول هر حلقه بسته صفر است. بنابراین در حلقه $A_1 A_2 A_3 A_4 A_5 A_6 A_1$ ، با جایگذاری مقادیر مقاومتها و ولتاژها داریم

$$I_1 + 9I_2 + 5I_3 = 100 \quad (5-3-6)$$

به طور مشابه در $A_1 A_2 A_3 A_4 A_5 A_6 A_1$ و $A_1 A_2 A_3 A_4 A_5 A_6 A_1$ ، به ترتیب داریم:

$$I_1 - 10I_2 + 5I_3 = 100 \quad (6-3-6)$$

$$9I_2 + 10I_3 = 0 \quad (7-3-6)$$

توجه کنید که $(5-3-6) + (7-3-6) = (6-3-6)$. بنابراین چهار معادله و چهار مجهول داریم

$$I_1 - I_2 + I_3 = 0 \quad (8-3-6)$$

$$I_2 - I_3 - I_4 = 0 \quad (9-3-6)$$

$$I_1 - 10I_2 + 5I_3 = 100 \quad (10-3-6)$$

$$9I_2 + 10I_3 = 0 \quad (11-3-6)$$

معادلات (۸-۳-۶) تا (۱۱-۳-۶) را می‌توان به صورت زیر نوشت:

$$\begin{pmatrix} 1 & -1 & 0 & 1 \\ 0 & 1 & -1 & -1 \\ 1 & 0 & 5 & -10 \\ 0 & 9 & 0 & 10 \end{pmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 100 \\ 0 \end{bmatrix}$$

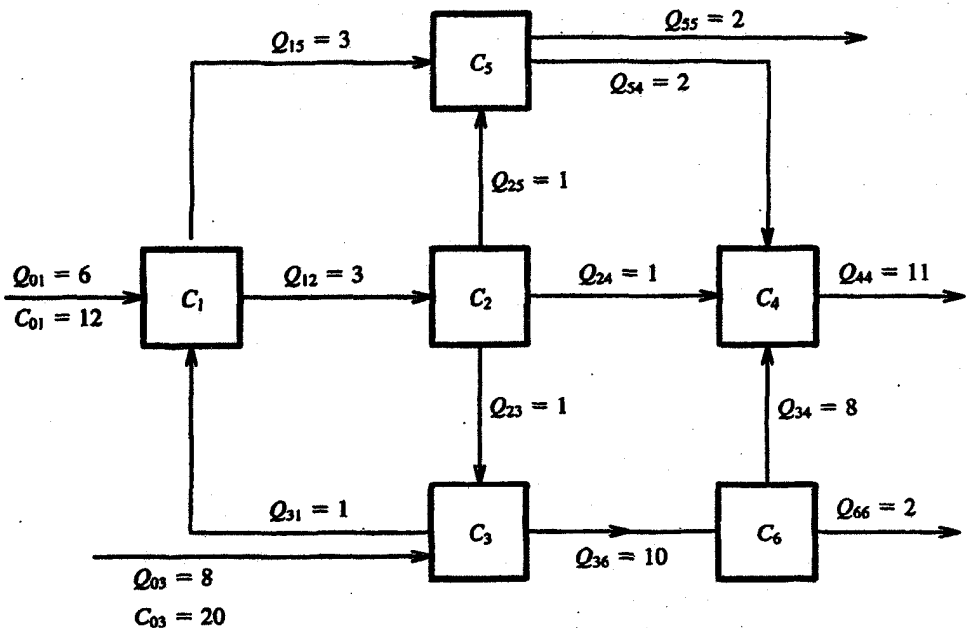
جواب این دستگاه جریان بین گره‌ها را ارائه می‌دهد.

۲-۳-۶ تحلیل یک دستگاه پردازش متشکل از راکتورهای بهم متصل

بسیاری از مدل‌های ریاضی بر قوانین بقاء نظیر بقاء جرم، بقاء اندازه حرکت، بقاء انرژی متبنی هستند. به عنوان مثال یک دستگاه پردازش شیمیایی متشکل از شش راکتور شیمیایی بهم متصل (شکل ۲-۶) را با نرخهای جریان جرمی مختلف از یک جزء از یک مخلوط به داخل و خارج راکتور در نظر بگیرید. ما علاقه‌مند به دانستن غلظت مخلوط در راکتورهای مختلف هستیم. مثال اینجا شبیه به مثال ارائه شده در مرجع (Chapra and Canale (1988, pp, 295-8) می‌باشد. کاربرد بقاء جرم برای همه این راکتورها منجر به یک دستگاه معادلات خطی به صورت زیر می‌شود که حاوی پنج معادله و پنج مجهول است. جواب دستگاه غلظت جزء را در هر یک از این راکتورها ارائه می‌دهد.

حالت پایا، راکتور مخلوط کامل. ابتدا یک راکتور با دو جریان به داخل و یک جریان به خارج مانند شکل ۳-۶ را در نظر بگیرید. کاربرد بقاء حالت پایا جرم برای یک جزء در مخلوط ارائه می‌دهد.

$$m_1 + m_2 = m_3 \tag{۱۲-۳-۶}$$



شکل ۲-۶

با توجه به این که

$$m_i = Q_i \cdot C_i$$

که در آن

m_i = نرخ جریان جرمی جزء در مقاطعهای مجرای ورودی و خروجی i , $i = 1, 2, 3$
بر حسب mg/min

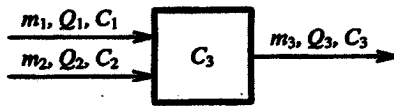
Q_i = نرخ جریان از لحاظ حجمی در مقطع i , $i = 1, 2, 3$ بر حسب m^3/min

C_i = چگالی یا غلظت در مقطع i , $i = 1, 2, 3$ بر حسب mg/m^3

از (۱۲-۳-۶) به دست می آوریم

$$Q_1 C_1 + Q_2 C_2 = Q_3 C_3 \quad (13-3-6)$$

برای نرخهای جریان مجرای ورودی مفروض و غلظتها، غلظت مجرای خروجی C_3 را می توان از معادله (۱۳-۳-۶) به دست آورد. تحت عمل حالت پایا، این غلظت مجرای خروجی همچنین غلظت همگن داخل راکتور را نشان می دهد. چنین اطلاعاتی برای طراحی راکتوری که مخلوطهایی را با یک غلظت مشخص تولید کند، ضروری است، برای جزئیات مرجع (Chapra and Canale (1988 را ملاحظه کنید.



شکل ۳-۶

اکنون با رجوع به شکل ۲-۶، که در آن دستگاهی متشکل از ۶ راکتور را ملاحظه می کنیم، معادلات زیر را داریم (به دست آوردن هر یک از معادلات نظیر به دست آوردن معادله (۱۳-۳-۶) است)
برای راکتور ۱:

$$6C_1 - C_3 = 72 \quad (14-3-6)$$

برای راکتور ۲:

$$3C_1 - 3C_2 = 0 \quad (15-3-6)$$

برای راکتور ۳:

$$-C_2 + 11C_3 = 160 \quad (16-3-6)$$

برای راکتور ۴:

$$C_2 - 11C_4 + 2C_5 + 8C_6 = 0 \quad (17-3-6)$$

برای راکتور ۵:

$$3C_1 + 2C_2 - 4C_5 = 0 \quad (18-3-6)$$

برای راکتور ۶:

$$10C_3 - 10C_6 = 0 \quad (19-3-6)$$

معادلات (۱۴-۳-۶) تا (۱۹-۳-۶) را می‌توان به شکل ماتریسی زیر نوشت

$$\begin{bmatrix} 6 & 0 & -1 & 0 & 0 & 0 \\ 3 & -3 & 0 & 0 & 0 & 0 \\ 0 & -1 & 11 & 0 & 0 & 0 \\ 0 & 1 & 0 & -11 & 2 & 8 \\ 3 & 1 & 0 & 0 & -4 & 0 \\ 0 & 0 & 10 & 0 & 0 & -10 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \end{bmatrix} = \begin{bmatrix} 72 \\ 0 \\ 160 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (20-3-6)$$

یا

$$AC = D$$

i امین مؤلفه بردار مجهول C غلظت مخلوط را در راکتور i دستگاه نشان می‌دهد. جواب دستگاه غلظتهای مورد نظر را ارائه می‌دهد.

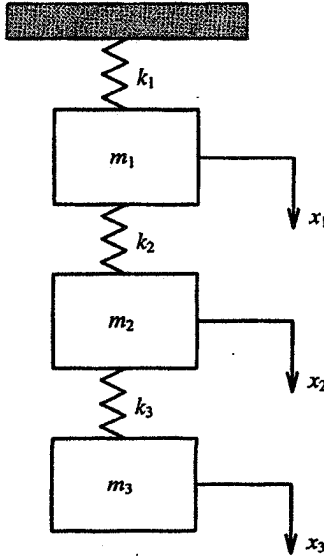
۳-۳-۶ دستگاههای خطی حاصل از معادلات دیفرانسیل معمولی (طرح تفاضل متناهی)

یک مطالعه موردی بر روی یک مسأله جرم فتر

یک دستگاه شامل سه جرم که به طور عمودی توسط یک سری فتر، به صورت نشان داده شده در

شکل ۴-۶، آویزان شده‌اند، در نظر بگیرید. در این شکل k_1 ، k_2 ، و k_3 ثابتهای فتر، و x_1 ، x_2 ، و x_3

تغییر مکانهای هر فتر از حالت تعادلشان می‌باشند.



شکل ۶-۴

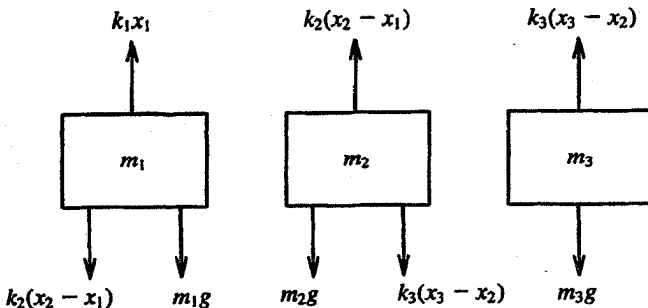
نمودار جسم - آزاد را برای این جرمها می توان به صورت شکل ۶-۵ نشان داد. با ارجاع به شکل ۶-۵، معادلات حرکت، بنابر قانون دوم نیوتن عبارتند از:

$$m_1 \frac{d^2 x_1}{dt^2} = k_2(x_2 - x_1) + m_1 g - k_1 x_1$$

$$m_2 \frac{d^2 x_2}{dt^2} = k_3(x_3 - x_2) + m_2 g - k_2(x_2 - x_1)$$

$$m_3 \frac{d^2 x_3}{dt^2} = m_3 g - k_3(x_3 - x_2)$$

فرض کنید ما علاقه مند به دانستن تغییر مکان این فنرها برای هنگامی هستیم که دستگاه سرانجام به حالت تعادل بر می گردد یعنی هنگامی که دستگاه ساکن می شود.



شکل ۶-۵

با قرار دادن مشتقهای مرتبه دوم برابر صفر، به دست می آوریم

$$k_1 x_1 + k_2 (x_1 - x_2) = m_1 g$$

$$k_2 (x_2 - x_1) + k_3 (x_2 - x_3) = m_2 g$$

$$k_3 (x_3 - x_2) = m_3 g$$

این دستگاه معادلات با سه مجهول x_1 ، x_2 و x_3 را می توان به شکل زیر نوشت

$$\begin{bmatrix} k_1 + k_2 & -k_2 & 0 \\ -k_2 & k_2 + k_3 & -k_3 \\ 0 & -k_3 & k_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} m_1 g \\ m_2 g \\ m_3 g \end{bmatrix}$$

یا

$$Kx = w$$

ماتریس

$$K = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 \\ -k_2 & k_2 + k_3 & -k_3 \\ 0 & -k_3 & k_3 \end{bmatrix}$$

ماتریس سختی نامیده می شود. جواب دستگاه تغییر مکانهای لازم را ارائه می دهد.

۶-۳-۴ دستگاههای خطی حاصل از معادلات دیفرانسیل جزئی

یک مطالعه موردی بر روی توزیع دما

بسیاری از مسائل مهندسی توسط معادلات دیفرانسیل جزئی مدل بندی می شوند. راههای عددی برای این معادلات نوعاً نیاز به گسسته سازی نقطه به نقطه یا قطعه به قطعه دارند. این فرایند گسسته سازی به نوبه خود موجب به وجود آمدن دستگاههای خطی از انواع بسیار جالب می شود. این مطلب را با یک مسأله از نظریه انتقال گرما توضیح می دهیم.

یک هدف اصلی در یک مسأله انتقال گرما تعیین توزیع دما $T(x, y, z, t)$ در یک محیط می باشد که از شرایط مرزی حاکم بر سطح محیط حاصل می گردد. به محض آن که این توزیع دما معلوم شود، شار گرما یا نرخ انتقال گرما برای هر واحد سطح در هر نقطه از محیط می تواند از قانون فوریه که به صورت زیر بیان می شود، محاسبه شود.

$$q_{x_i}'' = -K \frac{\partial T}{\partial x_i}$$

در اینجا x_i یا x یا y ، یا z است؛ q''_{x_i} شار گرما یا نرخ انتقال گرما، برای هر واحد سطح در جهت x_i می‌باشد؛ $\sigma T/\partial x_i$ گرادیان دما در جهت x_i است، و ثابت مثبت K میزان رسانایی حرارتی ماده نامیده می‌شود.

معادله پخش گرما یا معادله گرما که حاکم بر توزیع دما است با اعمال بقاء انرژی بر روی یک عنصر دیفرانسیلی بی‌نهایت کوچک به دست می‌آید، که از آن رابطه زیر را در مختصات دکارتی به دست می‌آوریم:

$$\frac{\partial}{\partial x}(K\frac{\partial T}{\partial x}) + \frac{\partial}{\partial y}(K\frac{\partial T}{\partial y}) + \frac{\partial}{\partial z}(K\frac{\partial T}{\partial z}) + q = \rho C_p \quad (21-3-6)$$

در اینجا ρ چگالی است، C_p گرمای ویژه، و \dot{q} انرژی تولید شده در واحد حجم می‌باشد. اغلب ممکن است با یک شکل ساده شده معادله (۲۱-۳-۶) کارکنیم. برای مثال، اگر میزان رسانایی حرارتی، K ، ثابت باشد، معادله حرارت به صورت زیر می‌باشد

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} + \frac{q}{K} = \frac{1}{\alpha} \frac{\partial T}{\partial t} \quad (22-3-6)$$

که در آن $\alpha = K/(\rho C_p)$ یک خاصیت ترموفیزیکی معروف به پخشندگی گرمایی است.

تحت شرایط حالت پایا، تغییرات ذخیره انرژی می‌تواند موجود نباشد؛ یعنی، جمله حالت ناپایا $\partial T/\partial t$ می‌تواند حذف شود و معادله (۲۱-۳-۶) به معادله پواسون سه بعدی تقلیل یابد:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} + \frac{q}{K} = 0 \quad (23-3-6)$$

اگر انتقال گرما دو بُعدی (مثلاً در جهتهای x و y) باشد و تولید انرژی وجود نداشته باشد، آنگاه معادله گرما به معادله لاپلاس تقلیل می‌یابد

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0 \quad (24-3-6)$$

اگر انتقال گرما ناپایا و یک بُعدی بدون تولید انرژی باشد، آنگاه معادله گرما به معادله زیر تقلیل می‌یابد

$$\frac{\partial^2 T}{\partial x^2} = \frac{1}{\alpha} \frac{\partial T}{\partial t} \quad (25-3-6)$$

بیشتر اوقات وضعیتهای عملی وجود دارند که در آنها شرایط هندسی یا مرزی به نحوی هستند که یک

جواب تحلیلی نمی‌توان به دست آورد. در چنین حالتی بهترین راه حلها روشهای تفاضل متناهی یا عنصر متناهی هستند که کاملاً برای کامپیوترها مناسب می‌باشند.

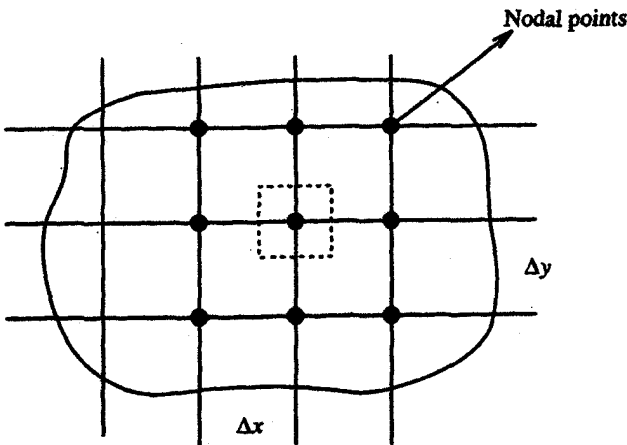
طرح تفاضل متناهی

یک طرح معروف برای حل یک معادله دیفرانسیل جزئی استفاده از تفاضلهای متناهی است. ایده این است که معادله دیفرانسیل جزئی را با قرار دادن تقریبهایی، یعنی تفاضلهای متناهی به جای مشتقهای جزئی، گسسته‌سازی کنیم. این طرح را با معادله لاپلاس در زیر توضیح می‌دهیم.

اجازه دهید که یک ناحیه دو بُعدی را به ناحیه‌های کوچک‌تر با نمو‌های مفروض Δx و Δy در جهت‌های x و y ، به صورت نشان داده شده در شکل ۶-۶، تقسیم کنیم. هر نقطه گره‌ای توسط یک طرح شماره‌گذاری i و j ، که در آن i نمو x و j نمو y را نشان می‌دهند، به صورت نشان داده شده در شکل ۶-۷، مشخص می‌گردد. فرض می‌شود که توزیع دما در محیط توسط دمای نقاط گره‌ای نمایش داده شوند. دما در هر نقطه گره‌ای (x_i, y_j) (که به طور نمادی توسط (i, j) ، همانند شکل ۶-۷)، نشان داده می‌شود (میانگین دما ناحیه احاطه شده توسط خط چین می‌باشد).

یک معادله تفاضلی متناهی مناسب برای گره‌های داخلی یک دستگاه دوبعدی پایا می‌تواند با ملاحظه معادله لاپلاس در نقطه گره‌ای i و j ، به صورت زیر به دست آید.

$$\frac{\partial^2 T}{\partial x^2} \Big|_{i,j} + \frac{\partial^2 T}{\partial y^2} \Big|_{i,j} = 0$$



شکل ۶-۶

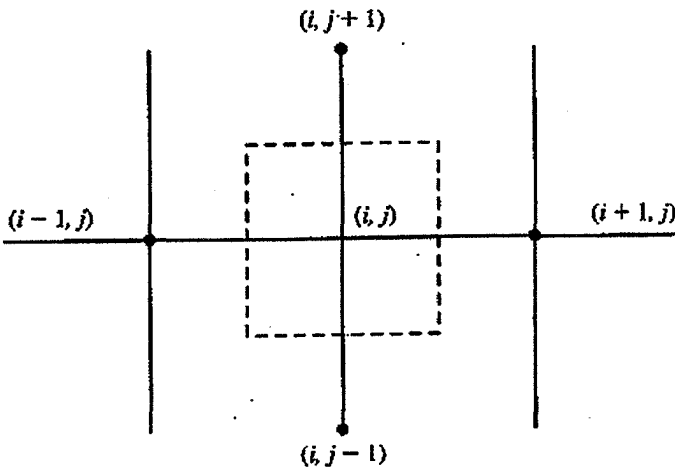
مشتقهای دوم در نقطه گره‌ای (i, j) می‌توانند به صورت زیر تقریب زده شوند (که از سری تیلور به دست می‌آیند).

$$\frac{\partial^2 T}{\partial x^2} \Big|_{i,j} \cong \frac{T_{i+1,j} - 2T_{i,j} + T_{i-1,j}}{(\Delta x)^2} \quad (27-3-6)$$

$$\frac{\partial^2 T}{\partial y^2} \Big|_{i,j} \cong \frac{T_{i,j+1} - 2T_{i,j} + T_{i,j-1}}{(\Delta y)^2} \quad (28-3-6)$$

سپس از (۲۶-۳-۷) نتیجه می‌گیریم

$$\frac{T_{i+1,j} - 2T_{i,j} + T_{i-1,j}}{(\Delta x)^2} + \frac{T_{i,j+1} - 2T_{i,j} + T_{i,j-1}}{(\Delta y)^2} = 0$$



شکل ۶-۷

با فرض $\Delta x = \Delta y$ ، تقریب تفاضل متناهی معادله لاپلاس برای ناحیه‌های داخلی می‌تواند به صورت زیر بیان شود.

$$T_{i,j+1} + T_{i,j-1} + T_{i+1,j} + T_{i-1,j} - 4T_{i,j} = 0 \quad (29-3-6)$$

$$4T_{i,j} - T_{i-1,j} - T_{i,j-1} - T_{i+1,j} - T_{i,j+1} = 0 \quad (30-3-6)$$

تقریبهای از مرتبههای بالاتر دقیقتر برای گرههای داخلی و گرههای مرزی نیز به طریق مشابه به دست می‌آیند.

مثال ۶-۳-۱ یک صفحه دو بُعدی مستطیلی ($0 \leq y \leq 1, 0 \leq x \leq 1$) تحت تأثیر شرایط مرزی دمای یکنواخت مانند شکل ۸-۶ قرار می‌گیرد (سطح بالایی در $100^\circ C$ و سایر سطوحها در $0^\circ C$ نگهداری می‌شوند)؛ یعنی $T(x, 0) = 0$ ، $T(1, y) = 0$ ، $T(0, y) = 0$ و $T(x, 1) = 100^\circ C$. فرض کنید ما فقط به مقادیر دما در نه نقطه گره‌ای داخلی (x_i, y_j) علاقمند باشیم، که در آن $x_i = i\Delta x$ و $y_j = j\Delta y$ و $i, j = 1, 2, 3$ با $\Delta x = \Delta y = \frac{1}{4}$. ولیکن برای ساده کردن مسأله، فرض می‌کنیم که مسأله متقارن باشد، یعنی فرض می‌کنیم که $T_{3,1} = T_{1,1}$ و $T_{3,2} = T_{1,2}$ ، $T_{3,3} = T_{1,3}$ که $(T_{1,1}, T_{1,2}, T_{1,3})$ و $(T_{3,1}, T_{3,2}, T_{3,3})$ از معادله (۶-۳-۳۰)، سپس داریم:

$$4T_{1,1} - 0 - 100 - T_{3,1} - T_{1,2} = 0$$

$$4T_{3,1} - T_{1,1} - 100 - T_{1,1} - T_{3,2} = 0$$

$$4T_{1,2} - 0 - T_{1,1} - T_{3,2} - T_{1,3} = 0$$

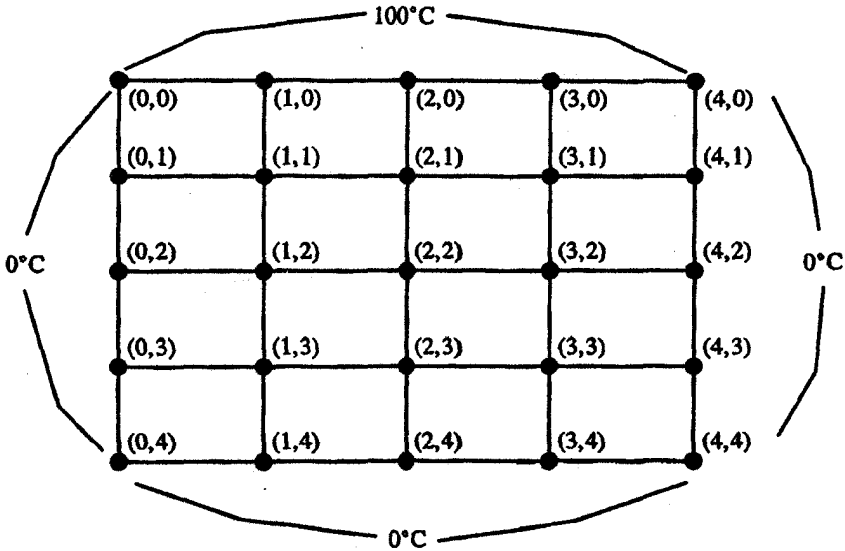
$$4T_{3,2} - T_{1,2} - T_{3,1} - T_{1,2} - T_{3,3} = 0$$

$$4T_{1,3} - 0 - T_{1,2} - T_{3,3} - 0 = 0$$

$$4T_{3,3} - T_{1,3} - T_{3,2} - T_{1,3} - 0 = 0$$

بعد از مرتب سازی مناسب، این معادلات می‌توانند به شکل زیر نوشته شوند.

$$\begin{bmatrix} 4 & -1 & -1 & 0 & 0 & 0 \\ -2 & 4 & 0 & -1 & 0 & 0 \\ -1 & 0 & 4 & -1 & -1 & 0 \\ 0 & -1 & -2 & 4 & 0 & -1 \\ 0 & 0 & -1 & 0 & 4 & -1 \\ 0 & 0 & 0 & -1 & -2 & 4 \end{bmatrix} \begin{bmatrix} T_{1,1} \\ T_{3,1} \\ T_{1,2} \\ T_{3,2} \\ T_{1,3} \\ T_{3,3} \end{bmatrix} = \begin{bmatrix} 100 \\ 100 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$



شکل ۶-۸

جواب این دستگاه، دما را در نقاط گره‌ای ارائه خواهد داد.

۵-۳-۶ دستگاه‌های خطی خاص مطرح در کاربردها

کاربردهای عملی زیادی منجر به دستگاه‌های خطی با خواص و ساختار خاص، نظیر سه‌قطری، غالب قطری، معین مثبت، و سه‌قطری بلوکی می‌شوند. روش‌های حل برای این دستگاه‌های خاص در بخش ۶-۴-۷ توصیف می‌شوند. ابتدا وضعیت‌ی را که منجر به یک دستگاه سه‌قطری می‌شود، بیان می‌کنیم.

دستگاه‌های سه‌قطری

هدایت پایای گرما را در حالت یک بُعدی در یک سیم، به صورت نشان داده شده در زیر، در نظر بگیرید. در یک چنین حالتی، دما نسبت به زمان ثابت می‌ماند. پس معادله عبارت است از:

$$\frac{\partial^2 T}{\partial x^2} = q/K = f(x)$$

$$T_0 = \alpha \quad \begin{array}{ccccccc} & \bullet & & \bullet & & \bullet & & \bullet & & \bullet \\ & x_0 & & x_1 & & x_2 & & x_3 & & x_4 \end{array} \quad T_4 = \beta$$

$|\Delta x|$

فرض کنید دامنه مسأله $0 \leq x \leq 1$ باشد. دامنه را به چهار قطعه با طول مساوی مثلاً Δx تقسیم می‌کنیم. بنابراین $\Delta x = 0.25$. T در $x = i\Delta x$ توسط T_i نمایش داده می‌شود. فرض کنید که مثلاً دما را

در نقاط انتهایی $x = 0$ و $x = 1$ بشناسیم؛ یعنی:

$$T_0 = \alpha$$

$$T_4 = \beta$$

در این صورت اینها شرایط مرزی مسأله هستند.

با استفاده از یک طرح شماره گذاری ماندقبل، دما در هر نقطه داده شده توسط رابطه زیر ارائه می شود.

$$T_{i+1} - 2T_i + T_{i-1} = (\Delta x)^2 f(x_i)$$

از این معادله، دما در هر گره $x_0 = 0$ ، $x_1 = \Delta x$ ، $x_2 = 2\Delta x$ ، $x_3 = 3\Delta x$ ، $x_4 = 1$ ، به صورت زیر محاسبه می شود

$$T_0 = \alpha \quad \text{در } x = 0 \quad \text{(مفروض است)}$$

$$T_0 + 2T_1 + T_2 = (\Delta x)^2 f(\Delta x) \quad \text{در } x = \Delta x$$

$$T_1 - 2T_2 + T_3 = (\Delta x)^2 f(2\Delta x) \quad \text{در } x = 2\Delta x$$

$$T_2 - 2T_3 + T_4 = (\Delta x)^2 f(3\Delta x) \quad \text{در } x = 3\Delta x$$

$$T_4 = \beta \quad \text{در } x = 1 \quad \text{(مفروض است)}$$

به شکل ماتریسی، این معادلات می توانند به صورت زیر نوشته شوند.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} T_0 \\ T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix} = \begin{bmatrix} \alpha \\ (\Delta x)^2 f(\Delta x) \\ (\Delta x)^2 f(2\Delta x) \\ (\Delta x)^2 f(3\Delta x) \\ \beta \end{bmatrix}$$

ماتریس این دستگاه سه قطری است.

دستگاههای متقارن سه قطری و غالب قطری

برای این که ببینیم که چگونه چنین دستگاههایی مطرح می شوند، اکنون هدایت ناپایا گرما را در نظر می گیریم. این شرط ایجاب می کند که دما T با زمان t تغییر کند. معادله گرما در این حالت به صورت زیر است.

$$\frac{1}{\alpha} \frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2}$$

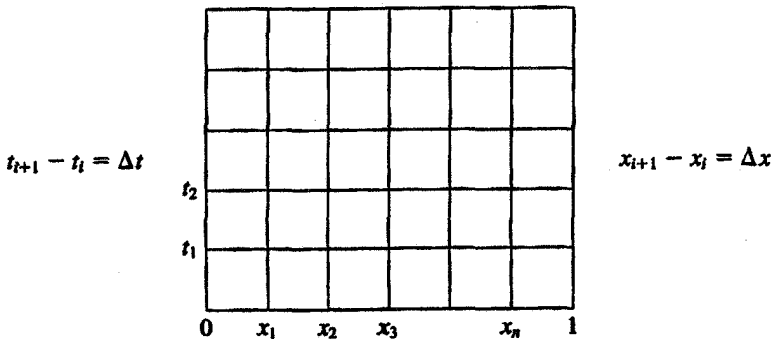
شبكة واقع در صفحه (x, t) را با فاصله‌های Δx در جهت محور x ها و با فاصله‌های Δt در جهت محور t ها، همانند شکل ۶-۹، تقسیم می‌کنیم. فرض کنید دما در نقطه گره‌ای $x_i = i\Delta x$ و $t_j = j\Delta t$ همانند قبل، با T_{ij} نمایش داده شود. با تقریب زدن $\frac{\partial T}{\partial t}$ و $\frac{\partial^2 T}{\partial x^2}$ با تفاضلات متناهی ضمنی

$$\frac{\partial T}{\partial t} \approx \frac{1}{\Delta t}(T_{i,j+1} - T_{i,j})$$

$$\frac{\partial^2 T}{\partial x^2} \approx \frac{1}{(\Delta x)^2}(T_{i+1,j+1} - 2T_{i,j+1} + T_{i-1,j+1})$$

معادله تفاضلی زیر را که مشابه معادله گرما می‌باشد، به دست می‌آوریم.

$$(1 + \gamma C)T_{i,j+1} - C(T_{i+1,j+1} + T_{i-1,j+1}) = T_{i,j}, \quad i = 1, 2, \dots, n$$



شکل ۶-۹

در اینجا

$$C = \alpha \frac{\Delta t}{(\Delta x)^2}$$

این معادلات ما را قادر می‌سازند که دما را در یک گام زمانی $k + 1 = j$ با دانستن دما در گام زمانی قبلی $k = j$ تعیین کنیم.

$$(1 + \gamma C)T_{1,k+1} - CT_{2,k+1} = CT_{0,k+1} + T_{1,k} \quad \text{به ازای } i = 1, j = k$$

$$(1 + \gamma C)T_{2,k+1} - CT_{3,k+1} - CT_{1,k+1} = T_{2,k} \quad \text{به ازای } i = 2, j = k$$

⋮

$$(1 + \gamma C)T_{n,k+1} - CT_{n-1,k+1} = T_{n,k} + CT_{n+1,k+1} \quad \text{به ازای } i = n, j = k$$

اکنون فرض کنید دما در دو ضلع عمودی معلوم باشد، یعنی:

$$T_{0,t} = T_{W_1}$$

$$T_{n+1,t} = T_{W_2}$$

پس معادلات قبلی می‌توانند با نماد ماتریسی به صورت زیر نوشته شوند:

$$\begin{bmatrix} (1+2C) & -C & \cdot & \dots & \dots & \cdot \\ -C & (1+2C) & -C & \cdot & \dots & \cdot \\ \cdot & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & & \cdot \\ \vdots & & & & & -C \\ \cdot & \dots & \dots & \cdot & -C & (1+2C) \end{bmatrix} \begin{bmatrix} T_{1,k+1} \\ T_{2,k+1} \\ \vdots \\ \vdots \\ T_{n,k+1} \end{bmatrix}$$

$$= \begin{bmatrix} T_{1,k} + CT_{W_1} \\ T_{2,k} \\ \vdots \\ \vdots \\ T_{n,k} + CT_{W_1} \end{bmatrix}$$

واضح است که ماتریس این دستگاه متقارن، سه قطری و غالب قطری است (توجه کنید که $C > 0$).

برای مثال، هنگامی که $C = 1$ ، داریم:

$$\begin{bmatrix} 3 & -1 & & \cdot \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ \cdot & & -1 & 3 \end{bmatrix} \begin{bmatrix} T_{1,k+1} \\ T_{2,k+1} \\ \vdots \\ \vdots \\ T_{n,k+1} \end{bmatrix} = \begin{bmatrix} T_{1,k} + T_{W_1} \\ T_{2,k} \\ \vdots \\ \vdots \\ T_{n,k} + T_{W_2} \end{bmatrix}$$

یا

$$Ax = b$$

ماتریس A متقارن، سه قطری، غالب قطری، و معین مثبت است.

دستگاههای سه قطری بلوکی

برای این که ببینیم که چگونه دستگاههای سه قطری بلوکی در کاربردها مطرح می شوند، معادله پواسون دو بُعدی زیر را در نظر می گیریم.

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = f(x, y); \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1$$

شکل گسسته متناظر با این معادله، مشابه با معادله لاپلاس که قبلاً به دست آمده، به صورت زیر است

$$T_{i+1,j} + T_{i-1,j} + T_{i,j+1} + T_{i,j-1} - 4T_{i,j} = (\Delta x)^2 f_{i,j};$$

$$i = 1, 2, \dots, n, \quad j = 1, 2, \dots, n$$

این رابطه یک دستگاه خطی با $(n + 2)^2$ متغیر را به وجود خواهد آورد.

اکنون فرض کنید که مقادیر T در چهار ضلع مربع واحد معلوم باشند و مقادیر T در نقاط داخلی شبکه مورد نظر باشند، یعنی مقادیر

$$T_{0,j}; T_{n+1,j} \quad \text{و} \quad T_{i,0}; T_{i,n+1}$$

$$j = 0, 1, \dots, n + 1, \quad i = 0, 1, \dots, n + 1$$

معلوم باشند و مقادیر $T_{11}, T_{12}, \dots, T_{1n}; T_{21}, T_{22}, \dots, T_{2n}; \dots, T_{n1}, \dots, T_{nn}$ باید به دست آیند. پس ما یک دستگاه $(n^2 \times n^2)$ با n^2 مجهول داریم، که بعد از مرتب سازی مناسب یک دستگاه سه قطری بلوکی خواهد بود. برای مثال، هنگامی که $n = 3$ دستگاه به شکل زیر خواهد بود.

$$\begin{pmatrix} 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 4 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 4 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 4 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 \end{pmatrix} \times \begin{pmatrix} T_{11} \\ T_{21} \\ T_{31} \\ T_{12} \\ T_{22} \\ T_{32} \\ T_{13} \\ T_{23} \\ T_{33} \end{pmatrix}$$

$$= \begin{bmatrix} T_{\cdot,1} + T_{1\cdot} - (\Delta x)^{\vee} f_{11} \\ T_{2\cdot} - (\Delta x)^{\vee} f_{21} \\ T_{3\cdot} + T_{\cdot 3} - (\Delta x)^{\vee} f_{31} \\ T_{\cdot 2} - (\Delta x)^{\vee} f_{12} \\ -(\Delta x)^{\vee} f_{22} \\ T_{22} - (\Delta x)^{\vee} f_{32} \\ T_{12} + T_{\cdot 2} - (\Delta x)^{\vee} f_{12} \\ T_{22} - (\Delta x)^{\vee} f_{22} \\ T_{22} + T_{22} - (\Delta x)^{\vee} f_{22} \end{bmatrix}$$

در حالت کلی، دستگاه به شکل زیر خواهد بود.

$$\begin{bmatrix} A_n & -I_n \\ -I_n & \ddots & \ddots & \ddots \\ & \ddots & \ddots & \ddots \\ & & -I_n & A_n \end{bmatrix} \begin{bmatrix} T_{11} \\ T_{21} \\ \vdots \\ T_{n1} \\ T_{12} \\ \vdots \\ T_{n2} \\ \vdots \\ T_{nn} \end{bmatrix} = \begin{bmatrix} T_{\cdot,1} + T_{1\cdot} - (\Delta x)^{\vee} f_{11} \\ T_{2\cdot} - (\Delta x)^{\vee} f_{21} \\ \vdots \\ T_{n-1,\cdot} - (\Delta x)^{\vee} f_{n-1,1} \\ T_{n+1,1} + T_{n,\cdot} - (\Delta x)^{\vee} f_{n,2} \\ \vdots \\ T_{\cdot 2} - (\Delta x)^{\vee} f_{12} \\ \vdots \end{bmatrix} \quad (31-3-6)$$

که در آن

$$A_n = \begin{bmatrix} 4 & -1 & \cdot & \cdot \\ -1 & \ddots & \ddots & \cdot \\ \cdot & \ddots & \ddots & -1 \\ \cdot & \cdot & -1 & 4 \end{bmatrix} \quad (32-3-6)$$

دستگاه ماتریسی فوق سه قطری بلوکی است و هر ماتریس A_n بر روی قطر، متقارن، سه قطری، و معین مثبت است.

۶-۳-۶ دستگاه خطی حاصل از روشهای عنصر منتهای

در چند بخش آخر ملاحظه نمودیم که چگونه گسسته‌سازی معادلات دیفرانسیل با استفاده از تفاضلات منتهای به انواع مختلفی از دستگاههای خطی منجر می‌شود. روش عنصر منتهای، یک طریق متداول دیگر برای گسسته‌سازی معادلات دیفرانسیل و به دست آوردن دستگاههای معادلات خطی است. برای درک بهتر، این ایده را با استفاده از یک معادله دیفرانسیل ساده توضیح می‌دهیم. خوانندگان علاقه‌مند می‌توانند به برخی از کتابهای معروف در این زمینه نظیر (Celia and Gray (1992)، Ames و Reddy (1993), Becker, Carey, and Oden (1981)، Strang and Fix (1973) (1992) مراجعه نمایند.

فرمول‌بندی تغییراتی یک مسأله مقدار مرزی دو نقطه‌ای اجازه دهید مسأله مقدار مرزی دو نقطه‌ای زیر را در نظر بگیریم

$$-u'' + u = f(x), \quad 0 < x < 1 \quad (33-3-6)$$

$$u(0) = u(1) = 0 \quad (34-3-6)$$

که در آن $u' = du/dx$ و f یک تابع پیوسته بر روی $[0, 1]$ است. بعلاوه فرض کنید که f به قسمی باشد که مسأله (۳۳-۳-۶) و (۳۴-۳-۶) دارای یک جواب منحصر به فرد باشد.

فضای زیر را معرفی می‌کنیم

$V = \{v \mid v \text{ تابع پیوسته بر روی } [0, 1] \text{ است و } v' \text{ قطعه به قطعه پیوسته و } v(0) = v(1) = 0\}$

اکنون اگر معادله $u'' + u = f(x)$ را در یک تابع دلخواه $v \in V$ ضرب کنیم (v یک تابع آزمون نامیده می‌شود)، و از طرف چپ جزء به جزء انتگرال‌گیری کنیم، به دست می‌آوریم

$$\int_0^1 (-u''(x) + u(x))v(x) dx = \int_0^1 f(x)v(x) dx$$

یعنی

$$\int_0^1 (-u'v' + uv) dx = \int_0^1 f(x)v(x) dx \quad (35-3-6)$$

رابطه (۳۵-۳-۶) را به صورت زیر می‌نویسیم

$$a(u, v) = (f, v), \quad v \in V \text{ به ازای هر } v$$

که در آن

$$a(u, v) = \int_0^1 (u'v' + uv) dx$$

و

$$(f, v) = \int_0^1 f(x)v(x) dx$$

توجه کنید که شکل $a(., .)$ متقارن است (یعنی، $a(u, v) = a(v, u)$) و دو خطی است. این دو خاصیت بعداً مورد استفاده قرار خواهند گرفت. می توان نشان داد که u یک جواب (۳-۳-۳۵) است اگر و فقط اگر u یک جواب (۳-۳-۳۳) و (۳-۳-۳۴) باشد.

مسأله گسسته

اکنون مسأله (۳-۳-۳۵) را گسسته سازی می کنیم. با ساختن یک زیر فضای با بُعد متناهی مانند V_n از V شروع می کنیم.

در اینجا فقط حالت ساده ای را که در آن V_n از توابع خطی قطعه به قطعه پیوسته تشکیل می شود، در نظر می گیریم. برای این منظور فرض می کنیم که $0 = x_0 < x_1 < x_2 < \dots < x_n < x_{n+1} = 1$ با طول $I_j = [x_{j-1}, x_j]$ به زیر بازه های $[0, 1]$ به زیر بازه های $j = 1, 2, \dots, n+1$ ، $h_j = x_j - x_{j-1}$ باشد. با این افزایش مجموعه V_n را با مجموعه همه توابع $v(x)$ که بر روی بازه $[0, 1]$ پیوسته هستند، و در هر زیر بازه I_j ، $j = 1, \dots, n+1$ خطی می باشند، و در شرایط مرزی $v(0) = v(1) = 0$ صدق می کند، مرتبط می سازیم.

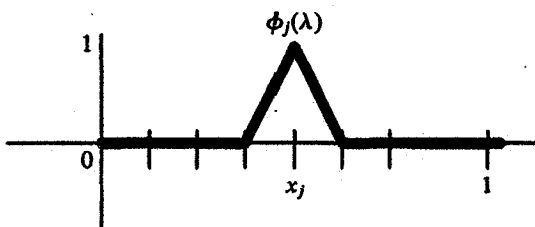
اکنون توابع پایه ای $\{\phi_1, \phi_2, \dots, \phi_n\}$ از V_n را معرفی می کنیم. $\phi_j(x)$ را به صورت زیر تعریف می کنیم.

$$\phi_j(x_i) = \begin{cases} 1 & \text{اگر } i = j \\ 0 & \text{اگر } i \neq j \end{cases} \quad .1$$

.2 $\phi_j(x)$ یک تابع قطعه به قطعه پیوسته است.

$\phi_j(x)$ می تواند به صورت صریح بیان و محاسبه شود:

$$\phi_j(x) = \begin{cases} \frac{x - x_{j-1}}{h_j} & \text{اگر } x_{j-1} \leq x \leq x_j \\ \frac{x_{j+1} - x}{h_{j+1}} & \text{اگر } x_j \leq x \leq x_{j+1} \end{cases}$$



شکل ۶-۱۰

شکل ۶-۱۰ را ملاحظه کنید. چون ϕ_1, \dots, ϕ_n توابع پایه‌ای هستند، هر تابع $v \in V_n$ می‌تواند به طور منحصر به فرد به صورت زیر نوشته شود.

$$v(x) = \sum_{i=1}^n v_i \phi_i(x), \quad \text{که در آن } v_i = v(x_i)$$

به آسانی ملاحظه می‌کنیم که $V_n \subset V$.

سپس گسسته متناظر (۶-۳-۳۵) بیان می‌کند: $u_n \in V_n$ را طوری پیدا کنید که

$$a(u_n, v) = (f, v) \quad \text{به ازای هر } v \in V_n \quad (۶-۳-۳۶)$$

اکنون اگر قرار دهیم $u_n = \sum_{i=1}^n c_i \phi_i(x)$ و توجه کنیم که معادله (۶-۳-۳۶) بویژه برای هر تابع $\phi_j(x)$ ، $n, \dots, 1, j = 1, \dots, n$ درست است، معادله خطی به صورت زیر به دست می‌آوریم

$$a\left(\sum_{i=1}^n c_i \phi_i, \phi_j\right) = (f, \phi_j), \quad \text{به ازای هر } j = 1, 2, \dots, n$$

با استفاده از خطی بودن $a(\cdot, \phi_j)$ ، n معادله و n مجهول زیر نتیجه می‌شود:

$$\sum_{i=1}^n c_i a(\phi_i, \phi_j) = (f, \phi_j), \quad \text{به ازای هر } j = 1, 2, \dots, n$$

این رابطه را می‌توان به شکل ماتریسی به صورت زیر نوشت

$$Ac = (f_n)$$

که در آن $(f_n)_i = (f, \phi_i)$ و $A = (a_{ij})$ یک ماتریس متقارن ارائه شده به صورت زیر است.

$$a_{ij} = a_{ji} = a(\phi_i, \phi_j) \quad \text{و} \quad c = (c_1, \dots, c_n)^T$$

عناصر ماتریس A می‌توانند به صورت صریح محاسبه شوند: ابتدا توجه می‌کنیم که

$$a_{ij} = a_{ji} = a(\phi_i, \phi_j) = 0 \quad \text{اگر } |i - j| \geq 2$$

(این به واسطه آن است که محمل تابع $\phi_j(x)$ موضعی است). اکنون یک محاسبه مستقیم منجر به رابطه زیر می شود

$$a_{j,j} = a(\phi_j, \phi_j) = \int_{x_{j-1}}^{x_j} \left[\frac{1}{h_j^2} + \frac{(x-x_{j-1})^2}{h_j^3} \right] dx + \int_{x_j}^{x_{j+1}} \left[\frac{1}{h_{j+1}^2} + \frac{(x_{j+1}-x)^2}{h_{j+1}^3} \right] dx$$

$$= \left[\frac{1}{h_j} + \frac{1}{h_{j+1}} \right] + \frac{1}{3} [h_j + h_{j+1}]$$

$$a_{j,j-1} = \int_{x_{j-1}}^{x_j} \left[-\frac{1}{h_j^2} + \frac{(x_j-x)(x-x_{j-1})}{h_j^3} \right] dx = -\frac{1}{h_j} + \frac{h_j}{6}$$

بنابراین دستگاه (۶-۳-۳۷) را می توان به صورت زیر نوشت

$$\begin{bmatrix} a_1 & b_1 & & & 0 \\ b_1 & \ddots & \ddots & & \\ & \ddots & \ddots & b_{n-1} & \\ 0 & & b_{n-1} & a_n & \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} (f_n)_1 \\ (f_n)_2 \\ \vdots \\ (f_n)_n \end{bmatrix}$$

که در آن $a_j = 1/h_j + 1/h_{j+1} + \frac{1}{3}[h_j + h_{j+1}]$ و $b_j = -1/h_j + h_j/6$. در حالت خاص شبکه یکنواخت $h_j = h = 1/(n+1)$ ، ماتریس A شکل زیر را اختیار می کند.

$$A = \frac{1}{h} \begin{bmatrix} 2 & -1 & & & 0 \\ -1 & 2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ 0 & & & -1 & 2 \end{bmatrix} + \frac{h}{6} \begin{bmatrix} 4 & 1 & & & 0 \\ 1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & 1 \\ 0 & & \ddots & \ddots & \\ & & & 1 & 4 \end{bmatrix} \quad (۶-۳-۳۸)$$

۶-۳-۷ تقریب یک تابع توسط یک چندجمله ای: دستگاه هیلبرت

در بخش ۳-۶ یک دستگاه خطی بدووضع با ماتریس هیلبرت را ذکر کردیم. در این بخش نشان می دهیم که چگونه یک چنین دستگاهی مطرح می شود. این بحث از کتاب Forsythe and Moler (1967, 80-81) گرفته شده است.

فرض کنید که یک تابع پیوسته $f(x)$ و تعریف شده بر بازه $0 \leq x \leq 1$ باید توسط یک چندجمله ای

درجه $n - 1$ مانند

$$\sum_{i=1}^n p_i x^{i-1}$$

تقریب زده شود به قسمی که خطای

$$E = \int_0^1 \left[\sum_{i=1}^n p_i x^{i-1} - f(x) \right]^2 dx$$

می‌نیمم شود. ضرایب p_i چند جمله‌ای به آسانی با قرار دادن

$$\frac{\partial E}{\partial p_i} = 0, \quad i = 1, \dots, n$$

تعیین می‌شوند (توجه کنید که خطای یک تابع مشتق‌پذیر از مجهولهای p_i است و بعلاوه هنگامی که مشتقهای جزئی برابر صفر می‌شوند یک می‌نیمم رخ می‌دهد.) بنابراین داریم

$$\frac{\partial E}{\partial p_i} = 2 \int_0^1 \left[\sum_{j=1}^n p_j x^{j-1} - f(x) \right] x^{i-1} dx = 0, \quad i = 1, \dots, n$$

یا

$$\sum_{j=1}^n \left[\int_0^1 x^{i+j-2} dx \right] p_j = \int_0^1 f(x) x^{i-1} dx, \quad i = 1, \dots, n$$

(برای به دست آوردن این شکل آخر، جای مجموع‌یابی و انتگرال‌گیری را تعویض کرده‌ایم.) با قرار دادن

$$h_{ij} = \int_0^1 x^{i+j-2} dx$$

و

$$b_i = \int_0^1 f(x) x^{i-1} dx, \quad i = 1, 2, \dots, n$$

داریم

$$\sum_{j=1}^n h_{ij} p_j = b_j, \quad i = 1, \dots, n$$

یعنی دستگاه خطی

$$Hp = b$$

را به دست می‌آوریم که در آن

$$H = (h_{ij}), \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

به آسانی می‌توان دریافت که H یک ماتریس هیلبرت است (بخش ۳-۶ ملاحظه شود)، زیرا

$$h_{ij} = \int_0^1 x^{i+j-2} dx = \frac{1}{i+j-1}$$

۴-۶ روشهای مستقیم

در این بخش روشهای مستقیم را برای حل مسأله $Ax = b$ مورد مطالعه قرار می‌دهیم. این روشها شامل روشهای زیر هستند.

- روش حذفی گاوس بدون محورگیری، مبتنی بر تجزیه LU ماتریس A (بخش ۶-۴-۲)
- روش حذفی گاوس با محورگیری جزئی، مبتنی بر تجزیه $MA = U$ ماتریس A (بخش ۶-۴-۳)
- روش حذفی گاوس با محورگیری کامل، مبتنی بر تجزیه $MAQ = U$ ماتریس A (بخش ۶-۴-۳)
- روش مبتنی بر تجزیه QR ماتریس A (بخش ۶-۴-۵)
- روش مبتنی بر تجزیه چولسکی یک ماتریس معین مثبت متقارن (بخش ۶-۴-۷)
- روش حذفی گاوس برای دستگاههای خاص: هسنبرگی، معین مثبت، سه قطری، و غالب قطری (بخش ۶-۴-۷).

برای مقایسه این روشها و مزایا و معایب آنها بخش ۶-۴-۹ و جدولهای همراه را ملاحظه کنید. ایده اصلی در پشت همه روشهای مستقیم این است که ابتدا دستگاه خطی $Ax = b$ ، با پیدا کردن عاملهای مثلثی ماتریس A ، به یک دستگاه مثلثی معادل تقلیل یابد و سپس دستگاه مثلثی که به مراتب برای حل ساده‌تر از دستگاه اصلی است، حل شود. در فصل ۵ روشهای گوناگونی را برای محاسبه عاملهای مثلثی A توصیف کردیم. در بخش ۳-۱ روش جایگزینی پسرو را برای حل یک دستگاه بالامثلثی توضیح دادیم. اکنون یک روش مشابه که حذفی پیشرو نامیده می‌شود، برای حل دستگاه پایین مثلثی بیان می‌نماییم.

۶-۴-۱ حل یک دستگاه پایین مثلثی

جواب $y = (y_1, y_2, \dots, y_n)^T$ دستگاه پایین مثلثی نامنفرد

$$Ly = b$$

می تواند به طریقی مشابه با دستگاه بالا مثلثی به دست آید. در اینجا γ_1 از اولین معادله به دست می آید، و سپس با قرار دادن مقدار آن در معادله دوم، γ_2 به دست می آید، و غیره. این فرایند حذفی پیشرو نامیده می شود.

الگوریتم ۱-۴-۶ حذفی پیشرو

For $i = 1, 2, \dots, n$ do

$$y_i = \frac{1}{l_{ii}} \left[b_i - \sum_{j=1}^{i-1} l_{ij} y_j \right]$$

تذکر: هنگامی که $i = 1$ مجموع یابی (Σ) وجود ندارد (مجموع صفر تعریف می شود). شمارش اعمال ممیز شناور و پایداری. الگوریتم حدوداً به $n^2/2$ عمل ممیز شناور (flops) نیاز دارد (۱) عمل ممیز شناور برای محاسبه γ_1 ، ۲ عمل ممیز شناور برای محاسبه γ_2 ، ۳ عمل ممیز شناور برای محاسبه γ_3 ، و غیره. الگوریتم به همان پایداری فرایند جایگزینی پسرو (الگوریتم ۳-۱-۳) است.

تذکر MATCOM: الگوریتم ۱-۴-۶ در برنامه FORELM در MATCOM پیاده سازی شده است.

۲-۴-۶ حل دستگاه $Ax = b$ با استفاده از روش حذفی گاوس بدون محورگیری
روش حذفی گاوس برای $Ax = b$ بر تجزیه LU ماتریس A مبتنی است. از بخش ۱-۲-۵ یادآور می شویم که مثلثی سازی با استفاده از روش حذفی گاوس بدون محورگیری، هنگامی که به طور کامل انجام شود، یک تجزیه LU از A را نتیجه می دهد. به محض این که این تجزیه A را داشته باشیم، دستگاه $Ax = b$ معادل با دو دستگاه زیر خواهد بود.

$$Ly = b$$

$$Ux = y$$

حل $Ax = b$ با استفاده از روش حذفی گاوس بدون محورگیری

حل دستگاه $Ax = b$ با استفاده از روش حذفی گاوس (بدون محورگیری) می تواند در دو گام

انجام شود:

گام ۱: پیدا کردن یک تجزیه LU برای A .

گام ۲: حل دو دستگاه مثلثی: ابتدا دستگاه پایین مثلثی $Ly = b$ ، و سپس دستگاه بالا مثلثی

$$Ux = y$$

شمارش اعمال ممیز شناور. چون یک تجزیه LU حدوداً به $n^3/3$ اعمال ممیز شناور نیاز دارد و حل هر دستگاه مثلثی فقط به $n^2/2$ عمل ممیز شناور نیاز دارد، کل تعداد اعمال ممیز شناور برای حل دستگاه با استفاده از روش حذفی گاوس در حدود $n^3/3 + n^2$ می باشد.

۳-۴-۶ حل دستگاه $Ax = b$ با استفاده از مثلثی سازی با محورگیری

حل با محورگیری جزئی

اگر روش حذفی گاوس با محورگیری جزئی برای مثلثی سازی A استفاده شود، آنگاه - همان طور که در بخش ۲-۲-۵ دیدیم - فرایند تجزیه رابطه

$$MA = U$$

را نتیجه خواهد داد. در این حالت، دستگاه $Ax = b$ با دستگاه مثلثی زیر معادل است

$$Ux = Mb = b'$$

حل $Ax = b$ با استفاده از روش حذفی گاوس با محورگیری جزئی

حل دستگاه $Ax = b$ با استفاده از روش حذفی گاوس با محورگیری جزئی در دو گام زیر

انجام شود:

گام ۱: پیدا کردن تجزیه $MA = U$ با الگوریتم مثلثی سازی و استفاده از محورگیری جزئی

(الگوریتم ۲-۲-۵).

گام ۲: حل دستگاه مثلثی زیر با استفاده از جایگزینی پسرو (الگوریتم ۳-۱-۳).

$$Ux = Mb = b'$$

پیاده‌سازی گام ۲. بردار

$$b' = Mb = M_{n-1}P_{n-1}M_{n-2}P_{n-2} \dots M_1P_1b$$

به صورت زیر ارائه می‌شود

Step 1: $s_1 = b = (b_1 \ b_2 \ \dots \ b_n)^T$

Step 2: For $k = 1, 2, \dots, n - 1$ do

$$s_{k+1} = M_k P_k s_k$$

تذکره‌های محاسباتی: الگوریتم ۲-۲-۵ ماتریسهای M_k و P_k را به طور صریح ارائه نمی‌دهد. در حقیقت به آنها نیاز نداریم. بردار s_{k+1} می‌تواند فوراً از روی s_k محاسبه شود اگر r_k اندیس سطر تعویض شده و ضریب‌های m_{ik} در گام k ذخیره شده باشند. همان طور که در بخش ۲-۲-۵ نشان داده شده است هیچ ضرب ماتریسی صریحی لازم نمی‌باشد. این مطلب با یک مثال 3×3 نشان داده می‌شود.

مثال ۱-۴-۶ فرض کنید

$$n = 3, \quad P_1 = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & \cdot & 1 \\ \cdot & 1 & \cdot \end{bmatrix}, \quad M_1 = \begin{bmatrix} 1 & \cdot & \cdot \\ m_{21} & 1 & \cdot \\ m_{31} & \cdot & 1 \end{bmatrix}$$

و فرض کنید که

$$s_2 = M_1 P_1 s_1 = \begin{bmatrix} s_1^{(2)} \\ s_2^{(2)} \\ s_3^{(2)} \end{bmatrix}$$

آنگاه داریم

$$P_1 s_1 = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

و عناصر s_2 توسط روابط زیر داده می‌شوند

$$s_1^{(2)} = b_1$$

$$s_2^{(2)} = m_{21} b_1 + b_2$$

$$s_3^{(2)} = m_{31} b_1 + b_2$$

حل با محورگیری کامل

روش حذفی گاوس با محورگیری کامل (بخش ۵-۲-۳) رابطه

$$MAQ = U$$

را ارائه می دهد. با استفاده از این تجزیه دستگاه $Ax = b$ را می توان به صورت

$$Uy = Mb = b'$$

نوشت، که در آن

$$y = Q^T x$$

بنابراین راه حل زیر را داریم

حل $Ax = b$ با استفاده از روش حذفی گاوس با محورگیری کامل
 برای حل $Ax = b$ با استفاده از محورگیری کامل گامهای زیر انجام می شود:
 گام ۱: پیدا کردن تجزیه $MAQ = U$ با الگوریتم تجزیه U و با استفاده از محورگیری کامل
 (الگوریتم ۵-۲-۳)
 گام ۲: حل دستگاه مثلثی (نسبت به U) (الگوریتم ۳-۱-۳): $Uy = b'$
 محاسبه b' به صورت قبلاً نشان داده شده:
 گام ۳: بالاخره به دست آوردن x از روی $y: x = Qy$.

پیاده سازی گام ۳. چون $Q = Q_1 Q_2 \dots Q_{n-1}$ ، $x = Qy$ ، طرح زیر می تواند برای محاسبه x از روی y مورد استفاده قرار گیرد.

Step 1: Set $w_n = y$

Step 2: For $k = n - 1, \dots, 2, 1$ do

$$w_k = Q_k w_{k+1}$$

Step 3: Then $x = w_1$

تذکر: چون Q_k یک ماتریس جایگشت است عناصر w_k همان عناصر w_{k+1} هستند که ترتیب آنها بر طبق اندیس جایگشت تغییر کرده اند.

مثال ۶-۴-۲ دستگاه $Ax = b$ را با

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix}$$

حل کنید. (الف) با استفاده از محورگیری جزئی و (ب) با استفاده از محورگیری کامل.

(الف) محورگیری جزئی: با نتایج به دست آمده از قبل (بخش ۵-۲-۲ مثال ۵-۲-۳)، محاسبه می‌کنیم

$$P_1 b = \begin{bmatrix} 6 \\ 2 \\ 3 \end{bmatrix}; \quad P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M_1 P_1 b = \begin{bmatrix} 6 \\ 2 \\ -3 \end{bmatrix}; \quad M_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

$$P_2 M_1 P_1 b = \begin{bmatrix} 6 \\ 2 \\ -3 \end{bmatrix}; \quad P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$b' = M_2 P_2 M_1 P_1 b = \begin{bmatrix} 6 \\ 2 \\ -1 \end{bmatrix}; \quad M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

جواب دستگاه

$$Ux = b'$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \\ -1 \end{bmatrix}$$

عبارت است از $x_1 = x_2 = x_3 = 1$.

(ب) محورگیری کامل: با استفاده از نتایج مثال ۵-۲-۵، داریم

$$P_1 b = \begin{bmatrix} 6 \\ 2 \\ 3 \end{bmatrix}, \quad P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M_1 P_1 b = \begin{bmatrix} 6 \\ 0 \\ 1 \end{bmatrix}, \quad P_2 M_1 P_1 b = \begin{bmatrix} 6 \\ 1 \\ 0 \end{bmatrix}$$

$$b' = M_2 P_2 M_1 P_1 b = \begin{bmatrix} 6 \\ 1 \\ 1 \\ \frac{1}{2} \end{bmatrix}$$

جواب دستگاه

$$Uy = b'$$

$$\begin{bmatrix} 3 & 1 & 2 \\ 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \\ \frac{1}{2} \end{bmatrix}$$

عبارت است از $y_1 = y_2 = y_3 = 1$ چون $\{x_k\}$ ، $k = 1, 2, 3$ ترتیب دیگری از $\{y_i\}$ است، داریم $x_1 = x_2 = x_3 = 1$.

برخی جزئیات پیاده‌سازی

توجه کنید که لازم نیست که بردارهای s_i و w_i به طور مجزا ذخیره شوند، زیرا همه آنچه که ما نیاز داریم بردار s_k برای محورگیری جزئی و w_1 برای محورگیری کامل است. بنابراین با شروع با $x = b$ ، هر بردار جدید می‌تواند به محض محاسبه در x ذخیره شود.

همچنین توجه کنید که اگر ما از الگوریتمهای عملی استفاده کنیم، ماتریسهای P_k ، Q_k ، و M_k به طور صریح موجود نیستند؛ آنها توسط اندیسهای r_k و s_k و ضریبهای m_{ik} ارائه می‌شوند. s_k ها و w_k ها باید از روی اینها محاسبه شوند.

شمارش تعداد اعمال ممیز شناور. در بخش ۵-۲-۱ ملاحظه کردیم که فرایند مثلثی سازی با استفاده

از ماتریسهای ابتدایی نیاز به $n^2/3$ عمل ممیز شناور (flops) دارد. دستگاه مثلثی $Ux = b'$ یا $Uy = b'$ می تواند با استفاده از جایگزینی پسرو با $n^2/2$ عمل ممیز شناور محاسبه شود و بردار b' می تواند با n^2 عمل ممیز شناور، با در نظر گرفتن ساختار خاص ماتریسهای M_k و P_k محاسبه شود. به دست آوردن x از روی y در گام ۳ فرایند محورگیری کامل نیاز به هیچ عمل ممیز شناوری ندارد. توجه کنید که x دقیقاً با جابه جا کردن عناصر y به دست می آید. بنابراین جواب دستگاه خطی $Ax = b$ با استفاده از روش حذفی گاوس با محورگیری جزئی یا کامل نیاز به $n^2/3 + O(n^2)$ عمل ممیز شناور دارد. ولیکن روش حذفی گاوس با محورگیری کامل برای مشخص کردن $(n - 1)$ عنصر محوری حدوداً نیاز به $n^3/3$ مقایسه دارد، در صورتی که با محورگیری جزئی فقط به $O(n^2)$ مقایسه نیاز دارد.

خاصیت گردکردن. در بخش ۵-۳ ما در مورد خاصیت گردکردن روش حذفی گاوس برای مثلثی سازی ماتریس A بحث کردیم. ما ملاحظه کردیم که عامل رشد پایداری فرایند مثلثی سازی را تعیین می کند. سؤال بعد این است که چگونه عامل رشد بر رویه حل $Ax = b$ با استفاده از یک چنین مثلثی سازی اثر می گذارد؟ جواب در زیر داده شده است.

نتیجه خطای گردکردن برای مسأله دستگاه خطی با روش حذفی گاوس

می توان نشان داد که جواب محاسبه شده \tilde{x} دستگاه خطی $Ax = b$ ، با استفاده از روش حذفی گاوس، در رابطه

$$(A + E)x = b$$

صدق می کند، که در آن $\mu \|A\|_\infty \leq c(n^3 + \delta n^2)\rho \|A\|_\infty$ و $\|E\|_\infty \leq c$ یک ثابت کوچک است. برای اثبات بخش ۱۱-۴ را ملاحظه کنید.

تبصره: از قبل ملاحظه کردیم که اگر عامل رشد ρ بزرگ باشد، باید انتظار یک خطای بزرگ را داشته باشیم.

۴-۴-۶ حل $Ax = b$ بدون تجزیه صریح

همان طور که اکنون ملاحظه کردیم، روش حذفی گاوس برای حل $Ax = b$ در دو مرحله انجام می شود. ابتدا ماتریس A به طور صریح تجزیه می شود.

$$A = LU \quad (\text{بدون محورگیری})$$

$$MA = U \quad (\text{با محورگیری جزئی})$$

$$MAQ = U \quad (\text{با محورگیری کامل})$$

سپس، تجزیه A برای حل $Ax = b$ مورد استفاده قرار می‌گیرد. ولیکن به آسانی ملاحظه می‌شود که این دو مرحله می‌توانند ترکیب شوند به نحوی که جواب بتواند با حل یک دستگاه بالا مثلثی با اعمال فرایند بر ماتریس A و بردار b به طور همزمان، به دست آید. در این حالت ماتریس افزوده (A, b) مثلثی می‌شود و سپس جواب با جایگزینی پسرو به دست می‌آید. ما این فرایند ضمنی را برای روش حذفی گاوس با محورگیری جزئی توضیح می‌دهیم.

الگوریتم ۶-۴-۲ حل $Ax = b$ با محورگیری جزئی بدون تجزیه صریح. با مفروض بودن یک ماتریس $n \times n$ نامنفرد مانند A و یک بردار b ، الگوریتم زیر تجزیه مثلثی ماتریس افزوده (A, b) را با استفاده از روش محورگیری جزئی محاسبه می‌کند. ماتریس مثلثی تبدیل یافته بر روی A و بردار تبدیل یافته بر روی b نوشته می‌شوند. ضربگرها در قسمت پایین مثلثی A ذخیره می‌شوند.

برای $k = 1, 2, \dots, n - 1$ انجام دهید

۱: عنصر با بزرگترین قدر مطلق را در ستون k و در زیر سطر $(k - 1)$ انتخاب کنید؛ آن را $a_{r_k, k}$ بنامید:

$$a_{r_k, k} = \max\{|a_{ik}| : i \geq k\}$$

اگر $a_{r_k, k} = 0$ ، متوقف شوید. در غیر این صورت به گام ۲ بروید.

۲: جای سطرهای k و r_k از A و جای عناصر k ام و r_k ام از b را عوض کنید.

$$a_{kj} \leftrightarrow a_{r_k, j} \quad (j = k, k + 1, \dots, n)$$

$$b_k \leftrightarrow b_{r_k}$$

۳: ضربگرها را تشکیل دهید:

$$a_{ik} \equiv m_{ik} = -\frac{a_{ik}}{a_{kk}} \quad (i = k + 1, \dots, n)$$

۴: عناصر A را به هنگام درآورید:

$$a_{ij} \equiv a_{ij} + m_{ik}a_{kj} \quad (i = k + 1, \dots, n; \quad j = k + 1, \dots, n)$$

۵: عناصر b را به هنگام درآورد:

$$b_i \equiv b_i + m_{ik}b_k \quad (i = k + 1, \dots, n)$$

تذکر MATCOM: الگوریتم ۶-۴-۲ در برنامه GAUSSWF در MATCOM پیاده‌سازی شده است.

مثال ۳-۴-۶

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 2 & 2 & 3 \\ 4 & 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix}$$

$k = 1$: عنصر محوری برابر $a_{11} = 4$ است و $r_1 = 3$.

سطرهای ۱ و ۳ از A و عناصر سوم و اول از b را تعویض می‌کنیم.

$$A \equiv \begin{bmatrix} 4 & 1 & 1 \\ 2 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix}, \quad b \equiv \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$$

$$m_{21} = -\frac{a_{21}}{a_{11}} = -\frac{1}{4}$$

$$A \equiv A^{(1)} = \begin{bmatrix} 4 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 \end{bmatrix}, \quad b \equiv b^{(1)} = \begin{bmatrix} 3 \\ \frac{9}{2} \\ 2 \end{bmatrix}$$

$k = 2$: عنصر محوری برابر $a_{22} = \frac{3}{2}$ است.

$$m_{32} = -\frac{a_{32}}{a_{22}} = -\frac{2}{3}$$

$$A \equiv A^{(2)} = \begin{bmatrix} 4 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 0 & -\frac{1}{3} \end{bmatrix}, \quad b \equiv b^{(2)} = \begin{bmatrix} 3 \\ \frac{9}{2} \\ -1 \end{bmatrix}$$

دستگاه بالا مثلثی کاهش یافته $A^{(2)}x = b^{(2)}$ به صورت زیر است

$$2x_1 + x_2 + x_3 = 3$$

$$\frac{3}{2}x_2 + \frac{5}{2}x_3 = \frac{9}{2}$$

$$-\frac{2}{3}x_3 = -1$$

جواب برابر است با

$$x_3 = \frac{3}{2}, \quad x_2 = \frac{1}{2}, \quad x_1 = \frac{1}{4}$$

۶-۴-۵ حل $Ax = b$ با استفاده از تجزیه صریح QR

اگر داشته باشیم

$$A = QR$$

آنگاه دستگاه $Ax = b$ به صورت زیر در می آید

$$QRx = b$$

یا

$$Rx = Q^T b = b'$$

بنابراین، به محض این که تجزیه QR ماتریس A را داشته باشیم، جواب دستگاه $Ax = b$ می تواند با حل دستگاه بالا مثلثی معادل آن، به دست آید:

$$Rx = b' = Q^T b$$

حل $Ax = b$ با استفاده از تجزیه QR

برای حل $Ax = b$ با استفاده از تجزیه QR: گامهای زیر انجام می شود

۱: پیدا کردن تجزیه QR ماتریس $A: A = QR$ (الگوریتم ۵-۴-۳ یا ۵-۵-۴).

۲: تشکیل $b' = Q^T b$.

۳: حل $Rx = b'$

تشکیل b'

برای محاسبه b' ، به ماتریس Q به صورت صریح نیاز نداریم. این بردار می تواند از روی شکل

حاصلضرب Q محاسبه شود. برای مثال، اگر تجزیه QR با استفاده از روش هاوس هولدر (بخش ۵-۴-۱) به دست آید، آنگاه

$$Q = H_{n-1}H_{n-2} \dots H_2H_1$$

و $b' = y_n = Q^T b$ می‌تواند به صورت زیر محاسبه شود

Step 1 : $y_1 = b$

Step 2 : For $k = 1, 2, \dots, n - 1$ do

$$y_{k+1} = H_k y_k$$

تبصره: با استفاده از الگوریتم ۵-۴-۲ بردارهای $y_{k+1} = H_k y_k$ را محاسبه می‌کنیم.

مثال ۶-۴-۴ ماتریس A و بردار b زیر را در نظر بگیرید

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix}$$

از مثال ۵-۴-۲، می‌دانیم که روش هاوس هولدر ماتریسهای زیر را ارائه می‌دهد.

$$R = \begin{bmatrix} -1/4142 & -2/1213 & -2/8284 \\ \cdot & 1/2247 & 1/6230 \\ \cdot & \cdot & -0/5774 \end{bmatrix}$$

$$H_1 = \begin{bmatrix} \cdot & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \cdot & -0/7071 & -0/7071 \\ -0/7071 & 0/5000 & -0/5000 \\ -0/7071 & -0/5000 & 0/5000 \end{bmatrix}$$

$$H_2 = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & -0/1691 & -0/9856 \\ \cdot & -0/9856 & 0/1691 \end{bmatrix}$$

محاسبه b' :

$$y_1 = b = \begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix}$$

$$y_2 = Hy_1 = \begin{bmatrix} -6/3640 \\ 0/858 \\ -2/9142 \end{bmatrix}$$

$$b' = y_3 = Hy_2 = \begin{bmatrix} -6/3640 \\ 2/8577 \\ -0/5774 \end{bmatrix}$$

(توجه کنید که b' فوق بدون تشکیل صریح ماتریس Q محاسبه شده است).حل $Rx = b'$:

$$x = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

شمارش اعمال ممیز شناور. اگر روش هاوس هولدر برای تجزیه A به QR مورد استفاده قرار گیرد، حل $Ax = b$ نیاز به $\frac{1}{3}n^3 + O(n^2)$ عمل ممیز شناور (flops) دارد (بخش ۵-۴-۱): از طرف دیگر روش دورانه‌های گیونز تقریباً به دو برابر این تعداد اعمال نیاز دارد (بخش ۵-۵-۱).

خاصیت گردکردن. از فصل ۵ می‌دانیم که برای تجزیه QR ماتریس A هر دو روش هاوس هولدر و گیونز به طور عددی پایدار هستند. فرایند جایگزینی پسرو برای حل یک دستگاه بالا مثلثی نیز به طور عددی پایدار است (بخش ۳-۱-۳). بنابراین روش حل دستگاه $Ax = b$ با استفاده از تجزیه QR احتمالاً باید پایدار باشد. در واقع چنین است.

نتیجه خطای گردکردن برای حل $Ax = b$ با استفاده از تجزیه QR

می‌توان نشان داد (مرجع (Lawson and Hanson (1974, 92) که جواب محاسبه شده \hat{x} ,

به دست آمده از روش QR هاوس هولدر، جواب واقعی دستگاه

$$(A + E)\hat{x} = b + \delta b$$

است که در آن $\|E\|_F \leq (\gamma n^2 + \epsilon n) \mu \|A\|_F + O(\mu^2)$ و

$$\|\delta b\| \leq (\gamma n^2 + \epsilon n) \mu \|b\| + O(\mu^2)$$

تبصره: در عبارت قبل «عامل رشد» وجود ندارد.

۴-۴-۶ حل یک دستگاه خطی با چند طرف راست

مسأله

$$AX = B$$

را در نظر بگیرید که در آن $B = (b_1, \dots, b_m)$ یک ماتریس $m \times n$ ($m \leq n$) است و $X = (x_1, x_2, \dots, x_m)$ در اینجا x_i و b_i ، $i = 1, \dots, m$ بردارهای n بعدی هستند.

مسائلی از این نوع در بیشتر کاربردها مطرح می‌شوند (یک چنین کاربردی در بخش ۶-۴-۷ صفحه ۳۳۱ ملاحظه شده است).

به محض آن‌که A با استفاده از یکی از روشهای توصیف شده در فصل ۵ تجزیه شود، تجزیه می‌تواند برای حل m دستگاه خطی فوق مورد استفاده قرار گیرد. ما روش را فقط با محورگیری جزئی بیان می‌کنیم.

الگوریتم ۳-۴-۶ حل $AX = B$: دستگاه خطی با چند طرف راست

گام ۱: تجزیه $MA = U$ ، با استفاده از روش حذفی گاوس با محورگیری جزئی

گام ۲: حل m دستگاه خطی

$$Ux_i = b_i' = Mb_i, \quad i = 1, \dots, m$$

شمارش اعمال ممیز شناور. الگوریتم حدوداً به $n^3/3 + mn^2$ عمل ممیز شناور (flops) نیاز دارد.

مثال ۵-۴-۶ دستگاه $AX = B$ را حل کنید که در آن

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$U = \begin{bmatrix} 7 & 8 & 9 \\ \cdot & \frac{6}{7} & \frac{19}{7} \\ \cdot & \cdot & \frac{-1}{7} \end{bmatrix}, \quad M = \begin{bmatrix} \cdot & \cdot & 1 \\ 1 & \cdot & -1 \\ -\frac{1}{7} & 1 & -\frac{1}{7} \end{bmatrix}$$

(مثال ۴-۲-۵ را ملاحظه کنید)

حل دستگاه:

$$Ux_1 = Mb_1 = \begin{bmatrix} 5 \\ 0.2857 \\ 0 \end{bmatrix}$$

$$x_1 = \begin{bmatrix} 0.2333 \\ 0.2333 \\ 0 \end{bmatrix}$$

حل دستگاه:

$$Ux_2 = Mb_2 = \begin{bmatrix} 6 \\ 1/1429 \\ 0 \end{bmatrix}, \quad x = \begin{bmatrix} -0.6667 \\ 1/2333 \\ 0 \end{bmatrix}$$

۶-۴-۷ دستگاههای خاص

در این زیر بخش دستگاههای خاص زیر را مطالعه می‌کنیم.

۱. دستگاههای معین مثبت متقارن

۲. دستگاههای هسنبرگی

۳. دستگاههای غالب قطری

۴. دستگاههای سه قطری و سه قطری بلوکی

در بخش ۶-۳ ملاحظه کردیم که این دستگاهها اغلب در کاربردهای عملی نظیر حل عددی معادلات دیفرانسیل جزئی رخ می‌دهند. در واقع اغلب توسط مهندسان با تجربه گفته می‌شود که به سختی دستگاههایی در کاربردهای عملی یافت می‌شوند که از نوع بالا نباشند. بنابراین، این دستگاهها استحقاق دارند که رفتارهای خاص بر روی آنها انجام پذیرد.

دستگاههای معین مثبت متقارن

ابتدا ما نشان می‌دهیم که برای یک ماتریس معین مثبت متقارن A ، یک تجزیه منحصر به فرد

$$A = HH^T$$

وجود دارد که در آن H یک ماتریس پایین مثلثی است. این تجزیه به نام مهندس فرانسوی چولسکی (۱۸۷۵-۱۹۱۸) تجزیه چولسکی نامیده می‌شود. وجود تجزیه چولسکی برای یک ماتریس معین مثبت متقارن A می‌تواند یا از روی تجزیه LU ماتریس A یا با پیدا کردن مستقیم ماتریس H از روی رابطه قبل دیده شود. در محاسبات عملی، از روش اخیر استفاده خواهیم کرد.

برای دیدن این امر از روی تجزیه LU ، توجه داریم که چون A یک ماتریس معین مثبت است و در نتیجه دارای کهدهای اصلی پیشرو مثبت است، بنابراین قضیه ۵-۲-۱ تجزیه

$$A = LU$$

منحصر به فرد می‌باشد. ماتریس بالا مثلثی U را می‌توان به صورت

$$U = DU_1$$

نوشت که در آن

$$\begin{aligned} D &= \text{diag}(u_{11}, u_{22}, \dots, u_{nn}) \\ &= \text{diag}(a_{11}, a_{22}^{(2)}, \dots, a_{nn}^{(n-1)}) \end{aligned}$$

و U_1 یک ماتریس بالا مثلثی واحد است. بنابراین

$$A = LDU_1$$

چون $A = A^T$ ، داریم

$$U_1^T D L^T = LDU_1$$

یا

$$D = (U_1^T)^{-1} LDU_1 (L^T)^{-1}$$

ماتریس $(U_1^T)^{-1} L$ یک ماتریس پایین مثلثی واحد، و ماتریس $U_1 (L^T)^{-1}$ یک ماتریس بالا مثلثی واحد است. بنابراین نتیجه می‌شود که

$$(U_1^T)^{-1} L = U_1 (L^T)^{-1} = I$$

یعنی

$$U_1 = L^T$$

بنابراین A می‌تواند به صورت

$$A = LDL^T$$

نوشته شود که در آن L یک ماتریس پایین مثلثی واحد است. چون کهدهای اصلی پیشرو ماتریس A ، $a_{11}, a_{22}^{(1)}, \dots, a_{nn}^{(n-1)}$ هستند، A معین مثبت است اگر و فقط اگر عناصر محوری

$a_{nn}^{(n-1)}, \dots, a_{22}^{(1)}, a_{11}$ مثبت باشند. یعنی هنگامی که A معین مثبت است، عناصر قطری D مثبت می‌باشند؛ بنابراین می‌توان نوشت

$$D = D^{1/2} \cdot D^{1/2}$$

که

$$D^{1/2} = \text{diag} \left[\sqrt{a_{11}}, \sqrt{a_{22}^{(1)}}, \dots, \sqrt{a_{nn}^{(n-1)}} \right] = \text{diag} (\sqrt{u_{11}}, \sqrt{u_{22}}, \dots, \sqrt{u_{nn}})$$

بنابراین

$$A = LDL^T = LD^{1/2} D^{1/2} L^T = HH^T$$

توجه کنید که عناصر قطری $H = LD^{1/2}$ مثبت هستند.

این بحث را می‌توان در قضیه زیر خلاصه کرد.

قضیه ۶-۴-۱ قضیه تجزیه چولسکی. فرض کنید A یک ماتریس معین مثبت متقارن باشد. آنگاه A می‌تواند به طور منحصر به فرد به شکل

$$A = HH^T$$

نوشته شود که در آن H یک ماتریس پایین مثلثی با عناصر قطری مثبت می‌باشد. یک عبارت صریح برای H به صورت

$$H = LD^{1/2}$$

ارائه می‌شود، که در آن L ماتریس پایین مثلثی واحد در تجزیه LU ماتریس A و به دست آمده توسط روش حذفی گاوس بدون محورگیری است و

$$D^{1/2} = \text{diag}(u_{11}^{1/2}, \dots, u_{nn}^{1/2})$$

روش حذفی گاوس برای ماتریس معین مثبت متقارن

اکنون با ارائه برخی خواص پایا و قابل توجه از ماتریسهای معین مثبت متقارن نشان می‌دهیم که روش حذفی گاوس بدون محورگیری برای ماتریسهای معین مثبت متقارن پایدار است. مثال زیر نشان می‌دهد حتی هنگامی که یک عنصر محوری کوچک وجود دارد، روش حذفی گاوس بدون محورگیری موجبات رشد را در عناصر ماتریسهای $A^{(k)}$ فراهم نمی‌سازد. فرض کنید

$$A = \begin{bmatrix} 0/00003 & 0/00500 \\ 0/00500 & 1/0000 \end{bmatrix}$$

فقط یک گام وجود دارد. عنصر محوری $0/00003$ کوچک است. مضرب m_{21} بزرگ است:

$$m_{21} = -\frac{a_{21}}{a_{11}} = -\frac{0/000500}{0/000003} = -\frac{500}{3}$$

اما

$$A^{(1)} = \begin{bmatrix} 0/00003 & 0/00500 \\ \cdot & 0/166667 \end{bmatrix}$$

عناصر $A^{(1)}$ رشد نمی‌کنند. در حقیقت $1 = \max(a_{ij}) < \max(a_{ij}^{(1)}) = 0/166667$. این پدیده جالب از روش حذفی گاوس بدون محورگیری و به کار برده شده در مورد مثال ساده 2×2 قبل (معین مثبت) را می‌توان به صورت زیر بیان کرد.

قضیه ۶-۴-۲ فرض کنید $A = (a_{ij})$ یک ماتریس معین مثبت متقارن $n \times n$ باشد و فرض کنید $A^{(k)} = (a_{ij}^{(k)})$ ماتریسهای کاهش یافته حاصل از به کار بردن روش حذفی گاوس بدون محورگیری بر A باشند. آنگاه

۱. برای هر $1 \leq k \leq n-1$ ، زیرماتریس به دست آمده از حذف k سطر اول و k ستون اول $A^{(k)}$ معین مثبت متقارن است.

$$2. \max |a_{ij}^{(k)}| \leq \max |a_{ij}^{(k-1)}|, \quad k = 1, 2, \dots, n-1.$$

اثبات: نتایج را فقط برای گام اول حذفی ثابت می‌کنیم. زیرا سپس نتایج برای گامهای دیگر به طور استقرایی ثابت می‌شوند.

بعد از اولین گام حذفی، داریم

$$A^{(1)} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \cdot & a_{22}^{(1)} & \dots & a_{2n}^{(1)} \\ \vdots & \vdots & & \vdots \\ \cdot & a_{n2}^{(1)} & \dots & a_{nn}^{(1)} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ \cdot & & & B \\ \vdots & & & \\ \cdot & & & \end{bmatrix}$$

برای اثبات این که B معین مثبت است، شکل درجه دوم زیر را ملاحظه کنید

$$\begin{aligned} \sum_{i=2}^n \sum_{j=2}^n a_{ij}^{(1)} x_i x_j &= \sum_{i=2}^n \sum_{j=2}^n \left[a_{ij} - \frac{a_i a_j}{a_{11}} \right] x_i x_j \\ &= \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j - a_{11} \left[x_1 + \sum_{i=2}^n \frac{a_{i1}}{a_{11}} x_i \right]^2 \end{aligned}$$

اگر B معین مثبت نباشد، آنگاه x_1, \dots, x_n وجود خواهند داشت به طوری که

$$\sum_{i=2}^n \sum_{j=2}^n a_{ij}^{(1)} x_i x_j \leq 0$$

اگر با این مقادیر x_2 تا x_n مقدار x_1 را به صورت:

$$x_1 = - \sum_{i=2}^n \frac{a_{i1}}{a_{11}} x_i$$

تعریف کنیم، آنگاه از عبارت قبل شکل درجه دوم

$$\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \leq 0$$

را داریم که با این حقیقت که A معین مثبت است، تناقض دارد. بنابراین B معین مثبت است.

همچنین، توجه داریم که

$$a_{ii}^{(1)} \leq a_{ii}, \quad i = 1, 2, \dots, n$$

زیرا

$$0 \leq a_{ii}^{(1)} = a_{ii} - \frac{a_{i1}^2}{a_{11}} \leq a_{ii} \quad (a_{11} > 0 \text{ زیرا})$$

بنابراین، هر عنصر قطری $A^{(1)}$ کوچکتر یا مساوی عنصر قطری متناظرش از A می‌باشد. چون بزرگترین عنصر (ازلحاظ قدر مطلق) یک ماتریس معین مثبت بر روی قطر قرار دارد، داریم

$$\max |a_{ij}^{(1)}| \leq \max |a_{ij}|$$

نتیجه‌ای از قضیه ۶-۴-۶

از قضیه ۶-۴-۶، بلافاصله نتیجه می‌گیریم که اگر $|a_{ij}| \leq 1$ ، آنگاه $|a_{ij}^{(k)}| \leq 1$ یعنی در این حالت عامل رشد ρ برابر ۱ است.

عامل رشد و پایداری روش حذفی گاوس برای یک ماتریس معین مثبت متقارن
عامل رشد ρ روش حذفی گاوس بدون محورگیری برای یک ماتریس معین مثبت متقارن
برابر ۱ است. بنابراین روش حذفی گاوس بدون محورگیری برای یک ماتریس معین مثبت
متقارن پایدار است.

مثال ۶-۴-۶

$$A = \begin{bmatrix} 5 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 5 \end{bmatrix}$$

$$A^{(1)} = \begin{bmatrix} 5 & 1 & 1 \\ 0 & \frac{4}{5} & \frac{4}{5} \\ 0 & \frac{4}{5} & \frac{24}{5} \end{bmatrix}$$

کهادهای اصلی پیشرو زیر ماتریس به دست آمده از حذف سطرها اول و ستون اول برابر $\frac{4}{5}$ و $\frac{16}{5}$ هستند. همچنین

$$\max(a_{ij}^{(1)}) = 5$$

$$A^{(2)} = \begin{bmatrix} 5 & 1 & 1 \\ 0 & \frac{4}{5} & \frac{4}{5} \\ 0 & 0 & 4 \end{bmatrix}$$

کهادهای اصلی پیشرو زیر ماتریس به دست آمده از حذف دو سطر اول و دو ستون اول برابر ۴ است. بعلاوه $\max(a_{ij}^{(2)}) = 5 = \max(a_{ij}^{(1)})$ عامل رشد برابر ۱ $\rho = \frac{5}{5}$ است.

محاسبه تجزیه چولسکی

قضیه ۶-۴-۱ پیشنهاد می کند که روش حذفی گاوس می تواند برای محاسبه تجزیه چولسکی یک ماتریس معین مثبت متقارن استفاده شود، و ملاحظه کردیم که روش حذفی گاوس بدون محورگیری برای یک چنین ماتریسی پایدار است، ولیکن در عمل روش حذفی گاوس برای محاسبه تجزیه چولسکی استفاده نمی شود. به جای آن می توان از یک روش کاراتر که الگوریتم چولسکی نامیده می شود، برای انجام آن استفاده کرد.

الگوریتم چولسکی

از

$$A = HH^T$$

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \begin{bmatrix} h_{11} & 0 & \dots & 0 \\ h_{21} & h_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \dots & h_{nn} \end{bmatrix} \begin{bmatrix} h_{11} & h_{21} & \dots & h_{n1} \\ 0 & h_{22} & \dots & h_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & h_{nn} \end{bmatrix}$$

داریم

$$h_{11} = \sqrt{a_{11}}, \quad h_{i1} = \frac{a_{i1}}{h_{11}}, \quad i = 2, \dots, n$$

$$\sum_{k=1}^i h_{ik}^2 = a_{ii}, \quad a_{ij} = \sum_{k=1}^j h_{ik} h_{jk}, \quad j < i$$

که منجر به الگوریتم چولسکی^۱ می‌شود.

الگوریتم ۶-۴-۴ الگوریتم چولسکی. با مفروض بودن یک ماتریس معین مثبت متقارن $n \times n$ مانند A ، الگوریتم زیر عامل چولسکی H را محاسبه می‌کند. ماتریس H سطر به سطر محاسبه می‌شود و در قسمت پایین مثلثی A ذخیره می‌شود.

For $k = 1, 2, \dots, n$ do

For $i = 1, 2, \dots, k - 1$ do

$$a_{ki} \equiv h_{ki} = \frac{1}{h_{ii}} \left[a_{ki} - \sum_{j=1}^{i-1} h_{ij} h_{kj} \right]$$

$$a_{kk} \equiv h_{kk} = \sqrt{a_{kk} - \sum_{j=1}^{k-1} h_{kj}^2}$$

تبصره‌ها:

۱. در شبه کد قبلی، $\sum_{j=1}^k () \equiv 0$. همچنین هنگامی که $k = 1$ ، حلقه داخلی انجام نمی‌شود.

۲. عامل چولسکی H سطر به سطر محاسبه می‌شود.

۳. معین مثبت بودن A موجب خواهد شد که مقدار زیر رادیکال مثبت باشد.

خاصیت گردکردن. فرض کنید عامل چولسکی محاسبه شده توسط \hat{H} نمایش داده شود. سپس می‌توان نشان داد (Demmel (1989)) که

$$A + E = \hat{H}(\hat{H})^T$$

$$\text{که در آن } |e_{ij}| \leq \frac{(n+1)\mu}{1 - (n+1)\mu} (a_{ii} a_{jj})^{1/2}$$

$$E = (e_{ij})$$

۱- این الگوریتم در برخی کاربردها (مثلاً آمار) به عنوان الگوریتم ریشه دوم معروف است. ولیکن یک الگوریتم بدون ریشه دوم نیز می‌توان تدوین کرد.

بنابراین الگوریتم تجزیه چولسکی پایدار است.

تذکر MATCOM : الگوریتم چولسکی (الگوریتم ۶-۴-۴) در برنامه CHOLES در MATCOM پیاده‌سازی شده است.

حل $Ax = b$ با استفاده از تجزیه چولسکی

با در دست داشتن تجزیه چولسکی $A = HH^T$ ، دستگاه خطی معین مثبت $Ax = b$ را می‌توان ابتدا با حل دستگاه پایین مثلثی $Hy = b$ و سپس حل دستگاه بالا مثلثی $H^T x = y$ ، حل کرد.

الگوریتم ۶-۴-۵ الگوریتم چولسکی برای دستگاه خطی معین مثبت $Ax = b$ گام ۱: پیدا کردن تجزیه چولسکی

$$A = HH^T$$

با استفاده از الگوریتم ۶-۴-۴

گام ۲: حل دستگاه پایین مثلثی نسبت به y

$$Hy = b$$

گام ۳: حل دستگاه بالا مثلثی نسبت به x

$$H^T x = y$$

مثال ۶-۴-۷ فرض کنید

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 5 & 5 \\ 1 & 5 & 14 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 11 \\ 20 \end{bmatrix}$$

گام ۱: پیدا کردن تجزیه چولسکی.

سطر اول ($k = 1$):

$$h_{11} = 1$$

سطر دوم ($k = 2$):

$$h_{21} = \frac{a_{21}}{h_{11}} = 1$$

$$h_{22} = \sqrt{a_{22} - h_{21}^2} = \sqrt{5 - 1} = \pm 2$$

(چون عناصر قطری H باید مثبت باشند، علامت + را اختیار می‌کنیم.)

سطر سوم ($k = 3$):

$$h_{31} = \frac{a_{31}}{h_{11}} = 1$$

$$h_{32} = \frac{1}{h_{22}}(a_{32} - h_{21}h_{31}) = \frac{1}{2}(5 - 1) = 2$$

$$h_{33} = \sqrt{a_{33} - (h_{31}^2 + h_{32}^2)} = \sqrt{14 - 5} = \sqrt{9}$$

(علامت + را اختیاری کنیم.)

$$h_{33} = +3$$

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{bmatrix}$$

گام ۲: حل $Hy = b$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 11 \\ 20 \end{bmatrix}$$

$$y_1 = 3,$$

$$y_2 = 4,$$

$$y_3 = 3$$

گام ۳: حل $H^T x = y$

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}$$

$$x_1 = 1,$$

$$x_2 = 1,$$

$$x_3 = 1$$

شمارش اعمال ممیز شناور. الگوریتم چولسکی برای محاسبه H به $n^3/6$ عمل ممیز شناور (flops) نیاز دارد که نصف تعداد اعمال ممیز شناور لازم برای انجام کار با استفاده از تجزیه LU می باشد. توجه کنید که فرایند همچنین به n ریشه دوم نیاز دارد. جواب هر دستگاه مثلثی $Hy = b$ و $H^T x = y$ نیاز به $n^2/2$ عمل ممیز شناور (flops) دارد. بنابراین جواب دستگاه معین مثبت $Ax = b$ با استفاده از الگوریتم چولسکی نیاز به $n^3/6 + n^2$ عمل ممیز شناور (flops) و n ریشه دوم دارد.

خاصیت گرد کردن. اگر \hat{x} جواب محاسبه شده دستگاه $Ax = b$ با استفاده از الگوریتم چولسکی

باشد، آنگاه می توان نشان داد که \hat{x} در رابطه

$$(A + E)\hat{x} = b$$

صدق می کند که در آن $\|E\|_2 \leq c\mu \|A\|_2$ می باشد و c یک ثابت کوچک است که بستگی به n دارد. بنابراین، الگوریتم چولسکی برای یک دستگاه معین مثبت متقارن کاملاً پایدار است.

خطای نسبی در جواب به دست آمده توسط الگوریتم چولسکی

اگر \hat{x} جواب محاسبه شده برای دستگاه معین مثبت $Ax = b$ با استفاده از الگوریتم چولسکی و به کمک حل دستگاههای مثلثی (به صورتی که هم اکنون توصیف شد) باشد، آنگاه می توان نشان داد که

$$\frac{\|x - \hat{x}\|}{\|\hat{x}\|} \leq O(\mu) \text{Cond}(A)$$

(یادآور می شویم که $\text{Cond}(A) = \|A\| \cdot \|A^{-1}\|$.)

دستگاههای نامعین متقارن

روش حذفی گاوس با محورگیری جزئی تقارن را حفظ نمی کند. بنابراین اگر روش حذفی گاوس با محورگیری جزئی برای یک دستگاه متقارن به کار برده شود حدوداً به $n^3/3$ عمل ممیز شناور نیاز دارد و $O(n^2)$ مقایسه نیاز خواهد داشت، و نتیجه تقارن را حفظ نخواهد کرد. الگوریتمهای کارا با حفظ تقارن نظیر الگوریتمهای تجزیه با محورگیری قطری وجود دارند. تجزیه با محورگیری قطری یک تجزیه به شکل $A = UDU^T$ است، که در آن U یک حاصلضرب از ماتریسهای بالا مثلثی واحد مقدماتی و جایگشت است و D یک ماتریس قطری بلوکی متقارن با بلوکهای از مرتبه ۱ و ۲ می باشد. تجزیه با محورگیری قطری حدوداً نیاز به $n^3/6$ عمل ممیز شناور (flops) دارد که نصف روش حذفی گاوس است. تجزیه همچنین در عمل پایدار است. این تجزیه عموماً به تجزیه های بانج - پارلت یا بانج - کوفمن معروف است. جزئیات آن را می توان در مراجع (Bunch (1971)، Bunch and Parlett (1971)، Bunch and Kaufman (1977) پیدا کرد. با مفروض بودن چنین تجزیه ای یک دستگاه متقارن می تواند با $O(n^2)$ عمل ممیز شناور حل شود.

دستگاه هستبرگی

دستگاه خطی

$$Ax = b$$

را در نظر بگیرید که در آن A یک ماتریس بالا هسنبرگی از مرتبه n می باشد. اگر روش حذفی گاوس با محورگیری جزئی برای مثلثی سازی A استفاده شود، و اگر $|a_{ij}| \leq 1$ باشد، آنگاه داریم

$$\left| a_{ij}^{(k+1)} \right| \leq k + 1 \quad (\text{Wilkinson (1965, 218)})$$

بنابراین می توان مطلب زیر را بیان کرد.

عامل رشد و پایداری روش حذفی گاوس برای یک دستگاه هسنبرگی عامل رشد برای یک ماتریس هسنبرگی با استفاده از روش حذفی گاوس با محورگیری جزئی توسط n کراندار می گردد. بنابراین یک دستگاه هسنبرگی می تواند با استفاده از محورگیری جزئی با اطمینان حل شود.

شمارش اعمال ممیز شناور. روش حذفی گاوس با محورگیری جزئی فقط به n^2 عمل ممیز شناور (flops) برای حل یک دستگاه هسنبرگی نیاز دارد، که به طور بامعنایی کمتر از $n^3/3$ عمل ممیز شناوری می باشد که برای حل یک دستگاه با یک ماتریس دلخواه مورد نیاز است. این وضعیت اتفاق می افتد زیرا در هر گام از روش حذفی گاوس در طی فرایند مثلثی سازی فقط یک عنصر لازم است حذف شود؛ چون $n - 1$ گام وجود دارد، فرایند مثلثی سازی حدوداً به $n^2/2$ عمل ممیز شناور (flops) نیاز دارد. به محض داشتن تجزیه

$$MA = U$$

دستگاه بالا مثلثی

$$Ux = Mb = b'$$

می تواند با $n^2/2$ عمل ممیز شناور (flops) حل شود. بنابراین یک دستگاه بالا هسنبرگی می تواند با فقط n^2 عمل ممیز شناور به طریقی پایدار با استفاده از روش حذفی گاوس با محورگیری جزئی حل شود.

مثال ۶-۴-۸ ماتریس

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 0 & 5 & 6 \end{bmatrix}$$

را با استفاده از محورگیری جزئی به شکل مثلثی درمی آوریم.

گام ۱:

$$P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P_1 A = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 3 \\ 0 & 5 & 6 \end{bmatrix}$$

$$\hat{M}_1 = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M_1 P_1 A = A^{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 3 \\ 0 & 5 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 3 & 4 \\ 1 & \frac{1}{2} & 3 \\ 0 & \boxed{5} & 6 \end{bmatrix}$$

گام ۲:

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$P_2 A^{(1)} = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & \frac{1}{2} & 1 \end{bmatrix}$$

$$\hat{M}_2 = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{5} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{10} & 1 \end{bmatrix}$$

$$U = A^{(2)} = M_2 P_2 A^{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{10} & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & \frac{1}{2} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 0 & \frac{2}{5} \end{bmatrix}$$

$$M = M_2 P_2 M_1 P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -\frac{1}{2} & -\frac{1}{10} \end{bmatrix}$$

عامل رشد، ρ ، را محاسبه می‌کنیم

$$\rho = \frac{\max(6, 6, 6)}{6} = 1$$

کاربردی از یک دستگاه هسنبرگی

محاسبه ماتریس پاسخ فرکانس

در نظریه کنترل اغلب لازم است برای مطالعه پاسخ یک دستگاه کنترل ماتریس

$$G(j\omega) = C(j\omega I - A)^{-1}B$$

به ازای مقادیر مختلف زیادی از ω محاسبه شود. ماتریسهای A ، B ، و C ماتریسهای یک دستگاه

کنترل هستند و به ترتیب از مرتبه‌های $n \times n$ ، $n \times m$ ، و $r \times n$ ($m \leq n$) می‌باشند. ماتریس $G(j\omega)$

ماتریس پاسخ فرکانس نامیده می‌شود ($j = \sqrt{-1}$).

چون محاسبه $(j\omega I - A)^{-1}B$ معادل حل m دستگاه خطی

$$(j\omega I - A)x_i = b_i, \quad i = 1, \dots, m$$

است که در آن b_i ، i امین ستون B می‌باشد (بخش ۶-۵-۱ را ملاحظه کنید)؛ برای هر ω ، حدوداً به

عمل ممیز شناور (flops) برای محاسبه $G(j\omega)$ (با استفاده از روش حذفی گاوس با محورگیری جزئی) نیاز داریم. معمولاً، لازم است که $G(j\omega)$ برای یک تعداد خیلی زیادی ω محاسبه شود، بنابراین چنین محاسبه‌ای بسیار زیاد خواهد بود.

از طرف دیگر اگر A در ابتدا به یک ماتریس هسنبرگی توسط تشابه متعامد

$$A = PHPT^T$$

تبدیل شود، آنگاه

$$G(j\omega) = CP(j\omega I - H)^{-1}P^T B$$

محاسبه $(j\omega I - H)^{-1}P^T B$ اکنون نیاز به حل m دستگاه هسنبرگی خواهد داشت که هر کدام فقط به n^2 عمل ممیز شناور (flops) نیاز خواهند داشت. پس برای محاسبه $G(j\omega) = CP(j\omega I - H)^{-1}P^T B$ برای هر ω ، صرفه‌جویی از مرتبه $O(n)$ برای هر ω وجود خواهد داشت. این شمارش اعمال مربوط به کاهش به شکل هسنبرگی را در بر ندارد. توجه کنید که ماتریس A یک بار به یک ماتریس هسنبرگی تبدیل می‌شود و همان ماتریس هسنبرگی برای محاسبه $G(j\omega)$ برای هر ω مورد استفاده قرار می‌گیرد.

بنابراین، محاسبه‌ای که از یک کاهش اولیه A به یک ماتریس هسنبرگی استفاده می‌کند خیلی کارا تر است. بعلاوه همان طور که قبلاً مشاهده نموده‌ایم کاهش به یک شکل هسنبرگی و حل دستگاه‌های هسنبرگی با استفاده از محورگیری جزئی هر دو محاسبات پایداری هستند. این راه در مرجع Laub (1981) پیشنهاد شده است.

تبصره‌ها: کاهش A به یک ماتریس هسنبرگی به عنوان شروعی برای محاسبات اصلی در نظریه کنترل مورد استفاده قرار می‌گیرد. الگوریتم‌های از نظر عددی توانا برای مسائل کنترل نظیر، کنترل‌پذیری، مشاهده‌پذیری، معادلات ماتریسی (لیاپانوف، سیلوستر، سیلوستر - ایزرور^۱، ریکاتی و غیره) به طور معمول قبل از شروع محاسبات اصلی ماتریس A را به یک ماتریس هسنبرگی تبدیل می‌کنند. خواننده‌های آشنا و علاقه‌مند به این مسائل می‌توانند به کتاب زیر جهت شرحی از این روشها مراجعه کنند.

Numerical Methods in Control Theory by B. N. Datta

برای مراجع مربوط به مقاله‌های فردی فصل ۸ را ملاحظه کنید. کتاب زیر نیز حاوی همه مقاله‌های

مناسب می باشد.

Numerical Linear Algebra Techniques for Systems and Control, edited by R. V. Patel, Alan Laub, and Paul Van Dooren (1994)

همچنین تز دکترای آرنولد ((Arnold (1992)) و مقاله مروری داتا ((Datta (1994)) را ملاحظه کنید.

دستگاههای غالب قطری

یادآور می شویم که یک ماتریس $A = (a_{ij})$ غالب قطری ستونی است اگر

$$|a_{11}| > |a_{21}| + |a_{31}| + \dots + |a_{n1}|$$

$$|a_{22}| > |a_{12}| + |a_{32}| + \dots + |a_{n2}|$$

⋮

$$|a_{nn}| > |a_{1n}| + |a_{2n}| + \dots + |a_{n-1,n}|$$

یک ماتریس غالب قطری، نظیر یک ماتریس معین مثبت متقارن، دارای این خاصیت جالب توجه است که تعویضهای سطرهای در هیچ گامی از فرایند مثلثی سازی برای روش حذفی گاوس با محورگیری جزئی لازم نیستند. عنصر محوری قبلاً در مکان درست قرار گرفته است. بنابراین در گام اول، a_{11} از لحاظ اندازه از همه عناصر ستون اول بزرگتر است، تعویض سطری لازم نیست. سپس در انتهای گام اول، داریم:

$$A^{(1)} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \cdot & a_{22}^{(1)} & \dots & a_{2n}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & a_{n2}^{(1)} & \dots & a_{nn}^{(1)} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \cdot & & & \\ \vdots & & & \\ \cdot & & & \end{bmatrix} \quad A'$$

می توان نشان داد (تمرین ۱۶ الف) که A' دوباره یک ماتریس غالب قطری است، بنابراین $a_{22}^{(1)}$ عنصر محوری برای گام دوم می باشد. این فرایند می تواند به طور واضح ادامه یابد و این نشان می دهد که محورگیری برای ماتریسهای غالب قطری ستونی لازم نمی باشد. بعلاوه مطلب زیر به سادگی ثابت می شود.

عامل رشد و پایداری روش حذفی گاوس برای ماتریسهای غالب قطری
 عامل رشد ρ برای یک ماتریس غالب قطری ستونی با استفاده از روش حذفی گاوس بدون
 محورگیری توسط ۲ کراندار می شود (تمرین ۱۶): $\rho \leq 2$ ، بنابراین برای دستگاههای غالب
 قطری ستونی، روش حذفی گاوس بدون محورگیری پایدار است.

مثال ۶-۴-۹

$$A = \begin{bmatrix} \boxed{5} & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 5 \end{bmatrix}$$

گام ۱:

$$A^{(1)} = \begin{bmatrix} 5 & \boxed{1} & 1 \\ 0 & \frac{24}{5} & \frac{4}{5} \\ 0 & \frac{4}{5} & \frac{24}{5} \end{bmatrix}$$

گام ۲:

$$A^{(2)} = \begin{bmatrix} 5 & 1 & 1 \\ 0 & \frac{24}{5} & \frac{4}{5} \\ 0 & 0 & \frac{14}{3} \end{bmatrix}$$

عامل رشد $\rho = 1$ است. (توجه کنید که برای این مثال A ماتریس غالب قطری ستونی و معین مثبت است؛ بنابراین $\rho = 1$).

مثال بعد نشان می دهد که عامل رشد روش حذفی گاوس برای یک ماتریس غالب قطری ستونی می تواند بزرگ تر از ۱ باشد، اما همواره کوچک تر یا مساوی ۲ است.

مثال ۶-۴-۱۰

$$A = \begin{bmatrix} 5 & -8 \\ 1 & 10 \end{bmatrix}$$

$$A^{(1)} = \begin{pmatrix} 5 & -8 \\ 0 & 11/6 \end{pmatrix}$$

عامل رشد عبارت است از

$$\rho = \frac{\max(10, 11/6)}{10} = \frac{11/6}{10} = 1/16$$

دستگاههای سه قطری

تجزیه LU یک ماتریس سه قطری T ، اگر وجود داشته باشد، ممکن است L و U بی نتیجه دهد که دارای ساختارهای ساده خیلی خاص باشند؛ هر دو دو قطری، L دارای ۱ در امتداد قطر اصلی و عناصر قطر فوقانی U همان عناصر قطر فوقانی T می باشند. بویژه اگر بنویسیم

$$T = \begin{pmatrix} a_1 & b_1 & & \\ c_2 & \ddots & \ddots & \\ & \ddots & \ddots & b_{n-1} \\ & & c_n & a_n \end{pmatrix}$$

$$= \begin{pmatrix} 1 & & & \\ l_2 & \ddots & \ddots & \\ & \ddots & \ddots & \\ & & \cdot & l_n \end{pmatrix} \begin{pmatrix} u_1 & b_1 & & \\ & \ddots & \ddots & \\ & & \ddots & b_{n-1} \\ & & \cdot & u_n \end{pmatrix}$$

با مساوی قرار دادن عناصر نظیر به نظیر ماتریسهای دو طرف تساوی، ملاحظه می کنیم که

$$a_1 = u_1$$

$$c_i = l_i u_{i-1}, \quad i = 2, \dots, n$$

$$a_i = u_i + l_i b_{i-1}, \quad i = 2, \dots, n$$

از این روابط $\{u_i\}$ و $\{l_i\}$ می توانند به سادگی محاسبه شوند.

محاسبه تجزیه LU یک ماتریس سه قطری

$$u_1 = a_1$$

For $i = 2, \dots, n$ do

$$l_i = \frac{c_i}{u_{i-1}}$$

$$u_i = a_i - l_i b_{i-1}$$

شمارش اعمال ممیز شناور. فرایند قبلی فقط با $2n - 2$ عمل ممیز شناور (flops) انجام می شود.

حل یک دستگاه سه قطری

به محض این که تجزیه ساده قبلی T را داشته باشیم، جواب دستگاه سه قطری $Tx = b$

می تواند با حل دو دستگاه دو قطری زیر انجام شود

$$Ly = b$$

و

$$Ux = y$$

شمارش اعمال ممیز شناور. جوابهای این دو دستگاه دو قطری همچنین به $2n - 2$ عمل ممیز شناور نیاز دارند.

بنابراین یک دستگاه سه قطری می تواند با این رویه فقط با $4n - 4$ عمل ممیز شناور حل شود که به راستی روشی بسیار ارزان است.

پایداری فرایند. متأسفانه این روش تجزیه با شکست مواجه می شود اگر هر u_i برابر صفر باشد. حتی اگر همه u_i ها از لحاظ نظری مخالف صفر باشند، پایداری فرایند به طور کلی نمی تواند تضمین شود. ولیکن در بسیاری از وضعیتهای عملی، نظیر گنسته سازی معادله پواسون، ماتریسهای سه قطری معین مثبت متقارن هستند، که در این حالتها، همان طور که قبلاً ملاحظه نمودیم، رویه فوق کاملاً پایدار است. در حقیقت، در حالت معین مثبت متقارن، این رویه بر روش تجزیه چولسکی ترجیح داده می شود زیرا حاوی محاسبه هیچ ریشه دومی نیست. درست است که تجزیه چولسکی یک ماتریس سه قطری

معین مثبت متقارن می تواند همچنین با $O(n)$ عمل ممیز شناور محاسبه شود، ولیکن n ریشه دوم نیز باید محاسبه شوند.

در حالت کلی، برای حفظ پایداری باید از روش حذفی گاوس با محورگیری جزئی استفاده شود. اگر $|a_i|, |b_i|, |c_i| \leq 1$ ، $(i = 1, \dots, n)$ ، آنگاه می توان نشان داد (Wilkinson (1965, 219)) که عناصر $A^{(k)}$ در هر گام توسط ۲ کراندار خواهند شد.

عامل رشد و پایداری روش حذفی گاوس برای یک ماتریس سه قطری

عامل رشد برای روش حذفی گاوس با محورگیری جزئی برای یک ماتریس سه قطری توسط ۲ کراندار می گردد:

$$\rho \leq 2$$

بنابراین روش حذفی گاوس با محورگیری جزئی برای یک دستگاه سه قطری خیلی پایدار است.

در این حالت تعداد اعمال ممیز شناور کمی بیشتر است؛ در حدود $7n$ عمل ممیز شناور برای حل دستگاه $Tx = b$ لازم است ($3n$ عمل ممیز شناور برای تجزیه و $4n$ عمل ممیز شناور برای حل دو دستگاه مثلثی)، اما هنوز هم یک رویه با $O(n)$ عمل ممیز شناور است.

اگر T متقارن باشد، طبیعی است که بخواهیم از خاصیت تقارن استفاده کنیم؛ در این حال از الگوریتمهای تجزیه با محورگیری قطری می توان استفاده کرد. این الگوریتمها می توانند به صورتی مرتب شوند که تعداد اعمال ممیز شناور قابل مقایسه با تعداد اعمال ممیز شناور روش حذفی گاوس با محورگیری جزئی باشد و حافظه کمتری جهت ذخیره سازی لازم باشد.

مثال ۶-۴-۱۱ با استفاده از (الف) فرمول $A = LU$ و (ب) روش حذفی گاوس با محورگیری جزئی ماتریس زیر را مثلثی سازی کنید.

$$A = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.8 & 0.5 & 0.1 \\ 0 & 0.1 & 0.5 \end{bmatrix}$$

(الف) تشکیل $A = LU$

$$u_1 = 0/9$$

$i = 2$:

$$l_2 = \frac{c_2}{u_1} = \frac{0/8}{0/9} = \frac{8}{9} = 0/8889$$

$$u_2 = a_2 - l_2 b_1 = 0/5 - \frac{8}{9} \times 0/1 = 0/4111$$

$i = 3$:

$$l_3 = \frac{c_3}{u_2} = \frac{0/1}{0/41} = 0/2432$$

$$u_3 = a_3 - l_3 b_2 = 0/5 - 0/24 \times 0/1 = 0/4757$$

$$L = \begin{bmatrix} 1 & \cdot & \cdot \\ 0/8889 & 1 & \cdot \\ \cdot & 0/2432 & 1 \end{bmatrix}$$

$$U = \begin{bmatrix} 0/9 & 0/1 & \cdot \\ \cdot & 0/4111 & 0/1 \\ \cdot & \cdot & 0/4757 \end{bmatrix}$$

(ب) با استفاده از روش حذفی گاوس و محورگیری جزئی، داریم:

گام ۱:

$$\text{ضریبگر } m_{21} = -\frac{0/8}{0/9} = -0/8889$$

$$A^{(1)} = \begin{bmatrix} 0/9 & 0/1 & \cdot \\ \cdot & 0/4111 & 0/1 \\ \cdot & 0/1 & 0/5 \end{bmatrix}$$

گام ۲:

$$\text{ضریبگر } m_{32} = -\frac{0/1}{0/4111} = -0/2432$$

$$A^{(2)} = \begin{bmatrix} 0/9 & 0/1 & \cdot \\ \cdot & 0/4111 & 0/1 \\ \cdot & \cdot & 0/4757 \end{bmatrix} = U$$

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -m_{21} & 1 & 0 \\ 0 & -m_{32} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0.8889 & 1 & 0 \\ 0 & 0.2432 & 1 \end{bmatrix}$$

دستگاههای سه قطری بلوکی

در این بخش حل دستگاه سه قطری بلوکی

$$Tx = b$$

را ملاحظه می‌کنیم که در آن T یک ماتریس سه قطری بلوکی و $b = (b_1, b_2, \dots, b_N)^T$ یک بردار بلوکی است. تعداد مؤلفه‌های بردار بلوکی b_i برابر بعد ماتریس قطری بلوکی i ام در T است.

تجزیه LU بلوکی

روش تجزیه ارائه شده در ابتدای این بخش می‌تواند به راحتی برای ماتریس سه قطری بلوکی زیر گسترش یابد.

$$T = \begin{bmatrix} A_1 & B_1 & & & \\ C_2 & \ddots & \ddots & & 0 \\ & \ddots & \ddots & \ddots & \\ & & 0 & \ddots & B_{N-1} \\ & & & C_N & A_N \end{bmatrix}$$

بنابراین، اگر T دارای تجزیه LU بلوکی

$$T = \begin{bmatrix} I & & & & \\ L_2 & \ddots & & & \\ & \ddots & \ddots & & \\ & & 0 & \ddots & \\ & & & L_N & I \end{bmatrix} \begin{bmatrix} U_1 & B_1 & \dots & & 0 \\ & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & B_{N-1} \\ 0 & \dots & \dots & & U_N \end{bmatrix} = LU$$

باشد، آنگاه ماتریسهای L_i ، $i = 2, \dots, N$ و U_i ، $i = 1, \dots, N$ به صورت زیر محاسبه می‌شوند.

الگوریتم ۶-۴-۶ تجزیه LU بلوکی

Set $U_1 = A_1$

For $i = 2, \dots, N$ do

Step 1: Solve for L_i :

$$L_i U_{i-1} = C_i$$

Step 2: Compute U_i :

$$U_i = A_i - L_i B_{i-1}$$

حل دستگاههای بلوکی

به محض داشتن تجزیه قبلی، می توان جواب دستگاه

$$Tx = b$$

را با حل متوالی $Ly = b$ و $Ux = y$ پیدا کرد. جواب $Ly = b$ می تواند توسط روش حذفی پیشرو بلوکی به دست آید، و جواب $Ux = y$ می تواند توسط روش جایگزینی پسرو بلوکی حاصل شود.

الگوریتم ۷-۴-۶ روش حذفی پیشرو بلوکی

Set $L_1 y_0 = 0$

For $i = 1, \dots, n$ do

$$y_i = b_i - L_i y_{i-1}$$

الگوریتم ۸-۴-۶ روش جایگزینی پسرو بلوکی

Set $B_N x_{N+1} = 0$

For $i = N, \dots, 1$ do

Solve for x_i : $U_i x_i = y_i - B_i x_{i+1}$

مثال ۱۲-۴-۶

$$A = \begin{bmatrix} 4 & -1 & 1 & 0 \\ -1 & 4 & 0 & 1 \\ 1 & 0 & 2 & -1 \\ 0 & 1 & -1 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 4 \\ 2 \\ 2 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$b_1 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \quad b_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

تجزیه LU بلوکی:

$$U_1 = A_1 = \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix}$$

$i = 2$

۱. محاسبه L_2 :

$$L_2 U_1 = I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$L_2 = U_1^{-1} = \begin{bmatrix} 0.2667 & 0.0667 \\ 0.6667 & 0.2667 \end{bmatrix}$$

۲. محاسبه U_2 :

$$U_2 = A_2 - L_2 B_1$$

$$= \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} - \begin{bmatrix} 0.2667 & 0.0667 \\ 0.6667 & 0.2667 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1.7333 & -1.0667 \\ -1.0667 & 1.7333 \end{bmatrix}$$

حذفی پیشرو بلوکی:

$$y_1 = b_1 - L_2 y_0 = b_1 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

$$y_2 = b_2 - L_2 y_1 = \begin{bmatrix} 0.6667 \\ 0.6667 \end{bmatrix}$$

جایگزینی پسرو بلوکی:

$$U_2 x_2 = y_2 - B_2 x_2 = y_2 = \begin{bmatrix} 0.6667 \\ 0.6667 \end{bmatrix} \quad (B_2 x_2 = 0)$$

$$x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$U_1 x_1 = y_1 - B_1 x_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

کاهش دوری بلوکی

اغلب در عمل ماتریس سه قطری بلوکی یک دستگاه ممکن است دارای برخی خواص خاص باشد که بتوان از آنها بهره‌برداری کرد و با استفاده از روشی که کاهش دوری بلوکی نامیده می‌شود، دستگاه را به یک دستگاه با مرتبه کمتر کاهش داد. برای جزئیات بیشتر کتاب Golub and Van Loan (1989, 173-6) و مرجعهای داخل آن را ملاحظه کنید.

۶-۴-۸ مقیاس کردن

اگر عناصر ماتریس A زیاد تفاوت داشته باشند، آنگاه ممکن است لازم باشد در طول فرایند حذفی یک عدد کوچک به یک عدد خیلی بزرگ اضافه شود. یک چنین وضعیتی می‌تواند تأثیر زیادی بر دقت بگذارد، زیرا مقدار بزرگتر می‌تواند مقدار کوچک‌تر را از بین ببرد. برای پیشگیری از این مشکل اغلب پیشنهاد می‌شود که قبل از شروع فرایند حذفی سطرهای A به طور مناسب مقیاس شوند. مثال ساده زیر این مطلب را توضیح می‌دهد.

دستگاه

$$\begin{bmatrix} 1.0 & 1.0^6 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1.0^6 \\ 2 \end{bmatrix}$$

را در نظر بگیرید. روش حذفی گاوس با محورگیری جزئی را به کار می‌بریم. چون 1.0 بزرگ‌ترین عنصر در ستون اول است، تعویض لازم نیست. بعد از گام اول روش حذفی و با استفاده از حساب 4 رقمی داریم:

$$\begin{bmatrix} 1.0 & 1.0^6 \\ 0 & -1.0^5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1.0^6 \\ -1.0^5 \end{bmatrix}$$

که نتیجه می‌دهد $x_2 = 1$ ، $x_1 = 0$. ولیکن جواب واقعی تقریباً برابر است با

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

بنابراین حتی انتخاب عنصر محوری کاذب 1.0 نمی‌تواند به ما کمک کند. ولیکن اگر عناصر سطر اول

ماتریس را با تقسیم بر 10^6 مقیاس کنیم و سپس دستگاه را (پس از تغییر عنصر اول b به طور مناسب) با استفاده از محورگیری جزئی حل کنیم، آنگاه جواب دقیق را، به همان گونه که قبلاً دیده‌ایم خواهیم داشت.

مقیاس کردن سطرهاى یک ماتریس A معادل با پیدا کردن یک ماتریس قطری معکوس‌پذیر D_1 است به طوری که بزرگ‌ترین عنصر (از لحاظ اندازه) در هر سطر $D_1^{-1}A$ تقریباً هم‌اندازه باشند. به محض این که چنین D_1 ای پیدا شود، جواب دستگاه

$$Ax = b$$

با حل دستگاه مقیاس شده

$$\tilde{A}x = \tilde{b}$$

پیدا می‌شود که در آن

$$\tilde{A} = D_1^{-1}A, \quad \tilde{b} = D_1^{-1}b$$

فرایند می‌تواند به طور ساده توسعه یابد و هم سطرها و هم ستونهای A مقیاس شوند. از لحاظ ریاضی این معادل با پیدا کردن ماتریسهای قطری نامنفرد D_1 و D_2 است به قسمی که بزرگ‌ترین عنصر (از لحاظ اندازه) در هر سطر و ستون ماتریس $D_1^{-1}AD_2$ بین دو عدد ثابت مثلاً $1/\beta$ و 1 قرار گیرد که در آن β پایه دستگاه اعداد است. به محض این که چنین D_1 و D_2 ای پیدا شدند جواب دستگاه

$$Ax = b$$

با حل دستگاه معادل

$$\tilde{A}y = \tilde{b}$$

و سپس محاسبه

$$x = D_2y$$

به دست می‌آید که در آنها

$$\tilde{A} = D_1^{-1}AD_2 \quad \text{و} \quad \tilde{b} = D_1^{-1}b$$

این فرایند به متعادل نمودن معروف است (Forsythe and Moler (1967, 44-45)).

در پایان متذکر می شویم، که مقیاس کردن و متعادل نمودن در حالت کلی توصیه می شود، اما فقط باید در یک پایه موردی استفاده شود، و بستگی به داده های مسئله دارد. «تحلیل خطای گردگردن برای روش حذفی گاوس هنگامی که یک ماتریس متعادل می شود»، مؤثرترین نتایج را ارائه می دهد (Forsyth and Moler (1967, 44)).

۹-۴-۶ LU در برابر QR و جدولی از مقایسه ها

هم اکنون ملاحظه نمودیم که روش حل $Ax = b$ با استفاده از تجزیه QR تقریباً دو برابر گرانتر از روش حذفی گاوس با محورگیری است اگر روش هاوس هولدر برای تجزیه A استفاده شود و تقریباً چهار برابر گرانتر از این روش است اگر از دورانهای گیونز استفاده شوند. از طرف دیگر روش تجزیه QR بدون قید و شرط پایدار است، در حالی که از نقطه نظر نظری برای روش حذفی گاوس با محورگیری جزئی و کاملاً همیشه خطرانی وجود دارد. بنابراین اگر پایداری عامل اصلی باشد، و هزینه عامل مهمی نباشد، مسلماً باید از روش تجزیه QR استفاده کرد. ولیکن در نظر گرفتن کارایی، و پایداری هر دو از نقطه نظر عملی، اغلب این توافق حاصل می شود که روش حذفی گاوس با محورگیری جزئی عملی ترین راه است. اگر حقیقتاً خواسته باشیم از یک روش متعادل سازی استفاده کنیم، مسلماً روش هاوس هولدر بر روش گیونز ترجیح داده می شود. بعلاوه روش حذفی گاوس بدون محورگیری نباید استفاده شود مگر A ماتریس متقارن معین مثبت یا غالب قطری باشد.

بحث قبلی را در جدولهای ۱-۶ و ۲-۶ خلاصه می کنیم.

۱-۶ جدول مقایسه روشهای متفاوت برای حل مسئله دستگاه خطی با ماتریسهای دلخواه

روش	تعداد اعمال ممیز شناور (flops) (به طور تقریبی)	عامل رشد ρ	پایداری
حذفی گاوس بدون محورگیری	$\frac{n^3}{3}$	دلخواه	ناپایدار
روش حذفی گاوس با محورگیری جزئی	$\frac{n^3}{3} (+ O(n^2))$ (مقایسه +)	$\rho \leq 2^{n-1}$	پایدار در عمل
روش حذفی گاوس با محورگیری کامل	$\frac{n^3}{3} (+ O(n^2))$ (مقایسه +)	$\rho \leq \{n^2, 3^{1/2}, \dots, n^{(1/n-1)}\}^{1/2}$	پایدار
تجزیه QR با استفاده از تبدیلات هاوس هولدر	$\frac{2n^3}{3} (+ n)$ (ریشه دوم +)	هیچ	پایدار
تجزیه QR با استفاده از دورانهای گیونز	$\frac{4n^3}{3} (+ \frac{2n^2}{3})$ (ریشه دوم +)	هیچ	پایدار

۲-۶ مقایسه روشهای متفاوت برای حل مسأله دستگاه خطی با ماتریسهای خاص

پایداری	عامل رشد ρ	تعداد اعمال ممیز شناور (به طور تقریبی)	روش	نوع ماتریس
پایدار	$\rho = 1$	$\frac{n^3}{3}$	۱- روش حذفی گاوس بدون محورگیری	معین مثبت متقارن
پایدار	هیچ	$\frac{n^3}{6} (n + 2)$	۲- چولسکی	
پایدار	$\rho \leq 2$	$\frac{n^3}{3}$	روش حذفی گاوس بدون محورگیری	غالب قطری
پایدار	$\rho \leq n$	n^2	روش حذفی گاوس با محورگیری جزئی	هسنبرگی
پایدار	$\rho \leq 2$	$O(n)$	روش حذفی گاوس با محورگیری جزئی	سه قطری

۵-۶ معکوسها و درمینانها

مسائل پیدا کردن درمینان، معکوس، و کهادهای اصلی مقدم ماتریس A به مسأله حل دستگاه خطی $Ax = b$ مربوط می شوند. در این بخش ملاحظه خواهیم کرد که چگونه این مسائل می توانند با استفاده از روشهای مختلف تجزیه که قبلاً بحث شده اند، حل شوند.

۱-۵-۶ پرهیز از محاسبه صریح معکوسها

معکوس یک ماتریس A بندرت لازم است که به طور صریح محاسبه شود. بسیاری از مسائل محاسباتی حاوی معکوسها می توانند دوباره بر حسب جواب دستگاههای خطی فرمول بندی شوند. برای مثال، مسائل زیر را ملاحظه کنید.

$$1. \quad A^{-1}b \text{ (معکوس ضرب در یک بردار)}$$

$$2. \quad A^{-1}B \text{ (معکوس ضرب در یک ماتریس)}$$

$$3. \quad b^T A^{-1}c \text{ (بردار ضرب در معکوس ضرب در یک بردار)}$$

اولین مسأله، محاسبه $A^{-1}b$ معادل با حل دستگاه خطی

$$Ax = b$$

است. به طور مشابه مسأله دوم می تواند بر حسب حل مجموعه ای از معادلات خطی فرمول بندی

شود. بنابراین اگر A از مرتبه $n \times n$ و B از مرتبه $n \times m$ باشند، آنگاه می‌نویسیم $C = A^{-1}B = (c_1, c_2, \dots, c_m)$. ملاحظه می‌کنیم که ستونهای c_1 تا c_m می‌توانند از حل دستگاههای

$$Ac_i = b_i, \quad i = 1, \dots, m$$

به دست آیند، که در آنها b_i ، $i = 1, \dots, m$ ، ستونهای ماتریس B هستند.

محاسبه $b^T A^{-1} c$ می‌تواند در دو مرحله به صورت زیر انجام شود:

۱. پیدا کردن $A^{-1}c$ ؛ یا حل دستگاه خطی $Ax = c$.

۲. محاسبه $b^T x$.

همان گونه که در این بخش ملاحظه خواهیم کرد، محاسبه A^{-1} سه برابر گران‌تر از حل دستگاه $Ax = b$ است. بنابراین مسائلی نظیر مسائل ذکر شده در فوق را می‌توان به جای حل طبیعی آنها با استفاده از ماتریس معکوس، به طور کاراتری با فرمول‌بندی آنها بر حسب دستگاههای خطی حل کرد.

هرگاه ممکن باشد باید از محاسبه صریح معکوس پرهیز شود. یک دستگاه خطی هرگز نباید با استفاده از محاسبه صریح معکوس ماتریس دستگاه حل شود.

اهمیت معکوس یک ماتریس در کاربردهای عملی

گفته شد که اغلب مسائل محاسباتی حاوی معکوس می‌توانند بر حسب دستگاههای خطی دوباره فرمول‌بندی شوند، متذکر می‌شویم که برخی کاربردهای عملی وجود دارند که برای آنها لازم است که معکوس ماتریس به طور صریح محاسبه شود، زیرا عناصر ماتریسهای معکوس در این کاربردها دارای معنای فیزیکی هستند.

مثال ۶-۵-۱ یک بار دیگر مسأله جرم فنر بحث شده در بخش ۶-۳-۳ را ملاحظه کنید. عنصر (i, j) معکوس ماتریس سختی K بیان می‌کند که تغییر مکان جرم i چقدر خواهد بود اگر یک واحد نیروی خارجی بر جرم j تحمیل شود. بنابراین عناصر K^{-1} بیان خواهند کرد که چگونه مؤلفه‌های دستگاه نسبت به نیروهای خارجی عکس‌العمل نشان خواهند داد.

اجازه دهید که یک مثال خاص را در نظر بگیریم که ثابتهای فنر مساوی می‌باشند:

$$k_1 = k_2 = k_3 = 1$$

$$K = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}, \quad K^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$$

چون عناصر ستون اول K^{-1} همگی برابر ۱ هستند، نتیجه گرفته می‌شود که یک واحد بار به سوی پایین به اولین جرم باعث یک تغییر مکان به سوی پایین همه جرمها به اندازه ۱ واحد می‌شود. تعبیرهای مشابهی می‌توانند برای عناصر ستونهای دیگر K^{-1} ارائه شوند.

برخی معکوسهایی که به سادگی قابل محاسبه‌اند

قبل از بحث در مورد محاسبه A^{-1} برای یک ماتریس دلخواه، معکوس برخی از ماتریسهای معروف می‌توانند به طور ساده محاسبه شوند.

۱. معکوس ماتریس پایین مثلثی مقدماتی $M = I - me_k^T$ توسط $M^{-1} = I + me_k^T$ ارائه می‌شود.

۲. معکوس یک ماتریس متعامد Q ترانژاده آن Q^T است (توجه کنید که یک ماتریس هاوس هولدر و یک ماتریس جایگشت ماتریسهای متعامد هستند).

۳. معکوس ماتریس مثلثی T از یک نوع، مجدداً یک ماتریس مثلثی از همان نوع است، عناصر قطری معکوسهای عناصر قطری T هستند.

۶-۵-۲ فرمولهای شرمن - موریسون و وودبری^۱

در بسیاری از کاربردها، با مفروض بودن معکوس یک ماتریس A ، ما باید معکوس یک ماتریس دیگر مانند B را پیدا کنیم که با A فقط در یک اختلال رتبه - یک تفاوت دارد. طبیعتاً این سؤال مطرح می‌شود که آیا معکوس B را می‌توان بدون شروع مجدد همه محاسبات به دست آورد، یعنی آیا معکوس B را می‌توان با استفاده از معکوس A که قبلاً محاسبه شده است، به دست آورد. فرمول شرمن - موریسون نشان می‌دهد که چگونه این کار می‌تواند انجام شود.

فرمول شرمن - موریسون

اگر v و u دو بردار و A یک ماتریس نامنفرد باشد آنگاه

$$(A - uv^T)^{-1} = A^{-1} + \alpha(A^{-1}uv^TA^{-1})$$

که در آن

$$\alpha = \frac{1}{(1 - v^TA^{-1}u)}, \quad v^TA^{-1}u \neq 1 \text{ اگر}$$

تبصره‌ها:

۱. بند ۱ در فهرست قبل از بخش ۶-۵-۲ یک حالت خاص از این فرمول است.
 ۲. فرمول شرمن - موریسون نشان می‌دهد که چگونه معکوس یک ماتریس به دست آمده از ماتریس A توسط یک تغییر رتبه یک را می‌توان بدون محاسبه صریح معکوس ماتریس جدید محاسبه کرد، وقتی که معکوس A محاسبه شده باشد.
- فرمول شرمن - موریسون را می‌توان به حالتی که در آن U و V نیز ماتریس هستند، توسعه داد. این تعمیم معروف به فرمول وودبری می‌باشد.

فرمول وودبری

$$(A - UV^T)^{-1} = A^{-1} + A^{-1}U(I - V^TA^{-1}U)^{-1}V^TA^{-1},$$

اگر $I - V^TA^{-1}U$ نامنفرد باشد.

تذکر MATCOM : فرمول شرمن - موریسون در برنامه SHERMOR در MATCOM پیاده‌سازی شده است.

مثال ۶-۵-۲ با مفروض بودن

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} -3 & -1 & 1 \\ 14 & 2 & -3 \\ -10 & -1 & 2 \end{bmatrix}$$

ماتریس $(A - uv^T)^{-1}$ را پیدا کنید، که در آن $u = v = (1, 0, 0)^T$

$$\alpha = \frac{1}{(1 - v^T A^{-1} u)} = \frac{1}{4}$$

$$A^{-1} + \alpha A^{-1} u v^T A^{-1} = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{4} & \frac{1}{4} \\ \frac{7}{2} & -\frac{3}{2} & \frac{1}{2} \\ -\frac{5}{2} & \frac{3}{2} & -\frac{1}{2} \end{bmatrix}$$

بنابراین

$$(A - uv^T)^{-1} = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{4} & \frac{1}{4} \\ \frac{7}{2} & -\frac{3}{2} & \frac{1}{2} \\ -\frac{5}{2} & \frac{3}{2} & -\frac{1}{2} \end{bmatrix}$$

۳-۵-۶ محاسبه معکوس یک ماتریس

محاسبه معکوس یک ماتریس A معادل حل مجموعه دستگاههای خطی زیر است

$$Ax_i = e_i, \quad i = 1, \dots, n$$

اکنون این n دستگاه خطی را می توان با استفاده از هر یک از روشهای بحث شده قبلی حل کرد. ولیکن در عمل روش حذفی گاوس با محورگیری جزئی استفاده می شود.

چون ماتریس A برای همه دستگاهها یکسان است، A فقط یک بار تجزیه می شود. بنابراین اگر روش حذفی گاوس با محورگیری جزئی استفاده شود، آنگاه با جایگزین کردن e_i به جای b در الگوریتم ۳-۴-۶ و نمایش Me_i توسط e'_i ، الگوریتم زیر را داریم.

الگوریتم ۱-۵-۶ محاسبه معکوس A با استفاده از روش حذفی گاوس با محورگیری جزئی

گام ۱: A را با استفاده از الگوریتم ۲-۲-۵ به صورت $MA = U$ تجزیه کنید.

گام ۲: n دستگاه بالا مثلثی $Me_i = e'_i = Ux_i$ ، $i = 1, \dots, n$ را حل کنید.

تذکر محاسباتی در مورد گام ۲: توجه کنید که $e'_i = Me_i = M_{n-1}P_{n-1} \dots M_1P_1M_1P_1$

بنابراین e'_i را می توان با شناخت ضریبهای m_{ik} و اندیسهای تعویض سطری r_k از الگوریتم ۲-۲-۵

محاسبه کرد. ضرب ماتریسی صریح لازم نیست.

مثال ۶-۵-۳

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

از مثال ۵-۲-۴ داریم

$$U = \begin{bmatrix} 7 & 8 & 9 \\ \cdot & \frac{6}{7} & \frac{19}{7} \\ \cdot & \cdot & -\frac{1}{7} \end{bmatrix}, \quad M = \begin{bmatrix} \cdot & \cdot & 1 \\ 1 & \cdot & -\frac{1}{7} \\ -\frac{1}{7} & 1 & -\frac{1}{7} \end{bmatrix}$$

$$x_1 = (1, -2, 1)^T, \quad x_2 = (-4/66667, 6/33333, 2/0000)^T$$

$$x_3 = (2/66667, -3/33333, 1)^T$$

بنابراین

$$A^{-1} = X = \begin{bmatrix} 1 & -4/66667 & 2/66667 \\ -2 & 6/33333 & -3/33333 \\ 1 & -2/0000 & 1 \end{bmatrix}$$

شمارش اعمال ممیز شناور. الگوریتم حدوداً به $5n^3/6$ عمل ممیز شناور (flops) نیاز دارد ($n^3/3$) عمل ممیز شناور برای محاسبه تجزیه مثلثی و $n^2/2$ برای حل هر دستگاه مثلثی).

۴-۵-۶ محاسبه دترمینان یک ماتریس

می دانیم که $\det(A) \neq 0$ اگر و فقط اگر A معکوس پذیر باشد. ولیکن در پیاده سازی کامپیوتری همان گونه که در بخش ۶-۷-۴ ملاحظه خواهیم کرد دترمینان یک اندازه قابل اطمینان از نامنفرد بودن نمی باشد. بنابراین در عمل بندرت دترمینان یک ماتریس مورد نیاز می باشد. ولیکن اگر کسی به محاسبه دترمینان A نیاز داشته باشد، آنگاه روش حذفی گاوس معمولی (با محورگیری جزئی و کامل) یا هر روش مثلثی سازی متعامد می تواند مورد استفاده قرار گیرد.

اگر روش حذفی گاوس بدون محورگیری استفاده شود، داریم

$$A = LU, \quad \det(A) = \det(L) \cdot \det(U)$$

U یک ماتریس بالا مثلثی است، بنابراین $\det(U) = u_{11}u_{22}\dots u_{nn} = a_{11}a_{22}^{(1)}\dots a_{nn}^{(n-1)}$ ؛ یک L

ماتریس پایین مثلثی واحد است بنابراین $\det(L) = 1$.

محاسبه دترمینان A از تجزیه LU

$$\det(A) = a_{11}^{(1)} a_{22}^{(1)} \dots a_{nn}^{(n-1)} = \text{حاصلضرب عناصر محوری}$$

اگر روش حذفی گاوس با محورگیری استفاده شود، داریم

$$MA = U$$

آنگاه $\det(M)\det(A) = \det(U)$. اکنون $M = M_{n-1}P_{n-1} \dots M_2P_2M_1P_1$. چون دترمینان هر یک از ماتریسهای پایین مثلثی برابر ۱ است و دترمینان هر کدام از ماتریسهای جایگشت برابر ± 1 می باشد، داریم:

$$\det(M) = (-1)^r$$

که در آن r تعداد تعویضهای سطری در طی فرایند محورگیری است. بنابراین داریم

$$\begin{aligned} \det(A) &= \frac{1}{\det(M)} \det(U) \\ &= (-1)^r \cdot u_{11} \cdot u_{22} \dots u_{nn} \\ &= (-1)^r a_{11} a_{22}^{(1)} \dots a_{nn}^{(n-1)} \end{aligned}$$

محاسبه $\det(A)$ از تجزیه $MA = U$

$$\det(A) = (-1)^r a_{11} a_{22}^{(1)} \dots a_{nn}^{(n-1)}$$

که در آن r تعداد تعویضها است.

اگر روش حذفی گاوس با محورگیری کامل استفاده شود، داریم

$$MAQ = U$$

در نتیجه

$$\det(M) \cdot \det(A) \cdot \det(Q) = \det(U)$$

فرض کنید r و s به ترتیب تعداد تعویضهای سطری و ستونی باشند، آنگاه

$$\det(M) = (-1)^r$$

$$\det(Q) = (-1)^s$$

بنابراین می‌توان مطلب زیر را بیان کرد

محاسبه $\det(A)$ از تجزیه $MAQ = U$

$$\det(A) = (-1)^{r+s} \det(U)$$

$$= (-1)^{r+s} a_{11} a_{22} \dots a_{nn}^{(n-1)}$$

که در آن r و s تعداد تعویضهای سطری و ستونی هستند.

مثال ۴-۵-۶

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

با استفاده از روش حذفی گاوس با محورگیری جزئی داریم (مثال ۳-۲-۵)،

$$U = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

فقط یک تعویض رخ داده است؛ $r = 1$ ، بنابراین، $\det(A) = (-1)\det(U) = (-1)(-1) = 1$

با استفاده از روش حذفی گاوس با محورگیری کامل (مثال ۵-۲-۵) داریم

$$U = \begin{bmatrix} 3 & 1 & 2 \\ 0 & \frac{2}{3} & 1 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$$

در گام اول، یک تعویض سطری و یک تعویض ستونی انجام شده است. در گام دوم، یک تعویض

سطری و یک تعویض ستونی انجام شده است.

بنابراین، $2, 2 = 2, 2 = 2$ و $s = 2$

$$\det(A) = (-1)^{r+s} \det(U) = (-1)^{2+2} \cdot \frac{2}{3} \cdot \frac{1}{2} = 1$$

۶-۶ تحلیل اختلال مسأله دستگاه خطی

داده‌های ورودی A و b ممکن است در عمل به خطا آلوده شوند. این خطا ممکن است تجربی باشد، ممکن است از فرایند گسسته‌سازی نتیجه شود، و غیره. به منظور برآورد دقت جواب محاسبه شده، خطا در داده‌ها باید به حساب آید. همان گونه که در فصل ۳ ملاحظه کردیم، مسائلی وجود دارند که جوابهای آنها ممکن است با تغییرات کوچکی در داده‌ها به طور جدی تغییر کنند، و این پدیده بدوضعی مستقل از الگوریتمهای استفاده شده برای حل این مسائل می‌باشد. ما بدوضعی چندین مسأله را در فصل ۳ مورد بحث قرار دادیم. اجازه دهید به مثال ساده دیگری از دستگاه خطی بدوضعی نگاهی بیان‌دازیم.

دستگاه خطی زیر را در نظر بگیرید

$$x_1 + 2x_2 = 3$$

$$2x_1 + 3/999x_2 = 5/999$$

جواب واقعی برابر $x_1 = x_2 = 1$ است. اکنون یک اختلال کوچک در طرف راست ایجاد کرده و دستگاه زیر را به دست می‌آوریم

$$x_1 + 2x_2 = 3$$

$$2x_1 + 3/999x_2 = 6$$

اکنون جواب دستگاه اختلال یافته و به دست آمده با روش حذفی گاوس با محورگیری جزئی (که ملاحظه شد در عمل یک روش پایدار می‌باشد) عبارت است از

$$x_1 = 3, \quad x_2 = 0$$

بنابراین یک تغییر خیلی کوچک در طرف راست کاملاً جواب را تغییر داد.

در این بخش اثر اختلالهای کوچک داده‌های ورودی A و b را بر روی جواب محاسبه شده دستگاه $Ax = b$ مطالعه می‌کنیم. این مطالعه خیلی مفید خواهد بود. این مطالعه نه تنها در تشخیص یک مقدار خطا در جواب دستگاه اختلال یافته، بدون توجه به الگوریتم استفاده شده، به ما کمک خواهد کرد، بلکه

همچنین وقتی که نتیجه یک تحلیل اختلال با نتیجه آنالیز خطای پسرو یک الگوریتم خاص ترکیب می‌شود به ما کمک خواهد کرد که یک کران خطا برای جواب محاسبه شده توسط الگوریتم به دست آوریم.

چون A و b داده‌های ورودی در مسأله دستگاه خطی $Ax = b$ هستند، می‌تواند در b یا A یا هر دو آلودگی وجود داشته باشد. بنابراین ما اثر اختلالها را بر روی جواب در هر یک از سه حالت به طور جداگانه مطالعه خواهیم کرد.

۱-۶-۶ اثر اختلال در بردار طرف راست b

در اینجا فرض خواهیم کرد که آلودگی در b وجود داشته باشد، اما ماتریس A دقیق باشد.

قضیه ۱-۶-۶ قضیه اختلال طرف راست. اگر δb و δx به ترتیب اختلالهای b و x در دستگاه خطی $Ax = b$ باشند و فرض شود که A نامنفرد و $b \neq 0$ باشد، آنگاه

$$\frac{\|\delta b\|}{\text{cond}(A) \|b\|} \leq \frac{\|\delta x\|}{\|x\|} \leq \text{Cond}(A) \frac{\|\delta b\|}{\|b\|}$$

اثبات: داریم

$$Ax = b$$

و

$$A(x + \delta x) = b + \delta b$$

معادله آخر می‌تواند به صورت زیر نوشته شود

$$Ax + A\delta x = b + \delta b$$

یا

$$A\delta x = \delta b, \quad Ax = b \text{ زیرا}$$

یعنی

$$\delta x = A^{-1}\delta b$$

با گرفتن یک نرم ماتریس - بردار وابسته، به دست می‌آوریم

$$\|\delta x\| \leq \|A^{-1}\| \|\delta b\| \quad (1-6-6)$$

دوباره، با گرفتن همان نرم بر روی دو طرف $Ax = b$ ، به دست می آوریم

$$\|Ax\| = \|b\|$$

یا

$$\|b\| = \|Ax\| \leq \|A\| \|x\| \quad (۲-۶-۶)$$

از ترکیب (۱-۶-۶) و (۲-۶-۶) داریم

$$\frac{\|\delta x\|}{\|x\|} \leq \|A\| \|A^{-1}\| \frac{\|\delta b\|}{\|b\|} \quad (۳-۶-۶)$$

از طرف دیگر، $A\delta x = \delta b$ نتیجه می دهد

$$\|\delta x\| \geq \frac{\|\delta b\|}{\|A\|} \quad (۴-۶-۶)$$

همچنین از $Ax = b$ داریم

$$\frac{1}{\|x\|} \geq \frac{1}{\|A^{-1}\| \|b\|} \quad (۵-۶-۶)$$

از ترکیب (۴-۶-۶) و (۵-۶-۶) داریم

$$\frac{\|\delta x\|}{\|x\|} \geq \frac{\|\delta b\|}{\|A\| \|A^{-1}\| \|b\|}$$

از فصل ۳ یادآور می شویم که $\|A\| \|A^{-1}\|$ عدد وضعیت A است و توسط $\text{Cond}(A)$ نمایش داده می شود. بنابراین قضیه ثابت می شود. \square

تعریف قضیه ۶-۶-۱. درک کامل مفهوم قضیه ۶-۶-۱ اهمیت دارد. قضیه ۶-۶-۱ بیان می کند که یک تغییر نسبی در جواب می تواند به بزرگی $\text{Cond}(A)$ ضربدر تغییر نسبی در بردار b باشد. بنابراین اگر عدد وضعیت خیلی بزرگ نباشد، آنگاه یک اختلال کوچک در بردار b دارای اثر خیلی کوچکی بر روی جواب خواهد بود. از طرف دیگر اگر عدد وضعیت بزرگ باشد، آنگاه حتی یک اختلال کوچک در b ممکن است جواب را به طور جدی تغییر دهد.

مثال ۶-۶-۱ یک مسأله بدوضع

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4/0.001 & 2/0.02 \\ 1 & 2/0.02 & 2/0.04 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ \sqrt{0.021} \\ 5/0.06 \end{bmatrix}$$

جواب واقعی برابر است با

$$x = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

b را به

$$b' = \begin{bmatrix} 4 \\ 1/0.020 \\ 5/0.061 \end{bmatrix}$$

تغییر می‌دهیم. آنگاه تغییر نسبی در b برابر است با

$$\frac{\|b' - b\|}{\|b\|} = \frac{\|\delta b\|}{\|b\|} = 1/379 \times 10^{-5} \quad (\text{کوچک})$$

اگر دستگاه $Ax' = b'$ را حل کنیم، به دست می‌آوریم

$$x' = x + \delta x = \begin{bmatrix} 3/0.850 \\ -0/0.436 \\ 1/0.022 \end{bmatrix}$$

(x' کاملاً متفاوت با x است.)

$$\text{تذکر: } \frac{\|\delta x\|}{\|x\|} = 1/3461$$

به آسانی تحقیق می‌شود که نامساوی قضیه ۶-۶-۱ برقرار است.

$$\text{cond}(A) \cdot \frac{\|\delta b\|}{\|b\|} = 4/4434$$

ولیکن تغییر پیش‌بینی شده به طور زیادی برآورد شده است.

مثال ۶-۶-۲ مسأله خوش‌وضع زیرا در نظر می‌گیریم

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$$

جواب واقعی برابر است با

$$x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

فرض کنید

$$b' = b + \delta b = \begin{bmatrix} 3/0001 \\ 7/0001 \end{bmatrix}$$

تغییر نسبی در b برابر است با

$$\frac{\|b' - b\|}{\|b\|} = 1/175 \times 10^{-5} \quad (\text{کوچک})$$

$$\text{Cond}(A) = 14/9330 \quad (\text{کوچک})$$

بنابراین یک تغییر جدی در جواب x انتظار نمی‌رود. در حقیقت بردار x' که در معادله

$$Ax' = b'$$

صدق می‌کند برابر است با

$$x' = \begin{bmatrix} 0/9999 \\ 1/0001 \end{bmatrix} \approx x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\text{تذکر: } \frac{\|\delta x\|}{\|x\|} = 10^{-5}$$

۶-۶-۲ اثر اختلال در ماتریس A

در اینجا فرض می‌کنیم که آلودگی فقط در A وجود داشته باشد، در نتیجه $A + \Delta A$ را در دست داریم، اما b دقیق است.

قضیه ۶-۶-۲ قضیه اختلال در چپ. فرض کنید A نامنفرد بوده و $b \neq 0$ باشد. فرض کنید که ΔA و δx به ترتیب اختلالهای A و x در دستگاه خطی

$$Ax = b$$

باشند. بعلاوه فرض کنید که ΔA به گونه‌ای باشد که $\|\Delta A\| \leq \frac{1}{\|A^{-1}\|}$. آنگاه

$$\frac{\|\delta x\|}{\|x\|} \leq \text{Cond}(A) \frac{\|\Delta A\|}{\|A\|} / (1 - \text{Cond}(A) \frac{\|\Delta A\|}{\|A\|})$$

اثبات: داریم

$$(A + \Delta A)(x + \delta x) = b$$

یا

$$(A + \Delta A)x + (A + \Delta A)\delta x = b \quad (6-6-6)$$

چون

$$Ax = b$$

از (۶-۶-۶) داریم

$$(A + \Delta A)\delta x = -\Delta Ax \quad (7-6-6)$$

یا

$$\delta x = -A^{-1}\Delta A(x + \delta x) \quad (8-6-6)$$

با گرفتن نرم از هر دو طرف این رابطه داریم

$$\|\delta x\| \leq \|A^{-1}\| \|\Delta A\| (\|x\| + \|\delta x\|) \quad (9-6-6)$$

$$= \frac{\|A^{-1}\| \|A\| \|\Delta A\|}{\|A\|} (\|x\| + \|\delta x\|)$$

در نتیجه

$$\left(1 - \frac{\|A\|^{-1} \|A\| \|\Delta A\|}{\|A\|}\right) \|\delta x\| \leq \frac{\|A\| \|A^{-1}\| \|\Delta A\|}{\|A\|} \|x\| \quad (10-6-6)$$

چون

$$\|A^{-1}\| \|\Delta A\| < 1$$

عبارت داخل پرانتز در طرف چپ مثبت است. بنابراین می‌توانیم هر دو طرف نامساوی را بر این عدد تقسیم کنیم بدون آن‌که نامساوی تغییر کند. پس از این اگر طرفین را بر $\|x\|$ نیز تقسیم کنیم، به دست می‌آوریم

$$\begin{aligned} \frac{\|\delta x\|}{\|x\|} &\leq \frac{\|A\| \|A^{-1}\| \frac{\|\Delta A\|}{\|A\|}}{\left[1 - \|A\| \|A^{-1}\| \frac{\|\Delta A\|}{\|A\|}\right]} \\ &= \frac{\text{Cond}(A) \frac{\|\Delta A\|}{\|A\|}}{\left(1 - \text{Cond}(A) \frac{\|\Delta A\|}{\|A\|}\right)} \end{aligned} \quad (11-6-6)$$

□ که این قضیه را ثابت می‌کند.

تبصره‌ها: به واسطه فرض $\|\Delta A\| \leq \frac{1}{\|A^{-1}\|}$ (که کاملاً معقول است)، مخرج طرف راست نامساوی قضیه ۶-۶-۲ کمتر از ۱ است. بنابراین حتی اگر $\frac{\|\Delta A\|}{\|A\|}$ کوچک باشد، می‌تواند یک تغییر جدی در

جواب وجود داشته باشد اگر $\text{Cond}(A)$ بزرگ باشد.

مثال ۳-۶-۶ مثال ۱-۶-۶ را یک بار دیگر ملاحظه کنید. a_{33} را به $۲/۰۰۲۱$ تغییر دهید؛ b را ثابت نگه دارید آنگاه

$$\Delta A = -۱۰^{-۲} \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & ۱ \\ \cdot & \cdot & \cdot \end{bmatrix} \quad (\text{کوچک})$$

اکنون دستگاه $(A + \Delta A)x' = b$ را حل می‌کنیم:

$$x' = \begin{bmatrix} ۳/۰۸۵۲ \\ -۰/۰۴۳۷ \\ ۱/۰۰۲۱ \end{bmatrix}$$

$$\delta x = x' - x = \begin{bmatrix} ۲/۰۸۵۲ \\ -۱/۰۴۳۷ \\ ۰/۰۰۲۱ \end{bmatrix}$$

$$\frac{\|\delta x\|}{\|x\|} = \frac{\|\delta x\|}{\|x\|} = ۱/۳۴۶۳ \quad (\text{کاملاً بزرگ})$$

۳-۶-۶ اثر اختلالها هم در ماتریس A و هم در بردار b

بالاخره اکنون فرض خواهیم کرد که داده‌های ورودی A و b دارای خطاهایی باشند. در نتیجه دستگاهی با $A + \Delta A$ به عنوان ماتریس و $b + \delta b$ به عنوان طرف راست خواهیم داشت.

قضیه ۳-۶-۶ قضیه اختلال کلی. فرض کنید که A نامنفرد باشد، $b \neq 0$ و $\|\Delta A\| < \frac{1}{\|A^{-1}\|}$. آنگاه

$$\frac{\|\delta x\|}{\|x\|} \leq \left[\frac{\text{Cond}(A)}{1 - \text{Cond}(A) \frac{\|\Delta A\|}{\|A\|}} \right] \left[\frac{\|\Delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|} \right]$$

اثبات: باکم کردن

$$Ax = b$$

از

$$(A + \Delta A)(x + \delta x) = b + \delta b$$

داریم

$$(A + \Delta A)(x + \delta x) - Ax = \delta b$$

یا

$$(A + \Delta A)(x + \delta x) - (A + \Delta A)x + (A + \Delta A)x - Ax = \delta b$$

یا

$$(A + \Delta A)(\delta x) + \Delta Ax = \delta b$$

یا

$$A(I - A^{-1}(-\Delta A))\delta x = \delta b - \Delta Ax \quad (۱۲-۶-۶)$$

فرض کنید $A^{-1}(-\Delta A) = F$. آنگاه

$$\|F\| = \|A^{-1}(-\Delta A)\| \leq \|A^{-1}\| \|\Delta A\| < 1 \quad (\text{بنابا فرض})$$

چون $\|F\| < 1$ ، $I - F$ معکوس پذیر است، و

$$\|(I - F)^{-1}\| \leq \frac{1}{1 - \|F\|} \quad (\text{قضیه ۱-۷-۷ را ملاحظه کنید}) \quad (۱۳-۶-۶)$$

سپس از (۱۲-۶-۶) داریم

$$\delta x = (I - F)^{-1}A^{-1}(\delta b - \Delta Ax)$$

یا

$$\|\delta x\| \leq \frac{\|A^{-1}\|}{(1 - \|F\|)} (\|\delta b\| + \|\Delta A\| \|x\|)$$

یا

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\|A^{-1}\|}{(1 - \|F\|)} \left(\frac{\|\delta b\|}{\|x\|} + \|\Delta A\| \right) \quad (۱۴-۶-۶)$$

$$\leq \frac{\|A^{-1}\|}{(1 - \|F\|)} \left(\frac{\|\delta b\| \|A\|}{\|b\|} + \|\Delta A\| \right) \left(\frac{1}{\|x\|} \leq \frac{\|A\|}{\|b\|} \right) \quad (\text{توجه کنید که})$$

در نتیجه

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\|A^{-1}\| \|A\|}{(1 - \|F\|)} \left[\frac{\|\delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|} \right] \quad (۱۵-۶-۶)$$

دوباره

$$\|F\| = \|A^{-1}(-\Delta A)\| \leq \|A^{-1}\| \|\Delta A\| = \frac{\|A^{-1}\| \|A\|}{\|A\|} \cdot \|\Delta A\| \quad (۱۶-۶-۶)$$

چون $\|F\| < 1$ ، از (۱۵-۶-۶) و (۱۶-۶-۶) می توان نوشت

$$\begin{aligned} \frac{\|\delta x\|}{\|x\|} &\leq \left[\frac{\|A^{-1}\| \|A\|}{\left[1 - \left(\frac{\|A^{-1}\| \|A\|}{\|A\|} \right) \cdot \|\Delta A\| \right]} \right] \left[\frac{\|\delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|} \right] \\ &= \left[\frac{\text{Cond}(A)}{\left(1 - \frac{\text{Cond}(A)}{\|\Delta A\|} \cdot \|\Delta A\| \right)} \right] \left[\frac{\|\delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|} \right] \quad (17-6-6) \end{aligned}$$

□

تبصره‌ها: مجدداً از (۱۷-۶-۶) ملاحظه می‌کنیم که اگر اختلالهای

$$\frac{\|\delta b\|}{\|b\|} \quad \text{و} \quad \frac{\|\Delta A\|}{\|A\|}$$

کوچک باشند، ممکن است یک تغییر جدی در جواب یافت شود اگر $\text{Cond}(A)$ بزرگ باشد. بنابراین $\text{Cond}(A)$ نقش بسیار قاطعی را در حساسیت جواب بازی می‌کند.

یک قرارداد

هنگامی که می‌نویسیم $\text{Cond}(A)$ ، منظورمان $\text{Cond}_2(A)$ - یعنی عدد وضعیت A نسبت به نرم ۲ - خواهد بود، در غیر این صورت ذکر خواهد شد. عدد وضعیت یک ماتریس A نسبت به یک p - نرم وابسته $(p = 1, 2, \infty)$ توسط $\text{Cond}_p(A)$ نمایش داده خواهد شد؛ یعنی $\text{Cond}_1(A)$ برای نشان دادن عدد وضعیت A نسبت به ۱ - نرم به کار برده خواهد شد.

۷-۶ عدد وضعیت و دقت جواب

مطالب زیر برخی خواص مهم (اما ساده از نظر اثبات) از عدد وضعیت یک ماتریس هستند.

۱. $\text{Cond}(A)$ نسبت به هر p - نرمی حداقل برابر ۱ است.

۲. اگر A یک ماتریس متعامد باشد، آنگاه $\text{Cond}(A)$ نسبت به ۲ - نرم برابر ۱ است. (توجه کنید

که این خاصیت یک ماتریس متعامد A ماتریس را برای محاسبات عددی جذاب می‌سازد.)

$$3. \quad \text{Cond}(A^T A) = (\text{Cond}(A))^2$$

$$4. \quad \text{Cond}(A) = \text{Cond}(A^T)$$

$$5. \text{Cond}(AB) \leq \text{Cond}(A)\text{Cond}(B)$$

$$6. \text{Cond}(\alpha A) = \text{Cond}(A), \text{ که در آن } \alpha \text{ یک عدد مخالف صفر است.}$$

$$7. \text{Cond}(A) \geq |\lambda_1|/|\lambda_n|, \text{ که در آن } |\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|, \text{ و } \lambda_1, \lambda_2, \dots, \lambda_n \text{ مقادیر ویژه } A \text{ هستند.}$$

$$8. \text{Cond}_r(A) = \sigma_1/\sigma_n, \text{ که در آن } \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \text{ مقادیر تکین } A \text{ هستند.}$$

اکنون به طور رسمی بدو وضعی و خوش وضعی را بر حسب عدد وضعیت بیان می‌کنیم.

تعریف ۶-۷-۱ دستگاه $Ax = b$ بدوضع است اگر $\text{Cond}(A)$ کاملاً بزرگ باشد. در غیر این صورت خوش وضع می‌باشد.

تبصره: توجه کنید که یک دستگاه می‌تواند به واسطه مقیاس کردن ضعیف به طور ساختگی بدوضع باشد. بنابراین قبل از تصمیم‌گیری در مورد این که آیا یک دستگاه حقیقتاً بدوضع است یا خیر، عناصر آن باید به طور صحیح مقیاس شوند. (بخش ۶-۷-۷ را برای اطلاعات بیشتر در مورد اعداد وضعیت و مقیاس کردن ملاحظه کنید).

بد وضعی نسبت به نرم‌های مختلف

اگر چه عدد وضعیت، به صورتی که در بالا تعریف شده، بستگی به نرم دارد، ولیکن اعداد وضعیت نیز نسبت به دو نرم متفاوت وابسته هستند. (برای مثال می‌توان نشان داد که اگر A یک ماتریس $n \times n$ باشد، آنگاه $n \leq \text{Cond}_r(A)/\text{Cond}_\infty(A) \leq \frac{1}{n}$). به طور کلی، اگر یک ماتریس نسبت به یک نرم خوش وضع یا بدوضع باشد، آنگاه نسبت به نرم‌های دیگر نیز خوش وضع یا بدوضع خواهد بود.

مثال ۶-۷-۱ ماتریس زیر را در نظر بگیرید

$$A = \begin{bmatrix} 1 & 0.9999 \\ 0.9999 & 1 \end{bmatrix}$$

$$A^{-1} = 1.2 \begin{bmatrix} 0.0003 & -0.99997 \\ -0.99997 & 0.0003 \end{bmatrix}$$

اعداد وضعیت نسبت به نرم بی‌نهایت و ۱ - نرم عبارتند از

$$\|A\|_\infty = \|A\|_1 = 1.9999$$

$$\|A^{-1}\|_\infty = \|A^{-1}\|_1 = 1.2$$

$$\text{Cond}_{\infty}(A) = \text{Cond}_1(A) = 1/9999 \times 10^4$$

عدد وضعیت نسبت به ۲- نرم عبارت است از

$$\|A\|_2 = \rho(A) = 1/9999$$

$$\|A^{-1}\|_2 = \rho(A^{-1}) = 10^4$$

$$\text{Cond}_2(A) = 1/9999 \times 10^4$$

تبصره: برای مثال ۶-۷-۱ عدد وضعیت نسبت به هر نرمی یکسان است. ولیکن این مطلب همواره درست نیست. گرچه در حالت کلی، اعداد وضعیت به طور نزدیکی وابسته هستند. (عدد وضعیت ماتریس هیلبرت زیر را نسبت به نرمهای متفاوت ملاحظه کنید).

۶-۷-۱ برخی ماتریسهای بدوضع معروف

ماتریسهای زیر ماتریسهای بدوضع معروف هستند.

۱. ماتریس هیلبرت

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \dots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \dots & \frac{1}{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} & \dots & \frac{1}{2n-1} \end{bmatrix}$$

برای $n = 10$

$$\text{Cond}_2(A) = 1/6025 \times 10^{13}$$

$$\text{Cond}_{\infty}(A) = 3/5353 \times 10^{13}$$

$$\text{Cond}_1(A) = 3/5353 \times 10^{13}$$

۲. ماتریس پای A (Pie) با $a_{ij} = 1$ ، $a_{ii} = \alpha$ ، به ازای $i \neq j$. این ماتریس هنگامی که α نزدیک به

۱ و یا $n - 1$ می باشد بدوضع می شود. برای مثال وقتی که $\alpha = 0/9999$ و $n = 5$ ،

$$\text{cond}_2(A) = 5 \times 10^4$$

۳. ماتریس دوقطری ویلکینسون از مرتبه ۲۰ (فصل ۳ را ملاحظه کنید).

۲-۷-۶ اثر عدد وضعیت بر روی دقت جواب محاسبه شده

وقتی که یک جواب \hat{x} از دستگاه $Ax = b$ محاسبه می‌شود، طبیعی است که آزمون کنیم که جواب محاسبه شده چه دقتی دارد. اگر جواب واقعی x معلوم باشد، آنگاه البته می‌توان برای آزمون نمودن \hat{x} خطای نسبی $\|x - \hat{x}\|/\|x\|$ را محاسبه کرد. ولیکن در بیشتر وضعیتهای عملی، جواب واقعی معلوم نیست. در چنین حالت‌هایی واضح‌ترین راه برای انجام آن، محاسبه مانده $r = b - A\hat{x}$ و دیدن میزان کوچک بودن خطای نسبی $\|r\|/\|b\|$ می‌باشد. جالب توجه است که، جواب به دست آمده توسط فرایند حذفی گاوس، در حالت کلی، یک مانده کوچک تولید می‌کند. (چرا؟) متأسفانه یک مانده نسبی کوچک دقت جواب را تضمین نمی‌کند. مثال زیر این حقیقت را توضیح می‌دهد.

مثال ۲-۷-۶ فرض کنید

$$A = \begin{bmatrix} 1/0001 & 1 \\ & 1 \end{bmatrix}$$

$$b = \begin{bmatrix} 2/0001 \\ 2 \end{bmatrix}$$

فرض کنید

$$\hat{x} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

آنگاه

$$r = b - A\hat{x} = \begin{bmatrix} 0/0001 \\ 0 \end{bmatrix}$$

توجه کنید که r کوچک است. ولیکن \hat{x} به جواب واقعی زیر در هیچ جا نزدیک نیست.

$$x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

پدیده قبل را می‌توان توسط قضیه زیر توضیح داد. اثبات می‌تواند به سادگی انجام شود.

قضیه ۱-۷-۶ قضیه مانده

$$\frac{\|\hat{x} - x\|}{\|x\|} \leq \text{Cond}(A) \frac{\|r\|}{\|b\|}$$

تفسیر قضیه ۶-۷-۱

قضیه ۶-۷-۱ بیان می‌کند که خطای نسبی در \hat{x} نه تنها به مانده نسبی بلکه به عدد وضعیت A نیز بستگی دارد. دقیق بودن یک جواب محاسبه شده می‌تواند تنها وقتی تضمین شود که حاصلضرب $\text{Cond}(A)$ و مانده نسبی کوچک باشد. توجه کنید که در مثال ۶-۷-۲ داریم

$$\text{Cond}(A) = 4/0.002 \times 10^2 \quad (\text{بزرگ})$$

۶-۷-۳ اندازه عدد وضعیت لازم برای بدو وضعی

سؤالی که اغلب مطرح می‌شود این است که، چقدر $\text{Cond}(A)$ باید بزرگ باشد تا این که دستگاه $Ax = b$ بدو وضع در نظر گرفته شود؟ قضیه ۶-۶-۳ را دوباره برای پاسخ دادن به این سؤال مطرح می‌کنیم.

$$\frac{\|\delta x\|}{\|x\|} \leq \left[\frac{\text{Cond}(A)}{1 - \text{Cond}(A) \frac{\|\Delta A\|}{\|A\|}} \right] \left[\frac{\|\Delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|} \right]$$

برای سادگی فرض کنید که

$$\frac{\|\Delta A\|}{\|A\|} = \frac{\|\delta b\|}{\|b\|} = 10^{-d} \quad \text{و} \quad \text{Cond}(A) \frac{\|\Delta A\|}{\|A\|} \ll 1$$

آنگاه $\|\delta x\|/\|x\|$ به طور تقریبی کمتر یا مساوی با $10^{-d} \times \text{Cond}(A) \times 2$ خواهد بود. بنابراین اگر داده‌ها دارای یک خطای نسبی 10^{-d} باشند و اگر خطای نسبی در جواب باید تضمین شود که کوچک‌تر یا مساوی 10^{-t} باشد، آنگاه $\text{Cond}(A)$ باید کوچک‌تر یا مساوی $10^{d-t} \times \frac{1}{2}$ باشد. بنابراین بدو وضعی یا خوش‌وضعی یک دستگاه به دقت داده‌ها و این که چه مقدار خطا در جواب می‌تواند وجود داشته باشد بستگی دارد.

برای مثال فرض کنید که داده‌ها دارای خطای نسبی در حدود 10^{-5} باشند و دقتی در حدود 10^{-3} مورد نظر باشد؛ آنگاه $50 = 10^2 \times \frac{1}{2} \leq \text{Cond}(A)$. از طرف دیگر اگر دقتی در حد 10^{-2} مورد نظر باشد، آنگاه $500 = 10^3 \times \frac{1}{2} \leq \text{Cond}(A)$. بنابراین در حالت اول دستگاه خوش‌وضع است اگر $\text{Cond}(A)$ کوچک‌تر یا مساوی ۵۰ باشد، در حالی که در حالت دوم، دستگاه خوش‌وضع خواهد بود اگر $\text{Cond}(A)$ کوچک‌تر یا مساوی ۵۰۰ باشد.

برآورد دقت از روی عدد وضعیت

به طور کلی، اگر داده‌ها تقریباً دقیق باشند، و اگر $\text{Cond}(A) = 10^5$ ، آنگاه فقط دقتی در حدود $(t - 5)$ رقم بامعنا در جواب محاسبه شده وجود خواهد داشت هنگامی که جواب در حساب t - رقمی محاسبه شود.

برای فهم بهتر وضعیت، پایداری و دقت ما دوباره خواننده را به مرجع Bunch (1987) ارجاع می‌دهیم.

۴-۷-۶ عدد وضعیت و نزدیکی به منفرد بودن

وقتی که ماتریس A به طور نسبی به یک ماتریس منفرد نزدیک باشد، عدد وضعیت این امر را نشان می‌دهد. اگر $\text{Cond}(A)$ بزرگ باشد، A نزدیک به منفرد است.

این معیار نزدیکی به منفرد بودن یک معیار دقیق‌تر از دترمینان A است. برای مثال ماتریس مثلثی $n \times n$ معروف زیر را در نظر بگیرید.

$$A = \begin{bmatrix} 1 & -1 & -1 & \dots & -1 & -1 \\ 0 & 1 & -1 & \dots & -1 & -1 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & -1 \\ 0 & \dots & \dots & \dots & 0 & 1 \end{bmatrix}$$

این ماتریس دارای دترمینان ۱ است؛ ولیکن برای n بزرگ نزدیک به منفرد است.

$$\text{Cond}_{\infty}(A) = n^{2^{n-1}}$$

به طور مشابه، کوچک بودن دترمینان یک ماتریس لزوماً بدین معنا نیست که A نزدیک به یک ماتریس منفرد است. ماتریس $A = \text{diag}(0/1, 0/1, \dots, 0/1)$ از مرتبه ۱۰۰۰ را در نظر بگیرید؛ $\det(A) = 10^{-1000}$ ، که یک عدد کوچک است. ولیکن A کاملاً نامنفرد در نظر گرفته می‌شود، زیرا $\text{Cond}_p(A) = 1$.

۵-۷-۶ بدوضعی و محورگیری

طبیعی است متعجب شویم اگر بدوضعی بتواند در طی فرایند مثلثی‌سازی با استفاده از روش حذفی

گاس با محورگیری جزئی کشف شود. در اینجا منظور ما از یک ماتریس نرمال شده ماتریسی است که $\|A\|_2 = 1$. فرض کنید که A و b نرمال شده باشند. آنگاه چندین نشان برای بدوضعی وجود دارد. سه تا از آنها را در زیر ارائه می‌دهیم.

نشانهایی برای بدوضعی

۱. یک عنصر محوری کوچک
۲. یک جواب محاسبه شده بزرگ
۳. یک بردار مانده بزرگ

توضیح: فرض کنید یک عنصر محوری کوچک یافت شود. آنگاه معکوس ماتریس U در تجزیه مثلثی A بزرگ خواهد بود. (الگوریتم ۵-۲-۲ یا ۵-۲-۳ را ملاحظه کنید)، و این U^{-1} بزرگ A^{-1} را بزرگ خواهد ساخت (توجه کنید که اگر محورگیری جزئی استفاده شود آنگاه $A^{-1} = U^{-1}M$). به طور مشابه، اگر جواب محاسبه شده \bar{x} بزرگ باشد، آنگاه از $A\bar{x} = \bar{b}$ داریم $\|A^{-1}\| \|\bar{b}\| \leq \|A^{-1}\bar{b}\| = \|\bar{x}\|$ ، و این رابطه نشان می‌دهد که A^{-1} ممکن است بزرگ باشد. البته $\|A^{-1}\|$ بزرگ به معنای بدوضعی است، زیرا در آن صورت $\|A^{-1}\| \|A\| = \text{Cond}(A)$ بزرگ خواهد بود.

تبصره: ماتریسهایی یافت می‌شوند که هیچ یک از این نشانه‌ها را ندارند اما باز هم بدوضع هستند (Wilkinson (1965, 254-255)).

۶-۷-۶ بدوضعی و مسأله مقدار ویژه

اگر یک ماتریس نرمال شده دارای یک مقدار ویژه کوچک باشد، باید بدوضع باشد. اگر $\lambda_1, \lambda_2, \dots, \lambda_n$ مقدارهای ویژه A باشند، آنگاه می‌توان نشان داد که

$$\|A^{-1}\|_2 \geq | \text{ماکزیمم مقدار ویژه } A^{-1} | = \frac{1}{\min |\lambda_i|}$$

مثال ۶-۷-۳ دستگاه خطی

$$Ax = b$$

را با

$$A = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & \cdot / \dots \cdot 1 & \cdot \\ \cdot & \cdot & \cdot / \dots \cdot 1 \end{bmatrix}$$

$$b = \begin{bmatrix} 0/1 \\ 0/1 \\ 0/1 \end{bmatrix}$$

$$x = 10^2 \begin{bmatrix} 0/00001 \\ 1 \\ 1 \end{bmatrix}$$

در نظر بگیرید که نسبتاً بزرگ است.

مقادیر ویژه A عبارتند از ۱، $0/00001$ ، و $0/00001$. بنابراین A دارای یک مقدار ویژه کوچک است.

$$A^{-1} = 10^5 \begin{bmatrix} 0/00001 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

که بزرگ است. بنابراین برای این مثال (۱) جواب محاسبه شده بزرگ است، (۲) A دارای یک مقدار ویژه کوچک است، و (۳) A^{-1} بزرگ است. بنابراین A باید احتمالاً بدوضع باشد. در حقیقت این مطلب درست است زیرا

$$\text{Cond}(A) = 10^5$$

۷-۷-۶ بدوضعی و مقیاس کردن

در بخش ۶-۴-۸ در مورد مقیاس کردن بحث کردیم، پیام این بود که به طور کلی مقیاس کردن توصیه می‌شود. اگر عناصر ماتریس A به طور وسیعی تغییر کنند. مقیاس کردن و به دنبال آن یک استراتژی محورگیری مفید خواهد بود. چیزی که ما روشن نساخته‌ایم این است که آیا مقیاس کردن بر روی عدد وضعیت ماتریس اثر خواهد داشت. برای مثال دوباره مثال بخش ۶-۴-۸ را ملاحظه کنید

$$A = \begin{bmatrix} 10 & 10^6 \\ 1 & 1 \end{bmatrix}$$

$$\text{Cond}(A) = 10^6$$

ولیکن اگر اولین سطر A مقیاس شود خواهیم داشت:

$$\tilde{A} = \begin{bmatrix} 0/00001 & 1 \\ 1 & 1 \end{bmatrix}$$

آنگاه

$$\text{Cond}(\tilde{A}) = 2$$

این سؤال به طور طبیعی مطرح می‌شود، با مفروض بودن یک ماتریس A چگونه می‌توان ماتریسهای قطری D_1 و D_2 را انتخاب کرد به طوری که $\text{Cond}(D_1^{-1}AD_2)$ تا حد ممکن کوچک باشد؟ یک راه حل کلاسیک که توسط بوئر ((Bauer (1963)) پیشنهاد شده است، برای این مسأله وجود دارد. متأسفانه این راه حل عملی نیست. برای مثال راه حل نرم بی‌نهایت نیاز به شناخت مقدار ویژه با ماکزیمم قدر مطلق و بردار ویژه متناظرش از ماتریس نامنفی $C = |A| \cdot |A^{-1}|$ دارد و برای حل دستگاه $Ax = b$ ، ما از قبل A^{-1} را نمی‌شناسیم. برای جزئیات بیشتر روش مرجع Forysthe and Moler (1967, 43-44) را ملاحظه کنید.

تبصره: دمل ((Demmel (1989)) نشان داده است که به هنگام حل یک دستگاه معین مثبت متقارن با استفاده از الگوریتم چولسکی مقیاس کردن جهت بهبود عدد وضعیت لازم نیست. کران خطای به دست آمده برای جواب حاصل از الگوریتم برای دستگاه مقیاس نشده $Ax = b$ اغلب با کران خطای حاصل از دستگاه مقیاس شده با $\tilde{A} = D^{-1}AD^{-1}$ یکسان است، در اینجا $D = \text{diag}(\sqrt{a_{11}}, \dots, \sqrt{a_{nn}})$. مرجع Higham(1995) را ملاحظه کنید.

۶-۷-۸ محاسبه و برآورد عدد وضعیت

راه واضح و روشن برای محاسبه عدد وضعیت محاسبه آن از روی تعریف آن می‌باشد:

$$1. \text{ محاسبه } A^{-1}$$

$$2. \text{ محاسبه } \|A\| \text{ و } \|A^{-1}\|، \text{ و پیدا کردن حاصلضرب آنها.}$$

ملاحظه کرده‌ایم که محاسبه معکوس A^{-1} تقریباً به n^3 عمل ممیز شناور (flops) نیاز دارد. بنابراین این راه سه برابر گرانتر از پیدا کردن جواب برای دستگاه $Ax = b$ می‌باشد. از طرف دیگر برای محاسبه $\text{Cond}(A)$ فقط شناخت $\|A^{-1}\|$ ، نه خود معکوس، لازم می‌باشد. بعلاوه خود مقدار واقعی $\text{Cond}(A)$ بندرت لازم می‌باشد؛ یک برآورد کافی است. بنابراین سؤالی که مطرح می‌شود این است که آیا می‌توان یک برآورد معقول از $\|A^{-1}\|$ بدون محاسبه معکوس A به دست آورد.

در این زمینه، توجه داریم که اگر y یک بردار دلخواه n بعدی مخالف صفر باشد، آنگاه از

$$Az = y$$

داریم

$$\|z\| = \|A^{-1}y\| \leq \|A^{-1}\| \|y\|$$

پس

$$\|A^{-1}\| \geq \frac{\|z\|}{\|y\|}, \quad y \neq 0$$

بنابراین اگر γ را به گونه‌ای اختیار کنیم که $\|z\|/\|y\|$ کاملاً بزرگ باشد، می‌توانیم یک برآورد خوب و معقول از $\|A^{-1}\|$ داشته باشیم. رایس (Rice (1981, 93)) متذکر شده است که یک بحث ابتکاری وجود دارد که می‌گوید که اگر γ به طور تصادفی اختیار شود، آنگاه مقدار انتظار داشته $\|z\|/\|y\|$ تقریباً برابر $\|A^{-1}\|$ است.

یک راه اصولی و منظم برای انتخاب y در LINPACK(1979) ارائه شده است. این راه بر الگوریتمی مبتنی است که در مرجع (Cline, Moler, Stewart, and Wilkinson(1979)) ارائه شده است. فرایند حاوی حل دو دستگاه

$$A^T y = e$$

و

$$Az = y$$

است که در آن e مضرب عددی از یک بردار با مؤلفه‌های ± 1 و انتخاب شده به طریقی است که رشد ممکن ماکزیمم باشد. برای جزئیات بیشتر مرجع (Cline, Moler, Stewart, and Wilkinson (1979)) یا LINPACK(1979) را ملاحظه کنید.

برای پرهیز از سرریز روال SEGCO (زیر برنامه برآوردگر عدد وضعیت LINPACK) در واقع یک برآورد از $1/\text{Cond}(A)$ را به نام

$$\text{RCOND} = \frac{\|y\|}{\|A\| \|z\|}$$

محاسبه می‌کند.

یک روش بهینه‌سازی برای برآورد $\|A^{-1}\|_1$ هاگر (Hager (1984)) یک روش برای برآورد $\|A^{-1}\|$ مبتنی بر یک روش بهینه‌سازی پیشنهاد کرده است. این روش به نظر می‌رسد که برای ماتریسهای تولید شده تصادفی کاملاً مناسب باشد.

فرض کنید $A^{-1} = B = (b_{ij})$. یک تابع $f(x)$ به صورت

$$f(x) = \|Bx\|_1 = \sum_{i=1}^n \left| \sum_{j=1}^n b_{ij} x_j \right|$$

تعریف می‌کنیم. پس

$$\|B\|_1 = \|A^{-1}\|_1 = \max\{f(x) : \|x\|_1 \leq 1\}$$

بنابراین مسأله پیدا کردن ماکزیمم تابع محدب $f(x)$ بر روی مجموعه محدب

$$S = \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$$

است. همان طور که می‌دانیم ماکزیمم یک تابع محدب (تعریف شده بر روی یک مجموعه فشرده محدب) در یک نقطه اکسترم به دست می‌آید. روش هاگر از پیدا کردن اصولی و منظم این ماکزیمم تشکیل می‌شود. هاگر متذکر می‌شود که الگوریتم معمولاً بعد از دو گام متوقف می‌شود. مروری عالی از برآوردگرهای مختلف عدد وضعیت، شامل روش هاگر و کارایی آنها در (Higham(1987 ارائه شده است.

الگوریتم ۶-۷-۱ برآوردگر عدد وضعیت ۱- نرم هاگر

$$\rho = \|A^{-1}\|_1 = 0 \quad \text{قرار دهید}$$

$$b = \begin{bmatrix} \frac{1}{n} \\ \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{bmatrix} \quad \text{قرار دهید}$$

گام ۱: دستگاه $Ax = b$ را حل کنید.

گام ۲: $\|x\|_1 \leq \rho$ را آزمون کنید اگر برقرار باشد به گام ۶ بروید. در غیر این صورت قرار دهید

$$\rho = \|x\|_1 \quad \text{و به گام ۳ بروید.}$$

گام ۳: دستگاه $A^T z = y$ را حل کنید، که در آن

$$y_i = 1 \quad \text{اگر } x_i \geq 0$$

$$y_i = -1 \quad \text{اگر } x_i < 0$$

گام ۴: قرار دهید $j = \arg \max\{|z_i|, i = 1 \text{ to } n\}$

گام ۵: اگر $|z_j| > z^T b$ ، بردار

$$b \equiv \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \text{ زامین عنصر}$$

را به هنگام درآوردن و به گام ۱ برگردید. در غیر این صورت به گام ۶ بروید.

گام ۶: قرار دهید $\rho \approx \|A^{-1}\|_1 \cdot \|A\|_1$ پس $\text{Cond}_1(A) = \rho$.

تذکره‌های LAPACK و MATCOM: عدد وضعیت هاگر در برنامه HAGCOND1 در MATCOM پیاده‌سازی شده است. الگوریتم ۶-۷-۱ در LAPACK نیز پیاده‌سازی شده است. زیر برنامه SGECON از LAPACK یک برآورد از معکوس عدد وضعیت ۱- نرم ماتریس را با استفاده از تغییر هایام از الگوریتم هاگر (Higham, 1988) حساب می‌کند.

مثال ۶-۷-۴ روش هاگر را به کمک یک ماتریس خیلی بدوضع توضیح می‌دهیم.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} \quad \text{cond}(A) = 3/3819 \times 10^{16}$$

تکرار ۱:

$$b = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

$$x = \begin{bmatrix} 1/0.895 \\ -2/5123 \\ 1/4228 \end{bmatrix}$$

$$\rho = \|x\|_1 = 5/0.245$$

$$y = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$z = 10^{16} \begin{bmatrix} 2/0.271 \\ -3/3785 \\ 1/3514 \end{bmatrix}$$

$$j = 2$$

$$|z_j| = 10^{16}(3/3785) > z^T b = -1/3340$$

به هنگام درمی آوریم

$$b \equiv \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

تکرار ۲:

$$b = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$x = 10^{17} \begin{bmatrix} -1/3564 \\ 2/7128 \\ -1/3564 \end{bmatrix}$$

$$\|x\|_1 = 5/4255 \times 10^{17}$$

چون $\|x\|_1 > \rho$ ، قرار می دهیم $\rho = 5/4255 \times 10^{17}$ و ادامه می دهیم.
توضیح: ثابت می شود که این مقدار فعلی یک برآورد عالی از $\|A^{-1}\|_1$ است.

برآورد عدد وضعیت از مثلثی سازی

اگر از تجزیه QR برای حل یک دستگاه خطی یا تجزیه چولسکی برای حل یک دستگاه معین مثبت متقارن استفاده کنیم، آنگاه به عنوان یک محصول فرعی از مثلثی سازی، یک کران بالا برای عدد وضعیت با کمی هزینه اضافی می توانیم به دست آوریم.

اگر از تجزیه QR با محورگیری ستونی استفاده شود، آنگاه از

$$Q^T A P = R$$

$$\text{Cond}_2(A) = \text{Cond}_2(R)$$

اگر از تجزیه چولسکی استفاده شود، آنگاه از

$$A = H H^T$$

داریم $\text{Cond}_p(A) = (\text{Cond}_p(H))^2$. بنابراین $\text{Cond}_p(A)$ می‌تواند تعیین شود اگر $\text{Cond}_p(R)$ یا $\text{Cond}_p(H)$ معلوم باشد. چون $\|R\|_p$ یا $\|H\|_p$ به راحتی محاسبه می‌شوند همه آنچه که مورد نیاز است الگوریتمی برای برآورد $\|R^{-1}\|_p$ یا $\|H^{-1}\|_p$ می‌باشد. چندین الگوریتم برای برآورد نرمهای معکوسهای ماتریسهای مثلثی وجود دارد. ما فقط یکی را از مقاله هایام (Higham (1987)) بیان می‌کنیم. برای جزئیات بیشتر آن مقاله را ملاحظه کنید.

الگوریتم ۶-۷-۲ برآورد عدد وضعیت یک ماتریس بالا مثلثی. با مفروض بودن یک ماتریس بالا مثلثی نامنفرد $T = (t_{ij})$ از مرتبه n ، الگوریتم زیر عدد C_E را به طریقی محاسبه می‌کند که $\|T^{-1}\|_\infty \leq C_E$ باشد.

step 1: Set $z_n = \frac{1}{t_{nn}}$

step 2: For $i = n - 1$ to 1 do

$$s \equiv 1$$

$$s \equiv s + |t_{ij}|z_j \quad (j = i + 1, \dots, n)$$

$$z_i = \frac{s}{|t_{ii}|}$$

step 3: Compute $C_E = \|z\|_\infty$, where $z = (z_1, z_2, \dots, z_n)^T$

شمارش اعمال ممیز شناور. الگوریتم نیاز به $\frac{n^2}{4}$ عمل ممیز شناور (flops) دارد.

تبصره: به محض این که $\|T^{-1}\|_\infty$ توسط الگوریتم ۶-۷-۲ برآورد شود، $\|T^{-1}\|_p$ می‌تواند از رابطه

$$\|T^{-1}\|_p \leq (\|M(T)^{-1}\|_1 C_E)^{1/2}$$

برآورد گردد که در آن $M(t) = (m_{ij})$ به صورت زیر تعریف می‌شود

$$m_{ij} = \begin{cases} |t_{ii}|, & i=j \\ -|t_{ij}|, & i \neq j \end{cases}$$

$\|M(T)^{-1}\|_1$ می‌تواند با استفاده از الگوریتم هاگر (الگوریتم ۶-۷-۱) برآورد شود.

۸-۶ اختلاهای مؤلفه‌ای و خطاها

اگر کرانه‌های مؤلفه‌ای اختلالها معلوم باشند، آنگاه نتیجه اختلال زیر، به دست آمده توسط اسکیل

(Skeel (1979)) برقرار می‌باشد. در زیر $\| \cdot \|$ برای بیان نرم بی‌نهایت به کار برده شده است.

قضیه ۱-۸-۶ (Skeel (1979)) فرض کنید $Ax = b$ و $(A + \Delta A)(x + \delta x) = b + \delta b$ فرض کنید $|\Delta A| \leq \varepsilon |A|$ و $|\delta b| \leq \varepsilon |b|$. آنگاه رابطه

$$\frac{\|\delta x\|}{\|x\|} \leq \varepsilon \frac{\| |A^{-1}| |A| |x| + |A^{-1}| |b| \|}{(1 - \varepsilon \| |A^{-1}| |A| \|) \|x\|}$$

برقرار می‌باشد به شرط آن که مخرج مثبت باشد.

تعریف ۱-۸-۶ عدد

$$\text{Cond}(A, x) = \frac{\| |A^{-1}| |A| |x| \|}{\|x\|}$$

را عدد وضعیت اسکیل و عدد $\text{Cond}_s(A) = \| |A^{-1}| |A| \|$ را کران بالای عدد وضعیت اسکیل خواهیم نامید.

تعریف ۱-۸-۶ برای برخی توصیفها به کار برده می‌شود. از قضیه ۱-۸-۶ می‌توان نشان داد که عبارت

$$\frac{\| |A^{-1}| |A| |x| + |A^{-1}| |b| \|}{\|x\|}$$

عدد وضعیت دستگاه $Ax = b$ است هنگامی که داده‌های A و b هر دو تحت اختلال قرار گیرند، و $\text{Cond}(A, x)$ عدد وضعیت است هنگامی که فقط A تحت اختلال قرار گیرد. ولیکن اسکیل (Skeel (1979)) نشان داده است که $\text{Cond}(A, x)$ کافی است هنگامی که A و b هر دو تحت اختلال قرار می‌گیرند.

در نوشته‌ها، $\text{Cond}(A, x)$ معمولاً توسط $\text{Cond}(A)$ نمایش داده می‌شود و عدد وضعیت معمولی $\|A\| \|A^{-1}\|$ توسط $\kappa(A)$ نمایش داده می‌شود. خواننده باید به این مطلب توجه داشته باشد تا از اشتباه جلوگیری به عمل آید. بنابراین در نمادگذاریهایمان داریم:

$$\text{Cond}(A) = \|A\| \|A^{-1}\|$$

$$\text{Cond}(A, x) = \frac{\| |A^{-1}| |A| |x| \|}{\|x\|}$$

$$\text{Cond}_s(A) = \| |A| |A^{-1}| \|$$

یک خاصیت مهم $\text{Cond}(A, x)$ این است که عدد وضعیت اسکیل تحت مقیاس کردن سطری

تغییرناپذیر است. بنابراین می‌تواند خیلی کوچک‌تر از عدد وضعیت معمولی $\text{Cond}(A)$ باشد. $\text{Cond}(A, x)$ هنگامی مفید است که نرم ستونهای A^{-1} به طور وسیعی تغییر کنند. چاندرازکاران و ایپسن ((Chandrasekaran and Ipsen (1994b)) اخیراً تحلیلی از این که چگونه مؤلفه‌های فردی بردار جواب x تحت تأثیر قرار می‌گیرند هنگامی که داده‌ها اختلال می‌یابند، ارائه کرده‌اند. این منجر به اعداد وضعیت مؤلفه‌ای شده است که حساسیت هر مؤلفه جواب را در برابر اختلال اندازه‌گیری می‌کنند. تحلیل آنها می‌تواند به طور مؤثری با نتایج اسکیل ترکیب شود اگر اختلالهای داده‌ها معلوم باشند. اکنون قضیه زیر را برای اعداد وضعیت مؤلفه‌ای ارائه می‌دهیم.

قضیه ۶-۸-۲ قضیه اختلال مؤلفه‌ای. فرض کنید $(A + \Delta A)\bar{x} = b$ و $|\Delta A| \leq \varepsilon |A|$. آنگاه

$$\frac{|\bar{x}_i - x_i|}{|x_i|} \leq \varepsilon \frac{|r_i^T| |A| |\bar{x}|}{|x_i|}$$

که در آن $r_i^T = e_i^T A_i^{-1}$.

بنابراین اختلالهای مؤلفه‌ای در عبارتهای خطا منجر به گونه مؤلفه‌ای عدد وضعیت اسکیل شده‌اند. نتایج مشابهی برای اختلال طرف راست نیز وجود دارد. برای جزئیات بیشتر مرجع Chandrasekaran and Ipsen (1994b) را ملاحظه کنید.

۶-۹ تصفیه تکراری

فرض کنید که یک جواب محاسبه شده \bar{x} از دستگاه $Ax = b$ قابل قبول نباشد. آنگاه طبیعی است متعجب شویم اگر \bar{x} بتواند به طور ارزانی با استفاده از مثلثی‌سازی ماتریس A که از قبل در دست می‌باشد، تصفیه شود.

فرایند زیر معروف به تصفیه تکراری، می‌تواند به طور تکراری برای تصفیه \bar{x} تا رسیدن به دقت مورد نظر استفاده شود.

الگوریتم تصفیه تکراری

این فرایند بر ایده ساده زیر مبتنی است: فرض کنید که \bar{x} یک جواب محاسبه شده برای دستگاه

$$Ax = b$$

باشد. اگر \bar{x} یک جواب واقعی می‌بود، آنگاه

$$r = b - A\bar{x}$$

صفر می‌شد. اما در عمل چنین انتظاری را نخواهیم داشت. اکنون اجازه دهید سعی کنیم دستگاه را دوباره با مانده محاسبه شده $r (r \neq 0)$ حل کنیم؛ یعنی فرض کنید c در دستگاه

$$Ax = r$$

صدق کند، پس $y = \bar{x} + c$ جواب واقعی $Ax = b$ خواهد بود به شرط آن که c جواب واقعی $Ac = r$ باشد، زیرا

$$Ay = A(\bar{x} + c) = A\bar{x} + Ac = b - r + r = b$$

درست است که c در عمل جواب واقعی $Ac = r$ نخواهد بود؛ ولیکن بحث قبل پیشنهاد می‌کند که y باید تقریب بهتری از \bar{x} باشد. اگر چنین باشد، می‌توان فرایند را تا رسیدن به یک دقت دلخواه ادامه داد.

الگوریتم ۱-۹-۶ تصفیه تکراری

$$x^{(0)} = \bar{x}$$

برای $k = 0, 1, 2, \dots$ انجام دهید:

گام ۱: بردار مانده $r^{(k)}$ را محاسبه کنید

$$r^{(k)} = b - Ax^{(k)}$$

گام ۲: بردار تصحیح $c^{(k)}$ را با حل دستگاه

$$Ac^{(k)} = r^{(k)}$$

حساب کنید. برای این منظور از مثلثی‌سازی حاصل از به دست آوردن جواب اولیه استفاده کنید.

گام ۳: تشکیل دهید $x^{(k+1)} = x^{(k)} + c^{(k)}$

گام ۴: اگر $\frac{\|c^{(k)}\|}{\|x^{(k+1)}\|}$ کوچکتر از تورلانس از پیش تعیین شده ϵ باشد، یا اگر تعداد تکرارها از تعداد از

پیش تعیین شده تجاوز کند متوقف شوید.

تذکر MATCOM: الگوریتم در برنامه ITERREF در MATCOM پیاده‌سازی شده است.

مثال ۱-۹-۶

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} 0/0001 \\ 0/0001 \\ -1/6666 \end{bmatrix}$$

جواب واقعی عبارت است از

$$x = \begin{bmatrix} -0/2777 \\ 0/2778 \\ -0/5555 \end{bmatrix} \quad (\text{تا چهار رقم اعشار صحیح است})$$

$$x^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

: $k = 0$

$$r^{(0)} = b - Ax^{(0)} = \begin{bmatrix} -1/9999 \\ -2/9999 \\ -4/6666 \end{bmatrix}$$

جواب واقعی $Ac^{(0)} = r^{(0)}$ به صورت زیر است

$$c^{(0)} = \begin{bmatrix} -1/2777 \\ -0/7222 \\ -1/5555 \end{bmatrix}$$

$$x^{(1)} = x^{(0)} + c^{(0)} = \begin{bmatrix} -0/2777 \\ 0/2778 \\ -0/5555 \end{bmatrix}$$

توجه کنید که $\text{Cond}(A) = 3/8078$. A خوش وضع است.

دقت به دست آمده توسط تصفیه تکراری

فرض کنید که تکرار همگرا گردد. آنگاه خطا در $(k + 1)$ امین گام کمتر از خطا در k امین گام خواهد بود.

دقت نسبی از تصفیه تکراری

فرض کنید

$$\frac{\|\hat{x} - x^{(k+1)}\|}{\|\hat{x}\|} \leq c \frac{\|\hat{x} - x^{(k)}\|}{\|\hat{x}\|}$$

آنگاه اگر $c \approx 10^{-s}$ باشد، به طور تقریبی s رقم در هر تکرار سود خواهیم برد.

شمارش اعمال ممیز شناور. روش نسبتاً ارزان است. زیرا A قبلاً برای حل دستگاه اصلی $Ax = b$ مثلثی شده است، هر تکرار فقط به $O(n^2)$ عمل ممیز شناور (flops) نیاز دارد.

مفید بودن تصفیه تکراری و چند تذکر

تصفیه تکراری یک روش خیلی مفید است. روش حذفی گاوس با محورگیری و به دنبال آن استفاده از تصفیه تکراری عملی‌ترین راه برای حل یک دستگاه خطی به طور دقیق می‌باشد. اسکیل (Skeel (1980)) نشان داده است که «فقط یک تصفیه تکراری با تنها انباشتن مانده‌ها با دقت ساده کافی خواهد بود تا روش حذفی گاوس را پایدار سازد». این عبارت به نظر می‌رسد که با اعتقاد همیشگی که «مطلقاً ضروری است که مانده‌ها با دقت بالاتری از بقیه محاسبات انجام شوند» تناقض دارد (Forsythe and Moler (1967, 49)). ولیکن اسکیل (Skeel (1980)) نشان داده است که انباشتن مانده‌ها با دقت مضاعف (یعنی محاسبه مانده‌ها با دقت مضاعف و گردکردن آنها به دقت ساده) دارای مزیت مهمی برای مسائلی است که هم به طور ضعیف مقیاس شده‌اند و هم بدوضع هستند.

حقیقت امر این است که می‌توان نشان داد که اگر $\text{Cond}(A) \leq \frac{1}{\mu}$ و اگر مانده‌ها در دقت مضاعف انباشته شوند، آنگاه روش تصفیه تکراری نهایتاً به جواب با دقت کامل $Ax = b$ همگرا می‌گردد. یعنی اگر همه محاسبات بجز مانده‌ها با استفاده از حساب ϵ رقمی انجام شوند، آنگاه روش تصفیه تکراری نهایتاً جوابی تولید خواهد که دارای دقتی با ϵ رقم خواهد بود. بعلاوه جان کوفسکی و وزینا کوفسکی (Jankowski and Wozniakowki (1977)) ثابت کرده‌اند که هر حل‌کننده دستگاه خطی، حاوی روش حذفی گاوس، می‌تواند با استفاده از تصفیه تکراری پایدار گردد. مقاله جدید هایام (Higham (1991)) را نیز که در مورد تصفیه تکراری روشهای تجزیه QR برای حل دستگاههای خطی است، ملاحظه نمایید. جنبه مهم دیگر تصفیه تکراری این است که همگرایی آهسته فرایند، همان گونه که نتیجه زیر نشان می‌دهد، نشانی از بدوضعی مسأله است.

برآورد $\text{Cond}(A)$ از تصفیه تکراری

یک برآورد خیلی خام از $\text{Cond}(A)$ ممکن است از روش تصفیه تکراری (الگوریتم ۶-۹-۱) به دست آید. فرض کنید k تعداد تکرارهای لازم برای روش تصفیه تکراری برای همگرایی باشد و ϵ تعداد ارقام استفاده شده در حساب باشند؛ آنگاه یک برآورد خام از $\text{Cond}(A)$ برابر $10^{\epsilon(1-1/k)}$ است (Rice (1981, 89)). بنابراین، اگر روش تصفیه تکراری به طور خیلی آهسته همگرا گردد، آنگاه A بدوضع است.

۶-۱۰ روشهای تکراری

در این بخش روشهای تکراری را که معمولاً برای حل یک دستگاه خطی تنک و خیلی بزرگ $Ax = b$ و حاصل از کاربردهای مهندسی استفاده می‌شوند، مطالعه می‌کنیم. این روشها شامل موارد زیر هستند:

۱. روش ژاکوبی (بخش ۶-۱۰-۱)
۲. روش گاوس - سایدل (بخش ۶-۱۰-۲)
۳. روش فوق تخفیف متوالی (بخش ۶-۱۰-۴)
۴. روش گرادیان مزدوج با و بدون پیش شرط (بخش ۶-۱۰-۵)
۵. روش GMRES (بخش ۶-۱۰-۶)

روش گاوس - سایدل تغییری از روش ژاکوبی است و یک حالت خاص از روش فوق تخفیف (SOR) است. روش گرادیان مزدوج اصولاً برای حل دستگاههای معین مثبت متقارن مورد استفاده قرار می‌گیرد. روشهای ژاکوبی و گاوس - سایدل برای ماتریسهای غالب قطری همگرا هستند؛ بعلاوه روش گاوس - سایدل همگرا می‌گردد، اگر A معین مثبت متقارن باشد. یادآور می‌شویم که ماتریسهای غالب قطری و معین مثبت متقارن در میان مهم‌ترین رده‌های ماتریسهایی قرار می‌گیرند که در کاربردهای عملی به وجود می‌آیند.

روشهای مستقیم مبتنی بر مثلثی‌سازی ماتریس A از جهت زمان کامپیوتری و ذخیره‌سازی توصیه نمی‌شوند اگر ماتریس A نسبتاً بزرگ باشد. از طرف دیگر وضعیتهای عملی دیگر نظیر گسسته‌سازی معادلات مشتقات جزئی وجود دارند که در آنها اندازه ماتریس می‌تواند به بزرگی چند صد هزار باشد. برای چنین مسأله‌هایی روشهای مستقیم غیرعملی می‌شوند. برای مثال اگر A از مرتبه 10000×10000 باشد آنگاه ممکن است ۲ یا ۳ روز برای یک IBM ۳۷۰ طول بکشد تا دستگاه $Ax = b$ با استفاده از روش حذفی گاوس یا روشهای متعامدسازی هاوس هولدر و گیونز حل شود. بعلاوه بیشتر مسائل تنک هستند و تنک بودن در طی فرایند مثلثی‌سازی از بین می‌رود و به طور قابل ملاحظه‌ای توسعه می‌یابد. بنابراین در انتها با یک ماتریس خیلی بزرگ با عناصر مخالف صفر بسیار زیادی سر و کار خواهیم داشت و ذخیره‌سازی یک عمل بسیار سخت می‌شود. برای چنین مسائلی، توصیه می‌شود از رده‌ای از روشهایی که روشهای تکراری نامیده می‌شوند، استفاده گردد که هرگز ماتریس A را تغییر نمی‌دهند و فقط به ذخیره‌سازی تعدادی بردار با طول n در یک زمان نیاز دارند.

ایده اساسی

ایده اساسی در پشت یک روش تکراری این است که ابتدا دستگاه $Ax = b$ به شکل معادل زیر نوشته شود:

$$x = Bx + d \quad (۱-۱۰-۶)$$

سپس با شروع از یک تقریب اولیه $x^{(۱)}$ از بردار جواب x ، یک دنباله از تقریبهای $\{x^{(k)}\}$ که به صورت

$$x^{(k+۱)} = Bx^{(k)} + d, \quad k = ۱, ۲, \dots \quad (۲-۱۰-۶)$$

تعریف می‌شوند، تولید گردند، با این امید که تحت برخی شرایط معتدل، دنباله $\{x^{(k)}\}$ هنگامی که $k \rightarrow \infty$ به جواب همگرا گردد.

بنابراین برای حل دستگاه خطی $Ax = b$ به طور تکراری و با استفاده از این ایده، لازم است که

بدانیم

۱. چگونه دستگاه خطی $Ax = b$ را به شکل (۱-۱۰-۶) بنویسیم، و

۲. چگونه باید $x^{(۱)}$ انتخاب شود تا تکرار (۲-۱۰-۶) به حد همگرا گردد یا تحت چه نوع فرضیهایی

تکرار با هر انتخاب دلخواه $x^{(۱)}$ به حد همگرا می‌گردد.

معیارهای توقف برای تکرار (۲-۱۰-۶)

طبیعی است که کنجکاو باشیم که چه موقع تکرار (۲-۱۰-۶) می‌تواند پایان یابد. چه موقع همگرایی رخ می‌دهد، $x^{(k+۱)}$ تقریب بهتری از $x^{(k)}$ است، بنابراین یک معیار توقف طبیعی به صورت زیر است.

معیار توقف ۱

تکرار (۲-۱۰-۶) را متوقف می‌سازیم اگر رابطه

$$\frac{\|x^{(k+۱)} - x^{(k)}\|}{\|x^{(k)}\|} < \varepsilon$$

به ازای یک عدد مثبت کوچک از پیش تعیین شده ε برقرار باشد (ε باید مطابق با دقت مورد نظر انتخاب شود) یا اگر تعداد تکرارها از یک تعداد از پیش تعیین شده تجاوز کند.

یک معیار توقف بر حسب مانده نیز به صورت زیر می‌باشد

معیار توقف ۲

تکرار (۶-۱۰-۲) را متوقف می‌سازیم اگر رابطه

$$\|r^{(k+1)}\| \leq \varepsilon (\|A\| \|x^{(k+1)}\| + \|b\|)$$

برقرار باشد که در آن

$$r^{(k+1)} = Ax^{(k+1)} - b$$

یا تعداد تکرارها از یک تعداد از پیش تعیین شده تجاوز کند. ε تولرانسی است که باید به قسمی انتخاب شود که $1 < \varepsilon < \mu$ ، μ دقت ماشین است.

تبصره: معیار توقف ۲ و برخی دیگر در کتاب Templates (1994) ارائه شده‌اند. برای جزئیات بیشتری در مورد این معیارها و توجیه‌های آنها کتاب Templates (1994, 53-54) را ملاحظه کنید.

۶-۱۰-۱ روش ژاکوبی

دستگاه

$$Ax = b$$

یا

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

⋮

$$a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n$$

می‌تواند به صورت زیر (بافرض $a_{ii} \neq 0$, $i = 1, \dots, n$) دوباره نوشته شود.

$$x_1 = \frac{1}{a_{11}}(b_1 - a_{12}x_2 - \dots - a_{1n}x_n)$$

$$x_2 = \frac{1}{a_{22}}(b_2 - a_{21}x_1 - a_{23}x_3 - \dots - a_{2n}x_n)$$

⋮

$$x_n = \frac{1}{a_{nn}}(b_n - a_{n1}x_1 - \dots - a_{n, n-1}x_{n-1})$$

در نمادگذاری ماتریسی داریم:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \cdot & -\frac{a_{12}}{a_{11}} & \dots & \dots & -\frac{a_{1n}}{a_{11}} \\ \frac{a_{21}}{a_{22}} & \cdot & -\frac{a_{23}}{a_{22}} & \dots & -\frac{a_{2n}}{a_{22}} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \frac{a_{n1}}{a_{nn}} & \dots & \dots & -\frac{a_{n,n-1}}{a_{nn}} & \cdot \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \frac{b_1}{a_{11}} \\ \frac{b_2}{a_{22}} \\ \vdots \\ \frac{b_n}{a_{nn}} \end{bmatrix}$$

یا

$$x = Bx + d$$

اگر ماتریس A را به شکل

$$A = L + D + U$$

بنویسیم که در آن

$$L = \begin{bmatrix} \cdot & \cdot & \cdot & \dots & \cdot \\ a_{21} & \cdot & \cdot & \dots & \cdot \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \cdot \\ a_{n1} & \dots & \dots & a_{n,n-1} & \cdot \end{bmatrix}$$

$$D = \text{diag}(a_{11}, \dots, a_{nn})$$

و

$$U = \begin{bmatrix} \cdot & a_{12} & a_{13} & \dots & a_{1n} \\ \cdot & \cdot & a_{23} & \dots & a_{2n} \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & a_{n-1,n} \\ \cdot & \dots & \dots & \dots & \cdot \end{bmatrix}$$

آنگاه به آسانی می‌توان ملاحظه کرد که

$$B = -D^{-1}(L + U) = (I - D^{-1}A)$$

$$d = D^{-1}b$$

(توجه کنید که به واسطه فرضمان که $a_{ii} \neq 0, i = 1, \dots, n$ ، D نامنفرد است).
ماتریس

$$B = -D^{-1}(L + U) = (I - D^{-1}A)$$

را ماتریس تکرار ژاکوبی می‌نامیم و آن را توسط B_J نمایش می‌دهیم. به طور مشابه بردار $D^{-1}b$ را به b_J نمایش داده و آن را بردار ژاکوبی می‌نامیم.

ماتریس تکرار ژاکوبی و بردار ژاکوبی

فرض کنید $A = L + D + U$. آنگاه

$$B_J = -D^{-1}(L + U)$$

$$b_J = -D^{-1}b$$

با ماتریس تکرار ژاکوبی و بردار ژاکوبی تعریف شده، تکرار (۲-۱۰-۶) به صورت زیر در می‌آید.

تکرار ژاکوبی

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^{(k)} \right], \quad i = 1, \dots, n$$

بنابراین رویه تکراری زیر را که روش ژاکوبی نامیده می‌شود داریم.

الگوریتم ۶-۱۰-۱ روش ژاکوبی

$$x^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \vdots \\ x_n^{(1)} \end{bmatrix} \quad \text{گام ۱: انتخاب کنید}$$

گام ۲: برای $k = 1, 2, \dots$ تا برقراری یک معیار توقف انجام دهید

$$x^{(k+1)} = B_J x^{(k)} + b_J \quad (۳-۱۰-۶)$$

یا

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^{(k)} \right], \quad i = 1, \dots, n \quad (۴-۱۰-۶)$$

تذکر: در پیاده‌سازی عملی از (۴-۱۰-۶) استفاده خواهیم کرد.

تذکر MATCOM: الگوریتم ۶-۱۰-۱ در برنامه JACOBI در MATCOM پیاده‌سازی شده است.

مثال ۶-۱۰-۱

$$A = \begin{bmatrix} 5 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 5 \end{bmatrix}, \quad b = \begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix}$$

$$x^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$B_J = \begin{bmatrix} \cdot & -\cdot/2 \dots & -\cdot/2 \dots \\ -\cdot/2 \dots & \cdot & -\cdot/2 \dots \\ -\cdot/2 \dots & -\cdot/2 \dots & \cdot \end{bmatrix}, \quad b_J = \begin{bmatrix} 1/4 \dots \\ 1/4 \dots \\ 1/4 \dots \end{bmatrix}$$

: $k = 1$

$$x^{(2)} = B_J x^{(1)} + b_J = \begin{bmatrix} 1/4 \dots \\ 1/4 \dots \\ 1/4 \dots \end{bmatrix}$$

$$:k = 2$$

$$x^{(2)} = B_J x^{(1)} + b_J = \begin{bmatrix} 0/8400 \\ 0/8400 \\ 0/8400 \end{bmatrix}$$

$$:k = 3$$

$$x^{(3)} = B_J x^{(2)} + b_J = \begin{bmatrix} 1/0640 \\ 1/0640 \\ 1/0640 \end{bmatrix}$$

$$:k = 4$$

$$x^{(4)} = B_J x^{(3)} + b_J = \begin{bmatrix} 0/9744 \\ 0/9744 \\ 0/9744 \end{bmatrix}$$

$$:k = 5$$

$$x^{(5)} = B_J x^{(4)} + b_J = \begin{bmatrix} 1/0102 \\ 1/0102 \\ 1/0102 \end{bmatrix}$$

$$:k = 6$$

$$x^{(6)} = B_J x^{(5)} + b_J = \begin{bmatrix} 0/9959 \\ 0/9959 \\ 0/9959 \end{bmatrix}$$

$$:k = 7$$

$$x^{(7)} = B_J x^{(6)} + b_J = \begin{bmatrix} 1/0016 \\ 1/0016 \\ 1/0016 \end{bmatrix}$$

۶-۱۰-۲ روش گاوس - سایدل

در روش ژاکوبی، برای محاسبه مؤلفه‌های بردار $x^{(k+1)}$ ، فقط مؤلفه‌های بردار $x^{(k)}$ استفاده می‌شوند، توجه کنید که برای محاسبه $x_i^{(k+1)}$ ما می‌توانستیم از $x_1^{(k+1)}$ تا $x_{i-1}^{(k+1)}$ که از قبل برای ما موجود هستند، استفاده کنیم. بنابراین یک تغییر طبیعی از روش ژاکوبی دوباره نوشتن تکرار ژاکوبی (۶-۱۰-۲) به شکل زیر خواهد بود.

تکرار گاوس - سایدل

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right], \quad i = 1, 2, \dots, n \quad (5-10-6)$$

ایده این است که از هر مؤلفه جدید، به محض موجود بودن آن، در محاسبه مؤلفه بعدی استفاده شود.

تکرار (۵-۱۰-۶) معروف به تکرار گاوس - سایدل است و روش تکراری مبتنی بر این تکرار روش گاوس - سایدل نامیده می‌شود.

با نمادگذاری قبلی، تکرار گاوس - سایدل به صورت زیر است

$$x^{(k+1)} = -(D + L)^{-1} U x^{(k)} + (D + L)^{-1} b$$

(توجه کنید که ماتریس $D + L$ یک ماتریس پایین مثلثی با مقادیر قطری a_{11}, \dots, a_{nn} می‌باشد؛ چون فرض کرده‌ایم که این عناصر مخالف صفر هستند، ماتریس $(D + L)$ نامنفرد است).
ماتریس

$$B = -(D + L)^{-1} U$$

را ماتریس گاوس - سایدل می‌نامیم و آن را با نماد B_{GS} نمایش می‌دهیم. به طور مشابه بردار گاوس - سایدل $(D + L)^{-1} b$ توسط b_{GS} نمایش داده می‌شود.

ماتریس گاوس - سایدل و بردار گاوس - سایدل

فرض کنید $A = L + D + U$ آنگاه

$$B_{GS} = -(D + L)^{-1} U$$

$$b_{GS} = (D + L)^{-1} b$$

الگوریتم ۶-۱۰-۲ روش گاوس - سایدل

گام ۱: یک تقریب اولیه $x^{(1)}$ انتخاب کنید.

گام ۲: برای $k = 1, 2, \dots$ تا برقراری یک معیار توقف انجام دهید

$$x^{(k+1)} = B_{GS}x^{(k)} + b_{GS} \quad (6-10-6)$$

یا

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)} \right], \quad i = 1, \dots, n \quad (7-10-6)$$

تذکر: در عمل از (۶-۱۰-۷) استفاده خواهیم کرد.

تذکر MATCOM: الگوریتم ۶-۱۰-۲ در برنامه GAUSED در MATCOM پیاده‌سازی شده است.

مثال ۶-۱۰-۲

$$A = \begin{bmatrix} 5 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 5 \end{bmatrix}, \quad b = \begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix}$$

$$B_{GS} = \begin{bmatrix} \cdot & -0.2 & -0.2 \\ \cdot & \cdot/0.4 & -0.16 \\ \cdot & \cdot/0.32 & \cdot/0.72 \end{bmatrix}, \quad b_{GS} = \begin{bmatrix} 1/4 \dots \\ 1/12 \dots \\ \cdot/896 \dots \end{bmatrix}$$

$k = 1$

$$x^{(2)} = B_{GS}x^{(1)} + b_{GS} = \begin{bmatrix} 1/4 \dots \\ 1/12 \dots \\ \cdot/896 \dots \end{bmatrix}$$

$k = 2$

$$x^{(3)} = B_{GS}x^{(2)} + b_{GS} = \begin{bmatrix} \cdot/9968 \\ 1/0.214 \\ \cdot/9964 \end{bmatrix}$$

$k = 3$

$$x^{(4)} = B_{GS}x^{(3)} + b_{GS} = \begin{bmatrix} \cdot/9964 \\ 1/0.14 \\ 1/0.04 \end{bmatrix}$$

$$:k = ۴$$

$$x^{(۵)} = B_{GS}x^{(۴)} + b_{GS} = \begin{bmatrix} ۰/۹۹۹۶ \\ ۱/۰۰۰۰ \\ ۱/۰۰۰۱ \end{bmatrix}$$

$$:k = ۵$$

$$x^{(۶)} = B_{GS}x^{(۵)} + b_{GS} = \begin{bmatrix} ۱/۰۰۰۰ \\ ۱/۰۰۰۰ \\ ۱/۰۰۰۰ \end{bmatrix}$$

پیاده‌سازیهای کامپیوتری

در پیاده‌سازیهای کامپیوتری واقعی مسلماً استفاده از روابط (۶-۱۰-۴) و (۶-۱۰-۷) اقتصادی است. استفاده از (۶-۱۰-۳) و (۶-۱۰-۶) مستلزم ذخیره‌سازی D ، L ، و U است. که یک ذخیره‌سازی زائد است.

۶-۱۰-۳ همگرایی روشهای تکراری

اغلب ساختن یک حدس خوب از تقریب اولیه $x^{(۱)}$ دشوار است. بنابراین داشتن شرایطی که همگرایی تکرار (۶-۱۰-۲) را برای هر انتخاب دلخواه از تقریب اولیه تضمین کند، مطلوب خواهد بود. چنین شرطی را در زیر به دست می‌آوریم.

قضیه ۶-۱۰-۱ قضیه همگرایی تکرار. تکرار

$$x^{(k+۱)} = Bx^{(k)} + c$$

به یک حد با یک انتخاب دلخواه از تقریب اولیه $x^{(۱)}$ ، همگرا می‌گردد اگر و فقط اگر ماتریس $B^k \rightarrow 0$ هنگامی که $k \rightarrow \infty$ ؛ یعنی B یک ماتریس همگرا باشد.

اثبات: معادله

$$x = Bx + c$$

و

$$x^{(k+۱)} = Bx^{(k)} + c$$

را تشکیل می‌دهیم. داریم:

$$x - x^{(k+۱)} = B(x - x^{(k)}) \quad (۶-۱۰-۸)$$

چون این رابطه برای هر مقدار k درست است، می‌توانیم بنویسیم:

$$x - x^{(k)} = B(x - x^{(k-1)}) \quad (۹-۱۰-۶)$$

با جایگذاری (۹-۱۰-۶) در (۸-۱۰-۶)، داریم:

$$x - x^{(k+1)} = B^2(x - x^{(k-1)}) \quad (۱۰-۱۰-۶)$$

با ادامه این فرایند k بار، می توانیم بنویسیم:

$$x - x^{(k+1)} = B^k(x - x^{(1)})$$

این نشان می دهد که $\{x^{(k)}\}$ به جواب x برای هر انتخاب دلخواه از $x^{(1)}$ همگرا می گردد اگر و فقط اگر $B^k \rightarrow 0$ هنگامی که $k \rightarrow \infty$ (مگر این که البته حدس اولیه $x^{(1)}$ همان جواب واقعی باشد). □

اکنون از فصل ۱ یادآور می شویم (قضیه ۱-۷-۴) که B یک ماتریس همگرا است اگر و فقط اگر شعاع طیفی B ، $\rho(B)$ ، کمتر از ۱ باشد. اکنون $\rho(B) = \max\{|\lambda_i|, i = 1, \dots, n\}$ ، که در آن λ_1 تا λ_n مقادیر ویژه B هستند، چون $\|B\| \geq |\lambda_i|$ به ازای هر i و هر نرم ماتریسی وابسته؛ بویژه داریم $\rho(B) \leq \|B\|$. بنابراین یک راه خوب برای این که ببینیم آیا B همگرا است این است که $\|B\|$ را با یک نرم مجموع سطری یا مجموع ستونی حساب کنیم و ببینیم که آیا این نرم کوچک تر از ۱ است یا خیر (توجه کنید که عکس آن درست نیست).

نتیجه قضیه ۶-۱۰-۱ را با مشاهده ای که هم اکنون انجام شد در زیر ترکیب می کنیم

شرایط برای همگرایی تکرار (۶-۱۰-۲)

یک شرط لازم و کافی برای همگرایی (۶-۱۰-۲) به ازای هر انتخاب دلخواهی از $x^{(1)}$ این است که $\rho(B) < 1$ باشد. یک شرط کافی این است که برای نرم دلخواهی $\|B\| < 1$ باشد.

اکنون این نتیجه را برای شناخت رده ای از ماتریسها که برای روشهای ژاکوبی و یا گاوس - سایدل به ازای هر انتخابی از تقریب اولیه $x^{(1)}$ همگرا می گردند، به کار می بریم.

روشهای ژاکوبی و یا گاوس - سایدل برای ماتریسهای غالب قطری

نتیجه ۶-۱۰-۱ اگر A غالب قطری سطری باشد، آنگاه روش ژاکوبی به ازای هر انتخاب دلخواه از تقریب اولیه $x^{(1)}$ همگرا می گردد.

اثبات: چون A غالب قطری سطری است، برطبق تعریف داریم

$$|a_{ii}| > \sum_{\substack{j=1 \\ i \neq j}}^n |a_{ij}|, \quad i = 1, 2, \dots, n \quad (11-10-6)$$

یادآور می‌شویم که ماتریس تکرار ژاکوبی B_J به صورت زیر می‌باشد

$$B_J = \begin{bmatrix} 0 & -\frac{a_{12}}{a_{11}} & \dots & \dots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & -\frac{a_{23}}{a_{22}} & \dots & -\frac{a_{2n}}{a_{22}} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & & & & -\frac{a_{n-1,n}}{a_{n-1,n-1}} \\ -\frac{a_{n1}}{a_{nn}} & \dots & \dots & -\frac{a_{n,n-1}}{a_{nn}} & 0 \end{bmatrix}$$

بنابراین از (11-10-6) نتیجه گرفته می‌شود که قدر مطلق مجموع سطری (یعنی مجموع سطری اختیار شده از قدر مطلق مقادیر) هر سطر کمتر از ۱ است، که به معنای آن است که

$$\|B_J\|_{\infty} < 1$$

بنابراین بر طبق قضیه ۱-۱۰-۶ نتیجه ۱-۱۰-۶ را داریم. □

نتیجه ۲-۱۰-۶ اگر A یک ماتریس غالب قطری سطری باشد، آنگاه روش گاوس - سایدل به ازای هر انتخاب دلخواه از تقریب اولیه $x^{(1)}$ همگرا می‌گردد.

اثبات: ماتریس تکرار گاوس سایدل به صورت زیر می‌باشد

$$B_{GS} = -(D + L)^{-1}U$$

فرض کنید λ یک مقدار ویژه این ماتریس و $x = (x_1, \dots, x_n)^T$ بردار ویژه نظیر آن باشد. آنگاه

$$B_{GS}x = \lambda x$$

یا

$$-Ux = (D + L)\lambda x$$

داریم

$$-\sum_{j=i+1}^n a_{ij}x_j = \lambda \sum_{j=1}^i a_{ij}x_j, \quad 1 \leq i \leq n$$

که می‌توان آن را به صورت زیر نوشت

$$\lambda a_{ii}x_i = -\lambda \sum_{j=1}^{i-1} a_{ij}x_j - \sum_{j=i+1}^n a_{ij}x_j, \quad 1 \leq i \leq n$$

فرض کنید x_k بزرگ‌ترین مؤلفه (دارای اندازه ۱) بردار x باشد. آنگاه از رابطه قبل داریم:

$$|\lambda| |a_{kk}| \leq |\lambda| \sum_{j=1}^{k-1} |a_{kj}| + \sum_{j=k+1}^n |a_{kj}| \quad (12-10-6)$$

یعنی

$$|\lambda| \left[|a_{kk}| - \sum_{j=1}^{k-1} |a_{kj}| \right] \leq \sum_{j=k+1}^n |a_{kj}| \quad (13-10-6)$$

یا

$$|\lambda| \leq \frac{\sum_{j=k+1}^n |a_{kj}|}{\left(|a_{kk}| - \sum_{j=1}^{k-1} |a_{kj}| \right)} \quad (14-10-6)$$

چون A غالب قطری سطری است

$$|a_{kk}| > \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}|$$

یا

$$|a_{kk}| - \sum_{j=1}^{k-1} |a_{kj}| > \sum_{j=k+1}^n |a_{kj}|$$

بنابراین از (۱۴-۱۰-۶)، نتیجه می‌گیریم که $|\lambda| < 1$ ؛ یعنی

$$\rho(B_{GS}) < 1$$

بنابراین از قضیه ۱-۱۰-۶ نتیجه ۲-۱۰-۶ را داریم.

□

یک تبصره در مورد ۶-۱۰-۱. معمولاً این درست است که بزرگتر بودن غالب قطری موجب سریع تر شدن همگرایی روش ژاکوبی می‌گردد. ولیکن مثالهای نقض ساده‌ای یافت می‌شوند که نشان می‌دهند که این همواره اتفاق نمی‌افتد.

مثال ساده 2×2 زیر جهت توضیح این مطلب در مرجع Golub and Van Loan (1989, 514) ارائه شده است. این مثال توسط ریچارد اس وارگا^۱ ارائه شده است.

$$A_1 = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & -\frac{3}{4} \\ -\frac{1}{2} & 1 \end{bmatrix}$$

به آسانی تحقیق می‌شود که $\rho(B_1)$ از A_1 بزرگتر از $\rho(B_2)$ از A_2 است.

اکنون همگرایی روش گاوس - سایدل را برای رده خیلی مهم دیگری از ماتریسها، یعنی ماتریسهای معین مثبت متقارن، مورد بحث قرار می‌دهیم.

روش گاوس - سایدل برای یک ماتریس معین مثبت متقارن ما نشان می‌دهیم که روش گاوس - سایدل، با یک انتخاب دلخواه $x^{(1)}$ ، برای یک ماتریس معین مثبت متقارن همگرا می‌گردد.

قضیه ۶-۱۰-۲ فرض کنید A یک ماتریس معین مثبت متقارن باشد. آنگاه روش گاوس - سایدل به‌ازای هر انتخاب دلخواه از تقریب اولیه $x^{(1)}$ همگرا می‌گردد.

اثبات: چون A متقارن است، داریم

$$A = L + D + L^T$$

بنابراین $B_{GS} = -(D + L)^{-1}L^T$. اکنون نشان می‌دهیم که $\rho(B_{GS}) < 1$.

فرض کنید λ یک مقدار ویژه ماتریس B_{GS} و u بردار ویژه نظیر آن باشد. پس

$$(D + L)^{-1}L^T u = \lambda u$$

یا

$$L^T u = \lambda(D + L)u$$

بنابراین

$$u^*L^T u = \lambda u^*(D + L)u \quad (u^* = (\bar{u})^T)$$

یا

$$u^*Au - u^*(L + D)u = \lambda u^*(D + L)u$$

(توجه کنید که $(L^T = A - (L + D))$ یا

$$u^*Au = (1 + \lambda)u^*(L + D)u \quad (15-10-6)$$

با گرفتن ترانواده مزدوج از هر دو طرف تساوی، داریم

$$u^*Au = (1 + \bar{\lambda})u^*(L^T + D^T)u \quad (16-10-6)$$

از (15-10-6) و (16-10-6) به دست می آوریم.

$$\left[\frac{1}{(1+\lambda)} + \frac{1}{(1+\bar{\lambda})} \right] u^*Au = u^*(L + D)u + u^*(L^T + D^T)u \quad (17-10-6)$$

$$= u^*(L + D + L^T + D^T)u \quad (18-10-6)$$

$$= u^*(A + D^T)u \quad (19-10-6)$$

$$= u^*(A + D)u > u^*Au \quad (20-10-6)$$

(توجه کنید که A معین مثبت است، بنابراین D نیز معین مثبت متقارن است؛ بنابراین $u^*Du > 0$.)

از تقسیم طرفین (20-10-6) بر $u^*Au (> 0)$ ، داریم

$$\frac{1}{(1+\lambda)} + \frac{1}{(1+\bar{\lambda})} > 1$$

یا

$$\frac{2+\lambda+\bar{\lambda}}{(1+\lambda)(1+\bar{\lambda})} > 1 \quad (21-10-6)$$

فرض کنید که $\lambda = \alpha + i\beta$. آنگاه $\bar{\lambda} = \alpha - i\beta$. سپس از (21-10-6) داریم

$$\frac{2(1+\alpha)}{(1+\alpha)^2 + \beta^2} > 1$$

□ از این رابطه نتیجه می شود که $1 < \alpha^2 + \beta^2$ ؛ یعنی $\rho(B_{GS}) < 1$ زیرا $\sqrt{\alpha^2 + \beta^2} = |\lambda|$.

نرخهای همگرایی و یک مقایسه بین روشهای گاوس سایدل و ژاکوبی

ملاحظه کردیم که برای ماتریسهای غالب قطری هر دو روش ژاکوبی و گاوس سایدل برای هر بردار دلخواه $x^{(1)}$ همگرا می گردند. سؤالی که به طور طبیعی مطرح می شود این است که: آیا این مطلب برای برخی ماتریسهای دیگر نیز درست است. همچنین هنگامی که دو روش همگرا می گردند، سؤال دیگری که مطرح می شود این است: کدام یک سریع تر همگرا می گردند؟

نتیجه ای از قضیه معروف استین - رزنبرگ^۱ (برای صورت و اثبات این قضیه مرجع Varga(1962) را ملاحظه کنید) این است که اگر A یک ماتریس دارای عناصر قطری مثبت و عناصر غیر قطری منفی باشد، آنگاه

۱. روش ژاکوبی و گاوس سایدل هر دو همگرا و یا واگرا می گردند.

۲. هنگامی که هر دو روش همگرا می گردند، روش گاوس - سایدل سریعتر از ژاکوبی همگرا می گردد.

توجه کنید که نرخ همگرایی مجانبی به صورت زیر تعریف می شود.

تعریف ۶-۱۰-۱ مقدار $\text{Imp}(B)$ - نرخ مجانبی همگرایی تکرار (۶-۱۰-۲) نامیده می شود.

تبصره ها: متأسفانه در حالت کلی چنین گزاره هایی را در مورد همگرایی و نرخهای همگرایی مجانبی دو روش تکراری نمی توان بیان کرد. در حقیقت مثالهایی وجود دارند که در آنها یک روش همگرا می گردد اما دیگری واگرا می گردد (مثال زیر را مشاهده کنید). ولیکن هنگامی که دو روش گاوس - سایدل و ژاکوبی همگرا می گردند، به واسطه نیاز به ذخیره سازی کمتر و نرخهای همگرایی مجانبی، روش گاوس - سایدل می باید بر روش ژاکوبی ترجیح داده شود.

مثال ۶-۱۰-۳ مثال زیر نشان می دهد که روش ژاکوبی می تواند همگرا شود وقتی که روش گاوس - سایدل همگرا نگردد.

$$A = \begin{bmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{bmatrix}$$

$$b = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$

$$B_J = \begin{bmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{bmatrix}$$

$$B_{GS} = \begin{bmatrix} 0 & -2 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 2 \end{bmatrix}$$

$$\rho(B_{GS}) = 2$$

$$\rho(B_J) = 6/7815 \times 10^{-6}$$

تعداد کمی تکرار با گاوس - سایدل

تکرارها نشان می دهند که گاوس - سایدل به طور واضح و اگر می گردد.

$$x^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad x^{(2)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad x^{(3)} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$$

$$x^{(4)} = \begin{bmatrix} 7 \\ -8 \\ 7 \end{bmatrix}, \quad x^{(5)} = \begin{bmatrix} 31 \\ -36 \\ 15 \end{bmatrix}$$

جواب واقعی برابر است با

$$x = \begin{bmatrix} 7 \\ -4 \\ -1 \end{bmatrix}$$

از طرف دیگر، با روش ژاکوبی فقط با دو تکرار همگرایی را داریم:

$$x^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$x^{(2)} = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$

$$x^{(3)} = \begin{bmatrix} 7 \\ -4 \\ -1 \end{bmatrix}$$

۴-۱۰-۶ روش فوق تخفیف متوالی (SOR)

هنگامی که $\rho(B_{GS})$ نزدیک به واحد می باشد روش گاوس - سایدل به طور ناامیدکننده ای آهسته است. ولیکن نرخ همگرایی گاوس - سایدل می تواند، در حالت های خاصی، با وارد کردن یک پارامتر ω ، معروف به پارامتر تخفیف، بهبود یابد. اگر $\omega > 1$ باشد تکرار گاوس - سایدل اصلاح شده زیر به تکرار فوق تخفیف متوالی یا به طور خلاصه تکرار SOR معروف است.

تکرار SOR

$$x_i^{(k+1)} = \frac{\omega}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right] + (1-\omega) x_i^k,$$

$$i = 1, 2, \dots \quad (22-10-6)$$

از (۲۲-۱۰-۶) مطالب زیر را ذکر می کنیم:

۱. هنگامی که $\omega = 1$ تکرار SOR به تکرار گاوس - سایدل کاهش می یابد.
 ۲. هنگامی که $\omega > 1$ ، در محاسبه $(k+1)$ امین تقریب وزن بیشتری به جاری ترین مقدار داده می شود تا هنگامی که $\omega < 1$ ، با این امید که همگرایی سریعتر خواهد بود.
- نماد ماتریسی تکرار SOR عبارت است از:

$$x^{(k+1)} = (D + \omega L)^{-1} [(1 - \omega)D - \omega U] x^{(k)} + \omega(D + \omega L)^{-1} b,$$

$$k = 1, 2, \dots \quad (23-10-6)$$

(توجه کنید که چون $a_{ii} \neq 0$ ، $i = 1, 2, \dots$ ، ماتریس $(D + \omega L)$ نامنفرد است.)

ماتریس $(D + \omega L)^{-1} [(1 - \omega)D - \omega U]$ ماتریس SOR نامیده می شود و توسط B_{SOR} نمایش داده می شود. به طور مشابه بردار $(D + \omega L)^{-1} b$ به b_{SOR} نمایش داده می شود.

ماتریس SOR و بردار SOR

$$B_{SOR} = (D + \omega L)^{-1} [(1 - \omega)D - \omega U]$$

$$b_{SOR} = \omega(D + \omega L)^{-1} b$$

در نمادگذاری ماتریسی الگوریتم تکرار SOR به صورت زیر است

الگوریتم ۳-۱۰-۶ روش فوق تخفیف متوالی

گام ۱: $x^{(1)}$ را انتخاب کنید

گام ۲: برای $k = 1, 2, \dots$ تا برقراری یک معیار توقف انجام دهید.

$$x^{(k+1)} = B_{SOR} x^{(k)} + b_{SOR}$$

یا $x^{(k+1)}$ را با استفاده از (۶-۱۰-۲۲) محاسبه کنید.

تذکر MATCOM: الگوریتم ۳-۱۰-۶ در برنامه SUCOV در MATCOM پیاده‌سازی شده است.

مثال ۴-۱۰-۶ فرض کنید A و b همانند مثال ۲-۱۰-۶ باشند. فرض کنید

$$x_1^{(1)} = x_2^{(1)} = x_3^{(1)} = 0, \quad \omega = 1/2$$

آنگاه

$$B_{SOR} = \begin{bmatrix} -0./2000 & -0./2400 & -0./2400 \\ -0./480 & -0./1424 & -0./1824 \\ -0./365 & -0./918 & -0./986 \end{bmatrix}, \quad b_{SOR} = \begin{bmatrix} 1/6800 \\ 1/2768 \\ -0./9704 \end{bmatrix}$$

: $k = 1$

$$x^{(2)} = \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{bmatrix} = B_{SOR} x^{(1)} + b_{SOR} = \begin{bmatrix} 1/6800 \\ 1/2768 \\ -0./9704 \end{bmatrix}$$

: $k = 2$

$$x^{(3)} = \begin{bmatrix} x_1^{(3)} \\ x_2^{(3)} \\ x_3^{(3)} \end{bmatrix} = B_{SOR} x^{(2)} + b_{SOR} = \begin{bmatrix} 0./8407 \\ -0./9986 \\ 1/0531 \end{bmatrix}$$

: $k = 3$

$$x^{(4)} = \begin{bmatrix} x_1^{(4)} \\ x_2^{(4)} \\ x_3^{(4)} \end{bmatrix} = B_{SOR} x^{(3)} + b_{SOR} = \begin{bmatrix} 1/0266 \\ -0./9811 \\ -0./9875 \end{bmatrix}$$

انتخاب ω در همگرایی تکرار SOR

طبیعی است که برد ω که برای آن SOR همگرا می‌گردد و انتخاب بهینه ω مورد توجه باشد. برای پی بردن به این مطلب ابتدا نتیجه مهم زیر را که اثر ویلیام کاهان^۱ می‌باشد، ثابت می‌کنیم.

قضیه ۶-۱۰-۳ کاهان. اگر ω خارج بازه $(0, 2)$ قرار گیرد، آنگاه تکرار SOR نمی‌تواند برای یک تقریب اولیه دلخواه همگرا گردد.

اثبات: یادآور می‌شویم که ماتریس تکرار SOR، B_{SOR} ، به صورت

$$B_{SOR} = (D + \omega L)^{-1}[(1 - \omega)D - \omega U]$$

ارائه می‌شود که در آن $A = L + D + U$.

ماتریس $(D + \omega L)^{-1}$ یک ماتریس پایین مثلثی با عناصر قطری $1/a_{ii}$ ، $i = 1, \dots, n$ ، و ماتریس $(1 - \omega)D - \omega U$ یک ماتریس بالامثلثی با عناصر قطری $(1 - \omega)a_{ii}$ ، $i = 1, \dots, n$ ، است. بنابراین

$$\det(B_{SOR}) = (1 - \omega)^n$$

چون دترمینان یک ماتریس برابر حاصلضرب مقادیر ویژه آن می‌باشد، نتیجه می‌گیریم که

$$\rho(B_{SOR}) \geq |1 - \omega|$$

که در آن $\rho(B_{SOR})$ شعاع طیفی ماتریس B_{SOR} می‌باشد.

چون بر طبق قضیه ۶-۱۰-۱، برای همگرایی هر روش تکراری باید شعاع طیفی ماتریس تکرار کوچک‌تر از ۱ باشد، نتیجه می‌گیریم که ω باید در بازه $(0, 2)$ قرار گیرد. □

قضیه بعد معروف به قضیه استروفسکی-ریچ^۲، نشان می‌دهد که شرط قبل در حالتی که ماتریس A متقارن و معین مثبت باشد، شرط کافی نیز هست.

قضیه برای تکرار گاوس-سایدل ($\omega = 1$) در سال ۱۹۴۹ توسط ریچ ثابت شد و پس از آن توسط استروفسکی در سال ۱۹۵۴ گسترش یافت.

قضیه را بدون اثبات بیان می‌کنیم. برای اثبات مرجع (Varga(1962 بخش ۳-۴ و یا مرجع Ortega(1990, 123) را ملاحظه کنید. قضیه استروفسکی-ریچ تعمیمی از قضیه ۶-۱۰-۳ برای ماتریسهای معین مثبت متقارن است.

قضیه ۶-۱۰-۴ استروفسکی-ریچ. فرض کنید A یک ماتریس معین مثبت متقارن باشد و فرض کنید $0 < \omega < 2$. آنگاه روش SOR برای هر انتخاب دلخواه $x^{(1)}$ همگرا خواهد بود.

انتخاب بهینه ω در تکرار SOR همگرا

اکنون به سؤال انتخاب بهینه ω برمی گردیم. باز هم برای هر ماتریس دلخواه هیچ معیاری تاکنون تدوین نشده است. ولیکن برای یک رده از ماتریسهای خیلی مفید که در کاربردهای عملی زیادی مطرح می شوند، یک معیار مشخص گردیده است. ماتریسهای این رده ماتریسهای به طور سازگار مرتب شده نامیده می شوند.

تعریف ۶-۱۰-۲ یک ماتریس $A = L + D + U$ به طور سازگار مرتب شده نامیده می شود اگر مقادیر ویژه ماتریس

$$C(\mu) = -D^{-1} \left(\frac{1}{\mu} L + \mu U \right)$$

به μ ، $\mu \neq 0$ ، بستگی نداشته باشند.

یانگ^۱ یک ماتریس به طور سازگار مرتب شده را طور دیگری تعریف کرده است (Young(1971)). تعریفی که او ارائه داده است به مقادیر ویژه بستگی ندارد.

تعریف ۶-۱۰-۳ ماتریس A ۲-دوری است اگر یک ماتریس جایگشت P یافت شود به قسمی که

$$PAP^T = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

که در آن A_{11} و A_{22} قطری هستند.

دیوید یانگ چنین ماتریسی را به عنوان یک ماتریس باخاصیت A تعریف کرده است. این تعریف می تواند به ماتریسهای بلوکی که در آن ماتریسهای قطری A_{11} و A_{22} ماتریسهای قطری بلوکی هستند تعمیم داده شود. چنین ماتریسهایی را ماتریسهای ۲-دوری بلوکی می نامیم. یک مثال معروف از یک ماتریس ۲-دوری بلوکی به طور سازگار مرتب شده ماتریس سه قطری بلوکی زیر است.

$$A = \begin{bmatrix} T & -I & & 0 \\ -I & \ddots & \ddots & \\ & \ddots & \ddots & -I \\ 0 & & -I & T \end{bmatrix}$$

که در آن

$$T = \begin{bmatrix} 4 & -1 & \dots & 0 \\ -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ 0 & \dots & -1 & 4 \end{bmatrix}$$

یادآور می‌شویم که این ماتریس از گسسته‌سازی معادله پواسون

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = f$$

بر روی مربع واحد به دست می‌آید. در حقیقت می‌توان نشان داد (تمرین ۳۹ ب) که هر ماتریس سه قطری بلوکی با بلوکهای قطری نامنفرد به طور سازگار مرتب شده و ۲-دوری است.

قضیه مهم و معروف زیر در مورد انتخاب بهینه برای ماتریسهای به طور سازگار مرتب شده اثر دیوید یانگ ((David Young (1971)) می‌باشد.

قضیه ۶-۱۰-۵ یانگ (Young). فرض کنید A به طور سازگار مرتب شده و ۲-دوری و با عناصر قطری مخالف صفر باشد. آنگاه

$$\rho(B_{GS}) = (\rho(B_J))^2$$

بعلاوه اگر مقادیر ویژه B_J حقیقی باشند و $\rho(B_J) < 1$ ، آنگاه انتخاب بهینه برای ω به منظور تولید کوچک‌ترین شعاع طیفی در SOR، که با ω_{opt} نشان داده می‌شود، به صورت زیر می‌باشد.

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - \rho(B_J)^2}}$$

$$\rho(B_{SOR}) = \omega_{opt} - 1 \quad \text{و}$$

به عنوان یک نتیجه مهم نتیجه زیر را به دست می‌آوریم.

نتیجه ۶-۱۰-۳ برای ماتریسهای ۲-دوری و به طور سازگار مرتب شده اگر روش ژاکوبی همگرا گردد، روش گاوس-سایدل نیز همگرا می‌گردد و روش گاوس-سایدل دو برابر سریع‌تر از روش ژاکوبی همگرا می‌گردد.

مثال ۶-۱۰-۵

$$A = \begin{bmatrix} 4 & -1 & 0 & -1 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 \\ -1 & 0 & 0 & 4 & -1 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 \\ 0 & 0 & -1 & 0 & -1 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

مقادیر ویژه B_J ، عبارتند از، $0/1036$ ، $0/2500$ ، $-0/1036$ ، $-0/2500$ ، $0/6036$ و $-0/6036$.

$$\rho(B_J) = 0/6036$$

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - (0/6036)^2}} = 1/1128$$

$$\rho(B_{GS}) = 0/3643$$

با پنج تکرار روش SOR با ω_{opt} و با بردار آغازین

$$x_{SOR}^{(1)} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

به جواب واقعی زیر (با چهار رقم با معنی) همگرا می‌گردد

$$x_{SOR}^{(5)} = \begin{bmatrix} 0/2948 \\ 0/0932 \\ 0/0282 \\ 0/0861 \\ 0/0497 \\ 0/0195 \end{bmatrix}$$

با همان بردار آغازین $x^{(1)}$ روش گاوس - سایدل به ۹ تکرار نیاز دارد. (امتحان کنید!) همچنین ملاحظه خواهید کرد که ۱۸ تکرار برای روش ژاکوبی لازم خواهد بود.

مقایسه نرخهای همگرایی روشهای SOR و گاوس - سایدل

قضیه زیر که همچنین اثر یانگ ((Young (1971)) می‌باشد نرخهای همگرایی روش گاوس - سایدل و روش SOR با ضریب تخفیف بهینه، ω_{opt} ، را برای ماتریسهای به طور سازگار مرتب شده با یکدیگر

مرتبط می‌سازد.

قضیه ۶-۱۰-۶ یانگ. فرض کنید A یک ماتریس به طور سازگار مرتب شده با عناصر قطری مخالف صفر باشد. فرض کنید مقادیر ویژه B_J همگی حقیقی باشند و $\rho(B_J) < 1$. فرض کنید R_{GS} و R_{SOR} به ترتیب نرخهای مجانبی همگرایی روشهای گاوس - سایدل و SOR با عامل تخفیف بهینه ω_{opt} را نمایش دهند، آنگاه

$$\gamma \rho(B_J) R_{GS}^{1/2} \leq R_{SOR} \leq R_{GS} + \gamma [R_{GS}]^{1/2}$$

دومین نامساوی برقرار است اگر $R_{GS} \leq 3$ باشد.

تبصره: قضیه ۶-۱۰-۶ اساساً بیان می‌کند که روش SOR با عامل تخفیف بهینه خیلی سریع‌تر از روش گاوس - سایدل همگرا می‌گردد (همان طور که انتظار می‌رود) هنگامی که نرخ همگرایی مجانبی روش گاوس - سایدل کوچک است.

مثال ۶-۱۰-۶ مجدداً ماتریس حاصل از گسسته‌سازی معادله پواسون با اندازه شبکه h (ماتریس $(6-3-31)$) را در نظر بگیرید. برای این ماتریس به راحتی تحقیق می‌شود که

$$\rho(B_J) = \cos(\pi h)$$

بعلاوه می‌دانیم که $\rho(B_{GS}) = (\rho(B_J))^2$. بنابراین

$$\begin{aligned} R_{GS} &= \gamma(-\ln \cos \pi h) \\ &= \gamma \frac{\pi^2 h^2}{2} + O(h^4) \\ &= \pi^2 h^2 + O(h^4) \end{aligned}$$

برای h کوچک R_{GS} کوچک است و توسط نامساوی دوم قضیه ۶-۱۰-۶ داریم:

$$R_{SOR} \approx \gamma R_{GS}^{1/2} \approx \gamma \pi h$$

و

$$\frac{R_{SOR}}{R_{GS}} \approx \frac{\gamma}{(\pi h)}$$

بنابراین هنگامی که h کوچک است نرخ همگرایی مجانبی روش SOR با عامل تخفیف بهینه خیلی سریع‌تر از روش گاوس - سایدل است. برای مثال، هنگامی که $h = \frac{1}{50}$ آنگاه

$$\frac{R_{SOR}}{R_{GS}} \equiv \frac{100}{\pi} = 31/8$$

بنابراین در این حالت روش SOR ، 31 برابر سریعتر از روش گاوس - سایدل همگرا می‌گردد. بعلاوه نرخ همگرایی هنگامی که h کاهش می‌یابد بزرگ‌تر می‌شود؛ هنگامی که h خیلی خیلی کوچک می‌باشد بهبودی حقیقتاً قابل توجه است.

۶-۱۰-۵ روش گرادیان مزدوج

در اینجا یک روش معروف به روش گرادیان مزدوج (CG) را برای حل یک دستگاه معین مثبت متقارن

$$Ax = b$$

توضیح می‌دهیم.

روش در اصل توسط هستنز و استایفل (Hestenes and Stiefel (1952)) ارائه شده است و اکنون به طور گسترده‌ای برای حل دستگاه‌های معین مثبت متقارن تنک و بزرگ استفاده می‌شود. از لحاظ نظری این روش یک روش مستقیم است. این روش بر نتیجه معروف زیر مربوط به بهینه سازی مبتنی است (Luenberger (1973)).

قضیه ۶-۱۰-۷ اگر A یک ماتریس معین مثبت متقارن حقیقی باشد، آنگاه حل

$$Ax = b$$

معادل با می‌نیم کردن تابع درجه دوم

$$\Phi(x) = \frac{1}{2}x^T Ax - x^T b$$

می‌باشد. بعلاوه می‌نیم مقدار تابع $\Phi(x)$ برابر $-\frac{1}{2}b^T A^{-1}b$ می‌باشد و با انتخاب $x = A^{-1}b$ به دست می‌آید.

تعداد زیادی روشهای تکراری در نوشتجات بهینه‌سازی برای حل این دستگاه می‌نیم سازی وجود دارند (Luenberger (1973)). در این روشهای تکراری تقریبهای متوالی x_k برای بردار جواب x به‌طور بازگشتی محاسبه می‌شوند؛

$$x_{k+1} = x_k + \alpha_k p_k$$

که در آن بردارهای $\{p_k\}$ بردارهای جهت هستند و اسکالره‌های α_k برای می‌نیم سازی $\Phi(x)$ در جهت‌های p_k انتخاب می‌شوند؛ یعنی برای می‌نیم کردن تابع $\Phi_\alpha(x_k + \alpha p_k)$ انتخاب می‌شود. می‌توان نشان داد (Ortega (1988a, 186)) که می‌نیم کردن رخ می‌دهد اگر انتخاب کنیم

$$\alpha = \alpha_k = \frac{p_k^T (b - Ax_k)}{p_k^T A p_k}$$

$$= p_k^T r_k / p_k^T A p_k$$

$$r_k = b - Ax_k \text{ که در آن}$$

(در الگوریتم ۶-۱۰-۴ از یک عبارت کمی متفاوت ولیکن معادل برای α استفاده خواهیم کرد.)

انتخاب بردارهای جهت

سؤال بعدی این است که چگونه بردارهای جهت p_k را انتخاب کنیم. روش گرادیان مزدوج (CG) روشی است که به طور خودکار بردارهای جهت را تولید می‌کند. بردار جهت لازم در هر گام در گام قبلی تولید می‌شود، علاوه بر این بردارهای جهت p_k خاصیت قابل توجهی دارند که

$$(p_j)^T A p_i = 0, \quad 0 \leq j < i, \quad i = 1, 2, \dots, n-1$$

یعنی، این بردارها نسبت به ضرب داخلی $x^T A y$ تعریف شده توسط A متعامد هستند. بردارهای جهت p_k که در خاصیت فوق صدق می‌کنند بردارهای مزدوج نامیده می‌شوند.

الگوریتم ۶-۱۰-۴ الگوریتم گرادیان مزدوج پایه‌ای (روش گرادیان مزدوج کلاسیک)

step 1: Choose x_0 and ε . Set $p_0 = r_0 = b - Ax_0$

step 2: For $i = 1, 2, \dots$, do

$$w = A p_i$$

$$\alpha_i = \|r_i\|_Y^2 / p_i^T w$$

$$x_{i+1} = x_i + \alpha_i p_i$$

$$r_{i+1} = r_i - \alpha_i w$$

Test for convergence: If $\|r_{i+1}\|_Y^2 \geq \varepsilon$, continue.

$$\beta_i = \|r_{i+1}\|_Y^2 / \|r_i\|_Y^2$$

$$p_{i+1} = r_{i+1} + \beta_i p_i$$

تذکر MATCOM: الگوریتم ۶-۱۰-۴ در برنامه CONGRAD در MATCOM پیاده‌سازی شده

است.

مثال ۶-۱۰-۷

$$A = \begin{bmatrix} 5 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 5 \end{bmatrix}, \quad b = \begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix}$$

$$x_0 = (0, 0, 0)^T, \quad p_0 = r_0 = b - Ax^{(0)} = b = \begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix}$$

: $i = 0$

$$w = Ap_0 = \begin{bmatrix} 49 \\ 49 \\ 49 \end{bmatrix}$$

$$\alpha_0 = \frac{\|r_0\|_2^2}{p_0^T w} = 0.1429$$

$$x_1 = x_0 + \alpha_0 p_0 = \begin{bmatrix} 1/0.1429 \\ 1/0.1429 \\ 1/0.1429 \end{bmatrix}$$

$$r_1 = r_0 - \alpha_0 w = \begin{bmatrix} -0.1429 \\ -0.1429 \\ -0.1429 \end{bmatrix}$$

$$\beta_0 = 9 \times 10^{-8}$$

$$p_1 = r_1 + \beta_0 p_0 = \begin{bmatrix} -0.1429 \\ -0.1429 \\ -0.1429 \end{bmatrix}$$

: $i = 1$

$$w = Ap_1 = \begin{bmatrix} -0.1429 \\ -0.1429 \\ -0.1429 \end{bmatrix}$$

$$\alpha_1 = 0.1429, \quad x_2 = x_1 + \alpha_1 p_1 = \begin{bmatrix} 1/0.1429 \\ 1/0.1429 \\ 1/0.1429 \end{bmatrix}$$

همگرایی

در صورتی که خطای گرد کردن وجود نداشته باشد روش گرادیان مزدوج در حداکثر n تکرار همگرا می‌گردد. در حقیقت می‌توان نشان داد که در هر گام خطا کاهش می‌یابد. بویژه می‌توان ثابت کرد (Ortega (1988a, 277)) که قضیه ۶-۱۰-۸ برقرار است.

قضیه ۶-۱۰-۸ رابطه

$$\|x - x_k\|_2 < \|x - x_{k-1}\|_2$$

برقرار است مگر $x_{k-1} = x$. در اینجا x جواب واقعی است.

ولیکن همگرایی می‌تواند بر اثر بدوضعی A بی‌نهایت آهسته باشد. این مطلب را می‌توان از روی نتیجه زیر ملاحظه کرد.

همگرایی روش گرادیان مزدوج

تعریف می‌کنیم $\|s\|_A = \sqrt{s^T A s}$. آنگاه برآوردی از نرخ همگرایی عبارت است از

$$\|x_k - x\|_A \leq \alpha^k \|x_0 - x\|_A$$

که در آن

$$\alpha = (\sqrt{\kappa} - 1) / (\sqrt{\kappa} + 1)$$

و

$$\kappa = \text{Cond}(A) = \|A\|_2 \|A^{-1}\|_2 = \lambda_n / \lambda_1$$

در اینجا λ_1 و λ_n بزرگ‌ترین و کوچک‌ترین مقادیر ویژه ماتریس معین مثبت متقارن A هستند (توجه کنید مقادیر ویژه A مثبت هستند).

تذکر: $\alpha = 0$ اگر $\text{Cond}(A) = 1$ باشد. هنگامی که $\alpha \rightarrow 1$, $\text{Cond}(A) \rightarrow \infty$. بنابراین برای

$\text{Cond}(A)$ بزرگ‌تر، نرخ همگرایی آهسته‌تر است. (برای اثبات (Luenberger (1973, 187)) را ملاحظه کنید).

پیش شرط سازی

چون یک عدد وضعیت بزرگ A سرعت همگرایی روش گرادیان مزدوج را کم می‌کند، طبیعی است ببینیم که آیا می‌توان عدد وضعیت A را قبل از آن که روش به کار برده شود بهبود بخشید؛

در این حالت ما قادر خواهیم بود که روش گرادیان مزدوج را در مورد یک دستگاه با پیش شرط به کار ببریم. در حقیقت استفاده از یک پیش شرط خوب نرخ همگرایی را اساساً سرعت می بخشد.

یک ایده اساسی از پیش شرط سازی این است که یک ماتریس نامنفرد S بیابیم به قسمی که $\text{Cond}(\tilde{A}) < \text{Cond}(A)$ ، که در آن $\tilde{A} = SAS^T$. وقتی که چنین S یی پیدا شد، آنگاه می توان $\tilde{A}\tilde{x} = \tilde{b}$ را حل کرد که در آن $\tilde{x} = (S^{-1})^T x$ و $\tilde{b} = Sb$ ، و سپس بردار x را می توان از روی \tilde{x} و با استفاده از $x = S^T \tilde{x}$ به دست آورد.

برای سادگی ماتریس S معمولاً توسط

$$(S^T S)^{-1} = M$$

تعریف می شود. توجه کنید که M معین مثبت متقارن است و یک پیش شرط نامیده می شود.

الگوریتم ۶-۱۰-۵ روش گرادیان مزدوج با پیش شرط (PCG)

Find a preconditioner M .

Choose x_0 and ε .

Set $r_0 = b - Ax_0$. $p_0 = y_0 = M^{-1}r_0$.

For $i = 0, 1, 2, 3, \dots$ do

(a) $w = A p_i$.

(b) $\alpha_i = y_i^T r_i / p_i^T w$

(c) $x_{i+1} = x_i + \alpha_i p_i$

(d) $r_{i+1} = r_i - \alpha_i w$

(e) $y_{i+1} = M^{-1}r_{i+1}$; that is, solve $M y_{i+1} = r_{i+1}$

(f) Test for convergence: If $r_{i+1}^T y_{i+1} \geq \varepsilon$, then if $\|r_{i+1}\| \geq \varepsilon$ continue.

(g) $\beta_i = y_{i+1}^T r_{i+1} / y_i^T r_i$

(h) $p_{i+1} = y_{i+1} + \beta_i p_i$

تذکر: اگر $M = I$ ، آنگاه روش گرادیان مزدوج با پیش شرط به روش گرادیان مزدوج پایه ای کاهش می یابد.

بنابراین سؤال بعد این است که چگونه یک پیش شرط M پیدا کنیم. چندین امکان در نوشتجات ارائه گردیده‌اند.

دو روش از میان آنها عبارتند از:

۱. پیش شرط سازی چندجمله‌ای

۲. تجزیه چولسکی ناقص (ICF)

ICF را در زیر توضیح خواهیم داد. برای توصیف پیش شرط سازی چندجمله‌ای مرجع Ortega (1988a, 206-208) را ملاحظه کنید.

تجزیه چولسکی ناقص

چون A معین مثبت متقارن است، تجزیه

$$A = LDL^T$$

را می‌پذیرد که در آن L یک ماتریس پایین مثلثی واحد و D قطری است (قضیه ۶-۴-۱). اگر A تنک باشد، L معمولاً کمتر از A تنک می‌باشد، چون پر شدن می‌تواند رخ دهد. ولیکن می‌توان پر شدن را نادیده گرفت و آن چه را که معروف به تجزیه چولسکی ناقص می‌باشد، به دست آورد.

قاعده کلی پایه‌ای تجزیه چولسکی ناقص $A = (a_{ij})$ عبارت است از:

اگر $a_{ij} \neq 0$ ، l_{ij} را محاسبه می‌کنیم

اگر $a_{ij} = 0$ ، قرار می‌دهیم $l_{ij} = 0$

الگوریتم ۶-۱۰-۶ تجزیه چولسکی ناقص (Ortega (1988a, 212))

$$\text{Set } l_{11} = \sqrt{a_{11}}$$

For $i = 2, \dots, n$ do

For $j = 1, 2, \dots, i - 1$ do

If $a_{ij} = 0$, then $l_{ij} = 0$ else

$$l_{ij} = \frac{1}{l_{ij}} \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk} \right)$$

$$l_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2}$$

تبصره: الگوریتم ۶-۱۰-۶ به محاسبه ریشه‌های دوم نیاز دارد. بنابراین ممکن است به طور کامل انجام نشود. ولیکن می‌توان یک تجزیه LDL^T ناقص بدون پر شدن از A به دست آورد که از محاسبه ریشه‌های دوم پرهیز نماید.

الگوریتم ۷-۱۰-۶ LDL^T ناقص بدون پر شدن

$$\text{Set } d_{11} = a_{11}$$

For $i = 2, \dots, n$ do

For $j = 1, 2, \dots, i - 1$ do

$$\text{If } a_{ij} = 0, l_{ij} = 0$$

else

$$l_{ij} = \left[a_{ij} - \sum_{k=1}^{j-1} l_{ik} d_{kk} l_{jk} \right] / d_{jj}.$$

$$d_{ii} = a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2 d_{kk}$$

تذکرات $MATCOM$: الگوریتمهای ۶-۱۰-۶ و ۷-۱۰-۶ به ترتیب در برنامه‌های $ICHOLES$ و $NICHOL$ در $MATCOM$ پیاده‌سازی شده‌اند.

استفاده از تجزیه چولسکی ناقص در پیش شرط سازی توجه کنید که الگوریتم تجزیه چولسکی ناقص از لحاظ ریاضی تجزیه‌ای از A را به شکل زیر ارائه می‌دهد.

$$A = LL^T + R$$

که در آن $R \neq 0$. چون بهترین انتخاب برای یک پیش شرط M خود ماتریس A می‌باشد، بعد از این که ماتریس L توسط تجزیه چولسکی ناقص به دست آمد، پیش شرط M به صورت زیر اختیار می‌شود:

$$M = LL^T$$

در الگوریتم گرادیان مزدوج با پیش شرط (الگوریتم ۵-۱۰-۶) لازم است که یک دستگاه معین مثبت متقارن به شکل:

$$My = r$$

در هر تکرار با M به عنوان ماتریس ضرایب حل شود (گام e). چون $M = LL^T$ ، این معادل با حل

$$Lx = r, \quad L^T y = x$$

می باشد. چون ماتریس ضرایب در هر گام یکسان است، تجزیه چولسکی ناقص فقط یک بار انجام خواهد شد. اگر از LDL^T ناقص بدون پرس شدن استفاده شود، آنگاه از نظر ریاضی داریم:

$$A = LDL^T + R$$

در این حالت پیش شرط M را به شکل

$$M = LDL^T$$

اختیار می کنیم. سپس در هر تکرار الگوریتم گرادیان مزدوج با پیش شرط نیاز به حل یک دستگاه به شکل

$$My = r$$

داریم که معادل با $Lx = r$ ، $Dz = x$ و $L^T y = z$ می باشد. مجدداً L و D باید فقط یک بار محاسبه شوند.

۶-۱۰-۶ فرایند آرنولدی و GMRES

در چند سال اخیر، روشی که GMRES نامیده می شود به طور وسیعی توجه متخصصین جبر خطی عددی را در زمینه حل دستگاههای تنک و بزرگ به خود جلب کرده است. روش مبتنی بر طرح کلاسیکی است که اثر آرنولدی ((Arnoldi, 1951) بوده و یک پایه یکا متعامد از یک فضا که زیر فضای کرلیف نامیده می شود، می سازد. زیر فضای کرلیف از بعد m و تعریف شده توسط یک ماتریس $n \times n$ مانند A و یک بردار v_1 به صورت زیر ارائه می شود

$$K_m(A, v_1) = \text{span}\{v_1, Av_1, \dots, A^{m-1}v_1\}$$

روش آرنولدی می تواند فقط با استفاده از ضرب ماتریس بردار پیاده سازی شود، و بنابراین برای ماتریسهای تنک مناسب است، زیرا عناصر صفر حفظ می شوند.

ایده اصلی در پشت سر حل یک مسأله بزرگ و تنک با استفاده از روش آرنولدی تصویر کردن مسأله بر روی زیر فضای کرلیف با بعد $m < n$ با استفاده از پایه متعامد ساخته شده توسط روش، حل مسأله m بعدی با استفاده از یک راه استاندارد، و سپس بازیافتن جواب مسأله اصلی از روی جواب مسأله

تصویر شده می باشد.

اکنون اصول اصلی روش آرنولدی را به طور خلاصه بیان می کنیم، و به دنبال آن یک توصیف الگوریتمی از GMRES را بیان می نمایم.

Step 1 : Start: Choose a vector of norm 1 and an integer $m \leq n$

Step 2 : Iterate: For $j = 1, 2, \dots, m$

$$h_{ij} = (v_i^T A v_j), i = 1, 2, \dots, j \text{ do}$$

$$\tilde{v}_{j+1} = A v_j - \sum_{i=1}^j h_{i,j} v_i$$

$$h_{j+1,j} = \|\tilde{v}_{j+1}\|_2$$

and

$$v_{j+1} = \tilde{v}_{j+1} / h_{j+1,j}$$

تذکرها:

۱. اسکالرها h_{ij} به قسمی انتخاب شده اند که بردارهای v_j یکا متعامد باشند.
۲. فرض کنید V_m ماتریس $n \times m$ یی باشد که ستون j ام آن بردار ستونی v_j باشد؛ یعنی $V_m \equiv (v_1, \dots, v_m)$. آنگاه بردارها در V_m تشکیل یک پایه یکا متعامد از زیر فضای کرلیف $\kappa_m(A, v_1)$ را می دهند.
۳. ماتریس \tilde{H}_m را به صورت یک ماتریس $(m+1) \times m$ تعریف می کنیم که عناصر مخالف صفر آن ضرایب h_{ij} هستند و H_m ماتریس $m \times m$ به دست آمده از \tilde{H}_m با حذف سطر آخر آن می باشد. آنگاه ماتریس H_m به قسمی است که

$$A V_m - V_m H_m = h_{m+1, m} [0, 0, \dots, 0, v_{m+1}] \quad (۶-۱۰-۲۴)$$

یا به طور معادل

$$A V_m = V_{m+1} \tilde{H}_m$$

که در آن $V_{m+1} = (v_1, v_2, \dots, v_{m+1})$.

۴. یک راه عددی معقول برای پیاده سازی روش آرنولدی استفاده از روش گرام اشمیت اصلاح شده یا دوباره متعامد سازی کامل یا جزیی در گام ۲ می باشد. گرام اشمیت اصلاح شده در فصل ۷

بحث می شود. همچنین مرجع (Saad (1992) را ملاحظه کنید.

مثال ۶-۱۰-۸

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}, \quad m=2$$

$$v_1 = (1, 0, 0)^T$$

: $j = 1$

$$i = 1, \quad h_{11} = 1$$

$$\tilde{v}_2 = Av_1 - h_{11}v_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$h_{21} = 1/4141$$

$$v_2 = \frac{\tilde{v}_2}{h_{21}} = \begin{bmatrix} 0 \\ \cdot/\sqrt{0.71} \\ \cdot/\sqrt{0.71} \end{bmatrix}$$

: $j = 2$

$$i = 1,$$

$$h_{12} = 3/5255,$$

$$h_{22} = 3/5000$$

تشکیل ماتریسهای V_2 و H_2

$$V_2 = \begin{bmatrix} 1 & \cdot \\ \cdot & \cdot/\sqrt{0.71} \\ \cdot & \cdot/\sqrt{0.71} \end{bmatrix}, \quad H_2 = \begin{bmatrix} 1 & 3/5255 \\ 1/4142 & 3/5000 \end{bmatrix}$$

می توان تحقیق کرد که

$$AV_2 - V_2H_2 = \begin{bmatrix} \cdot & \cdot \\ \cdot & 1/0.67 \\ \cdot & -1/0.67 \end{bmatrix}$$

روش GMRES

روش GMRES (مانده مینیمال تعمیم یافته^۱) جهت می نیمم سازی نرم بردار مانده $b - Ax$ بر روی همه بردارهای زیر فضای آفین $x_0 + \mathcal{K}_m(A, r_0)$ طرح ریزی شده است، در اینجا x_0 بردار اولیه دلخواه و

است. $r_0 = b - Ax_0$.

با مفروض بودن یک تقریب اولیه x_0 از جواب x دستگاه $Ax = b$ ، روش مانده مینیمال تعمیم یافته (GMRES) یک جواب بهبود یافته به شکل

$$x_1 = x_0 + z \in x_0 + \kappa_m(A, r_0)$$

که در آن

$$r_0 = b - Ax_0.$$

به قسمی محاسبه می‌کند که

$$\|b - Ax_1\| = \min_{z \in \kappa_m(A, r_0)} \|b - A(x_0 + z)\|$$

فرض کنید m گام از روش آرنولد را با $v_1 = \frac{r_0}{\|r_0\|}$ اجرا کنیم؛ آنگاه در انتهای m گام، داریم

$$AV_m = V_{m+1} \tilde{H}_m$$

و

$$V_{m+1}^T r_0 = \|r_0\| e_1$$

بنابراین

$$\begin{aligned} \min_{z \in \kappa_m(A, r_0)} \|r_0 - Az\| &= \min_{y \in \mathbb{R}^m} \|r_0 - AV_m y\| \\ &= \min_{y \in \mathbb{R}^m} \|r_0 - V_{m+1} \tilde{H}_m y\| \\ &= \min_{y \in \mathbb{R}^m} \|V_{m+1}^T (r_0 - V_{m+1} \tilde{H}_m y)\| \\ &= \min_{y \in \mathbb{R}^m} \|\|r_0\| e_1 - \tilde{H}_m y\| \end{aligned}$$

وقتی که مسأله کمترین توانهای دوم

$$\min_{y \in \mathbb{R}^m} \|\|r_0\| e_1 - \tilde{H}_m y\|$$

حل شود و یک بردار $y_0 \in \mathbb{R}^m$ به دست آید، داریم

$$x_1 \approx x_0 + V_m y_0.$$

چندین روش برای حل مسأله کمترین توانهای دوم از جمله روشی که مبتنی بر تجزیه QR ماتریس \tilde{H}_m می‌باشد (و توسط سد و شولتز (Saad and Schultz (1986) استفاده شده است) در فصل ۷ توصیف

خواهیم کرد.

بحثهای فوق به الگوریتم زیر منجر می شود.

الگوریتم ۶-۱۰-۹ روش مانده مینیمال تعمیم یافته (GMRES) (Saad and Schultz (1986))

گام ۱: شروع: یک بردار اولیه x_0 و یک بعد m از زیر فضای کرلیف اختیار کنید. $r_0 = b - Ax_0$ را محاسبه کنید.

گام ۲: فرایند آرنولدی: m گام الگوریتم آرنولدی و شروع شده با $v_1 = r_0 / \|r_0\|$ را انجام دهید تا ماتریس \tilde{H}_m و ماتریس متعامد V_m تولید شوند.

گام ۳: تشکیل جواب تقریبی:

• بردار y را پیدا کنید که تابع $J(y) = \| \beta e_1 - \tilde{H}_m y \|$ را بر روی همه بردارهای y متعلق به R^m می نیمم کند، در اینجا $e_1 = [1, 0, \dots, 0]^T$ و $\beta = \|r_0\|$.

• بردار $x_1 = x_0 + V_m y$ را محاسبه کنید.

تبصره ها:

۱. همان گونه که قبلاً ذکر شد، پیاده سازی اصلی روش GMRES که در سال ۱۹۸۶ توسط سد و شولتز پیشنهاد شده است از روش گرام اشمیت اصلاح شده (که در فصل ۷ توصیف می شود) برای محاسبه پایه یکا متعامد V_m استفاده می کند.

پس از آن چندین پیاده سازی دیگر پیشنهاد شده اند. در میان آنها، آنهایی که جالب توجه هستند عبارتند از: الگوریتم ارائه شده توسط والکر (Walker (1988)) و الگوریتم ارائه شده توسط بای، هو، و ریشل (Bai, Hu, Reichel (1994)). پیاده سازی اخیر برای کامپیوترهای برداری و موازی مناسب تر است.

۲. اغلب در عمل روش GMRES نیاز دارد که پس از انجام به ازای یک مقدار انتخاب شده مناسب m ، دوباره شروع شود؛ یا همانند روش گرادیان مزدوج لازم است از یک پیش شرط مناسب استفاده شود. بحثهایی از این نوع خارج از محدوده این کتاب هستند. برای جزئیات بیشتر کتابهای جدید (Barrett, et al, (1994)) و (Saad (1995)) را ملاحظه کنید.

حل یک دستگام انتقال یافته. یک مشاهده مهم این است که پایه آرنولدی V_m تحت یک انتقال از A پایا است: اگر در فرایند آرنولدی به جای A ماتریس $A - \sigma I$ را استفاده کنیم همان دنباله $\{v_1, \dots, v_m\}$ را به دست می آوریم. این به واسطه آن است که زیر فضای کرلیف κ_m برای A و $A - \sigma I$ یکسان است،

به شرط آن که بردار v_1 اولیه یکسان باشد. توجه کنید که از (۶-۱۰-۲۴) داریم

$$(A - \sigma I)V_m = V_m(H - \sigma I) + h_{m+1,m}v_{m+1}e_m^T$$

که بدین معنی است که اگر فرایند آرنولدی را با $A - \sigma I$ اجرا کنیم، همان ماتریس V_m را به دست می آوریم، ولیکن ماتریس H_m دارای قطری خواهد بود که انتقال یافته توسط σI می باشد.

حل دستگاههای انتقال یافته در بسیاری از کاربردها مطرح می شوند، نظیر محاسبه ماتریس پاسخ فرکانس یک دستگاه کنترل (بخش ۶-۴-۷ را ملاحظه کنید). برخی روشهای کارای محاسبه ماتریس پاسخ فرکانس برای دستگاههای کنترل خطی تنک بزرگ که از GMRES با شروع دوباره همراه با تجزیه ناقص LU استفاده می کنند در رساله Ph.D اخیر شودهاری (Choudhury (1994)) ارائه شده اند.

رابطه بین روش GMRES و روش گرادیان مزدوج

از توضیح روش گرادیان مزدوج روشن نیست که روش گرادیان مزدوج نیز یک روش زیر فضای کرلیف دیگری است. همان گونه که روش GMRES نرم مانده $r_0 = b - Ax_0$ را بر روی همه بردارهای زیر فضای کرلیف آفین $(x_0 + \kappa_m(A, r_0))$ می نیمم می سازد، تکرار گرادیان مزدوج x_{k+1} خطا را (در A -نرم) بر روی همه بردارهای واقع در زیر فضای کرلیف $(x_0 + \kappa_{k+1}(A, p_0))$ می نیمم می سازد. همچنین توجه کنید که $\kappa_{k+1}(A, p_0) = \kappa_{k+1}(A, r_0)$.

بویژه قضیه زیر را از مرجع Ortega (1988a) بدون اثبات بیان می کنیم. اثبات آن را می توان در مرجع Ortega (1988a, 273-4) مشاهده کرد.

قضیه ۹-۱۰-۶ هر تکرار گرادیان مزدوج x_{k+1} ($k = 0, 1, \dots, n-1$) $\|x - \bar{x}\|_A$ را بر روی $(x_0 + \kappa_{k+1}(A, p_0))$ می نیمم می سازد که در آن p_0 بردار جهت اولیه است.

روش QMR

یک روش مشابه با روش GMRES که روش مانده شبه مینیمال^۱ (QMR) نامیده می شود. توسط فروند و ناشتیگال (Freund and Nachtigal (1991)) تدوین شده است و در سالهای اخیر در زمینه حل دستگاههای خطی نامتقارن تنک متداول گشته است. روش دستگاه خطی پر را به یک دستگاه سه قطری تقلیل می دهد و سپس دستگاه تقلیل یافته را در جهت کمترین توانهای دوم حل می کند. از تغییری از روش معروف لانزوس که لانزوس پیش نگرش نامیده می شود و برای جلوگیری از شکست در

فرایند لائزوس به کار می‌رود، استفاده می‌کند. فرایند لائزوس به طور کامل در فصل ۸ این کتاب توصیف خواهد شد. برای جزئیات بیشتر در مورد روش QMR مرجع (Freund and Nachtigal 1991) را ملاحظه کنید.

روشهای تکراری بحث نشده در این کتاب

علاوه بر روش QMR برخی روشهای مهم دیگر که در این کتاب بحث نشده‌اند عبارتند از: گرادیان دو مزدوجی (BiCG) ، گرادیان دو مزدوجی پایدار شده (Bi-CGSTAB) ، گرادیان مزدوج مربع شده (CGS) ، تکرار چیبیشف، MINRES و SYMMLQ. در حالی که روشهای Bi-CGSTAB ، Bi-CGSTAB ، CGS ، و چیبیشف همگی قابل اعمال بر روی دستگاههای نامتقارن هستند، روشهای MINRES و SYMMLQ تغییرهایی از روش گرادیان مزدوج هستند و برای حل دستگاههای خطی تنک و بزرگ نامعین متقارن طرح ریزی شده‌اند. برای بحثهای اجمالی و پیاده‌سازی با جزئیات بیشتر و مطالعه‌های مقایسه‌ای، کتابهای (Templaes (Barret, et, al 1994) و Saad (1995) را ملاحظه کنید.

۱۱-۶ مرور و خلاصه

مهمترین نتایج بحث شده در این فصل را در زیر مرور می‌کنیم.

۱. روشهای عددی برای دستگاههای خطی دلخواه: دو نوع روش - مستقیم و تکراری - بحث شدند.

(الف) روشهای مستقیم عبارتند از: روشهای حذفی گاوس و تجزیه QR.

• حذفی گاوس بدون تعویض سطرها، هنگامی که وجود دارد، یک تجزیه $A = LU$ ارائه می‌دهد. سپس دستگاه $Ax = b$ ابتدا با حل دستگاه پایین مثلثی $Ly = b$ و سپس با حل دستگاه بالا مثلثی $Ux = y$ حل می‌شود.

روش نیاز به $n^3/3$ عمل ممیز شناور (flops) دارد. برای ماتریسهای دلخواه ناپایدار است و برای استفاده‌های عملی توصیه نمی‌شود مگر ماتریس A معین مثبت باشد. عامل رشد ρ می‌تواند برای یک ماتریس دلخواه بزرگ شود.

• روش حذفی گاوس با محورگیری جزئی یک تجزیه از A را ارائه می‌دهد:

$$MA = U$$

با داشتن این تجزیه، می‌توانیم دستگاه $Ax = b$ را با حل دستگاه بالا مثلثی $Ux = b'$ حل کنیم، که در آن $b' = Mb$. فرایند به $n^3/3$ عمل ممیز شناور (flops) و $O(n^2)$ مقایسه نیاز دارد. در نظری برخی خطرها می‌تواند وجود داشته باشد، اما در عمل این الگوریتم پایدار است. این الگوریتم متداول‌ترین الگوریتم عملی مورد استفاده برای حل یک دستگاه خطی

متراکم است.

- روش حذفی گاوس با محورگیری کامل تجزیه

$$MAQ = U$$

را ارائه می دهد. وقتی که این تجزیه را داشته باشیم می توانیم دستگاه $Ax = b$ را ابتدا با حل دستگاه

$$Uy = b',$$

حل کنیم که در آن $y = Q^T x$ و $b' = Mb$. سپس x را از $x = Qy$ بازیابی کنیم. این فرایند نیاز به $n^3/3$ عمل ممیز شناور (flops) و $O(n^3)$ مقایسه دارد. بنابراین گران تر از روش حذفی گاوس با محورگیری جزئی است، اما پایدارتر است (عامل رشد ρ در این حالت توسط تابع رشد آهسته ای از n کراندار می باشد، در حالی که عامل رشد ρ روش حذفی گاوس با محورگیری جزئی می تواند به بزرگی 2^{n-1} باشد).

- روشهای مثلثی سازی متعامد بر تجزیه QR ماتریس A مبتنی هستند:

$$A = QR$$

وقتی که این تجزیه را داشته باشیم می توانیم $Ax = b$ را با حل دستگاه بالا مثلثی $Rx = b'$ حل کنیم، که در آن $b' = Q^T b$. می توانیم یا از روش هاوس هولدر یا روش گیونز برای رسیدن به چنین تجزیه ای استفاده کنیم. روش هاوس هولدر اقتصادی تر از روش گیونز است ($2n^3/3$ عمل ممیز شناور در برابر $4n^3/3$ عمل ممیز شناور). هر دو روش از پایداری تضمین شده ای برخوردار هستند.

(ب) روشهای تکراری روشهای ژاکوبی، گاوس سایدل، و SOR هستند که ما آنها را بحث کردیم. یک فرمول بندی کلی از یک روش تکراری به صورت زیر است

$$x^{(k+1)} = Bx^{(k)} + d$$

تفاوت روشها در انتخاب B و d است. با نوشتن $A = L + U + D$ داریم:

- برای روش ژاکوبی:

$$B = B_J = -D^{-1}(L + U)$$

$$d = b_J = -D^{-1}b$$

• برای روش گاوس - سایدل:

$$B = B_{GS} = -(D + L)^{-1}U$$

$$d = b_{GS} = (D + L)^{-1}b$$

• برای روش SOR :

$$B = B_{SOR} = (D + \omega L)^{-1}[(1 - \omega)D - \omega U]$$

$$d = d_{SOR} = \omega(D + \omega L)^{-1}b$$

(ω پارامتر تخفیف است).

• تکرار

$$x^{(k+1)} = Bx^{(k)} + d$$

برای هر انتخاب دلخواه تقریب اولیه $x^{(1)}$ همگرا می‌گردد اگر و فقط اگر شعاع طیفی ماتریس B کوچک‌تر از ۱ باشد (قضیه ۶-۱۰-۱). یک شرط کافی برای همگرایی این است که $\|B\| < 1$ (قضیه ۶-۱۰-۱).

• با استفاده از شرط کافی، نشان داده شد که هر دو روش ژاکوبی و گاوس سایدل همگرا هستند هنگامی که A یک ماتریس غالب قطری می‌باشد (نتیجه‌های ۶-۱۰-۱ و ۶-۱۰-۲).
• روش گاوس - سایدل همچنین همگرا است هنگامی که A معین مثبت متقارن می‌باشد (قضیه ۶-۱۰-۲).

• تکرار SOR نمی‌تواند برای هر انتخاب دلخواه تقریب اولیه همگرا باشد مگر پارامتر تخفیف ω در بازه $(0, 2)$ قرار گیرد (قضیه ۶-۱۰-۳).

• برای یک ماتریس دو دوری به طور سازگار مرتب شده با عناصر قطری مخالف صفر، انتخاب بهینه ω ، نمایش داده شده توسط ω_{opt} ، به صورت

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - \rho(B_T)^2}}$$

ارائه می‌شود با فرض این که مقادیر ویژه B_T حقیقی هستند و $\rho(A) \cdot \rho(B_T) < 1$ برای نمایش شعاع طیفی A به کار برده می‌شود (قضیه ۶-۱۰-۴).

۲. دستگاههای خاص: دستگاههای معین مثبت متقارن، غالب قطری، هسنبرگی، و سه قطری مورد بحث قرار گرفتند.

(الف) دوروش - روش حذفی گاوس بدون محورگیری و روش چولسکی - برای دستگاههای معین مثبت متقارن توصیف می شوند.

- روش حذفی گاوس با محورگیری جزئی یک تجزیه از A ارائه می دهد:

$$A = LDL^T$$

دستگاه $Ax = b$ ابتدا با حل دستگاه پایین مثلثی $LDy = b$ و سپس با حل دستگاه بالا مثلثی $L^T x = y$ حل می شود. روش نیاز به $n^3/3$ عمل ممیز شناور (flops) دارد و هیچ محاسبه ریشه دومی مورد نیاز نیست. روش پایدار است ($\rho = 1$).

- الگوریتم تجزیه چولسکی یک تجزیه از A به شکل $A = HH^T$ محاسبه می کند که در آن H یک ماتریس پایین مثلثی با عناصر قطری مثبت است. وقتی که این تجزیه را داشته باشیم می توانیم دستگاه $Ax = b$ را ابتدا با حل دستگاه پایین مثلثی $Hy = b$ ، که در آن $y = H^T x$ ، و سپس با حل دستگاه بالا مثلثی $H^T x = y$ حل کنیم. روش نیاز به $n^3/6$ عمل ممیز شناور (flops) و محاسبه n ریشه دوم دارد. روش پایدار است.

(ب) روش حذفی گاوس بدون محورگیری برای دستگاههای غالب قطری پایدار است ($\rho \leq 2$).

(ج) برای دستگاه هسنبرگی، روش حذفی گاوس با محورگیری جزئی فقط به $O(n^2)$ عمل ممیز شناور نیاز دارد. روش پایدار است

(د) برای یک دستگاه سه قطری، روش حذفی گاوس با محورگیری جزئی فقط به $O(n)$ عمل ممیز شناور نیاز دارد. روش پایدار است ($\rho \leq 2$).

۳. معکوس، دترمینان، کهادهای اصلی پیشرو: معکوس و دترمینان یک ماتریس را می توان با داشتن تجزیه ای از A محاسبه کرد.

(الف) ابتدا معکوس را ملاحظه می کنیم.

- وقتی که تجزیه ای از ماتریس A به دست آمد، معکوس می تواند با حل یک مجموعه دستگاههای خطی با طرفهای راست متفاوت محاسبه شود. همان تجزیه می تواند برای حل دستگاههای متفاوت مورد استفاده قرار گیرد (بخش ۶-۵ ملاحظه شود). توجه کنید که بیشتر مسائل حاوی معکوس می توانند طوری حل شوند که به محاسبه معکوس به طور صریح نیاز نداشته باشند. بعلاوه ماتریسهایی یافت می شوند (نظیر ماتریسهای مثلثی، و غیره) که معکوس آنها به سادگی محاسبه می شوند.

• معکوس یک ماتریس B که با ماتریس A فقط در یک اختلال رتبه یک تفاوت دارد می تواند به راحتی با استفاده از فرمول شرمز - موریسون^۱ محاسبه شود وقتی که معکوس A به دست آمده باشد: فرض کنید $B = A - uv^T$. آنگاه $B^{-1} = A^{-1} + \alpha(A^{-1}uv^TA^{-1})$ ، که در آن $\alpha = 1/(1 - v^TA^{-1}u)$. تعمیمی از این فرمول معروف به فرمول وودبری^۲ وجود دارد.

(ب) درمیتان بندرت در کاربردهای عملی مورد نیاز می باشد. ولیکن درمیتان A می تواند وقتی که یک تجزیه از A به دست آید، محاسبه شود.

• اگر از روش حذفی گاوس با محورگیری استفاده شود، داریم

$$MA = U$$

آنگاه

$$\det(A) = (-1)^r a_{11} a_{22}^{(1)} \dots a_{nn}^{(n-1)}$$

که در آن r تعداد تعویضهای سطری انجام شده در طی فرایند حذفی می باشد. a_{11} ، $a_{22}^{(1)}$ ، ...، $a_{nn}^{(n-1)}$ عناصر محوری هستند که به صورت عناصر قطری U ظاهر می شوند. به طور مشابه تجزیه های دیگر می توانند به راحتی برای پیدا کردن درمیتان مورد استفاده قرار گیرند.

۴. عدد وضعیت ودقت جواب: در مسأله دستگاه خطی $Ax = b$ داده های ورودی A و b هستند.

خطاهایی ممکن است در A یا b یا هر دو وجود داشته باشند.

تحلیل خطاها در هر سه حالت ارائه شده اند. این نتایج در قضیه های ۶-۶-۱، ۶-۶-۲، و

۶-۶-۳ بیان شده اند. قضیه ۶-۶-۳ کلی ترین قضیه است.

در هر سه حالت ثابت می شود که عدد وضعیت

$$\text{Cond}(A) = \|A\| \|A^{-1}\|$$

عامل تصمیم گیرنده است. اگر این عدد بزرگ باشد، آنگاه یک اختلال کوچک در داده های ورودی ممکن است موجب یک خطای بزرگ در جواب محاسبه شده گردد. در این حالت دستگاه یک دستگاه بد وضع نامیده می شود؛ در غیر این صورت خوش وضع است. ماتریس دارای یک عدد وضعیت بزرگ یک ماتریس بد وضع نامیده می شود.

برخی خواص مهم عدد وضعیت یک ماتریس فهرست شدند. برخی ماتریسهای بد وضع

معروف عبارتند از: ماتریس هیلبرت، ماتریس پای (Pie) ، و ماتریس دو قطری ویلکینسون.

البته عدد وضعیت اثر قابل توجهی بر روی دقت جواب دارد.

یک جواب محاسبه شده فقط هنگامی می تواند دقیق در نظر گرفته شود که هم $\text{Cond}(A)$ و هم مانده نسبی کوچک باشند (قضیه ۶-۷-۱). بنابراین یک مانده نسبی کوچک به تنهایی نمی تواند دقت جواب را تضمین نماید.

سؤالی که اغلب مطرح می شود این است: چقدر $\text{Cond}(A)$ باید بزرگ باشد قبل از آن که دستگاه به صورت یک دستگاه بد وضع در نظر گرفته شود؟

جواب بستگی به دقت داده های ورودی و سطح تolerانس خطا در جواب دارد.

به طور کلی، اگر داده ها به طور تقریبی دقیق باشند و اگر $\text{Cond}(A) = 10^8$ ، آنگاه دقتی در حدود $(t - 8)$ رقم با معنی در جواب وجود دارد اگر در حساب t رقمی محاسبه شود.

- محاسبه عدد وضعیت از روی تعریف به طور واضح گران است؛ این محاسبه شامل پیدا کردن نرم معکوس و پیدا کردن معکوس A بوده که در حدود سه برابر گران تر از حل خود دستگاه خطی می باشد.

- دو برآوردگر عدد وضعیت، برآوردگر عدد وضعیت LINPACK و برآوردگر عدد وضعیت ۱- نرم هاگر توصیف شده اند.

- علامتهای نشان دهنده ای در طی روش حذفی گاوس با محورگیری وجود دارند نظیر یک عنصر محوری کوچک، یک جواب محاسبه شده بزرگ، و یک مانده بزرگ، که صرفاً نشان می دهند اگر یک دستگاه بد وضع باشد، اما اینها آزمونهای مطمئنی نیستند.

- اختلالهای مؤلفه ای شناخته شده هستند. بویژه هنگامی که نرمهای ستونهای ماتریس معکوس به طور وسیعی تغییر می کنند عدد وضعیت اسکیل می تواند مفید باشد.

۵. تصفیه تکراری: برای یک جواب محاسبه شده یک راه کم خرج برای تصفیه کردن جواب به طور تکراری، معروف به روش تصفیه تکراری، توصیف شد (الگوریتم ۶-۹-۱).

۶. روشهای گرادیان مزدوج و $GMRES$: روش گرادیان مزدوج، هنگامی که با یک پیش شرط مناسب استفاده می شود، یکی از متداول ترین روشهای استفاده شده برای حل یک دستگاه خطی معین مثبت متقارن تنک و بزرگ می باشد.

فقط یک نوع پیش شرط، تجزیه چولسکی ناقص (ICF) در این کتاب توصیف شد. ایده اساسی ICF این است که تجزیه چولسکی A را فقط برای عناصر غیر صفر محاسبه کنیم و عناصر صفر به صورت صفر باقی گذاشته شوند.

ما فقط روشهای پایه‌ای آرنولدی و GMRES مبتنی بر آرنولدی را توصیف کردیم. در عمل فرایند گرام اشمیت اصلاح شده لازم است که در پیاده‌سازی روش آرنولدی استفاده شود، و لازم است که GMRES با یک پیش شرط مناسب استفاده گردد. از لحاظ نظری روش گرادیان مزدوج یک روش مستقیم است، ولیکن در عمل تکراری است. روش گرادیان مزدوج بی‌نهایت آهسته است هنگامی که A بدوضع است و یک پیش شرط برای تسریع روش لازم می‌باشد.

تمرینها

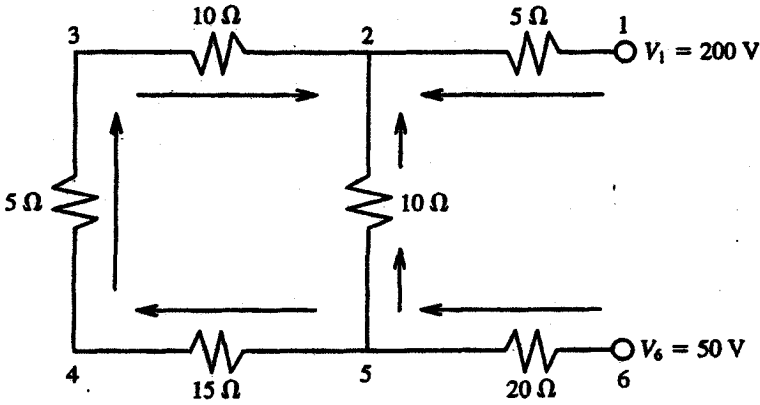
تذکر: از MATLAB هر جا لازم و مقتضی می‌باشد، استفاده کنید.

بخش ۳-۶

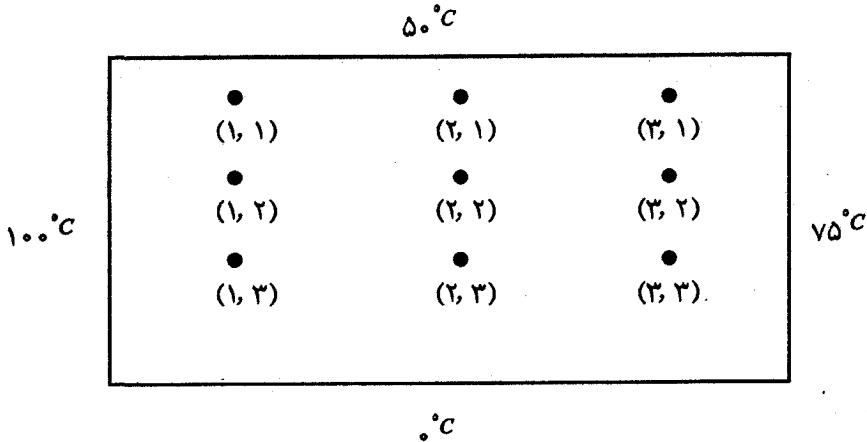
۱. یک مهندس به ۵۵۰۰، ۵۵۰۰، ۶۰۰۰ یارد مکعب از شن، سیمان، سنگ ریزه برای یک پروژه ساختمانی نیاز دارد. او مواد را از سه منبع دریافت می‌کند. توزیعی از هر یک از این مواد در این سه منبع در جداول زیر ارائه شده است.

منبع	% شن	% سیمان	% سنگ ریزه
۱	۶۰	۲۰	۲۰
۲	۴۰	۴۰	۲۰
۳	۲۰	۳۰	۵۰

- چند یارد مکعب باید مهندس از هر منبع دریافت کند تا نیاز او را برآورده سازند.
۲. اگر ورودی به راکتور ۱ در مسأله راکتور بخش ۳-۶-۲ به مقدار ۱۰٪ کاهش یابد، درصد تغییر در غلظت سایر راکتورها چقدر خواهد بود.
۳. نمودار جریان زیر را ملاحظه کنید. یک دستگاه خطی برای تعیین جریان بین گره‌ها تشکیل دهید.



۴. با استفاده از معادله تفاضلی (۶-۳-۳۰) یک دستگاه خطی برای توزیع گرما در نقاط داخلی نشان داده شده در شکل زیر از یک صفحه گرم شده بنویسید که دمای مرزهای آن ثابت نگه داشته شوند.



۵. دستگاه خطی برای تقریب تفاضل متناهی معادله بیضوی زیر به دست آورید.

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = f(x, y)$$

حوزه مقادیر داخل مربع واحد است، $\Delta x = \Delta y = 0/25$ ، و شرایط مرزی به صورت زیر ارائه می شوند

$$T(x, 0) = 1 - x$$

$$T(1, y) = y$$

$$T(0, y) = 1$$

$$T(x, 1) = 0$$

۶. برای مسأله قبل اگر داشته باشیم

$$f(x, y) = -\pi^2 \sin(\pi x) \sin(\pi y)$$

آنگاه جواب تحلیلی برای معادله بیضوی

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = f(x, y)$$

با همان شرایط مرزی تمرین ۵ به صورت زیر می‌باشد.

$$T(x, y) = 1 - x + xy + \left(\frac{1}{\pi}\right) \sin(\pi x) \sin(\pi y)$$

(مرجع Ceila and Gray (1992, 105-60) را ملاحظه کنید).

(الف) از طرح تفاضل متناهی بخش ۶-۳ برای تقریب زدن مقادیر T در نقاط داخلی با

$$n = 4, 8, 16, \Delta x = \Delta y = 1/n$$
 استفاده کنید.

(ب) مقادیر به دست آمده در (الف) را با جواب واقعی مقایسه کنید.

(ج) دستگاه حاصل از روش عناصر متناهی را برای جواب مسأله مقدار مرزی دو نقطه‌ای

$$u'' = -2u'' + 3u = x^2, u(0) = u(1) = 0, 0 \leq x \leq 1$$

همان توابع پایه‌ای $\phi_j(x)$ کتاب و شبکه یکنواخت استفاده کنید.

بخش ۶-۴

۷. دستگاه خطی $Ax = b$ را حل کنید که در آن b یک بردار با همه مؤلفه‌های برابر ۱ بوده و A هر

یک از ماتریسهای تمرین ۱۸ از فصل ۵ می‌باشد. برای این منظور از روشهای زیر استفاده کنید.

(الف) روش حذفی گاوس بدون محورگیری

(ب) روش حذفی گاوس با محورگیری جزئی و کامل

(ج) تجزیه QR

۸. هر یک از دستگاههای تمرین ۷ را با استفاده از محورگیری جزئی ولیکن بدون تجزیه صریح حل

کنید (بخش ۶-۴-۴).

۹. دستگاه

$$\begin{bmatrix} 0/0001 & 1 & 1 \\ 3 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2/0001 \\ 3 \\ 3 \end{bmatrix}$$

را با استفاده از روش حذفی گاوس بدون و با محورگیری جزئی حل کنید و جوابها را با هم مقایسه کنید.

۱۰. دستگاه خطی زیر را در نظر بگیرید.

$$Ax_i = b_i, \quad i = 1, 2, \dots, m$$

(الف) یک الگوریتم برای حل دستگاههای فوق با استفاده از روش حذفی گاوس با محورگیری کامل تدوین نمایید. الگوریتم شما باید از این حقیقت که همه m دستگاه دارای ماتریس ضرایب یکسان می باشند بهره مند گردد.

(ب) تعداد اعمال ممیز شناور الگوریتم خودتان را تعیین کنید.

(ج) الگوریتم خودتان را برای حالت خاصی که در آن b_i ستون i ام ماتریس واحد می باشد، به کار ببرید $i = 1, 2, \dots, n$.

(د) الگوریتم قسمت (ج) را به کار ببرید و نشان دهید که معکوس یک ماتریس $n \times n$ مانند A را می توانیم با استفاده از روش حذفی گاوس با محورگیری کامل و با حدود $(4/3)n^3$ عمل ممیز شناور محاسبه کنیم.

(ه) الگوریتم قسمت (ج) را برای محاسبه معکوس ماتریس زیر به کار ببرید.

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}$$

۱۱. دستگاه $Ax = b$ را در نظر بگیرید، که در آن هم A و هم b مختلط هستند. نشان دهید که چگونه دستگاه را می توان فقط با استفاده از حساب حقیقی حل کرد. تعداد اعمال ممیز شناور را در این حالت با آنچه که برای حل دستگاه با روش حذفی گاوس با استفاده از حساب مختلط لازم می باشد، مقایسه کنید.

۱۲. تجزیه چولسکی ماتریس

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1/0.01 & 1/0.01 \\ 1 & 1/0.01 & 2 \end{bmatrix}$$

را با استفاده از روشهای زیر محاسبه کنید

(i) روش حذفی گاوس با محورگیری

(ii) الگوریتم چولسکی

(ب) در قسمت (الف) تحقیق کنید که، $\max |a_{ij}^{(k)}| \leq \max |a_{ij}^{(k-1)}|$ ، $k = 1, 2$ ،

(ج) عامل رشد چیست؟

۱۳. با استفاده از همان نتایج تمرین ۱۲ دستگاه $Ax = b$ را حل کنید که در آن

$$b = \begin{bmatrix} 3 \\ 3/0.020 \\ 4/0.010 \end{bmatrix}$$

و A همان ماتریس تمرین ۱۲ می باشد.

۱۴. (الف) نشان دهید که ماتریس

$$A = \begin{bmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{bmatrix}$$

معین مثبت است. این کار را با و بدون پیدا کردن تجزیه چولسکی انجام دهید.

(ب) دستگاه زیر را حل کنید.

$$Ax = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}$$

۱۵. (الف) یک الگوریتم برای حل یک دستگاه سه قطری با استفاده از روش حذفی گاوس با

محورگیری جزئی تدوین نمایید.

(ب) نشان دهید که در این حالت عامل رشد توسط ۲ کسراندار می شود. (راهنمایی:

$$\max |a_{ij}^{(1)}| \leq 2 \max |a_{ij}|$$

۱۶. (الف) نشان دهید که روش حذفی گاوس اعمال شده بر یک ماتریس غالب قطری ستونی خاصیت غالب قطری را در هر گام کاهش حفظ می‌کند؛ یعنی اگر $A = (a_{ij})$ غالب قطری ستونی باشد، آنگاه $A^{(k)} = (a_{ij}^{(k)})$ ، $k = 1, 2, \dots, n - 1$ نیز غالب قطری ستونی هستند.

(ب) نشان دهید که عامل رشد با استفاده از روش حذفی گاوس بدون محورگیری توسط ۲ کراندار می‌باشد.

(ج) گزاره قسمت (الف) را در مورد ماتریس تمرین ۱۴ تحقیق کنید.

(د) یک ماتریس غالب قطری ستونی 2×2 بسازید که عامل رشد آن با روش حذفی گاوس بزرگ‌تر از ۱، اما کمتر یا مساوی ۲ باشد.

۱۷. دستگاه سه قطری زیر را حل کنید.

$$\begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

(الف) با استفاده از روش حذفی گاوس

(ب) با محاسبه تجزیه LU ماتریس A مستقیماً از $A = LU$

۱۸. دستگاه غالب قطری

$$\begin{pmatrix} 10 & 1 & 1 & 1 \\ 1 & 10 & 1 & 1 \\ -1 & 0 & 10 & 1 \\ -1 & -1 & -1 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 13 \\ 13 \\ 10 \\ 7 \end{pmatrix}$$

را با استفاده از روش حذفی گاوس با محورگیری حل کنید. عامل رشد را محاسبه کنید.

۱۹. (الف) الگوریتمهای کارایی برای مثلثی سازی (i) یک ماتریس بالا هسنبرگی و (ii) یک ماتریس پایین هسنبرگی تدوین نمایید. برای این منظور از روش حذفی گاوس با محورگیری جزئی استفاده کنید.

(ب) نشان دهید که اگر A بالا هسنبرگی باشد، روش حذفی گاوس با محورگیری جزئی نتیجه

می‌دهد

$$|a_{ij}^{(k)}| \leq k + 1, \quad |a_{ij}| \leq 1$$

بنابراین نتیجه بگیرید که در این حالت عامل رشد توسط n کراندار می‌شود.

(ج) الگوریتمهای خود را برای حل دستگاههای زیر به کار ببرید.

$$\begin{pmatrix} 1 & 2 & 2 & 3 \\ 3 & 4 & 5 & 7 \\ 0 & \dots\dots 1 & 2 & 3 \\ 0 & 0 & \dots\dots 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 6 \\ 7 \\ 8 \\ 9 \end{pmatrix} \quad : (i)$$

$$\begin{pmatrix} 1 & \dots\dots 1 & 0 \\ 0 & 2 & \dots\dots 1 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1/\dots\dots 1 \\ 2/\dots\dots 1 \\ 3/\dots\dots 1 \end{pmatrix} \quad : (ii)$$

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \quad : (iii)$$

(د) عامل رشد را در این حالت تعیین کنید.

(ه) فرض کنید داده‌ها در این مسأله تا چهار رقم دقیق باشند و دقتی با سه رقم برای جواب خود

جستجو می‌کنید. مشخص کنید که کدام یک از این مسائل بد وضع هستند.

۲۰. الف) تجزیه QR ماتریس زیر را به دست آورید.

$$A = \begin{pmatrix} 1 & 1 \\ 10^{-5} & 0 \\ 0 & 10^{-5} \end{pmatrix}$$

(ب) از نتایج قسمت الف) استفاده کنید، و دستگاه $Ax = b$ را حل کنید که در آن

$$b = \begin{pmatrix} 2 \\ 10^{-5} \\ 10^{-5} \end{pmatrix}$$

۲۱. تجزیه QR ماتریس

$$A = \begin{pmatrix} T & I & 0 \\ I & T & I \\ 0 & I & T \end{pmatrix}$$

را به دست آورید که در آن

$$T = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

از این تجزیه برای حل دستگاه $Ax = b$ استفاده کنید که در آن هر یک از عناصر b برابر ۲ می‌باشند.

بخش ۵-۶

۲۲. دترمینان و معکوس هر یک از ماتریسهای تمرین ۱۸ فصل ۵ را با استفاده از روشهای زیر پیدا کنید.

(الف) روش حذفی گاوس با محورگیری جزئی و کامل

(ب) تجزیه QR

۲۳. (الف) فرض کنید

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

بعلاوه فرض کنید که A_{11} و A_{22} مربعی باشند و A_{11} و $A_{12} - A_{21}A_{11}^{-1}A_{12}$ نامنفرد باشند. فرض کنید ماتریس

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

معکوس A باشد. نشان دهید که

$$B_{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$$

$$B_{12} = -A_{11}^{-1}A_{12}B_{22}$$

$$B_{21} = -B_{22}A_{21}A_{11}^{-1}$$

$$B_{11} = A_{11}^{-1} - B_{12}A_{21}A_{11}^{-1}$$

(ب) چند عمل ممیز شناور برای محاسبه A^{-1} با استفاده از نتایج قسمت (الف) لازم می‌باشد

اگر A_{11} و A_{22} به ترتیب ماتریسهای $m \times m$ و $p \times p$ باشند.

(ج) از نتایج خود برای محاسبه معکوس ماتریس زیر استفاده کنید.

$$A = \begin{bmatrix} 4 & 0 & -1 & -1 \\ 0 & 4 & -1 & -1 \\ -1 & -1 & 4 & 0 \\ -1 & 1 & 0 & 4 \end{bmatrix}$$

۲۴. فرض کنید

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4/0.001 & 2/0.002 \\ 1 & 2/0.002 & 2/0.004 \end{bmatrix} \quad \text{و} \quad B = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 4/0.001 & 2/0.002 \\ 1 & 2/0.002 & 2/0.004 \end{bmatrix}$$

۲۵. را به شکل $B = A - uv^T$ بنویسید؛ سپس B^{-1} را با استفاده از فرمول شرمین - موریسون محاسبه کنید در صورتی که بدانیم

$$A^{-1} = 10^4 \begin{bmatrix} 4/0.010 & -2/0.006 & 0/0.003 \\ -2/0.006 & 1/0.004 & -0/0.002 \\ 0/0.003 & -0/0.002 & 0/0.001 \end{bmatrix}$$

۲۵. فرض کنید شما یک دستگاه خطی با ماتریس ضرایب A را حل کرده باشید. سپس نشان دهید که چگونه می‌توانیم دستگاه افزوده

$$\begin{bmatrix} A & \delta \\ c & \alpha \end{bmatrix} \begin{bmatrix} x \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} b \\ b_{n+1} \end{bmatrix}$$

را با استفاده از جواب قبلاً به دست آمده حل کنیم در صورتی که A یک ماتریس نامنفرد $n \times n$ و δ ، b ، و c بردار باشند. نتایج خود را برای حل دستگاه زیر به کار ببرید.

$$\begin{bmatrix} 1 & 2 & 3 & 1 \\ 4 & 5 & 6 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} y = (6, 15, 3, 1)^T$$

بخشهای ۶-۶ و ۷-۶

۲۶. دستگاه مقارن $Ax = b$ را در نظر بگیرید که در آن

$$A = \begin{bmatrix} 0/4445 & 0/4444 & -0/2222 \\ 0/4444 & 0/4445 & -0/2222 \\ -0/2222 & -0/2222 & 0/1112 \end{bmatrix}, \quad b = \begin{bmatrix} 0/6667 \\ 0/6667 \\ -0/3332 \end{bmatrix}$$

جواب واقعی دستگاه عبارت است از:

$$x = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

(الف) یک اختلال کوچک δb در b ایجاد کنید، A را بدون تغییر حفظ کنید. دستگاه $Ax' = b + \delta b$ را حل کنید. x' را با x مقایسه کنید. $\text{Cond}(A)$ را محاسبه کنید و نامساوی مناسب در متن را تحقیق کنید.

(ب) یک اختلال کوچک ΔA در A به قسمی ایجاد کنید که $\|\Delta A\| \leq 1/\|A^{-1}\|$. دستگاه $(A + \Delta A)x' = b$ را حل کنید. x' را با x مقایسه کنید. و نامساوی مناسب در متن را تحقیق کنید (راهنمایی: $\|A^{-1}\|_2 = O(10^4)$).

۲۷. نامساوی زیر را ثابت کنید

$$\frac{\|\delta x\|}{\|x + \delta x\|} \leq \text{Cond}(A) \frac{\|\Delta A\|}{\|A\|}$$

در اینجا $Ax = b$ و $(A + \Delta A)(x + \delta x) = b$. نامساوی را برای دستگاه زیر تحقیق کنید.

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{و} \quad \Delta A = \begin{pmatrix} \cdot & \cdot & \cdot / \dots \dots \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$$

۲۸. (الف) چگونه $\text{Cond}(A)$ و $\text{Cond}(A^{-1})$ به یکدیگر وابسته اند؟

(ب) نشان دهید که

$$\text{Cond}(A) \geq 1 \quad (i)$$

$$\text{Cond}(A^T A) = (\text{Cond}(A))^2 \quad (ii)$$

۲۹. (الف) فرض کنید O یک ماتریس متعامد باشد. نشان دهید که $\text{Cond}(O)$ نسبت به ۲-نرم

برابر ۱ است.

(ب) نشان دهید که $\text{Cond}(A)$ نسبت به ۲ - نرم برابر ۱ است اگر و فقط اگر A مضرب اسکالری از یک ماتریس متعامد باشد.

۳۰. فرض کنید $U = (u_{ij})$ یک ماتریس بالا مثلثی نامنفرد باشد. اولاً نشان دهید که نسبت به بی‌نهایت - نرم داریم

$$\text{Cond}(U) \geq \frac{\max(u_{ij})}{\min(u_{ij})}$$

ثانیاً مثال ساده‌ای از یک ماتریس معین مثبت متقارن غیر قطری بدوضع بسازید.

۳۱. فرض کنید $A = LDL^T$ یک ماتریس معین مثبت متقارن باشد. فرض کنید $D = \text{diag}(D_{ii})$. اولاً نشان دهید که نسبت به ۲ - نرم داریم

$$\text{Cond}(A) \geq \frac{\max(d_{ii})}{\min(d_{ii})}$$

ثانیاً مثال ساده‌ای از یک ماتریس معین مثبت متقارن غیر قطری بدوضع بسازید.

۳۲. (الف) نشان دهید که برای هر ماتریس A ، $\text{Cond}(A)$ نسبت به ۲ - نرم به صورت

$$\text{Cond}(A) = \frac{\sigma_{\max}}{\sigma_{\min}}$$

ارائه می‌شود که در آن σ_{\max} و σ_{\min} به ترتیب بزرگ‌ترین و کوچک‌ترین مقادیر تکین A هستند.

(ب) از عبارت فوق برای $\text{Cond}(A)$ استفاده کنید و مثالی از یک ماتریس بدوضع به طریق زیر بسازید: دو ماتریس متعامد غیر قطری U و V و یک ماتریس قطری Σ با یک یا چند عنصر قطری کوچک انتخاب کنید (بویژه ماتریسهای U و V می‌توانند به صورت ماتریسهای هاوس هولدر انتخاب شوند). آنگاه ماتریس

$$A = U\Sigma V^T$$

دارای همان عدد وضعیت Σ خواهد بود و بدوضع می‌باشد.

۳۳. (الف) مثالی خودتان بسازید و نشان دهید که یک مانده کوچک لزوماً تضمین نمی‌کند که جواب دقیق باشد.

(ب) اثباتی برای قضیه ۶-۷-۱ (قضیه مانده) ارائه دهید.

(ج) با استفاده از قضیه مانده، ثابت کنید که اگر یک الگوریتم یک مانده کوچک برای هر

ماتریس خوش وضع تولید کند، آنگاه به طور ضعیف پایدار است.

۳۴. (الف) برای چه مقادیری از a ماتریس زیر بدوضع است؟

$$A = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}$$

(ب) فرض کنید $a = 0.999$ دستگاه

$$Ax = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

را با استفاده از روش حذفی گاوس بدون محورگیری حل کنید.

(ج) عدد وضعیت A چند است؟

بخش ۶-۹

۳۵. تصفیه تکراری را در مورد دستگاه تمرین ۹ با استفاده از

$$x^{(1)} = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

به کار ببرید. $\text{Cond}(A)$ را از روی $x^{(1)}$ و $x^{(2)}$ برآورد کنید و آن را با مقدار واقعی

$$\text{Cond}(A) = 24/6414 \text{ مقایسه کنید.}$$

بخش ۶-۱۰

۳۶. دستگاه خطی تمرین ۱۴(الف) را ملاحظه کنید.

(الف) چرا هر دو روش ژاکوبی و گاوس سایدل با هر انتخاب دلخواهی از تقریب اولیه این

دستگاه همگرا می گردند.

(ب) پنج تکرار از هر دو روش را با تقریب اولیه یکسان

$$x^{(1)} = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

انجام دهید و نرخهای همگرایی را مقایسه کنید.

۳۷. مثالی بسازید تا نشان دهد که همگرایی ژاکوبی لزوماً همگرایی روش گاوس سایدل را ایجاب

نخواهد کرد.

۳۸. فرض کنید ماتریس $n \times n$ بعدی A به شکل

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,N} \\ A_{2,1} & A_{2,2} & \dots & A_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N,1} & A_{N,2} & \dots & A_{N,N} \end{bmatrix}$$

افراز شود که در آن هر ماتریس قطری A_{ii} مربعی و نامنفرد باشد. دستگاه خطی $Ax = b$ را در نظر بگیرید که A به صورت فوق و x و b به طور مناسب افراز شده باشند.

(الف) تکرارهای ژاکوبی بلوکی، گاوس سایدل بلوکی، SOR بلوکی را برای دستگاه خطی $Ax = b$ بنویسید (راهنمایی: بنویسید $A = L + D + U$ ، که در آن $D = \text{diag}(A_{1,1}, \dots, A_{N,N})$ و L و U به ترتیب ماتریسهای بالا و پایین مثلثی بلوکی اکید هستند).

(ب) اگر A معین مثبت متقارن باشد، آنگاه نشان دهید که $U = L^T$ و D معین مثبت است. در این حالت با استفاده از نتایج متناظر در حالتهای اسکالر، ثابت کنید با یک انتخاب دلخواه از تقریب اولیه روش گاوس سایدل بلوکی همواره همگرا است و SOR بلوکی همگرا است اگر و فقط اگر $0 < \omega < 2$.

۳۹. دستگاه بلوکی به وجود آمده از گسسته‌سازی معادله پواسون $u_{xxx} + u_{yyy} = f$ را در نظر بگیرید:

$$\begin{bmatrix} T & -I & & & \\ -I & T & -I & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -I \\ & & & -I & T \end{bmatrix} x = b$$

که در آن

$$T = \begin{bmatrix} 4 & -1 & & & \\ -1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & -1 \\ & & & -1 & 4 \end{bmatrix}$$

(الف) نشان دهید که در این حالت تکرار ژاکوبی به صورت زیر است

$$Tx_i^{(k+1)} = x_{i+1}^{(k)} + x_{i-1}^{(k)} + b_i, \quad i = 1, \dots, N$$

تکرارهای گاوس سایدل بلوکی و SOR بلوکی را برای این دستگاه بنویسید.

(ب) نشان دهید که هر ماتریس سه قطری بلوکی با بلوکهای قطری نامنفرد به طور سازگار مرتب شده و ۲-دوری است. ماتریس جایگشت P را برای حالتی که ماتریس ضرایب آن به صورت قسمت (الف) می باشد پیدا کنید.

۴۰. (الف) ثابت کنید که $\sqrt{\lambda_1} \|x\|_2 \leq \|x\|_A \leq \sqrt{\lambda_n} \|x\|_2$ ، که در آن A یک ماتریس معین مثبت متقارن با مقادیر ویژه $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ است.

(ب) با استفاده از نتیجه قسمت (الف) ثابت کنید که برای روش گرادیان مزدوج داریم:

$$\|x - x_k\|_2 \leq \sqrt{\kappa} \alpha^k \|x - x_0\|_2$$

۴۱. نشان دهید که روش ژاکوبی برای یک دستگاه معین مثبت متقارن 2×2 همگرا است.

۴۲. برای دستگاه تمرین ۳۹، مقادیر $\rho(B_J)$ ، $\rho(B_{GS})$ و ω_{opt} را با $N = 50, 100, 1000$ محاسبه کنید. نرخ همگرایی تکرار SOR با استفاده از مقدار بهینه ω_{opt} را در هر حالت با نرخ همگرایی روش گاوس سایدل، مقایسه کنید بدون آن که تکرارها را حقیقتاً انجام دهید.

۴۳. دستگاه قطری بلوکی

$$\begin{bmatrix} A & \cdot & \cdot & \dots & \cdot \\ \cdot & A & \cdot & \dots & \cdot \\ \vdots & \vdots & A & \dots & \vdots \\ \cdot & \cdot & \vdots & A & \cdot \\ \cdot & \cdot & \dots & \cdot & A \end{bmatrix}_{25 \times 25} b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

را در نظر بگیرید که در آن

$$A = \begin{bmatrix} 2 & -1 & & \cdot \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ \cdot & & -1 & 2 \end{bmatrix}$$

$\rho(B_{GS})$ ، $\rho(B_J)$ را محاسبه کنید و پیدا کنید که چگونه به یکدیگر وابسته هستند. دستگاه را با استفاده از پنج تکرار گاوس سایدل و SOR با مقدار بهینه ω حل کنید. نرخهای همگرایی را مقایسه کنید.

۴۴. ثابت کنید که انتخاب

$$\alpha = p^T(Ax - b) / p^TAp$$

تابع درجه دوم

$$\begin{aligned}\Phi_a &= \Phi(x - \alpha p) \\ &= \frac{1}{2} (x - \alpha p)^T A (x - \alpha p) - b^T (x - \alpha p)\end{aligned}$$

را می‌نیمیم می‌سازد.

۴۵. نشان دهید که بردارهای ویژه ماتریس A بردارهای جهتی هستند که توسط روش گرادیان مزدوج به دست می‌آیند.

۴۶. الف) الگوریتم تجزیه چولسکی ناقص را برای یک ماتریس سه قطری کاهش نیافته T به کار ببرید و نشان دهید که نتیجه تجزیه چولسکی معمولی T است. گزاره فوق را با ماتریس زیر تحقیق کنید.

$$T = \begin{bmatrix} 4 & -1 & & & \\ -1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 4 \end{bmatrix}_{5 \times 5}$$

ب) تکرار SOR را در مورد ماتریس T قسمت الف) با $x^{(0)} = (0, 0, 0, 0, 0)^T$

و $\omega = 1/5$ به کار ببرید و جدولی با نتایج تکرارها بسازید. b را تصادفی انتخاب کنید.

۴۷. فرض کنید p_0, p_1, \dots, p_{n-1} بردارهای جهت تولید شده توسط الگوریتم گرادیان مزدوج

پایه‌ای باشند. فرض کنید $r_k = b - Ax_k, k = 0, 1, \dots, n-1$. سپس ثابت کنید که

$$k = 0, 1, \dots, n-1, r_k \in \text{span}(p_0, \dots, p_k) \quad \text{الف)}$$

$$i = 0, 1, \dots, n-1, \text{span}(p_0, \dots, p_i) = \text{span}(r_0, Ar_0, \dots, A^i r_0) \quad \text{ب)}$$

ج) r_0, \dots, r_{n-1} دو به دو متعامد هستند.

۴۸. (چند شکافی) تکرار

$$x^{(k+1)} = Bx^{(k)} + d$$

را در نظر بگیرید که در آن B به صورت زیر ارائه می‌شود

$$B = \sum_{i=1}^k D_i B_i^{-1} C_i, \quad d = \left[\sum_{i=1}^k D_i B_i^{-1} \right] b$$

و

$$A = B_i - C_i, \quad i = 1, \dots, k, \quad \sum_{i=1}^k D_i = I \quad (D_i \geq 0)$$

روشهای ژاکوبی، گاوس سایدل و SOR را بر اساس این چند شکافی تدوین نمایید.

۴۹. روشهای ژاکوبی، گاوس سایدل و SOR (با عامل تخفیف بهینه) را برای دستگاه مثال بعد از قضیه ۵-۱۰-۶ (مثال ۵-۱۰-۶) به کار ببرید و گزاره مربوط به تعداد تکرارهای بیان شده در آن جا را در مورد روشهای متفاوت تحقیق کنید.

۵۰. اثباتی برای قضیه ۵-۱۰-۶ ارائه دهید.

۵۱. قضیه ۵-۱۰-۶ را ثابت کنید.

برنامه‌های MATLAB و MATCOM و مسائل

تذکر: شما به برنامه‌های `hagcond1`، `compiv`، `givqr`، `incompiv`، `inparpiv`، `inlu`، `lugsel`، `nichol`، `sucov`، `gaused`، `jacobi`، `iterref`، `invuptr` و `forelm` به نام `forelm` و مبتنی بر الگوریتم ۱-۴-۶ بنویسید:

$$[y] = \text{forelm}(L, b)$$

این برنامه یک دستگاه پایین مثلثی نامنفرد $Ly = b$ را با استفاده از روش حذفی پیشرو حل خواهد کرد.

(ب) یک برنامه MATLAB به نام `backsub` و مبتنی بر الگوریتم ۳-۱-۳ بنویسید:

$$[x] = \text{backsub}(U, b)$$

این برنامه یک دستگاه بالا مثلثی نامنفرد $Ux = b$ را حل خواهد کرد.

از ماتریسهای آزمایشی تولید شده تصادفی استفاده کنید و ماتریسهای تولید شده L و U با یک یا چند عنصر قطری کوچک را آزمون کنید.

(تذکر: `forelm` و `backsub` همچنین در MATCOM یادر ضمیمه B هستند.)

ماتریسهای آزمون برای مسائل ۲ تا ۸

برای مسائل ۲ تا ۸ ماتریسهای زیر را به عنوان ماتریسهای آزمون در نظر بگیرید. هنگامی که مسأله دستگاه خطی $Ax = b$ است، یک بردار به قسمی تولید کنید که بردار جواب یک بردار با همه مؤلفه‌های برابر ۱ باشد.

۱. ماتریس هیلبرت از مرتبه ۱۰

۲. ماتریس پای (*Pie*)، از مرتبه ۱۰

۳. ماتریس هانکل از مرتبه ۱۰

۴. ماتریس تولید شده تصادفی از مرتبه ۱۰

$$A = \begin{bmatrix} 0./\dots\dots 1 & 1 & 1 \\ \cdot & 0./\dots\dots 1 & 1 \\ \cdot & \cdot & 0./\dots\dots 1 \end{bmatrix} \quad 5.$$

۶. ماتریس واندروموند از مرتبه ۱۰

۲. الف) از **lugsel** (از **MATCOM**)، **backsub** و **forelm** استفاده کنید و برنامه **MATLAB** زیر را بنویسید تا دستگاه $Ax = b$ را با استفاده از روش حذفی گاوس بدون محورگیری حل کند.

$$[x] = \text{linsyswp}(A, b)$$

عامل رشد، زمان اجرا، و تعداد اعمال ممیز شناور را برای هر دستگاه حساب کنید.

ب) برنامه **inlu** را از **MATCOM** اجرا کنید و نتیجه را در بردار b ضرب کنید تا بردار

جواب $x = A^{-1}b$ به دست آید. تعداد اعمال ممیز شناور را حساب کنید.

ج) جوابهای محاسبه شده و تعداد اعمال ممیز شناور قسمتهای الف) و ب) را مقایسه کنید.

۳. الف) از **parpiv** و **elmul** از فصل ۵ و **backsub** استفاده کنید و برنامه **MATLAB**

$$[x] = \text{linsyspp}(A, b)$$

را بنویسید تا $Ax = b$ را با استفاده از روش حذفی گاوس با محورگیری جزئی حل کند.

عامل رشد، زمان اجرا، و تعداد اعمال ممیز شناور را برای هر دستگاه حساب کنید.

ب) برنامه **inparpiv** را از **MATLAB** اجرا کنید و نتیجه را در بردار b ضرب کنید تا بردار

جواب $x = A^{-1}b$ به دست آید تعداد اعمال ممیز شناور را برای هر دستگاه حساب کنید.

(ج) جوابهای محاسبه شده و تعداد اعمال ممیز شناور و زمان اجرای (الف) و (ب) را مقایسه کنید.

۴. (الف) با استفاده از **compiv** از **MATCOM**، **elmul** از فصل ۵ و **backsub**، برنامه **MATLAB**

$$[x] = \text{linsyscp}(A, b)$$

را بنویسید تا $Ax = b$ را با استفاده از روش حذفی گاوس با محورگیری کامل حل کند. تعداد اعمال ممیز شناور، زمان اجرا، و عامل رشد را برای هر دستگاه حساب کنید.

(ب) برنامه **incompiv** را از **MATCOM** اجرا کنید و نتیجه را در بردار b ضرب کنید تا بردار جواب $x = A^{-1}b$ به دست آید تعداد اعمال ممیز شناور را برای هر دستگاه حساب کنید.

(ج) جوابهای محاسبه شده و تعداد اعمال ممیز شناور و زمان اجرای (الف) و (ب) را مقایسه کنید.

۵. (الف) الگوریتم ۶-۴-۲ را برای حل $Ax = b$ بدون تجزیه صریح با استفاده از محورگیری جزئی پیاده‌سازی کنید:

$$[x] = \text{linsyswf}(A, b)$$

(ب) A^{-1} را با استفاده از تجزیه صریح محاسبه کنید.

۶. (الف) با استفاده از **housqr** از فصل ۵ (یا تابع **qr** از **MATLAB**) و **backsub**، برنامه **MATLAB**

$$[x] = \text{linsysqrh}(A, b)$$

را بنویسید تا $Ax = b$ را با استفاده از تجزیه **QR** با ماتریسهای هاوس هولدر حل کند. تعداد اعمال ممیز شناور را برای هر دستگاه حساب کنید.

(ب) (الف) را با **givqr** به جای **housqr** تکرار کنید؛ یعنی یک برنامه **MATLAB** به نام

linsysqrg برای حل دستگاه $Ax = b$ با استفاده از روش گیونز برای تجزیه **QR** بنویسید.

۷. (هدف از این تمرین انجام یک مطالعه مقایسه‌ای از نظر دقت، زمان اجرا، تعداد اعمال

ممیز شناور و عامل رشد برای روشهای مختلف برای حل دستگاه $Ax = b$ می باشد. نتایج مسائل ۲ تا ۶ را به صورت جدول ۳-۶ جدول بندی کنید. برای هر ماتریس یک جدول تشکیل دهید. \hat{x} جواب واقعی را نشان می دهد.

جدول ۳-۶ مقایسه روشهای مختلف برای مسائل دستگاه خطی

روش	جواب محاسبه شده \hat{x}	خطای نسبی $\ x - \hat{x}\ / \ x\ $	مانده $\ b - A\hat{x}\ $	عامل رشد	زمان اجرا
linsyswp					
linsyspp					
linsyscp					
linsyswf					
linsysqrh					
linsysqrh					
$A^{-1}b$					

۸. (الف) یک برنامه MATLAB برای پیدا کردن معکوس A با استفاده از **housqr** (یا تابع **qr** از MATLAB) و **invuptr** (از MATCOM) بنویسید.

$$[A] = \text{invqrh}(A)$$

تعداد اعمال ممیز شناور را برای هر ماتریس حساب کنید.

(ب) قسمت (الف) را با استفاده از **givqr** و **invuptr** تکرار کنید.

$$[A] = \text{invqrg}(A)$$

تعداد اعمال ممیز شناور را برای هر ماتریس حساب کنید.

(ج) برنامه های **inlu**، **inparpiv**، **incompiv** از MATCOM را با هر کدام از ماتریسهای

داده شده اجرا کنید. یک جدول برای هر ماتریس A برای مقایسه روشهای مختلف برای پیدا کردن معکوس، از نظر دقت و تعداد اعمال ممیز شناور بسازید. معکوس محاسبه شده را توسط \hat{A} نمایش دهید. A^{-1} را با استفاده از دستور **inv(A)** از MATCOM به دست آورید. جدول ۳-۶ را ملاحظه کنید.

۹. (الف) یک برنامه MATLAB به نام **chlgau** برای پیدا کردن تجزیه چولسکی یک

ماتریس متقارن معین مثبت A با استفاده از روش حذفی گاوس با محورگیری بنویسید.

$$[H] = \text{chlgau}(A)$$

یک ماتریس پایین مثلثی 15×15 مانند L با عناصر قطری مثبت بسازید، برخی از عناصر قطری را به اندازه کافی کوچک اختیار کنید تا خیلی نزدیک به صفر باشند. آن را در L^T ضرب کنید و $A = LL^T$ را به عنوان ماتریس آزمایش خودتان اختیار کنید. H را محاسبه کنید. تعداد اعمال ممیز شناور را حساب کنید.

(ب) برنامه **chol** از **MATLAB** را بر روی همان ماتریس قسمت (الف) اجرا کنید و ترانهاده نتیجه را توسط \hat{H} نمایش دهید. تعداد اعمال ممیز شناور را حساب کنید.

(ج) نتایج (الف) و (ب) را مقایسه کنید. (توجه کنید که $\text{chol}(A)$ یک ماتریس بالا مثلثی H ارائه می‌دهد به طوری که $A = HH^T$).

۶-۴ مقایسه روشهای مختلف برای محاسبه معکوس

تعداد اعمال ممیز شناور	خطای نسبی $\ A^{-1} - (\hat{A})^{-1}\ $	روش
		inlu
		inparpiv
		incompiv
		invqrh
		invqrg

۱۰. برنامه **linsyswp** را با ماتریسهای غالب قطری، سه قطری متقارن، سه قطری بلوکی ارائه شده در بخش ۶-۳-۵ اجرا کنید. بردار b طرف راست را به گونه‌ای اختیار کنید که بردار جواب x از پیش معلوم باشد. جواب واقعی x را با جواب محاسبه شده \hat{x} مقایسه کنید. (هدف از این تمرین تحقیق این است که در حل یک دستگاه معین مثبت متقارن محورگیری لازم نیست تا پایداری را در روش حذفی گاوس تضمین نماید).

۱۱. (الف) یک برنامه **MATLAB** برای پیاده‌سازی الگوریتم ۶-۷-۲ بنویسید که یک کران بالا در ۲- نرم برای معکوس یک ماتریس بالا مثلثی U پیدا کند:

$$[CEBOUND] = \text{norminvtr}(U)$$

نتیجه خودتان را با تولید تصادفی یک ماتریس بالا مثلثی 10×10 با چندین عنصر قطری کوچک امتحان کنید. سپس نتیجه خود را با آنچه که با اجرای دستور MATLAB

$$\text{norm}(\text{inv}(U))$$

به دست می آورید، مقایسه کنید.

(ب) اکنون عدد وضعیت U را به صورت زیر محاسبه کنید.

$$\text{norm}(U) * \text{norminvtr}(U)$$

نتیجه خودتان را با آنچه با اجرای دستور MATLAB ، $\text{cond}(U)$ به دست می آورید، مقایسه کنید. تحقیق کنید که

$$\text{cond}(U) \geq \frac{\max(u_{ij})}{\min(u_{ij})}$$

(از همان ماتریس آزمایشی قسمت (الف) استفاده کنید).

۱۲. (الف) هدف از این تمرین مقایسه راههای مختلف برای برآورد عدد وضعیت یک ماتریس می باشد. عدد وضعیت هر یک از ماتریسهای A از مرتبه ۱۰ زیر را محاسبه و یا برآورد کنید: ماتریسهای هیلبرت، پای، تولید شده به طور تصادفی، و اندرموند، و هانکل. از راههای زیر استفاده کنید.

(i) تجزیه QR ماتریس A را با محورگیری ستونی پیدا کنید: $Q^T A P = R$ - نرم معکوس R را با اجرای برنامه norminvtr بر روی R برآورد کنید. اکنون $\text{norm}(R) * \text{norminvtr}(R)$ را حساب کنید. تعداد اعمال ممیز شناور را محاسبه کنید.

(ii) با استفاده از hagcond1 برآوردی در ۱ - نرم برای عدد وضعیت A حساب کنید. (ب) نتایج و تعداد اعمال ممیز شناور را مقایسه کنید.

در مسائل ۱۳، ۱۴، ۱۵، و ۱۸ بردار b را به طور تصادفی تولید کنید.

۱۳. (الف) برنامه تصفیه تکراری iterref را از MATCOM بر روی هر یک از دستگانههای 15×15

زیر اجرا کنید: هیلبرت، پای، و اندرموند، تولید شده به طور تصادفی، هانکل. از جواب به دست آمده از برنامه **linsyspp** (مسأله ۳) به عنوان تقریب اولیه $x^{(0)}$ استفاده کنید.

(ب) عدد وضعیت هر یک از ماتریسهای قسمت (الف) را با استفاده از روش تصفیه تکراری برآورد کنید.

(ج) نتایج خودتان را که مربوط به برآورد عدد وضعیت می باشد با نتایج به دست آمده در مسأله ۱۲ مقایسه کنید.

۱۴. برنامه های **gaused**، **jacobi**، و **suconv** از **MATCOM** را بر روی ماتریس 6×6 بعدی A مثال ۶-۱۰-۵ با همان بردار آغازین $x^{(0)} = (0, 0, \dots, 0)^T$ اجرا کنید. تعیین کنید که هر روش چند تکرار برای همگرایی اختیار خواهد کرد. اظهارات این مثال را تحقیق کنید که بیان می دارد که ۵ تکرار برای همگرایی **SOR** با ω_{opt} و ۱۸ تکرار برای ژاکوبی لازم می باشد.

۱۵. برنامه های **gaused** و **jacobi** از **MATCOM** را بر روی مثال ۶-۱۰-۳ اجرا کنید و اظهارات مثال را تحقیق کنید.

۱۶. برنامه **nichol** از **MATCOM** را که پیاده سازی تجزیه چولسکی ناقص بدون پر شدن می باشد، بر روی ماتریس معین مثبت متقارن سه قطری T از مرتبه ۲۰ و به وجود آمده از گسسته سازی معادله پواسون اجرا کنید. نتیجه خودتان را با آنچه که از اجرای **chol(T)** بر روی T به دست می آید مقایسه کنید.

۱۷. یک برنامه **MATLAB** به نام **arnoldi** مبتنی بر روش آرنولد بنویسید (الگوریتم ۶-۱۰-۸) از الگوریتم اصلاح شده گرام اشمیت استفاده کنید. (گرام اشمیت اصلاح شده در برنامه **mdgrsch** در **MATCOM** پیاده سازی شده است (فصل ۷ را ملاحظه کنید).

۱۸. از فرایند **arnoldi** و یک روش کمترین توانهای دوم از فصل ۷ استفاده کنید و یک برنامه **MATLAB** به نام **gmres** بنویسید که الگوریتم **GMRES** را پیاده سازی نماید (الگوریتم ۶-۱۰-۹).

پیشنهادها برای مطالعه بیشتر

کتابهای مربوط به روشهای عددی در رشته‌های مهندسی به طور معمول در مورد چگونه منجر شدن کاربردهای گوناگون مهندسی به مسائل دستگاههای خطی بحث می‌کنند. از دو کتاب زیر در بحثهایمان استفاده کرده‌ایم و آنها را مفید یافته‌ایم:

Chapra, Steven C., and Raymond P. Canale. *Numerical Methods for Engineers*, 2nd ed. New York: McGraw-Hill, 1988.

O'Neil, Peter V. *Advanced Engineering Mathematics*, 3d ed. Belmont, Calif.: Wadsworth, 1991.

برای شناخت مفاهیم و نتایج اساسی درباره معادلات دیفرانسیل معمولی و جزئی، هر کتاب درسی در مورد معادلات دیفرانسیل می‌تواند مطالعه شود. ما بویژه کتابهای زیر را توصیه می‌کنیم.

Braun, M. *Differential Equations and Their Applications*, New York: Springer Verlag, 1978.

Campbell, S. L. *An Introduction to Differential Equations and Their Applications*, 2nd ed. Belmont, Calif.: Wadsworth, 1990.

Weinberger, H. F. *A First Course in Partial Differential Equations*, New York: Wiley, 1965.

برای روشهای گسسته‌سازی معادلات دیفرانسیل جزئی با استفاده از تفاضلات متناهی و عناصر متناهی کتابهای زیر را مطالعه کنید.

Strang (1986), Strang and Fix (1973), Strikwerda (1989), Reddy (1993), Celia and Gray (1992), Chapra and Canale (1988), and Ames (1992).

روشهای مستقیم (نظیر روش حذفی گاوس، تجزیه QR، و غیره) برای دستگاههای خطی و مسائل وابسته، بحثهای مربوط به تحلیل اختلال و وضعیت مسائل دستگاههای خطی، تصفیه تکراری، و غیره را می‌توان در کتابهای درسی جبر خطی عددی استاندارد پیدا کرد. بویژه کتابهای زیر اکیداً توصیه می‌شوند.

Golub and Van Loan (1989), Stewart (1973).

بیشتر کتابهای درسی آنالیز عددی حاوی بحثهایی از این روشها هستند، اما هیچ کدام از کتابهای موجود بحثی کامل و عمیق را فراهم نمی‌نمایند. برای بحث در مورد جوابهای دستگاههای خطی با

ماتریسهای خاص، نظیر غالب قطری، هستبرگی، و معین مثبت کتاب Wilkinson (1965, 218-220) را ملاحظه کنید.

برای اثباتها و تحلیلهای خطای پسروی الگوریتمهای مختلف کتاب Wilkinson (1965) یک مرجع کلاسیک است. چندین مقاله که اخیراً توسط نیک هایام (Nick Higham 1987; 1990) برای برآوردگرهای عدد و وضعیت ارائه شده‌اند برای خواندن جالب می‌باشند. برای روشهای تکراری دو کتاب تقریباً کلاسیک و اساسی (تاکنون) عبارتند از:

Varga, Richard. *Matrix Iterative Analysis*, Englewood Cliffs, N. J.: Prentice Hall, 1962

Young, David. *Iterative Solution of Large Linear Systems*, New York: Academic Press, 1971.

کتاب مهم دیگر در این زمینه عبارت است از:

Applied Iterative Methods, by L. A. Hageman and D. M Young (1981).

جدیدترین کتاب در این زمینه عبارت است از:

Templates for the Solution of Linear Systems: Building Blocks For Iterative Methods by Richard Barrett and others (1994).

این کتاب جدیدترین روشهای محاسباتی و فن‌آوری را برای حل دستگاههای نامتقارن تنک و بزرگ در بر دارد. در این زمینه همچنین مقاله عالی و جدید (Freund, Golub, and Nachtigal (1992) و کتابهای (Saad (1995) و (Axelsson (1994) را ملاحظه کنید.

کتاب

Introduction to Parallel and Vector Solutions of Linear Systems and Numerical Analysis: A Second Course by James Ortega (1988a)

نیز حاوی توصیفهای خیلی روشن در مورد همگرایی روشهای تکراری است.

روش‌گردان مزدوج که در اصل توسط هستنز و استایفل (M.R.Hestenes and E. Stiefel (1952)) تدوین شده است، به طور قابل ملاحظه‌ای در زمینه حل دستگاههای معین مثبت تنک و بزرگ مورد توجه قرار گرفته است. قسمت قابل توجهی از کار توسط متخصصین جبر عددی خطی و محققان در زمینه‌های کاربردی (نظیر بهینه‌سازی) انجام شده است. کتابهای زیر حاوی برخی بحثهای عمیق هستند.

Luenberger, David G. *Introduction to linear and Nonlinear Programming*, New York:

Addison-Wesley, 1973.

Ortega, James. *Introduction to Parallel and Vector Solutions of Linear Systems*, New York: Plenum Press, 1988a.

Shewchuk, Jonathan R. *An Introduction to the Conjugate Gradient Method without the Agonizing Pain*, Technical Report CMU-CS-94-125 of Carnegie Mellon University, 1994.

در این زمینه همچنین مقاله مروری عالی Golub and O'Leary (1989) را برای مطالعه بیشتر توصیه می‌کنیم. مقاله مروری Axelsson (1985) را نیز ملاحظه کنید. برای تحلیل پایداری طرحهای تجزیه LU ناقص مرجع Elman (1986) را ملاحظه کنید.

مقاله مروری و جالب Young, Jea, and Mai (1988)، در کتاب زیر ارزش خواندن را دارد.

Linear Algebra in Signals, Systems, and Control

توسعه روشهای از نوع گرادیان مزدوج برای دستگاههای خطی نامتقارن و متقارن نامعین یکی از زمینه‌های فعال تحقیق می‌باشد.

یک بحث خوب و مفید در مورد مقیاس کردن در مرجع Forsythe and Moler (1967, chapter 11) ارائه شده است. مقاله Skeel (1979) را نیز برای روابط بین مقیاس کردن و پایداری روش حذفی گاوس و مقاله Edelman and Stewart (1993) را در مورد مقیاس کردن و تعامد ملاحظه کنید.

مقاله جدید Chandrasekaran and Ipsen (1994b) حساسیت مؤلفه‌های فردی بردار جواب را توصیف می‌کند هنگامی که داده‌ها تحت اختلال قرار می‌گیرند.

برای یادگیری بیشتر در مورد روشهای زیر فضای کريلف نظیر GMRES، QMR و گرادیان مزدوج مقاله‌های زیر را ملاحظه کنید.

Bai, Hu, and Reichel (1994), Bramley and Sameh (1992), Saad (1981, 1984, 1987), Saad and Schultz (1986), Sorensen (1992), Freund (1993), Freund and Nachtigal (1991), Gutknecht (1992), Elman, Saad, and Saylor (1986), Greenbaum and Strakos (1992), Nachtigal, Reichel, and Trefethen (1992), Parlett (1980a, 1989), Pierce and Plemmons (1988), Saylor, Ashby, and Manteuffel (1990), Walker (1988), Woznicki (1993), Young, Jea, and Mai (1988), Saad (1992), Barrett et al. (1994).

برای کاربردهای روشهای زیر فضای کريلف در مسائل کنترل مراجع زیر را ملاحظه کنید.

Boley and Golub (1984), Choudhary (1994), Datta and Saad (1991), Hu and Reichel (1992), Hodel and Poola (1992), and Saad (1988, 1990).

برای ماتریسهای چند شکافی و کاربردهای آنها در حل موازی دستگاههای خطی مراجع زیر را ملاحظه کنید.

Ortega (1988a), O'Leary and White (1985), Neumann and Plemmons (1987), and White (1990).

برای محاسبه ماتریسهای تنک بر روی MATLAB مقاله جدید Gilbert, Moler, and Schreiber (1992) را ملاحظه کنید. یک کتاب معتبر قدیمی تر در زمینه حل دستگاههای معین مثبت تنک و بزرگ کتاب George and Liu (1981) است. یک کتاب جدیدتر در مورد روشهای مستقیم برای محاسبات ماتریسی تنک کتاب Duff, Erisman, and Ried (1986) است. برای برآورد عدد و وضعیت ماتریسهای تنک مراجع زیر را ملاحظه کنید.

Bischof, Lewis, and pierce (1990), Ferng, Golub, and Plemmons (1991), and Pierce and Plemmons (1992).

برای وضعیت ماتریسهای ساخت یافته مرجع Gohberg and Van Dooren (1991) و Koltracht (1993) را ملاحظه کنید. برای مطالب بیشتر در مورد تحلیل اختلالها و وضعیت مسائل دستگاه خطی و تجزیه‌های ماتریسی مختلف مراجع زیر را ملاحظه کنید.

Stewart and Sun (1990), Demmel (1987b, 1989), Edelman (1988, 1992), McCarthy and Strang (1973), Higham (1991), Stewart (1977b, 1993a), Sun (1991), Zha (1993), O'Leary (1980), and Drmac, Omladic, and Veselic (1994).

مقاله جدید Eiermann and Varga (1993) را برای بحثی در مورد عامل تخفیف بهینه برای روش تکراری SOR ملاحظه کنید.

جوابهای کمترین توانهای دوم برای دستگاههای خطی

۱-۷ مقدمه

در فصل ۶ چندین روش را برای حل دستگاه خطی

$$Ax = b$$

مورد بحث قرار دادیم و فرض کردیم که A مربعی و نامنفرد است؛ و لیکن در برخی وضعیتهای عملی نظیر کاربردهای آماری، مدل‌سازی هندسی، و پردازش سیگنال، نیاز به حل یک دستگاه داریم که در آن ماتریس A نامربعی و / یا منفرد است. در چنین حالتی ممکن است به هیچ وجه جواب وجود نداشته باشد؛ در حالتی که جواب یافت می‌شود، ممکن است بی‌نهایت جواب وجود داشته باشد. برای مثال هنگامی که A یک ماتریس $m \times n$ با $m > n$ است، یک دستگاه فرامعین داریم؛ (یعنی تعداد معادلات، بزرگ‌تر از تعداد مجهولات است)، و یک دستگاه فرامعین معمولاً جواب ندارد. برعکس یک دستگاه فرامعین $m < n$ معمولاً یک تعداد نامتناهی جواب دارد.

در این حالتها بهترین وضعیتی که انتظارش را می‌توان داشت این است که یک بردار x پیدا کنیم که Ax را تا حد ممکن به بردار b نزدیک سازد. به بیان دیگر، یک بردار x جستجو می‌کنیم به قسمی که

$\|r(x)\| = \|Ax - b\|$ را کمینه سازد. هنگامی که نرم اقلیدسی $\|\cdot\|$ استفاده می‌شود به این جواب به‌عنوان یک جواب کم‌ترین توانهای دوم برای دستگاه $Ax = b$ اطلاق می‌گردد. عبارت جواب کم‌ترین توانهای دوم قابل توجیه است؛ زیرا جوابی است که نرم اقلیدسی بردار مانده را می‌نیمم می‌سازد و طبق تعریف توان دوم نرم اقلیدسی یک بردار دقیقاً برابر مجموع توانهای دوم مؤلفه‌های آن بردار می‌باشد. مسأله پیدا کردن جوابهای کم‌ترین توانهای دوم برای دستگاه $Ax = b$ به مسأله کم‌ترین توانهای دوم خطی (LSP) معروف است. مسأله کم‌ترین توانهای دوم به طور رسمی به صورت زیر تعریف می‌شود.

بیان مسأله کم‌ترین توانهای دوم

با مفروض بودن یک ماتریس حقیقی $m \times n$ مانند A و یک بردار b ، یک بردار حقیقی x به قسمی پیدا کنید که تابع $\|r(x)\| = \|Ax - b\|$ می‌نیمم شود.

اگر مسأله کم‌ترین توانهای دوم بیش از یک جواب داشته باشد، جوابی که دارای نرم اقلیدسی می‌نیمم می‌باشد جواب با طول می‌نیمم، یا جواب با می‌نیمم نرم، نامیده می‌شود. این فصل به مطالعه چنین مسائلی اختصاص داده شده است. تنظیم این فصل به صورت زیر است. در بخش ۷-۲ نشان می‌دهیم که چگونه یک کاربرد تجاری خیلی ساده به یک مسأله کم‌ترین توانهای دوم فرامعین تبدیل می‌شود. در این بخش به طور ساده مسأله را به یک مسأله کم‌ترین توانهای دوم فرمول‌بندی می‌کنیم؛ در بخش ۷-۸ یک جواب برای این مسأله با استفاده از معادلات نرمال ارائه خواهیم داد.

در بخش ۷-۳ قضیه‌ای در مورد وجود و یکتایی جواب یک مسأله کم‌ترین توانهای دوم ثابت می‌کنیم. برخی کاربردها که منجر به دستگاههای فرامعین می‌شوند در بخش ۷-۲ و ۷-۵ مورد بحث قرار می‌گیرند. در بخش ۷-۷ حساسیت مسائل کم‌ترین توانهای دوم ناشی از اختلال در داده‌ها را تجزیه و تحلیل خواهیم کرد. ما فقط در اینجا یک نتیجه ساده را ثابت می‌کنیم و سایر نتیجه‌ها را بدون اثبات بیان خواهیم کرد. در بخش ۷-۴ تعبیر هندسی از مسأله کم‌ترین توانهای دوم ارائه خواهیم داد. در بخش ۷-۶ در مورد رابطه بین مسأله کم‌ترین توانهای دوم و شبه معکوس یک ماتریس بحث خواهیم کرد.

بخش ۷-۸ به بحث در مورد روشهای محاسباتی برای هر دو مسأله فرامعین با رتبه کامل و یا رتبه ناقص می‌پردازد. در مورد روش معادلات نرمال و روشهای تجزیه QR با استفاده از تبدیلهای هاوس هولدر، متعامد سازی گرام اشمیت کلاسیک و گرام اشمیت اصلاح شده بحث خواهیم کرد.

مسائل دستگاههای فرومعین را در بخش ۷-۹ ملاحظه خواهیم کرد. در اینجا دوباره در مورد معادلات نرمال و روشهای QR برای یک مسأله فرومعین بحث خواهیم کرد. در بخش ۷-۱۰ یک رویه بهبود تکراری برای تصفیه جوابهای تقریبی ارائه می‌شود. در بخش ۷-۱۱ یک راه کار آرا برای محاسبه ماتریس وارینانس - کوواریانس از جواب کم‌ترین توانهای دوم، که برابر $(ATA)^{-1}$ می‌باشد، توصیف خواهیم کرد.

ابزار اصلی زیر و مفاهیم قبلاً ذکر شده برای یادگیری و فهم راحت مواد این فصل لازم خواهند بود: الگوریتم تجزیه چولسکی (الگوریتم ۶-۴-۴)، الگوریتمهای تجزیه QR با استفاده از ماتریسهای هاوس هولدر و گیونز (الگوریتم ۵-۴-۳ و ۵-۵-۴)، تجزیه QR یک ماتریس نامربعی (بخش ۵-۴)، محاسبات پایه‌های یکا متعامد و تصویرهای متعامد (بخش ۵-۶)، تجزیه QR با محورگیری ستونی (بخش ۵-۷)، الگوریتم تصفیه تکراری (الگوریتم ۶-۹-۱)، و نتایج تحلیل اختلال مسائل دستگاههای خطی (بخشهای ۶-۶ و ۷-۶).

۷-۲ یک کاربرد ساده و منجر به یک دستگاه فرامعین

فرض کنید که تعداد واحدهای b_i از یک کالای فروخته شده توسط یک شرکت در ناحیه i از یک شهر بستگی به جمعیت آن شهر a_{i1} (بر حسب هزار نفر) و درآمد سرانه a_{i2} (بر حسب دلار) دارد. جدول زیر گرفته شده از مرجع (Neter, Wasserman, and Kunter (1983) گردآوری شده توسط شرکت، فروشها در پنج ناحیه، همچنین جمعیت متناظر و درآمد سرانه را نشان می‌دهد.

ناحیه	فروشهای	جمعیت	درآمد سرانه
i	b_i	a_{i1}	a_{i2}
۱	۱۶۲	۲۷۴	۲۴۵۰
۲	۱۲۰	۱۸۰	۲۲۵۴
۳	۲۲۳	۳۷۵	۲۸۰۲
۴	۱۳۱	۲۰۵	۲۸۲۸
۵	۶۷	۸۶	۲۳۴۷

فرض کنید شرکت می‌خواهد از این جدول برای پیش‌بینی فروشهای آینده استفاده کند و باور دارد (از تجربه‌های گذشته) که رابطه زیر بین b_i ، a_{ij} و a_{ij} مناسب است.

$$b_i = x_1 + a_{i1}x_2 + a_{i2}x_3$$

اگر داده‌های جدول در این رابطه صدق نمایند، داریم

$$162 = x_1 + 274x_2 + 2450x_3$$

$$120 = x_1 + 180x_2 + 3254x_3$$

$$223 = x_1 + 375x_2 + 3802x_3$$

$$131 = x_1 + 205x_2 + 2838x_3$$

$$67 = x_1 + 86x_2 + 2347x_3$$

یا $Ax = b$ که در آن

$$A = \begin{pmatrix} 1 & 274 & 2450 \\ 1 & 180 & 3254 \\ 1 & 375 & 3802 \\ 1 & 205 & 2838 \\ 1 & 86 & 2347 \end{pmatrix}, \quad b = \begin{pmatrix} 162 \\ 120 \\ 223 \\ 131 \\ 67 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

این یک دستگاه فرامعین از پنج معادله و سه مجهول است.

۳-۷ وجود و یکتایی جواب

نظیر حالت حل دستگاه خطی، طبیعتاً سؤالهایی مطرح می‌شود

۱. آیا همیشه یک جواب کم‌ترین توانهای دوم برای $Ax = b$ وجود دارد؟

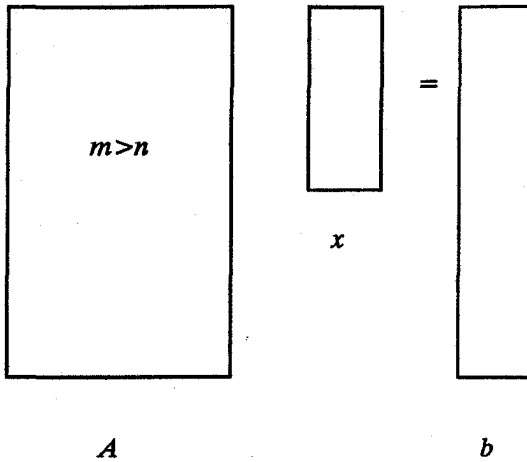
۲. اگر وجود دارد، آیا یکتا است؟

۳. چگونه چنین جوابی را به دست آوریم؟

قضیه زیر به سؤالهای وجود و یکتایی پاسخ می‌دهد. فرض کنید که دستگاه $Ax = b$ فرامعین یا

مربعی باشد، یعنی A از مرتبه $m \times n$ باشد، که در آن $m \geq n$. یک دستگاه فرامعین $Ax = b$ می‌تواند

به صورت نموداری به شکل ۷-۱ نشان داده شود.



شکل ۷-۱

حالت فرومعین در این فصل بعداً مورد بحث قرار خواهد گرفت.

قضیه ۷-۳-۱ قضیه وجود و یکتایی جواب. همیشه یک جواب برای مسأله کمترین توانهای دوم وجود دارد. این جواب یکتا است اگر و فقط اگر A دارای رتبه کامل باشد؛ یعنی $\text{rank}(A) = n$ اگر A دارای رتبه ناقص باشد، آنگاه مسأله کمترین توانهای دوم دارای جوابهای بیشماری خواهد بود.

در اینجا اثباتی برای حالت رتبه کامل، یعنی برای حالتی که A دارای رتبه کامل است، ارائه می‌دهیم. حالت رتبه ناقص بعداً در این فصل و فصل ۱۰ بررسی خواهد شد. ابتدا لم زیر را ملاحظه می‌کنیم.

لم ۷-۳-۱ یک جواب کمترین توانهای دوم x است اگر و فقط اگر در رابطه زیر صدق کند.

$$A^T A x = A^T b \quad (۷-۳-۱)$$

اثبات: مانده $r = b - Ax$ را توسط $r(x)$ نمایش می‌دهیم تا بر این تأکید نماییم که با داشتن A و b ، مانده r تابعی از x است. فرض کنید y یک بردار n بعدی باشد. آنگاه $r(y) = b - Ay = r(x) + Ax - Ay = r(x) + A(x - y)$ بنابراین

$$\|r(y)\|_p^2 = \|r(x)\|_p^2 + 2(x - y)^T A^T r(x) + \|A(x - y)\|_p^2$$

ابتدا فرض کنید که x در رابطه

$$A^T Ax = A^T b$$

صدق کند، یعنی $A^T r(x) = 0$. آنگاه از قبل داریم:

$$\|r(y)\|_V^2 = \|r(x)\|_V^2 + \|A(x-y)\|_V^2 \geq \|r(x)\|_V^2$$

این رابطه ایجاب می‌کند که x یک جواب کم‌ترین توانهای دوم باشد.

سپس فرض کنید که $A^T r(x) \neq 0$. قرار دهید $A^T r(x) = z \neq 0$. یک بردار y به قسمی تعریف می‌کنیم که

$$y = x + cz$$

آنگاه برای c به اندازه کافی کوچک خواهیم داشت:

$$r(y) = r(x) + A(x-y) = r(x) - cz$$

$$\begin{aligned} \|r(y)\|_V^2 &= \|r(x)\|_V^2 + c^2 \|Az\|_V^2 - 2cz^T A^T r(x) \\ &= \|r(x)\|_V^2 + c^2 \|Az\|_V^2 - 2c \|z\|_V^2 < \|r(x)\|_V^2 \end{aligned}$$

این ایجاب می‌کند که x یک جواب کم‌ترین توانهای دوم نباشد. □

اثبات قضیه ۷-۳-۱ (در حالت رتبه کامل): از لم ۷-۳-۱ می‌دانیم که x یک جواب کم‌ترین توانهای دوم برای $Ax = b$ است اگر و فقط اگر x در رابطه زیر صدق کند.

$$A^T Ax = A^T b$$

اکنون نشان می‌دهیم که در حالت رتبه کامل، این جواب یکتا است.

از قضیه ۶-۲-۱ می‌دانیم که جواب دستگاه $A^T Ax = A^T b$ یکتا است اگر و فقط اگر $A^T A$ نامنفرد باشد. بنابراین همه آنچه که لازم است انجام دهیم این است که نشان دهیم که $A^T A$ نامنفرد است اگر و فقط اگر A دارای رتبه کامل باشد. برای این مورد آخر، ما یک نتیجه قوی‌تر که $A^T A$ معین مثبت است اگر و فقط اگر A دارای رتبه کامل باشد را ثابت خواهیم کرد.

اگر A دارای رتبه کامل باشد، آنگاه $x \neq 0$ ایجاب می‌کند که $y = Ax \neq 0$. بنابراین $\langle y^T y \rangle = x^T A^T A x = y^T y > 0$. $A^T A$ معین مثبت باشد. از طرف دیگر اگر A دارای رتبه کامل نباشد، برای یک $x \neq 0$ خواهیم داشت $Ax = 0$ ، و در نتیجه $x^T A^T A x = 0$ ، که ایجاب می‌کند $A^T A$ معین مثبت نباشد.

۴-۷ تعبیر هندسی مسأله کمترین توانهای دوم

فرض کنید A یک ماتریس $m \times n$ با $m > n$ باشد. آنگاه A یک نگاشت خطی از $\mathbb{R}^n \rightarrow \mathbb{R}^m$ است. $\mathbb{R}(A)$ یک زیر فضا از \mathbb{R}^m است. هر بردار $u \in \mathbb{R}(A)$ می تواند به صورت $u = Ax$ به ازای برداری مانند $x \in \mathbb{R}^n$ نوشته شود. فرض کنید $b \in \mathbb{R}^m$. چون $\| \cdot \|_2$ یک نرم اقلیدسی است، $\| b - Ax \|_2$ فاصله بین نقاط انتهایی b و Ax می باشد. واضح است که این فاصله می نیمم است اگر و فقط اگر $b - Ax$ بر $\mathbb{R}(A)$ عمود باشد (شکل ۷-۲ را ملاحظه کنید). در این حالت $\| b - Ax \|_2$ فاصله از نقطه انتهایی b تا «صفحه» $\mathbb{R}(A)$ است.

از این تعبیر به آسانی می فهمیم که یک جواب مسأله کمترین توانهای دوم برای دستگاه خطی $Ax = b$ همواره وجود دارد. این جواب وجود دارد زیرا می توانیم b را بر روی صفحه $\mathbb{R}(A)$ تصویر کنیم و یک بردار $u \in \mathbb{R}(A)$ به دست آوریم، و یک بردار $x \in \mathbb{R}^n$ وجود دارد به قسمی که $u = Ax$. این x یک جواب است. چون $b - Ax$ بر $\mathbb{R}(A)$ عمود است و هر بردار در $\mathbb{R}(A)$ یک ترکیب خطی از بردارهای ستونی A است، آنگاه $b - Ax$ بر هر بردار ستونی A عمود است؛ یعنی

$$A^T(b - Ax) = 0.$$

یا

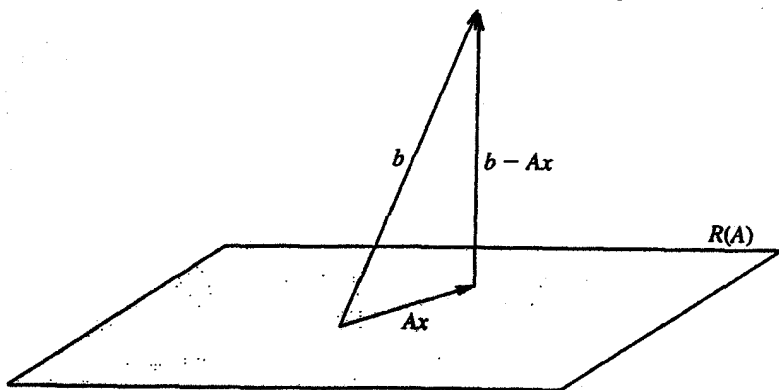
$$A^T Ax = A^T b$$

۵-۷ معادلات نرمال و برازش چند جمله ای

تعریف ۷-۵-۱ دستگاه معادلات

$$A^T Ax = A^T b \quad (۷-۵-۱)$$

معادلات نرمال نامیده می شود.



شکل ۷-۲

یک مثال آشنا از چگونگی مطرح شدن معادلات نرمال برازش یک چندجمله‌ای به یک مجموعه از داده‌های تجربی است.

مهندسان و دانشمندان از آزمایشها داده‌ها را جمع‌آوری می‌کنند. نمایش بامعنایی از داده‌های جمع‌آوری شده لازم می‌باشد تا تصمیمهای بامعنایی برای آینده گرفته شود.

فرض کنید $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ یک مجموعه از جفت مشاهدات باشند. فرض کنید چندجمله‌ای درجه m ($m \leq n$)

$$y(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m \quad (2-5-7)$$

بهترین برازش برای این مجموعه داده باشد. یک استراتژی برای بهترین برازش می‌نیم کردن مجموع توان دوم مانده‌هاست.

$$E = \sum_{i=1}^n (y_i - a_0 - a_1x_i - a_2x_i^2 - \dots - a_mx_i^m)^2 \quad (3-5-7)$$

سپس ما باید داشته باشیم

$$\frac{\partial E}{\partial a_i} = 0, \quad i = 1, \dots, m$$

اکنون

$$\begin{aligned} \frac{\partial E}{\partial a_0} &= -2 \sum_{i=1}^n (y_i - a_0 - a_1x_i - a_2x_i^2 - \dots - a_mx_i^m) \\ \frac{\partial E}{\partial a_1} &= -2 \sum_{i=1}^n x_i (y_i - a_0 - a_1x_i - a_2x_i^2 - \dots - a_mx_i^m) \\ &\vdots \\ \frac{\partial E}{\partial a_m} &= -2 \sum_{i=1}^n x_i^m (y_i - a_0 - a_1x_i - a_2x_i^2 - \dots - a_mx_i^m) \end{aligned} \quad (4-5-7)$$

با قرار دادن این معادلات برابر صفر خواهیم داشت

$$\begin{aligned} a_0n + a_1 \sum x_i + a_2 \sum x_i^2 + \dots + a_m \sum x_i^m &= \sum y_i \\ a_0 \sum x_i + a_1 \sum x_i^2 + \dots + a_m \sum x_i^{m+1} &= \sum x_i y_i \\ &\vdots \\ a_0 \sum x_i^m + a_1 \sum x_i^{m+1} + \dots + a_m \sum x_i^{2m} &= \sum x_i^m y_i \end{aligned} \quad (5-5-7)$$

(در اینجا Σ مجموع را از $i = 1$ تا n نشان می‌دهد).

با قرار دادن $\sum x_i^k = S_k$ ، $k = 0, 1, \dots, 2m$ ، و نمایش عناصر طرف راست به ترتیب توسط b_1, \dots, b_m دستگاه می‌تواند به صورت زیر نوشته شود.

$$\begin{bmatrix} S_0 & S_1 & \dots & S_m \\ S_1 & S_2 & \dots & S_{m+1} \\ \vdots & \vdots & & \vdots \\ S_m & S_{m+1} & \dots & S_{2m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{bmatrix} \quad (6-5-7)$$

(توجه کنید که $S_0 = n$). این دستگاه یک دستگاه $(m+1)$ معادله با $(m+1)$ مجهول a_0, a_1, \dots, a_m است.

این دستگاه در حقیقت یک دستگاه معادلات نرمال است. برای نشان دادن این امر، تعریف می‌کنیم

$$V = \begin{bmatrix} 1 & x_1 & \dots & x_1^m \\ 1 & x_2 & \dots & x_2^m \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \dots & x_n^m \end{bmatrix} \quad (7-5-7)$$

سپس این دستگاه برابر است با

$$V^T V a = b = V^T y \quad (8-5-7)$$

که در آن

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

این یک دستگاه معادلات نرمال است؛ بعلاوه اگر x_i ها همگی متمایز باشند، آنگاه ماتریس دارای رتبه کامل است.

این ماتریس V به ماتریس واندروموند معروف است. از مبحث بخش قبل ملاحظه می‌کنیم که a جواب کم‌ترین توانهای دوم دستگاه $(8-5-7)$ است.

مثال ۷-۵-۱ فرض کنید یک مهندس الکترونیک داده‌های تجربی زیر و متشکل از اندازه جریان در یک سیم الکتریکی برای ولتاژهای مختلف را جمع‌آوری کرده باشد.

$x =$ ولتاژ	۰	۲	۵	۷	۹	۱۳	۲۴
$y =$ جریان	۰	۶	۷/۹	۸/۵	۱۲	۲۱/۵	۳۵

می‌خواهیم معادلات نرمال را برای این داده‌ها متناظر با بهترین برازش داده‌ها توسط (الف) یک خط راست و (ب) چندجمله‌ای درجه دو به دست آوریم و مایل هستیم مقایسه‌ای از نتایج پیش‌بینی شده را با نتیجه واقعی هنگامی که $x = ۵$ ملاحظه کنیم.
(الف) برازش خط راست:

$$V = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 9 \\ 1 & 13 \\ 1 & 24 \end{bmatrix}, \quad y = \begin{bmatrix} 0 \\ 6 \\ 7/9 \\ 8/5 \\ 12 \\ 21/5 \\ 35 \end{bmatrix}$$

معادلات نرمال عبارتند از

$$V^T V a = V^T y = b$$

$$\begin{bmatrix} 7 & 60 \\ 60 & 904 \end{bmatrix} a = 10^3 \begin{bmatrix} 0.0906 \\ 1/3385 \end{bmatrix}$$

جواب این معادلات به صورت زیر است

$$a_0 = 0.6831$$

$$a_1 = 1/4353$$

مقدار $a_0 + a_1 x$ در $x = ۵$ برابر $۷/۸۵۹۶ = ۰.۶۸۳۱ + ۱/۴۳۵۳ \times ۵$ می‌باشد.

(ب) برازش درجه دوم:

$$V = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 5 & 25 \\ 1 & 7 & 49 \\ 1 & 9 & 81 \\ 1 & 13 & 169 \\ 1 & 24 & 576 \end{bmatrix}$$

معادلات نرمال عبارتند از

$$V^T V a = V^T y = b$$

$$\begin{bmatrix} 7 & 6. & 9.4 \\ 6. & 9.4 & 17/226 \\ 9.4 & 17/226 & 369/94. \end{bmatrix} a = 1.2 \begin{bmatrix} 0./0.091 \\ 0./1338 \\ 2/54.4 \end{bmatrix}$$

جواب این معادلات به صورت زیر است

$$a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0./8977 \\ 1/3695 \\ 0./0.27 \end{bmatrix}$$

مقدار $a_0 + a_1 x + a_2 x^2$ در $x = 5$ برابر $7/8127$ می باشد.

تبصرها: مجموعه چند جمله‌ایهای از درجه کم‌تر یا مساوی با m تشکیل یک فضای برداری $(m + 1)$ بُعدی می‌دهند. بنابراین اگر $\phi_0, \phi_1, \dots, \phi_m$ یک پایه از این فضای برداری باشند، آنگاه هر چند جمله‌ای $y(x)$ در این فضا می‌تواند به صورت زیر نوشته شود.

$$y(x) = \sum_{j=0}^m a_j \phi_j(x)$$

اگر توابع پایه‌ای $\phi_i(x)$ برابر $\phi_0(x) = 1, \phi_1(x) = x, \dots, \phi_m(x) = x^m$ انتخاب شوند، آنگاه $y = a_0 + a_1 x + \dots + a_m x^m$ را همانند توضیح فوق به دست می‌آوریم. به طور کلی چنین انتخاب تابعهای پایه‌ای در عمل استفاده نمی‌شوند؛ روش انتخاب تولید یک مجموعه چند جمله‌ایهای متعامد است. حداقلی که ما باید انجام دهیم این است که یک انتقال طوری طرح‌ریزی کنیم که تابعهای پایه‌ای حول وسط بازه متمرکز شوند. برای مثال، اگر از یک برازش خط راست در مثال ۷-۵-۱ استفاده کنیم، انتخاب $\phi_0(x) = 1, \phi_1(x) = \frac{1}{\sqrt{v}}(x - v)$ می‌تواند به صورت $i = 0, 1$ باشد. با این انتخاب، ماتریس V به شکل زیر می‌باشد.

$$V = \begin{bmatrix} 1 & -1 \\ 1 & -0./7143 \\ 1 & -0./2857 \\ 1 & 0 \\ 1 & 0./2857 \\ 1 & 0./8571 \\ 1 & 2/4286 \end{bmatrix}$$

توجه کنید که عدد وضعیت این ماتریس جدید و اندرموند $1/2537$ می باشد، در صورتی که عدد وضعیت ماتریس و اندرموند اصلی برابر $17/3845$ می باشد.

۶-۷ شبه معکوس و مسأله کم ترین توانهای دوم

تعریف ۱-۶-۷ ماتریس $A^+ = (A^T A)^{-1} A^T$ ، هنگامی که A یک ماتریس $m \times n$ ($m \geq n$) بوده و دارای رتبه n می باشد، یا ماتریس $A^+ = A^T (A A^T)^{-1}$ ، هنگامی که A یک ماتریس $(m \times n)$ ($m < n$) بوده و دارای رتبه m می باشد، شبه معکوس A نامیده می شوند. بنابراین از (۱-۳-۷) نتیجه می شود که جواب یکتای کم ترین توانهای دوم برای مسأله فرامعین $Ax = A^+b$ است.

جواب کم ترین توانهای دوم با استفاده از شبه معکوس
 جواب یکتای کم ترین توانهای دوم برای مسأله کم ترین توانهای دوم فرامعین با رتبه کامل $Ax = b$ توسط رابطه زیر ارائه می شود.

$$x = (A^T A)^{-1} A^T b = A^+ b$$

به طور واضح این تعریف شبه معکوس تعریف معمولی معکوس یک ماتریس مربعی را تعمیم می دهد. توجه کنید که اگر A مربعی و معکوس پذیر باشد، آنگاه

$$A^+ = (A^T A)^{-1} A^T = A^{-1} (A^T)^{-1} A^T = A^{-1}$$

این مفهوم مهم را با جزئیات بیشتر در فصل ۱۰ بحث خواهیم کرد. یک مرجع مهم در این مورد کتاب زیر می باشد.

Generalized Inverses of Matrices and Its Applications by C. R. Rao and S. K. Mitra (1971).

با تعریف معکوس تعمیم یافته یک ماتریس مستطیلی، اکنون عدد وضعیت یک ماتریس را به صورت $Cond(A) = \|A\| \|A^+\|$ تعریف می کنیم.

مثال ۱-۶-۷

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 4 & 5 \end{bmatrix}, \quad \text{rank}(A) = 2$$

بنابراین A دارای رتبه کامل است.

$$A^+ = (A^T A)^{-1} A^T = \begin{bmatrix} -1/2857 & -0/5714 & 0/8571 \\ 1 & 0/5000 & -0/5000 \end{bmatrix}$$

$$\text{Cond}_2(A) = \|A\|_2 \|A^+\|_2 = 7/6656 \times 2/0487 = 15/7047$$

ماتریس وایانس - کوواریانس

در برخی کاربردها، ماتریس $A^T A$ ماتریس اطلاعات نامیده می‌شود، زیرا اطلاعات موجود در آزمایش را اندازه‌گیری می‌کند و ماتریس $(A^T A)^{-1}$ به ماتریس واریانس - کوواریانس معروف است. یک الگوریتم برای محاسبه ماتریس واریانس - کوواریانس بدون محاسبه صریح معکوس در بخش ۷-۱۱ ارائه شده است.

۷-۲ حساسیت مسأله کمترین توانهای دوم

در این بخش حساسیت جواب مسأله کمترین توانهای دوم را برای اختلالات در داده‌ها مطالعه می‌کنیم؛ یعنی بررسی می‌کنیم که چگونه یک جواب کمترین توانهای دوم نسبت به تغییرهای کوچک در داده‌ها تغییر می‌کند. این مطالعه در فهم رفتارهای مختلف روشهای مختلف که برای مسائل کمترین توانهای دوم در بخش بعد بحث خواهند شد، مهم می‌باشد. دو حالت را در نظر می‌گیریم: اختلال در بردار b و اختلال در ماتریس A . نتایج در این بخش نتایج اختلال نرمی هستند. برای نتایج اختلال مؤلفه‌ای مراجع (1994) Björck و (1994b) Chandrasekaran and Ipsen را ملاحظه کنید.

حالت ۱: اختلال در بردار b

در اینجا فرض می‌کنیم که بردار b به صورت $\bar{b} = b + \delta b$ اختلال یافته باشد، اما A بدون تغییر باقی بماند. قضیه ۷-۷-۱ قضیه اختلال طرف راست کمترین توانهای دوم. فرض کنید x و \bar{x} به ترتیب جوابهای یکتای کمترین توانهای دوم برای مسائل اصلی و اختلال یافته باشند. آنگاه اگر $\|b\| \neq 0$ ،

$$\frac{\|\bar{x} - x\|}{\|x\|} \leq \text{Cond}(A) \frac{\|\delta b_R\|}{\|b_R\|}$$

در اینجا

$$\text{Cond}(A) = \|A\| \|A^+\|$$

و b_R و δb_R به ترتیب تصویرهای بردارهای b و δb بر روی $R(A)$ هستند.

اثبات: چون x و \hat{x} جوابهای یکتای کم‌ترین توانهای دوم مسائل اصلی و اختلال یافته هستند، داریم

$$x = A^+b$$

$$\hat{x} = A^+(b + \delta b)$$

بنابراین

$$\hat{x} - x = A^+b + A^+\delta b - A^+b = A^+\delta b$$

فرض کنید δb_N تصویر δb را بر روی متمم متعامد $R(A)$ نمایش دهد؛ یعنی

$$\delta b = \delta b_R + \delta b_N$$

چون δb_N در متمم متعامد $R(A)$ ، یعنی $N(A^T)$ قرار دارد، داریم $A^T(\delta b_N) = 0$. بنابراین

$$\hat{x} - x = A^+\delta b = A^+(\delta b_R + \delta b_N)$$

$$= A^+(\delta b_R) + A^+(\delta b_N) = A^+\delta b_R + (A^T A)^{-1} A^T \delta b_N = A^+\delta b_R \quad (1-7-7)$$

دوباره چون x جواب یکتای کم‌ترین توانهای دوم است، داریم

$$Ax = b_R$$

از اینجا (دوباره با نرم گرفتن از هر دو طرف) به دست می‌آوریم

$$\|x\| \geq \frac{\|b_R\|}{\|A\|} \quad (2-7-7)$$

با گرفتن نرم از هر دو طرف (1-7-7) و ترکیب آن با (2-7-7) قضیه را خواهیم داشت. \square

تعبیر قضیه 1-7-7. قضیه 1-7-7 بیان می‌کند که اگر فقط بردار b اختلال یابد، آنگاه همانند حالت دستگانه‌های خطی $\text{Cond}(A) = \|A\| \|A^+\|$ به عنوان عدد وضعیت در تحلیل حساسیت جواب یکتای کم‌ترین توانهای دوم به کار می‌رود. اگر این عدد بزرگ باشد، آنگاه حتی با یک خطای نسبی کوچک در تصویر b بر روی $R(A)$ ، ما ممکن است یک تغییر شدید در جواب کم‌ترین توانهای دوم داشته باشیم. از طرف دیگر، اگر این عدد کوچک باشد و خطای نسبی در تصویر b بر روی $R(A)$ نیز کوچک باشد، آنگاه جواب کم‌ترین توانهای دوم تغییر زیادی نخواهد کرد. توجه کنید که کوچک بودن

خطای نسبی در تصویر b بر روی $R(A)$ ، یعنی $\|\delta b_R\|/\|b_R\|$ است که در اینجا نقش بازی می‌کند نه صرفاً کوچک بودن $\|b_R\|$.

مثال ۷-۷-۱ یک مسأله غیر حساس

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \delta b = 10^{-4} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

تصویر A بر روی $P_A = R(A)$

$$P_A = \begin{bmatrix} 0.8333 & 0.3333 & 0.1667 \\ 0.3333 & 0.3333 & -0.3333 \\ 0.1667 & -0.3333 & 0.8333 \end{bmatrix}$$

(مثال ۵-۶-۲ را ملاحظه کنید)

$$b_R = P_A b = \begin{bmatrix} 1.3333 \\ 0.3333 \\ 0.6667 \end{bmatrix}$$

$$\delta b_R = P_A \delta b = 10^{-4} \begin{bmatrix} 0.13333 \\ 0.33333 \\ 0.66667 \end{bmatrix}$$

$$\text{Cond}(A) = 2/4495$$

چون $\|\delta b_R\|/\|b_R\| = 10^{-4}$ ، و یک عدد کوچک است و A خوش‌وضع است، از قضیه ۷-۷-۱ ما انتظار داریم که جواب کمترین توانهای دوم زیاد تغییر نکند. محاسبه‌های زیر این حالت را نشان می‌دهند.

$$x = A^+ b = \begin{bmatrix} 0.6667 \\ 0.3333 \end{bmatrix}$$

$$\hat{x} = A^+(b + \delta b) = \begin{bmatrix} 0.6667 \\ 0.3333 \end{bmatrix}$$

$$\frac{\|\hat{x} - x\|}{\|x\|} = 10^{-4}$$

مثال ۷-۷-۲ یک مسأله حساس

$$A = \begin{bmatrix} 1 & 1 \\ 10^{-4} & 0 \\ 0 & 10^{-4} \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 10^{-4} \\ 10^{-4} \end{bmatrix}$$

$$\delta b = 10^{-3} \begin{bmatrix} 1 \\ 0.1 \\ 0 \end{bmatrix}, \quad \text{Cond}(A) = O(10^4)$$

$$x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \hat{x} = \begin{bmatrix} 1/500.5 \\ 0/500.5 \end{bmatrix}$$

حالت ۲: اختلال در ماتریس A

تحلیل در این حالت به مراتب پیچیده تر از حالت قبل است. نتیجه و پیامدهای مهم آن را بدون اثبات بیان می‌کنیم. فرض کنید اختلال E ماتریس A به اندازه کافی کوچک باشد به طوری که

$$\text{rank}(A) = \text{rank}(A + E)$$

فرض کنید x و x̂ به ترتیب جوابهای یکتای کم‌ترین توانهای دوم را برای مسائل اصلی و اختلال یافته نمایش دهند. فرض کنید EA و EN به ترتیب تصویرهای E را بر روی R(A) و مکمل متعامد R(A) نمایش دهند. آنگاه اگر bR ≠ 0 باشد قضیه زیر را از مرجع Stewart (1973, 223) داریم.

قضیه ۷-۷-۲ قضیه اختلال طرف چپ کم‌ترین توانهای دوم. فرض کنید x و x̂ به ترتیب جوابهای کم‌ترین توانهای دوم برای Ax = b و (A + E)x̂ = b باشند، و rank(A + E) برابر rank(A) باشد. آنگاه

$$\frac{\|\hat{x} - x\|}{\|x\|} \leq \gamma \text{Cond}(A) \frac{\|E_A\|}{\|E\|} + \gamma (\text{Cond}(A))^2 \frac{\|E_N\|}{\|A\|} \frac{\|b_N\|}{\|b_R\|} + O\left(\frac{\|E_N\|}{\|A\|}\right)^2$$

تعبیر قضیه ۷-۷-۲. قضیه ۷-۷-۲ بیان می‌کند که در حالتی که ماتریس A اختلال می‌یابد حساسیت جواب یکتای کم‌ترین توانهای دوم، در حالت کلی، به توان دوم عدد وضعیت A بستگی دارد. ولیکن اگر \|EN\| یا \|bN\| صفر باشند، آنگاه حساسیت فقط به Cond(A) بستگی دارد. توجه کنید که مانده r = b - Ax صفر است اگر bN = 0 باشد.

دو مثال با حساسیتهای مختلف

اکنون دو مثال با یک ماتریس A اما با b های مختلف ارائه می دهیم تا حساسیتهای متفاوت مسأله کمترین توانهای دوم را در حالت های مختلف توضیح دهیم. در مثال اول $\text{Cond}(A)^2$ به عنوان عدد وضعیت مسأله و در مثال دوم $\text{Cond}(A)$ به عنوان عدد وضعیت به کار می روند.

مثال ۷-۷-۳ حساسیت به توان دوم عدد وضعیت بستگی دارد.

$$A = \begin{bmatrix} 1 & 1 \\ 0.0001 & 0 \\ 0 & 0.0001 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$b_R = P_A b = \begin{bmatrix} 1/0.0001 \\ 0.0001 \\ 0.0001 \end{bmatrix}, \quad b_N = \begin{bmatrix} -0.0001 \\ 0.9999 \\ 0.9999 \end{bmatrix}, \quad \frac{\|b_N\|}{\|b_R\|} = 1/4140.$$

(با استفاده از P_A مثال ۵-۶-۳). فرض کنید

$$E = 10^{-2} \begin{bmatrix} 0 & -0.0001 \\ 0 & 0.9999 \\ 0 & 0.9999 \end{bmatrix}$$

$$(A + E) = \begin{bmatrix} 1 & 1 \\ 0.0001 & 0.0001 \\ 0 & 0.0002 \end{bmatrix}, \quad E_N = 10^{-2} \begin{bmatrix} 0 & -0.0002 \\ 0 & 0.9999 \\ 0 & 0.9999 \end{bmatrix}$$

$$\frac{\|E_N\|}{\|A\|} = \frac{\|E\|}{\|A\|} = 9/9999 \times 10^{-5}$$

اگرچه حاصلضرب $\|E_N\|/\|A\|$ و $\|b_N\|/\|b_R\|$ نسبتاً کوچک است، چون $\text{Cond}(A) = 2 \times 10^8$ بزرگ است، ما انتظار یک انحراف شدید برای جواب اختلال یافته از جواب اصلی داریم. همان طور که محاسبات زیر نشان می دهند این در حقیقت درست است.

$$\hat{x} = 10^3 \begin{bmatrix} -4/999 \\ 5 \end{bmatrix}, \quad x = \begin{bmatrix} 0/5 \\ 0/5 \end{bmatrix}$$

$$\frac{\|\hat{x} - x\|}{\|x\|} = 9/999 \times 10^3 \quad (\text{بزرگ})$$

مثال ۷-۷-۴ حساسیت بستگی به عدد وضعیت دارد. فرض کنید

$$E = 10^{-4} \begin{bmatrix} 0 & -0.0001 \\ 0 & 0.9999 \\ 0 & 0.9999 \end{bmatrix} \quad (\text{نظیر مسأله ۷-۷-۳})$$

$$A = \begin{bmatrix} 1 & 1 \\ 0.0001 & 0 \\ 0 & 0.0001 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 0.0001 \\ 0.0001 \end{bmatrix}$$

در این حالت $b_R = b$ و

$$b_N = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

(مثال ۵-۶-۳ را ملاحظه کنید.)

بنابراین بر طبق قضیه ۷-۷-۲ توان دوم $\text{Cond}(A)$ هیچ اثری ندارد؛ جواب کم‌ترین توانهای دوم فقط تحت تاثیر $\text{Cond}(A)$ قرار می‌گیرد. ما این را به صورت زیر تحقیق می‌کنیم

$$x = A^+b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} 1/4999 \\ 0.5000 \end{bmatrix}$$

$$\text{Cond}(A) = 1/4142 \times 10^4$$

$$\frac{\|\hat{x} - x\|}{\|x\|} = 0.5000$$

حساسیت مانده. ملاحظه کردیم که حساسیتهای جوابهای کم‌ترین توانهای دوم ناشی از اختلالها در ماتریس A برای بردارهای طرف راست متفاوت b با هم تفاوت دارند، ولیکن قضیه زیر نشان می‌دهد که حساسیت مانده همیشه بستگی به عدد وضعیت ماتریس A دارد. ما نتیجه را به شکل ساده شده و خام آن بیان می‌کنیم. برای بیانی دقیق و اثبات یک نتیجه بر روی حساسیت مانده مرجع Golub and Van Loan (1989, 228-30) را ملاحظه کنید.

قضیه ۷-۷-۳ قضیه حساسیت مانده کم‌ترین توانهای دوم. فرض کنید \hat{x} و x به ترتیب مانده‌ها را

برای مسائل کمترین توانهای دوم اصلی و اختلال یافته نمایش دهند؛ یعنی

$$r = b - Ax$$

$$\hat{r} = b - (A + E)(\hat{x})$$

پس

$$\frac{\|\hat{r} - r\|}{\|b\|} \leq \varepsilon(1 + 2\text{Cond}(A)) + O(\varepsilon^2)$$

که در آن

$$\varepsilon = \frac{\|E\|}{\|A\|}$$

تعبیر قضیه ۷-۷-۳. این نتیجه بیان می‌کند که هنگامی که A اختلال می‌یابد حساسیت مانده حداکثر به عدد وضعیت A بستگی دارد.

مثال ۷-۷-۵

$$A = \begin{bmatrix} 1 & 1 \\ 10^{-4} & 0 \\ 0 & 10^{-4} \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$E = 10^{-4} \begin{bmatrix} 0 & -0.0001 \\ 0 & 0.9999 \\ 0 & 0.9999 \end{bmatrix}$$

$$x = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, \quad \hat{x} = 10^3 \begin{bmatrix} -4/9999 \\ 5 \end{bmatrix}$$

$$r = b - Ax = \begin{bmatrix} -0.0001 \\ 0.9999 \\ 0.9999 \end{bmatrix}$$

$$\hat{r} = \begin{bmatrix} -0.0001 \\ 0.9999 \\ 0 \end{bmatrix}$$

$$\frac{\|\hat{r} - r\|}{\|b\|} = 0.5773,$$

$$\text{Cond}(A) = 1/4142 \times 10^4$$

$$\text{Cond}(A) \cdot \frac{\|E_N\|}{\|A\|} = 1/4142$$

اکنون نامساوی قضیه ۷-۷-۳ به سادگی تحقیق می‌شود.

حساسیت شبه معکوس. نتیجه زیر، از ودین (Wedin (1973)) نشان می‌دهد که دوباره $\text{Cond}(A)$ است که یک نقش در تحلیل حساسیت شبه معکوس یک ماتریس بازی می‌کند.

قضیه ۷-۷-۴ قضیه حساسیت شبه معکوس. فرض کنید A یک ماتریس $m \times n$ با $m \geq n$ باشد. فرض کنید A^+ و \tilde{A}^+ به ترتیب شبه معکوسهای A و $\tilde{A} = A + E$ باشند. آنگاه به شرط آن که $\text{rank}(A) = \text{rank}(\tilde{A})$ داریم

$$\frac{\|\tilde{A}^+ - A^+\|}{\|\tilde{A}^+\|} \geq c \text{Cond}(A) \frac{\|E\|}{\|A\|}$$

که در آن c یک عدد کوچک است.

مثال ۷-۷-۶

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 4 & 5 \end{bmatrix}, \quad E = 10^{-2}A = \begin{bmatrix} 0.010 & 0.020 \\ 0.020 & 0.030 \\ 0.040 & 0.050 \end{bmatrix}$$

$$A^+ = \begin{bmatrix} -1/2857 & -0.5714 & 0.8571 \\ & 1 & -0.5000 \end{bmatrix}$$

$$A + E = \tilde{A} = \begin{bmatrix} 1.01 & 2.02 \\ 2.02 & 3.03 \\ 4.04 & 5.05 \end{bmatrix}$$

$$\tilde{A}^+ = \begin{bmatrix} -1/2844 & -0.5709 & 0.8563 \\ & 0.9990 & -0.4995 \end{bmatrix}$$

$$\frac{\|\tilde{A}^+ - A^+\|}{\|\tilde{A}^+\|} = 10^{-2} \approx \frac{\|E\|}{\|A\|}$$

توجه کنید که $\text{Cond}(A) = 15/7047$.

۷-۸ روشهای محاسباتی برای مسائل کمترین توانهای دوم فرامین

۷-۸-۱ روش معادلات نرمال

یکی از متداولترین راههای مورد استفاده (بویژه در آمار) برای محاسبه جواب کمترین توانهای دوم روش معادلات نرمال است. این روش بر جواب دستگاه معادلات نرمال

$$A^T A x = A^T b$$

مبتنی است. فرض می‌کنیم که A یک ماتریس $m \times n$ ($m > n$) با رتبه کامل باشد. چون در این حالت $A^T A$ متقارن و معین مثبت است، تجزیه چولسکی را می‌پذیرد.

$$A^T A = H H^T$$

بنابراین راه معادلات نرمال برای حل مسأله کمترین توانهای دوم می‌تواند به صورت زیر بیان شود.

الگوریتم ۷-۸-۱ جواب کمترین توانهای دوم با استفاده از معادلات نرمال. با مفروض بودن یک ماتریس $m \times n$ ($m > n$) مانند A و با $\text{rank}(A) = n$ ، الگوریتم زیر جواب کمترین توانهای دوم x را از معادلات نرمال و با استفاده از تجزیه چولسکی محاسبه می‌کند.

گام ۱: $c = A^T b$ را تشکیل دهید.

گام ۲: عامل چولسکی H ماتریس $A^T A$ را با بهره‌گیری از متقارن بودن $A^T A$ محاسبه کنید.

گام ۳: دستگاههای مثلثی زیر را به ترتیب حل کنید.

$$H y = c$$

$$H^T x = y$$

شمارش اعمال ممیز شناور. این روش برای حل مسأله کمترین توانهای دوم با رتبه کامل به حدوداً $\frac{mn^2}{2} + n^3/6$ عمل ممیز شناور نیاز دارد؛ $\frac{mn^2}{2}$ برای محاسبه $A^T A$ و $A^T b$ و $n^3/6$ برای محاسبه تجزیه چولسکی $A^T A$ و n^2 برای حل دو دستگاه مثلثی. بنابراین کاملاً کارا است.

تذکر MATCOM: الگوریتم ۷-۸-۱ در برنامه LSFRNME در MATCOM پیاده‌سازی شده است.

مثال ۷-۸-۱ حل معادلات نرمال مسأله مورد مطالعه

$$A = \begin{pmatrix} 1 & 274 & 245.0 \\ 1 & 18.0 & 3254 \\ 1 & 275 & 28.2 \\ 1 & 2.5 & 2838 \\ 1 & 86 & 2327 \end{pmatrix}, \quad b = \begin{pmatrix} 162 \\ 12.0 \\ 223 \\ 131 \\ 67 \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 5 & 112.0 & 14,691 \\ 112.0 & 297,522 & 3,466,402 \\ 14,691 & 3,466,402 & 44,608,873 \end{pmatrix} \quad \text{گام ۱: تشکیل}$$

$$A^T b = \begin{pmatrix} 7.3 \\ 182,23.0 \\ 2,164,253 \end{pmatrix} \quad \text{گام ۲: تشکیل}$$

گام ۳: حل معادلات نرمال

$$A^T A x = A^T b$$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 7/0.325 \\ 0/5.44 \\ 0/0.70 \end{pmatrix}$$

توجه کنید که دستگاه به طور ساختگی بدوضع می باشد زیرا ستونهای A خارج از مقیاس هستند.

$$\text{Cond}(A^T A) = 3/0.891 \times 10^8$$

برای دیدن این که چگونه جواب کم ترین توانهای دوم محاسبه شده با داده های جدول توافق دارد، محاسبات زیر را انجام می دهیم:

$$(1 \quad 274 \quad 245.0) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 162/4.43$$

$$(\text{جواب واقعی} = 162)$$

$$(1 \quad 18.0 \quad 3254) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 12.0/6153$$

$$(\text{جواب واقعی} = 12.0)$$

$$(1 \quad 375 \quad 3802) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 222/8193$$

(جواب واقعی = ۲۲۳)

$$(1 \quad 205 \quad 2838) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 130/3140$$

(جواب واقعی = ۱۳۱)

$$(1 \quad 86 \quad 2347) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 66/8471$$

(جواب واقعی = ۶۷)

فرض کنید که شرکت مایل باشد که با استفاده از این نتایج فروش را در یک ناحیه با جمعیت ۲۲۰۰۰۰ و در آمد سرانه ۲۵۰۰ دلار پیش بینی کند. آنگاه بهترین پیش بینی با استفاده از الگوری داده شده به صورت زیر است.

$$(1 \quad 220 \quad 2500) \begin{bmatrix} 7/0.325 \\ 0/5.44 \\ 0/0.070 \end{bmatrix} = 135/5142$$

تذکر: ستونهای A خارج از مقیاس هستند؛ به این دلیل A بدوضع است. در عمل آنها باید به قسمی مقیاس شوند که دارای طول یکسان باشند. شما می توانید تحقیق کنید که در این حالت داریم $\text{Cond}(A^T A) = 264/68$. بنابراین حقیقتاً مسأله بدوضع نیست.

مثال ۷-۸-۲

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 5 \\ 9 \end{bmatrix}$$

در اینجا $\text{rank}(A) = 2$ و $\text{rank}(A, b) = 3$. بنابراین دستگاه $Ax = b$ نمی تواند جوابی داشته باشد. بنابراین جواب کمترین توانهای دوم را محاسبه می کنیم.

$$c = A^T b = \begin{bmatrix} 40 \\ 57 \end{bmatrix} \quad 1.$$

۲. عامل چولسکی $A^T A$ به صورت زیر است.

$$H = \begin{bmatrix} ۳/۷۴۱۷ & ۰ \\ ۵/۳۴۵۲ & ۰/۶۵۴۷ \end{bmatrix}$$

۳. جواب دستگاههای مثلثی عبارتند از:

$$y = \begin{bmatrix} ۱۰/۶۹۰۴ \\ -۰/۲۱۸۲ \end{bmatrix}$$

$$x = \begin{bmatrix} ۳/۳۳۳۳ \\ -۰/۳۳۳۳ \end{bmatrix}$$

جواب یکتای کمترین توانهای دوم برابر $x = \begin{bmatrix} ۳/۳۳۳۳ \\ -۰/۳۳۳۳ \end{bmatrix}$ است.

توجه کنید که

$$A^+ = (A^T A)^{-1} A^T = \begin{bmatrix} -۱/۸۳۳۳ & -۰/۳۳۳۳ & ۱/۱۶۶۷ \\ ۱/۳۳۳۳ & ۰/۳۳۳۳ & -۰/۶۶۶۷ \end{bmatrix}$$

و

$$A^+ b = \begin{bmatrix} ۳/۳۳۳۳ \\ -۰/۳۳۳۳ \end{bmatrix}$$

مشکلات عددی با روش معادلات نرمال

اگرچه روش معادلات نرمال برای فهم و پیاده‌سازی آسان است، ولیکن ممکن است مشکلات عددی را در برخی از حالتها به وجود آورد.

اولاً ممکن است تعدادی از ارقام با معنی را در طی تشکیل صریح $A^T A$ از دست بدهیم و ماتریس محاسبه شده $A^T A$ ممکن است از معین مثبت دور باشد، از لحاظ محاسباتی حتی ممکن است منفرد باشد. در حقیقت توسط استوارت (Stewart (1973, 225-6)) نشان داده شده است، که بجز حالتی که $Cond(A)$ کم‌تر از $۱۰^{t/2}$ است، با فرض این که $A^T A$ به طور دقیق محاسبه شده باشد و سپس با t رقم گرد شده باشد، $A^T A$ ممکن است در معین مثبت بودن با شکست مواجه گردد و یا ممکن است حتی منفرد باشد. مثال زیر این حقیقت را توضیح می‌دهد.

مثال ۷-۸-۳ ماتریس

$$A = \begin{bmatrix} ۱ & ۱ \\ ۱۰^{-۴} & ۰ \\ ۰ & ۱۰^{-۴} \end{bmatrix}$$

را با $t = 8$ ملاحظه کنید، ستونهای A مستقل خطی هستند. اکنون محاسبه می‌کنیم

$$A^T A = \begin{bmatrix} 1 + 10^{-8} & 1 \\ 1 & 1 + 10^{-8} \end{bmatrix}$$

چون $t = 8$ ، به دست خواهیم آورد

$$A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

که منفرد است. توجه کنید که $10^4 > 10^{4/2} = 10^2$

ثانیاً راه معادلات نرمال ممکن است در برخی حالات خطاهای بیشتری از آنچه که در مسأله اجتناب‌ناپذیر است، وارد کند. این به صورت زیر دیده می‌شود.

از تحلیل اختلال انجام شده در فصل ۶ به راحتی ملاحظه می‌کنیم که اگر \hat{x} جواب محاسبه شده کمترین توانهای دوم به دست آمده توسط روش معادلات نرمال باشد، آنگاه

$$\frac{\|\hat{x} - x\|}{\|x\|} \approx \mu \text{Cond}(A^T A)$$

$$= \mu (\text{Cond}(A))^2 \quad (\text{تمرین ۵})$$

بنابراین دقت جواب کمترین توانهای دوم با استفاده از معادلات نرمال بستگی به توان دوم عدد وضعیت A دارد. ولیکن اکنون در بخش تحلیل اختلالهای مسأله کمترین توانهای دوم دیدیم که در برخی از حالتها نظیر هنگامی که مانده صفر می‌باشد، حساسیت مسأله فقط بستگی به عدد وضعیت A دارد. (قضیه‌های ۷-۷-۱ و ۷-۷-۲ را ملاحظه کنید). بنابراین در این حالتها روش معادلات نرمال خطاهای بیشتری در جواب از آنچه که داده‌ها ضمانت می‌کنند، وارد می‌کند.

یک تذکر خاص در مورد روش معادلات نرمال

با وجود اشکالهای روش معادلات نرمال، باید تاکید کنیم که روش باز هم به عنوان یک ابزار مفید برای حل مسأله کمترین توانهای دوم حداقل در حالتی که ماتریس A خوش‌وضع است، در نظر گرفته می‌شود. در حقیقت به طور معمول در کاربردهای عملی زیادی به کار برده می‌شود و به نظر می‌رسد که مهندسان تجربی و آماردانان کاملاً با آن مأنوس هستند. توجه کنید که در مثال ۷-۸-۳ اگر از یک دقت توسعه یافته استفاده می‌کردیم ماتریس محاسبه شده $A^T A$ می‌توانست به صورت یک ماتریس معین مثبت به دست آید و با وجود بدوضع معادلات نرمال یک جواب دقیق نتیجه دهد. محاسبات زیر این

حقیقت را نشان می دهند.

$$A^T A = \begin{bmatrix} 1/\dots\dots\dots 1 \\ \vdots \\ 1/\dots\dots\dots 1 \end{bmatrix}, \quad b = (\gamma, \dots, \dots, \dots)^T$$

$$c = A^T b = \begin{bmatrix} \gamma/\dots\dots\dots 1 \\ \vdots \\ \gamma/\dots\dots\dots 1 \end{bmatrix}$$

$$H = A^T A \text{ عامل چولسکی} = \begin{bmatrix} 1/\dots\dots\dots 5 & \cdot \\ \cdot/\dots\dots\dots 14142135651 & \cdot \end{bmatrix}$$

جواب $Hy = c$:

$$y = \begin{bmatrix} \gamma \\ \cdot/\dots\dots\dots 14142135651 \end{bmatrix}$$

جواب $H^T x = y$:

$$x = \begin{bmatrix} \cdot/\dots\dots\dots 14142135651 \\ \gamma/\dots\dots\dots 14142135651 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

(جواب واقعی)

۷-۲ روشهای تجزیه QR برای مسأله با رتبه کامل

فرض کنید

$$Q^T A = R = \begin{bmatrix} R_1 \\ \cdot \end{bmatrix}$$

تجزیه QR ماتریس A باشد. آنگاه چون طول یک بردار توسط ضرب ماتریس متعامد حفظ می شود، داریم:

$$\begin{aligned} \|Ax - b\|_2^2 &= \|Q^T Ax - Q^T b\|_2^2 \\ &= \|R_1 x - c\|_2^2 + \|d\|_2^2 \end{aligned}$$

که در آن $Q^T b = \begin{bmatrix} c \\ d \end{bmatrix}$ بنابراین $\|Ax - b\|_2^2$ می نیمم خواهد شد اگر x به قسمی انتخاب شود که

$$R_1 x - c = 0$$

آنگاه مانده متناظر به صورت زیر ارائه می شود

$$\|r\|_2 = \|d\|_2$$

این مشاهده فوراً راه QR زیر را برای حل مسأله کمترین توانهای دوم پیشنهاد می کند.

کمترین توانهای دوم با استفاده از QR

۱. تجزیه $A_{m \times n} = Q_{m \times n} R_{m \times n}$

۲. تشکیل $Q_{m \times m}^T b_{m \times 1} = \begin{bmatrix} c \\ d \end{bmatrix}$ که در آن c یک بردار $1 \times n$ و d یک بردار $1 \times (m - n)$ است.

۳. حل دستگاه بالا مثلثی $n \times n$ بعدی

$$R_1 x = c$$

که در آن

$$R = \begin{bmatrix} R_1 \\ \cdot \end{bmatrix}$$

مثال ۷-۸-۴

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 5 \\ 9 \end{bmatrix}$$

۱. $A = QR$

$$Q = \begin{bmatrix} -0.2673 & 0.8729 & 0.4082 \\ -0.5345 & 0.2182 & -0.8165 \\ -0.8018 & -0.4264 & 0.4082 \end{bmatrix}$$

$$R = \begin{bmatrix} -۳/۷۴۱۷ & -۵/۳۴۵۲ \\ \cdot & \cdot/۶۵۴۷ \\ \cdot & \cdot \end{bmatrix} = \begin{bmatrix} R_1 \\ \cdot \\ \cdot \end{bmatrix}$$

$$Q^T b = \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} -۱۰/۶۹۰۴ \\ -۰/۲۱۸۲ \\ \cdot/۸۱۶۵ \end{bmatrix} \quad .۲$$

۳. حل دستگاه $R_1 x = c$:

$$\begin{bmatrix} -۳/۷۴۱۷ & -۵/۳۴۵۲ \\ \cdot & \cdot/۶۵۴۷ \end{bmatrix} x = \begin{bmatrix} -۱۰/۶۹۰۴ \\ -۰/۲۱۸۲ \end{bmatrix}$$

جواب کم‌ترین توانهای دوم به صورت زیر است

$$x = \begin{bmatrix} ۳/۳۳۳۲ \\ -۰/۳۳۳۳ \end{bmatrix}$$

نرم مانده برابر است با $\|r\|_2 = \|d\| = ۰/۸۱۶۵$

استفاده از ماتریسهای هاوس هولدر

به محض آن که A به QR تجزیه شود و بردار $Q^T b$ تشکیل گردد جواب توسط جایگزینی پسرو به دست می‌آید. فرض کنید که روش هاوس هولدر برای تجزیه QR ماتریس A استفاده شود. آنگاه با دنبال کردن نمادگذاری فصل ۵ ملاحظه می‌کنیم که ضرب $Q^T b$ می‌تواند به صورت زیر انجام شود.

For $k = 1, 2, \dots, n$

$$b = H_k b$$

که در آن $H_k, k = 1, \dots, n$ ، ماتریسهای هاوس هولدر هستند به قسمی که

$$Q^T = H_n H_{n-1} \dots H_2 H_1$$

بنابراین ماتریس Q به طور صریح نباید تشکیل شود.

ایده حل مسأله کم‌ترین توانهای دوم مبتنی بر تجزیه QR ماتریس A با استفاده از تبدیل هاوس هولدر ابتدا توسط گولوب ((Golub (1965) پیشنهاد شده است.

الگوریتم ۷-۸-۲ روش گولوب - هاوس هولدر برای مسائل کم‌ترین توانهای دوم با رتبه کامل. با

مفروض بودن ماتریس A ، $n \times m$ بعدی ($m \geq n$) با $\text{rank}(A) = n$ و یک بردار m بعدی b ، الگوریتم زیر جواب کمترین توانهای دوم x را با استفاده از ماتریسهای هاوس هولدر H_1 تا H_n محاسبه می‌کند.

گام ۱: الگوریتم ۵-۴-۳ را در مورد A به کار ببرید. R_1 و بردارهای تعریف‌کننده ماتریسهای H_1 ، H_2 ، \dots ، H_n را به دست آورید.

گام ۲: بردار $\begin{bmatrix} c \\ d \end{bmatrix} = H_1 H_2 \dots H_n b$ را تشکیل دهید که در آن c یک بردار n بعدی است، برای این منظور از الگوریتم ۵-۴-۲ و از بردارهای محاسبه شده گام ۱ استفاده کنید.

گام ۳: دستگاه بالا مثلثی $R_1 x = c$ را حل کنید.

شمارش اعمال ممیز شناور. چون در هزینه الگوریتم هزینه تجزیه QR ماتریس A غالب می‌باشد، در کل تعداد اعمال ممیز شناور برای جواب کمترین توانهای دوم با رتبه کامل و با استفاده از روش گولوب - هاوس هولدر برابر است با

$$n^2(m - \frac{n}{3}) \quad (\text{تمرین ۱۷ ج})$$

بنابراین در حالت $m > n$ روش گران‌تر از روش معادلات نرمال است. توجه کنید که روش معادلات نرمال حدوداً به $\frac{n^3}{6} + \frac{n^2 m}{3}$ عمل ممیز شناور (flops) نیاز دارد.

خطای گردگردن و پایداری. روش پایدار است. در مرجع (Lawson and Hanson (1974, 90) نشان داده شده است که جواب محاسبه شده \tilde{x} به قسمی است که نرم

$$\|(A + E)x - (b + \delta b)\|_2$$

را می‌نیمم می‌سازد. در اینجا E و δb کوچک هستند. بویژه

$$\|E\|_F \leq c\mu n \|A\|_F + O(\mu^2),$$

$$\|\delta b\|_F \leq c\mu \|b\|_2 + O(\mu^2)$$

که در آن $c \approx (6m - 3n + 4)n$ و μ دقت ماشین است. یعنی، جواب محاسبه شده جواب کمترین توانهای دوم واقعی یک مسأله نزدیک است.

تذکر MATCOM : الگوریتم ۷-۸-۲ در برنامه LSFQRH در MATCOM پیاده‌سازی شده است.

مثال ۷-۸-۵ : جواب کم‌ترین توانهای دوم $Ax = b$ را با استفاده از روش گولوب - هاوس هولدر پیدا کنید.

$$A = \begin{bmatrix} 1 & & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

گام ۱ : با استفاده از الگوریتم ۵-۴-۳ بردارهای u_3 و u_2 را به قسمی پیدا می‌کنیم که

$$H_3 H_2 A = \begin{bmatrix} R_1 \\ \cdot \\ \cdot \end{bmatrix}$$

که در آن

$$H_1 = I - \gamma \frac{u_1 u_1^T}{u_1^T u_1}, \quad H_2 = \begin{bmatrix} 1 & & \cdot \\ \cdot & \hat{H}_2 & \\ \cdot & & \cdot \end{bmatrix}$$

$$\hat{H}_2 = I - \gamma \frac{u_2 u_2^T}{u_2^T u_2}$$

$$u_3 = \begin{bmatrix} 2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}, \quad u_2 = 10^{-3} \begin{bmatrix} -\cdot / 2414 \\ \cdot / 1000 \end{bmatrix}$$

$$R_1 = \begin{bmatrix} -1 & -1 \\ \cdot & \cdot / 0001 \end{bmatrix}$$

گام ۲ : با استفاده از الگوریتم ۵-۴-۲ تشکیل می‌دهیم

$$H_3 H_2 b = \begin{bmatrix} c \\ d \end{bmatrix}$$

$$H_1 b = \left(I - \gamma \frac{u_1 u_1^T}{u_1^T u_1} \right) b$$

$$= \begin{bmatrix} ۲ \\ \cdot/\cdot\cdot\cdot\cdot ۱ \\ \cdot/\cdot\cdot\cdot\cdot ۱ \end{bmatrix} - ۰/۵ \begin{bmatrix} ۸ \\ \cdot/\cdot\cdot\cdot\cdot ۴ \\ \cdot \end{bmatrix}$$

$$= \begin{bmatrix} -۲ \\ -\cdot/\cdot\cdot\cdot\cdot ۱ \\ \cdot/\cdot\cdot\cdot\cdot ۱ \end{bmatrix}$$

$$\begin{aligned} \hat{H}_\gamma \begin{bmatrix} -\cdot/\cdot\cdot\cdot\cdot ۱ \\ \cdot/\cdot\cdot\cdot\cdot ۱ \end{bmatrix} &= \left[I - \gamma \frac{u_\gamma u_\gamma^T}{u_\gamma^T u_\gamma} \right] \begin{bmatrix} -\cdot/\cdot\cdot\cdot\cdot ۱ \\ \cdot/\cdot\cdot\cdot\cdot ۱ \end{bmatrix} \\ &= \begin{bmatrix} -\cdot/\cdot\cdot\cdot\cdot ۱ \\ \cdot/\cdot\cdot\cdot\cdot ۱ \end{bmatrix} - ۱۰^{-۳} \begin{bmatrix} -\cdot/۲۴۱۴ \\ \cdot/۱۰۰۰ \end{bmatrix} \\ &= ۱۰^{-۳} \begin{bmatrix} \cdot/۱۴۱۴ \\ \cdot/\cdot\cdot\cdot\cdot \end{bmatrix} \end{aligned}$$

بنابراین

$$H_\gamma H_\gamma b = \begin{bmatrix} -۲ \\ ۱۰^{-۳}(\cdot/۱۴۱۴) \\ \cdot \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix}$$

در نتیجه

$$c = \begin{bmatrix} -۲ \\ \cdot/\cdot\cdot\cdot\cdot ۱ \end{bmatrix} \quad (\text{با } ۴ \text{ رقم حسابی})$$

گام ۳: $Rx = c$ را حل می‌کنیم

$$x = \begin{bmatrix} ۱ \\ ۱ \end{bmatrix}$$

استفاده از دورانهای گیونز

البته می‌توانیم از دورانهای گیونز برای تجزیه A به QR استفاده کنیم و سپس از این تجزیه برای حل مسائل کمترین توانهای دوم استفاده نماییم. ولیکن همان گونه که قبلاً دیدیم استفاده از دورانهای گیونز گران‌تر از استفاده از ماتریسهای هاوس هولدر خواهد بود. یادآور می‌شویم که محاسبات دورانهای گیونز نیاز به محاسبه ریشه‌های دوم دارند؛ ولیکن دورانهای گیونز مستقل از ریشه دوم توسط ژنتمن (Gentleman (1973) ارائه شده‌اند که می‌توانند برای حل مسأله کمترین توانهای دوم مورد استفاده قرار گیرند. دورانهای گیونز مستقل از ریشه دوم به دورانهای گیونز سریع معروف هستند.

استفاده از الگوریتمهای گرام اشمیت و گرام اشمیت اصلاح شده

روش دیگر، معروف به فرایند گرام اشمیت، می‌تواند برای تجزیه A به QR استفاده شود. یک فرایند گرام اشمیت که به طور صحیح پیاده‌سازی شده باشد، هنگامی که برای حل مسأله کم‌ترین توانهای دوم به کار می‌رود، دقیقاً کمی گران‌تر از روش هاوس هولدر می‌باشد. ولیکن به نظر می‌رسد برای حل مسأله کم‌ترین توانهای دوم از لحاظ عددی به همان مؤثری روش هاوس هولدر باشد. ابتدا این الگوریتم را برای تجزیه QR بیان می‌کنیم و سپس نشان می‌دهیم که چگونه الگوریتم گرام اشمیت می‌تواند برای حل مسأله کم‌ترین توانهای دوم مورد استفاده قرار گیرد.

الگوریتم ۷-۸-۳ گرام اشمیت کلاسیک (CGS) برای تجزیه QR . با مفروض بودن $A = (a_1, a_2, \dots, a_n)$ از مرتبه $m \times n$ با $\text{rank}(A) = n$ ، الگوریتم زیر یک ماتریس $Q_{m \times n} = (q_1, \dots, q_n)$ با ستونهای یکا متعامد و ماتریس بالا مثلثی $R = (r_{ij})_{n \times n}$ را به قسمی می‌سازد که $A = Q_{m \times n} R_{n \times n}$.

For $k = 1, 2, \dots, n$ do

For $i = 1, 2, \dots, k - 1$ do

$$r_{ik} = q_i^T a_k$$

$$q_k = a_k - \sum_{i=1}^{k-1} r_{ik} q_i$$

$$r_{kk} = \|q_k\|_2$$

$$q_k \equiv \frac{q_k}{r_{kk}}$$

الگوریتم به صورت طرح شده در اینجا معروف به داشتن مشکلات عددی جدی است. در طی محاسبه q_k ها حذفها می‌توانند صورت گیرند و به عنوان یک نتیجه، q_k های محاسبه شده می‌توانند از متعامد بودن دور باشند. (جزئیات را در بخش بعد ذکر خواهیم کرد).

و لیکن برای داشتن خواص عددی بهتر الگوریتم می‌تواند اصلاح شود. الگوریتم زیر، معروف به الگوریتم گرام اشمیت اصلاح شده تجزیه QR ماتریس A را محاسبه می‌کند، که در آن درگام k ام ستون k ام Q و سطر k ام ماتریس R محاسبه می‌شوند. (توجه کنید که الگوریتم گرام اشمیت ستونهای Q و R را درگام k محاسبه می‌کند).

الگوریتم ۷-۸-۴ گرام اشمیت اصلاح شده (MGS) برای تجزیه QR

Set $q_k = a_k, k = 1, 2, \dots, n$.

For $k = 1, 2, \dots, n$ do

$$r_{kk} = \|q_k\|_2$$

$$q_k \equiv \frac{q_k}{r_{kk}}$$

For $j = k + 1, \dots, n$ do

$$r_{kj} \equiv q_k^T q_j$$

$$q_j = q_j - r_{kj} q_k$$

این الگوریتم روش گرام اشمیت اصلاح شده سطرگرا است. گونه ستونگرا به طور مشابه می تواند تدوین یابد (تمرین ۱۷). دو گونه از لحاظ عددی معادل هستند.

تبصره: برای تجزیه QR ماتریس A الگوریتم گرام اشمیت اصلاح شده به رضایتبخشی روش هاوس هولدر یا روش گیونز نیست. می توان نشان داد (مرجع Bjorck (1994a, 1994b) را ملاحظه کنید) که Q محاسبه شده توسط MGS، نمایش داده شده توسط \tilde{Q} ، در رابطه $\tilde{Q}^T \tilde{Q} = I + E$ صدق می کند که در آن $\|E\| \approx \mu \text{Cond}(A)$ ، در حالی که روش هاوس هولدر $\tilde{Q}^T \tilde{Q} = I + E$ را تولید می کند که در آن $\|E\| \approx \mu$. (برای جزئیات بیشتر بحثهای بخش بعد را ملاحظه کنید).

شمارش اعمال ممیز شناور. تعداد اعمال ممیز شناور برای MGS برابر mn^2 می باشد که قابل مقایسه با $n^3/3 - mn^2$ عمل ممیز شناور لازم برای روش هاوس هولدر می باشد (توجه کنید که MGS با بردار ستونی با طول کامل در هر گام کار می کند، در صورتی که روش هاوس هولدر متوالیاً به ستونهای کوتاه تر می پردازد).

تذکر MATCOM: الگوریتم ۷-۸-۳ و ۷-۸-۴ در برنامه های CLGRSCH و MDGRSCH در MATCOM پیاده سازی شده اند.

مثال ۷-۸-۶ تجزیه QR ماتریس

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 6 & 7 \end{bmatrix}$$

را با استفاده از فرایند گرام اشمیت اصلاح شده (الگوریتم ۷-۸-۴) پیدا کنید.

$$: k = 1$$

$$r_{11} = \|q_1\|_2 = \|a_1\|_2 = 6/4.31$$

$$q_1 = \frac{a_1}{r_{11}} = (0/1562, 0/3123, 0/9370)$$

$$r_{12} = q_1^T q_2 = q_1^T a_2 = 7/8.87$$

$$q_2 \equiv q_2 - r_{12}q_1 = a_2 - r_{12}q_1 = \begin{bmatrix} 0/78.05 \\ 0/561.0 \\ -0/3171 \end{bmatrix}$$

$$: k = 2$$

$$r_{22} = 1/0.121$$

$$q_2 \equiv \frac{q_2}{r_{22}} = (0/7711, 0/5543, 0/3133)^T$$

بنابراین

$$Q = (q_1, q_2) = \begin{bmatrix} 0/1562 & 0/7711 \\ 0/3123 & 0/5543 \\ 0/9370 & -0/3133 \end{bmatrix}_{3 \times 2}$$

$$R = \begin{bmatrix} 6/4.31 & 7/8.87 \\ \cdot & 1/0.121 \end{bmatrix}_{2 \times 2}$$

تحقیق کنید که $A = QR$. توجه کنید که اگر شما QR هاوس هولدر را استفاده کرده بودید (الگوریتم ۳-۴-۵) نتایج به صورت زیر می‌بود.

$$Q = \begin{bmatrix} -0/1562 & -0/7711 & -0/6172 \\ 0/3123 & -0/5543 & 0/7715 \\ -0/9370 & 0/3133 & -0/5143 \end{bmatrix}_{3 \times 3}$$

$$R = \begin{bmatrix} -6/4.31 & -7/8.87 \\ \cdot & -1/0.121 \\ \cdot & \cdot \end{bmatrix}_{3 \times 3}$$

الگوریتم گرام اشمیت اصلاح شده در برابر الگوریتم گرام اشمیت کلاسیک از لحاظ ریاضی الگوریتمهای CGS و MGS معادل هستند. ولیکن همان گونه که قبلاً ملاحظه کردیم

خواص عددی آنها متفاوت است. برای مثال محاسبه q_2 را توسط روش CGS با مفروض بودن q_1 با $\|q_1\|_2 = 1$ ملاحظه کنید. داریم

$$q_2 = a_2 - r_{12}q_1$$

که در آن $r_{12} = q_1^T a_2$. سپس می توان نشان داد (Björck (1994b)) که

$$\|fl(q_2) - q_2\| < (1/06)(2m + 3)\mu \|a_2\|_2$$

چون $q_1^T q_2 = 0$ ، نتیجه می شود که

$$\|q_1^T fl(q_2)\| < (1/06)(2m + 3)\mu \|a_2\|_2$$

این نتیجه نشان می دهد که در CGS دوبردار محاسبه شده q_1 و q_2 می توانند از متعامد بودن دور باشند. از طرف دیگر می توان نشان داد (Björck (1994b)) که در MGS از دست دادن تعامد بستگی به عدد وضعیت ماتریس A دارد. بویژه می توان نشان داد که

$$\|I - \tilde{Q}^T \tilde{Q}\|_2 \leq \frac{c_1 \mu \text{Cond}_2(A)}{1 - c_2 \mu \text{Cond}_2(A)}$$

با فرض این که $c_2 \mu \text{Cond}_2(A) < 1$ ، در اینجا c_1 و c_2 ثابتهای کوچکی هستند.

چون در MGS از دست دادن تعامد بستگی به عدد وضعیت دارد، می توانیم از محورگیری ستونی استفاده کنیم تا تعامد تا حد ممکن حفظ شود. بنابراین تا آن جا که به پیدا کردن تجزیه QR ماتریس A مربوط می شود هیچ الگوریتمی نمی تواند بر روشهای هاوس هولدر یا گینوزن ترجیح داده شود. با CGS متعامد بودن Q می تواند به طور کامل از بین برود؛ با MGS متعامد بودن Q ممکن است قابل پذیرش نباشد هنگامی که A بدوضع است.

برای مثال برای یک ماتریس 3×4 از نوع ارائه شده در مثال ۷-۸-۵، ماتریس \tilde{Q} محاسبه شده توسط CGS به گونه ای است که $|q_1^T q_2| = \frac{1}{3}$ و برای MGS، $|q_1^T q_2| = \left(\frac{2}{3}\right)^{\frac{1}{2}} \times 10^{-4}$. مرجع Björck (1994b) را ملاحظه کنید.

برای مقایسه محاسبات با فرایندهای گرام اشمیت مختلف، ما همچنین تجزیه QR ماتریس 5×5 هیلبرت را با استفاده از روشهای گرام اشمیت، گرام اشمیت اصلاح شده، هاوس هولدر (که معروف به خیلی پایدار می باشد) محاسبه کرده ایم و نتایج $\|I - \tilde{Q}^T \tilde{Q}\|$ را با دقت توسعه یافته جدول بندی کرده ایم. نتایج در جدول ۷-۱ نمایش داده شده اند.

تبصره: جدول به طور واضح برتری روش هاوس هولدر را بر روشهای گرام اشمیت و گرام اشمیت اصلاح شده نشان می دهد. از این دو روش آخر MGS به طور واضح بر CGS ترجیح داده می شود.

جدول ۷-۱ مقایسه تجزیه QR باروشهای مختلف

روش	$\ I - \tilde{Q}^T \tilde{Q}\ $
گرام اشمیت	$1/178648780488241 \times 10^{-7}$
گرام اشمیت اصلاح شده	$2/504305729523455 \times 10^{-12}$
هاوس هولدر	$2/841449989971528 \times 10^{-16}$

استفاده از روش MGS در حل کم ترین توانهای دوم

اکنون دیدیم که روش MGS بر روش CGS برای تجزیه QR برتری دارد، اما باز هم کاملاً رضایتبخش نیست. ولیکن تا آنجا که به مسأله کم ترین توانهای دوم مربوط می شود، داستان متفاوت است. به نظر می رسد که گرایش زیادی به استفاده از MGS در حل مسائل کم ترین توانهای دوم وجود داشته باشد. بیورک و پیچ ((Björck and Paige (1992)) نشان داده اند که MGS از لحاظ عددی معادل با تجزیه QR هاوس هولدر به کار برده شده در مورد A افزایش یافته به یک ماتریس مربعی با عناصر صفر در بالا می باشد. ولیکن به هنگام استفاده از MGS در حل یک مسأله کم ترین توانهای دوم باید دقت خاصی صورت پذیرد. اگر $c = Q^T b$ با استفاده از Q به دست آمده از MGS محاسبه شود و این c برای حل $Rx = c$ استفاده شود، آنگاه جواب کم ترین توانهای دوم محاسبه شده ممکن است دقیق نباشد. MGS باید در مورد ماتریس افزوده (A, b) به کار رود تا تجزیه زیر به دست آید:

$$(A, b) = (Q_1, q_{n+1}) \begin{bmatrix} R & z \\ \cdot & \rho \end{bmatrix}$$

از این نتیجه می شود که

$$\begin{aligned} Ax - b &= (A, b) \begin{bmatrix} x \\ -1 \end{bmatrix} \\ &= (Q_1, q_{n+1}) \begin{bmatrix} R & z \\ \cdot & \rho \end{bmatrix} \begin{bmatrix} x \\ -1 \end{bmatrix} \\ &= Q_1 (Rx - z) - \rho q_{n+1} \end{aligned}$$

اگر q_{n+1} بر Q_1 متعامد باشد، آنگاه $\|Ax - b\|_2$ می‌نیمم خواهد بود هنگامی که $Rx = z$. بنابراین جواب کمترین توانهای دوم می‌تواند با حل

$$Rx = z$$

به دست آید و

مانده r توسط $r = \rho q_{n+1}$ ارائه شود.

جزئیات این مطالب را می‌توان در مرجع Björck (1994a or 1994b) پیدا کرد. بحث قبل به الگوریتم کمترین توانهای دوم زیر منجر می‌شود.

الگوریتم ۷-۸-۵ جواب کمترین توانهای دوم توسط MGS

گام ۱: MGS را در مورد $A_{m \times n}$ به کار ببرید و $Q = (q_1, \dots, q_n)$ و R را به دست آورید.
گام ۲:

For $k = 1, \dots, n$ do

$$\delta_k = q_k^T b$$

$$b \equiv b - \delta_k q_k$$

گام ۳: $Rx = (\delta_1, \dots, \delta_n)^T$ را حل کنید.

مثال ۷-۸-۷ حل مسأله کمترین توانهای دوم را با استفاده از MGS و با داده‌های زیر ملاحظه کنید.

$$A = \begin{bmatrix} 1 & 1 \\ \cdot/\cdot\cdot\cdot\cdot & \cdot \\ \cdot & \cdot/\cdot\cdot\cdot\cdot \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ \cdot/\cdot\cdot\cdot\cdot \\ \cdot/\cdot\cdot\cdot\cdot \end{bmatrix}$$

جواب واقعی $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ می‌باشد.

$$Q = \begin{bmatrix} 1 & \cdot \\ \cdot/\cdot\cdot\cdot\cdot & \cdot\cdot/\cdot\cdot\cdot\cdot\cdot \\ \cdot & \cdot/\cdot\cdot\cdot\cdot \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 1 \\ \cdot & \cdot/\cdot\cdot\cdot\cdot \end{bmatrix}$$

اکنون اگر $c = Q^T b$ را تشکیل دهیم و $Rx = c$ را حل کنیم جواب $x = \begin{bmatrix} ۲ \\ ۰ \end{bmatrix}$ را به دست می آوریم. از طرف دیگر اگر x را با استفاده از الگوریتم قبل به دست آوریم، $(\delta_1, \delta_2) = (۲, ۰/۰۰۰۱)$ را به دست آوریم، و جواب $Rx = (\delta_1, \delta_2)^T$ بردار $x \approx \begin{bmatrix} ۱ \\ ۱ \end{bmatrix}$ می باشد.

خاصیت گرد کردن و شمارش اعمال ممیز شناور. می توان نشان داد (Björck and Paige (1992)) که الگوریتم MGS برای روش تجزیه QR از لحاظ عددی معادل با روش هاوس هولدر به کار برده شده در مورد ماتریس افزوده زیر می باشد.

$$\begin{bmatrix} \cdot & \cdot \\ A & b \end{bmatrix}$$

یعنی

$$H_n H_{n-1} \dots H_2 H_1 \begin{bmatrix} \cdot & \cdot \\ A & b \end{bmatrix} = \begin{bmatrix} R & c_1 \\ \cdot & c_2 \end{bmatrix}$$

از این معادل بودن، نتیجه می شود که روش MGS برای جواب مسأله کم ترین توانهای دوم پایدار پسترو است. روش کمی گرانتر از روش هاوس هولدر است. حدوداً به mn^2 عمل ممیز شناور (flops) نیاز دارد که قابل مقایسه با $n^3/۳ - mn^2$ عمل ممیز شناوری است که برای روش هاوس هولدر لازم می باشد.

تذکر MATCOM : الگوریتم ۷-۸-۵ در برنامه LSFRMGS در MATCOM پیاده سازی شده است.

۷-۸-۳ روش تجزیه QR برای حالت رتبه ناقص

در این بخش مسأله کم ترین توانهای دوم با رتبه ناقص را در نظر می گیریم. همان گونه که در قضیه ۷-۳-۱ بیان شد در این حالت جوابهای بیشماری وجود دارند. مثالهایی وجود دارند که در آنها رتبه ناقص در حقیقت مطلوب است، زیرا رتبه ناقص یک خانواده غنی از جوابهایی را فراهم می سازد که می توانند برای بهینه سازی برخی جنبه های دیگر مسأله اصلی مورد استفاده قرار گیرند.

در حالتی که ماتریس A با مرتبه $m \times n$ ، $m \geq n$ ، دارای رتبه $r < n$ است، ماتریس R در تجزیه QR ماتریس A دارای رتبه ناقص است. ولیکن دیدیم که استفاده از تجزیه QR با محورگیری ستونی می تواند، از لحاظ نظری، رتبه ماتریس A را آشکار سازد. یادآور می شویم که روش هاوس هولدر با محورگیری ستونی نتیجه می دهد

$$Q^T A P = \begin{bmatrix} R_{11} & R_{12} \\ \cdot & \cdot \end{bmatrix}$$

که در آن P یک ماتریس جایگشت است، R_{11} یک ماتریس بالا مثلثی نامنفرد و R_{12} از مرتبه $r \times (n - r)$ است. این تجزیه می‌تواند به طور واضح برای حل مسأله کمترین توانهای دوم با رتبه ناقص به صورت زیر استفاده شود.

قرار می‌دهیم

$$PTx = y$$

$$AP = \tilde{A}$$

$$Q^T b = \begin{bmatrix} c \\ d \end{bmatrix}$$

سپس

$$\|Ax - b\|_2^2 = \|Q^T A P P^T x - Q^T b\|_2^2 \quad (2- \text{نرم تحت ضرب ماتریسی})$$

متعامد پایا باقی می‌ماند)

$$= \left\| \begin{bmatrix} R_{11} & R_{12} \\ . & . \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} - \begin{bmatrix} c \\ d \end{bmatrix} \right\|_2^2$$

$$= \|(R_{11}y_1 + R_{12}y_2 - c)\|_2^2 + \|d\|_2^2$$

بنابراین $\|Ax - b\|_2^2$ می‌نیمم خواهد بود اگر y_1 به قسمی انتخاب شود که

$$R_{11}y_1 = c - R_{12}y_2$$

بعلاوه نرم مانده در این حالت، به صورت زیر ارائه می‌شود

$$\|r\|_2 = \|b - Ax\|_2 = \|d\|_2$$

این مشاهده الگوریتم تجزیه QR زیر را برای جوابهای کمترین توانهای دوم با رتبه ناقص پیشنهاد می‌کند.

الگوریتم ۷-۸-۶ جوابهای کمترین توانهای دوم برای مسأله با رتبه ناقص با استفاده از تجزیه QR

گام ۱: با استفاده از تجزیه QR با محورگیری ستونی A را به صورت زیر تجزیه کنید:

$$Q^T A P = \begin{bmatrix} R_{11} & R_{12} \\ . & . \end{bmatrix}$$

گام ۲: $Q^T b = \begin{bmatrix} c \\ d \end{bmatrix}$ را تشکیل دهید.

گام ۳: یک بردار دلخواه y_2 را انتخاب کنید.

گام ۴: دستگاه بالا مثلثی نامنفرد $r \times r$ زیر را حل کنید.

$$R_{11} y_1 = c - R_{12} y_2$$

گام ۵: x را از رابطه زیر بازبازی کنید.

$$x = Py, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

تذکر محاسباتی: دوباره گام ۱ و گام ۲ باید بدون محاسبه صریح ماتریس Q پیاده‌سازی شوند.

جواب با می‌نیمم نرم

چون y_2 دلخواه است، همان‌گونه که قضیه ۷-۳-۱ بیان می‌کند، روش قبل تعداد بیشماری جواب کم‌ترین توانهای دوم را ارائه می‌دهد. جواب حاصل از قرار دادن $y_2 = 0$ جواب پایه‌ای نامیده می‌شود. در حالت $R_{12} = 0$ ، جواب پایه‌ای جواب کم‌ترین توانهای دوم با می‌نیمم نرم نامیده می‌شود، یعنی، در میان تعداد بیشمار جوابهای کم‌ترین توانهای دوم، این جواب دارای می‌نیمم نرم است. در حالت $R_{12} \neq 0$ ، جواب پایه‌ای نمی‌تواند جواب با می‌نیمم نرم باشد (تمرین ۲۰). در چنین حالتی جواب با می‌نیمم نرم با اختیار کردن y_2 به صورت جواب $\|Sy_2 - t\|_2 = \min_y \|Sy_2 - t\|_2$ حاصل می‌گردد، در اینجا

$$S = \begin{bmatrix} R_{11}^{-1} R_{12} \\ -I_{(n-r) \times (n-r)} \end{bmatrix}$$

$$t = \begin{bmatrix} R_{11}^{-1} c \\ \cdot \\ \cdot \end{bmatrix}$$

به روش دیگر می‌توانیم از تجزیه QR کامل با محورگیری ستونی استفاده کنیم. از فصل ۵ یادآوری می‌کنیم که تجزیه QR کامل یک ماتریس $m \times n$ مانند A به صورت زیر ارائه می‌شود.

$$Q^T A W = \begin{bmatrix} T & \cdot \\ \cdot & \cdot \end{bmatrix}$$

اگر محورگیری ستونی استفاده شود، آنگاه

$$AP = Q \begin{bmatrix} T & \cdot \\ \cdot & \cdot \end{bmatrix} W^T$$

که در آن P یک ماتریس جایگشت است.

سپس جواب با می نیمم نرم به صورت زیر ارائه خواهد شد.

$$x = PW \begin{bmatrix} T^{-1}c \\ \cdot \end{bmatrix}$$

که در آن c یک بردار متشکل از r عنصر اول $Q^T b$ می باشد (تمرین ۲۱ (الف)).

تبصره‌ها:

۱. یک تذکر در مورد استفاده از محورگیری ستونی: نشان داده‌ایم که محورگیری ستونی برای مسأله کمترین توانهای دوم با رتبه ناقص مفید است. ولیکن حتی در حالت رتبه کامل، استفاده از محورگیری ستونی پیشنهاد می شود.
۲. برای مسأله کمترین توانهای دوم با رتبه ناقص، قابل اطمینانترین راه استفاده از تجزیه مقدار تکین است (بخش ۷-۸-۴ و فصل ۱۰ را ملاحظه کنید).

خاصیت گرد کردن. می توان نشان داد (Lawson and Hanson (1974, 95)) که برای مسأله کمترین توانهای دوم با طول می نیمم، بردار محاسبه شده \hat{x} نزدیک به جواب واقعی یک مسأله اختلال یافته است. یعنی یک ماتریس E و بردارهای \hat{x} و δb وجود دارند به قسمی که \hat{x} جواب کمترین توانهای دوم با طول می نیمم دستگاه $(A + E)\hat{x} = b + \delta b$ می باشد، که در آن

$$\|E\|_F \leq (\rho_m + \rho_n - \rho_r - 3s + 84)s\mu \|A\|_F + O(\mu^2)$$

و $\|\delta b\|_F \leq (\rho_m - 3r + 40)r\mu \|b\| + O(\mu^2)$ که در آن r رتبه A و $s = \min(m, n)$. بعلاوه

$$\|\bar{x} - \hat{x}\| \leq (\rho_n - \rho_r + 43)r\mu \|\hat{x}\| + O(\mu^2)$$

تذکر: در نتایج قبل فرض کرده‌ایم که R_{rr} در

$$R = \begin{bmatrix} R_{11} & R_{12} \\ \cdot & R_{rr} \end{bmatrix}$$

صفر است. اما در محاسبات عملی مساوی صفر نخواهد بود. در آن حالت اگر $\widehat{R}_{\gamma\gamma}$ گونه محاسبه شده $R_{\gamma\gamma}$ باشد، آنگاه داریم

$$\|E\|_F \leq \|\widehat{R}_{\gamma\gamma}\|_F + (\epsilon m + \epsilon n - \epsilon r - \gamma s + \lambda^4) s \mu \|A\|_F + O(\mu^2)$$

تذکر MATCOM : الگوریتم ۷-۸-۶ در برنامه LSRDQRH در MATCOM با استفاده از تجزیه QR هاوس هولدر با محورگیری ستونی پیاده‌سازی شده است.

مثال ۷-۸-۸

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} \quad \text{rank}(A) = 1$$

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad AP = QR$$

$$Q = \begin{bmatrix} -0.4472 & -0.8944 & 0 \\ -0.8944 & 0.4472 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R = \begin{bmatrix} -2/2361 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$c = (-6/7082), \quad Q^T b = \begin{bmatrix} -6/7082 \\ 0 \\ 0 \end{bmatrix} \quad \text{گام ۲}$$

گام ۳: انتخاب $y_2 = 0$

گام ۴: حل دستگاه $R_{11}y_1 = c - R_{12}y_2$

$$y_1 = \frac{c}{R_{11}} = \frac{-6/7082}{-2/2361} = 3$$

جواب کم‌ترین توانهای دوم با می‌نیمم نرم به صورت زیر است

$$x = y = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

مثال ۷-۸-۹

$$A = \begin{bmatrix} 0.6667 & 0.3333 \\ -1/3333 & -0.6667 \\ -1/3333 & -0.6667 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

گام ۱:

$$AP = QR, \quad P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} -2 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

گام ۲:

$$Q^T b = \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} -1/000 \\ -1/4142 \\ 0 \end{bmatrix}$$

$$c = 1/000$$

گام ۳: چون $R_{12} = -1 \neq 0$ ، y_2 را به صورت جواب کمترین توانهای دوم با S و t که به صورت زیر تعریف می‌شوند، حساب می‌کنیم

$$S = \begin{bmatrix} 0.5000 \\ -1/000 \end{bmatrix}, \quad t = \begin{bmatrix} -0.5000 \\ 0.0000 \end{bmatrix}$$

$$y_2 = -0.2000$$

گام ۴: y_1 با حل دستگاه زیر محاسبه می‌شود

$$R_{11} y_1 = c - R_{12} y_2$$

$$y_1 = -0.4000$$

گام ۵: جواب کمترین توانهای دوم با می‌نیم نرم به صورت زیر است

$$x = y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -0.4000 \\ -0.2000 \end{bmatrix}$$

۷-۸-۴ جواب کمترین توانهای دوم با استفاده از SVD

با عمل نمودن بر طبق تبصره بخش قبل، به خاطر خواننده‌های کنجکاو و آنهایی که می‌دانند که چگونه

تجزیه مقدار تکین (SVD) یک ماتریس را محاسبه کنند (مثلاً با استفاده از بسته‌های نرم‌افزاری نظیر LAPACK، MATLAB، یا LINPACK) نتایج زیر را بیان می‌کنیم که نشان می‌دهند که چگونه از SVD می‌توان برای حل یک مسأله کم‌ترین توانهای دوم استفاده کرد. یک طرز عمل کامل در فصل ۱۰ ارائه می‌شود
فرض کنید

$$A = U\Sigma V^T$$

تجزیه مقدار تکین (SVD) A باشد، که در آن A از مرتبه $m \times n$ با $m \geq n$ است و $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ فرض کنید

$$b' = U^T b = \begin{bmatrix} b'_1 \\ \vdots \\ b'_m \end{bmatrix}$$

جواب کم‌ترین توانهای دوم با استفاده از SVD
یک خانواده از جوابهای کم‌ترین توانهای دوم به صورت
 $x = Vy$

ارائه می‌شود که در آن

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}; \quad y_i = \begin{cases} \frac{b'_i}{\sigma_i}, & \sigma_i \neq 0 \\ \text{دلخواه}, & \sigma_i = 0 \end{cases}$$

در حالت رتبه ناقص، یک جواب با می‌نیمم نرم عبارت است از

$$x = \sum_{i=1}^r \frac{u_i^T b}{\sigma_i} v_i$$

که در آن $r = \text{rank}(A)$ و u_i و v_i به ترتیب ستونهای U و V هستند.
(توجه کنید که تعداد مقادیر تکین مخالف صفر $(\text{rank}(A) = A$

مثال ۷-۸-۱۰

$$A = \begin{bmatrix} 1 & 1 \\ \cdot/\cdot\cdot\cdot\cdot 1 & \cdot \\ \cdot & \cdot/\cdot\cdot\cdot\cdot 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ \cdot/\cdot\cdot\cdot\cdot 1 \\ \cdot/\cdot\cdot\cdot\cdot 1 \end{bmatrix}$$

تجزیه $A = U\Sigma V^T$ ارائه می‌دهد

$$U = \begin{bmatrix} 1 & \cdot & -\cdot/\cdot\cdot\cdot\cdot 1 \\ \cdot & -\cdot/\cdot\cdot\cdot\cdot 1 & \cdot/\cdot\cdot\cdot\cdot 1 \\ \cdot & \cdot/\cdot\cdot\cdot\cdot 1 & \cdot/\cdot\cdot\cdot\cdot 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1/\cdot\cdot\cdot\cdot\cdot 2 & \cdot \\ \cdot & \cdot/\cdot\cdot\cdot\cdot 1 \\ \cdot & \cdot \end{bmatrix}$$

$$V = \begin{bmatrix} \cdot/\cdot\cdot\cdot\cdot 1 & -\cdot/\cdot\cdot\cdot\cdot 1 \\ \cdot/\cdot\cdot\cdot\cdot 1 & \cdot/\cdot\cdot\cdot\cdot 1 \end{bmatrix}$$

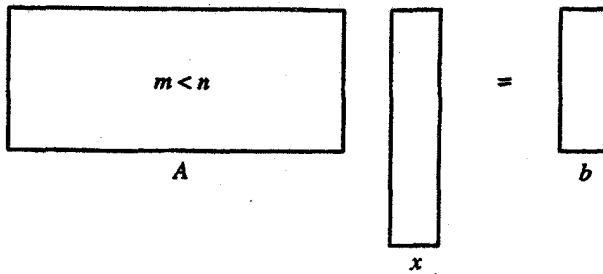
$$b' = U^T b = \begin{bmatrix} 2 \\ \cdot \\ \cdot \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1/\cdot\cdot\cdot\cdot\cdot 2 \\ \cdot \end{bmatrix}, \quad x = Vy = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

۷-۹ دستگاههای فرومعیین

فرض کنید A از مرتبه $m \times n$ با $m < n$ باشد. آنگاه در دستگاه

$$Ax = b$$

تعداد معادلات کم‌تر از تعداد مجهولات است. یک چنین دستگاهی یک دستگاه فرومعیین نامیده می‌شود. یک دستگاه فرومعیین را می‌توان به صورت نموداری نشان داده شده در شکل ۷-۳ نشان داد.



شکل ۳-۷

دستگاههای فرومعین در کاربردهای عملی متنوعی مطرح می‌شوند. متأسفانه دستگاههای فرومعین به طور وسیعی در نوشتجات بحث نشده‌اند. یک منبع عالی مقالهٔ مروری کلین و پلمونس (Cline and Plemmons (1976)) می‌باشد. یک دستگاه فرومعین یا جواب ندارد و یا تعداد بیشماری جواب دارد. این را می‌توان از قضیه زیر ملاحظه کرد.

قضیه ۷-۹-۱ فرض کنید $Ax = b$ یک دستگاه فرومعین باشد. هر جواب x این دستگاه را می‌توان به صورت زیر نمایش داد.

$$x = x_R + x_N$$

که در آن x_R می‌تواند هر جوابی باشد، یعنی

$$Ax_R = b$$

و x_N در فضای پوچ A است، یا

$$Ax_N = 0$$

تذکره: یک دستگاه فرومعین با رتبه کامل دارای تعداد بیشماری جواب است، یعنی در این حالت مجموعه جواب کامل به صورت زیر ارائه می‌شود

$$x = A^T(AA^T)^{-1}b + (I - A^T(AA^T)^{-1}A)y$$

که در آن y دلخواه است

در زیر دو رابطه برای محاسبه جواب با می نیمم نرم با فرض کامل بودن رتبه A ، مثلاً $\text{rank}(A) = m$ ، بیان می کنیم.

۱-۹-۷ جواب با می نیمم نرم برای یک مسأله فرومعین با رتبه کامل با استفاده از معادلات نورمال

قضیه ۲-۹-۷ فرض کنید A از مرتبه $m \times n$ با $m < n$ و دارای رتبه کامل باشد. آنگاه جواب با می نیمم نرم x برای دستگاه فرومعین

$$Ax = b \quad (1-9-7)$$

به صورت زیر ارائه می شود

$$x = A^T(AA^T)^{-1}b \quad (2-9-7)$$

اثبات: چون A دارای رتبه کامل است، AA^T نامنفرد است، و به طور واضح x ارائه شده توسط (۲-۹-۷) یک جواب دستگاه (۱-۹-۷) است. برای دیدن این که x در حقیقت جواب با می نیمم نرم است، اجازه دهید فرض کنیم که y جواب دیگری باشد و بردار

$$z = y - x$$

را تعریف می کنیم. آنگاه

$$\begin{aligned} \|y\|_2^2 &= \|x+z\|_2^2 \\ &= \|x\|_2^2 + \|z\|_2^2 + 2(x^T z) \\ &= \|x\|_2^2 + \|z\|_2^2 + 2(b^T (AA^T)^{-1} A(y-x)) \\ &= \|x\|_2^2 + \|z\|_2^2 \end{aligned}$$

(چون $(A(y-x) = Ay - Ax = b - b = 0$). این نشان می دهد که

$$\|y\|_2 \geq \|x\|_2$$

بنابراین x جواب با می نیمم نرم است.

تذکر: ماتریس $A^T(AA^T)^{-1}$ شبه معکوس A است.

قضیه ۷-۹-۲ فوراً الگوریتم زیر را پیشنهاد می دهد.

الگوریتم ۷-۹-۱ جواب با می نیمم نرم برای یک مسأله فرومعین با رتبه کامل با استفاده از معادلات نرمال

گام ۱: دستگاه $AA^T y = b$ را حل کنید.

گام ۲: جواب با می نیمم نرم $x = A^T y$ را تشکیل دهید.

تذکر MATCOM: الگوریتم ۷-۹-۱ در برنامه MNUDNME در MATCOM پیاده سازی شده است.

مثال ۷-۹-۱

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 6 \\ 9 \end{bmatrix}$$

گام ۱:

$$AA^T = \begin{bmatrix} 14 & 20 \\ 20 & 29 \end{bmatrix}$$

$$y = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

گام ۲:

$$x = A^T y = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

تبصره: همان مشکلاتی که در رابطه با معادلات نرمال در حالت دستگاه فرامعین ذکر کردیم در حالت معادلات فرومعین نیز می توانند مطرح شوند. برای مثال شکل صریح AA^T را با ε رقم در نظر بگیرید هنگامی که

$$A = \begin{bmatrix} 1 & \varepsilon & 0 \\ 1 & 0 & \varepsilon \end{bmatrix}$$

اگر ε به قسمی باشد که $10^{-1/2} < \varepsilon < 10^{-1}$ ، آنگاه ماتریس زیر منفرد است

$$AA^T = \begin{bmatrix} 1 + \varepsilon^2 & \varepsilon \\ \varepsilon & 1 + \varepsilon^2 \end{bmatrix}$$

۲-۹-۷ راه QR برای مسأله فرومعین با رتبه کامل

البته تجزیه QR ماتریس A می تواند برای حل یک دستگاه فرومعین، بویژه برای جواب با می نیمم نرم استفاده شود.

چون در این حالت ستونهای A بیشتر از سطرها A هستند، ما A^T را به جای A به QR تجزیه می کنیم. بنابراین داریم

$$Q^T A^T = \begin{bmatrix} R_1 \\ \cdot \end{bmatrix}$$

(توجه کنید A^T از مرتبه $m \times n$ با $n > m$ است.) بنابراین دستگاه $Ax = b$ به صورت زیر در می آید

$$(R_1^T, \cdot^T) Q^T x = b$$

قرار دهید $y = Q^T x$. آنگاه داریم

$$(R_1^T, \cdot^T) y = b$$

با افزایش y به طور مناسب، به دست می آوریم

$$(R_1^T, \cdot^T) \begin{bmatrix} y_R \\ y_N \end{bmatrix} = b$$

یا

$$R_1^T y_R = b$$

با قرار دادن

$$y_N = \cdot$$

جواب کمترین توانهای دوم با می نیمم نرم به دست می آید. بنابراین جواب با می نیمم نرم به صورت زیر ارائه می شود

$$x = Qy = (Q_1, Q_2) \begin{bmatrix} y_R \\ \cdot \end{bmatrix} = Q_1 y_R$$

بحث قبل الگوریتم زیر را پیشنهاد می دهد.

الگوریتم ۲-۹-۷ جواب با می نیمم نرم برای مسأله فرومعین با رتبه کامل با استفاده از

گام ۱: تجزیه QR ماتریس A^T را پیدا کنید:

$$Q^T A^T = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

گام ۲: ماتریس $Q = (Q_1, Q_2)$ را افراز کنید.

گام ۳: دستگاه پایین مثلثی نامنفرد $m \times m$ زیر را حل کنید:

$$R_1^T y_R = b$$

گام ۴: بردار $x = Q_1 y_R$ را تشکیل دهد.

تذکری در مورد پیاده‌سازی. توجه کنید که اگر از روش هاوس هولدر برای محاسبه تجزیه QR ماتریس A استفاده کنیم حاصلضرب $Q_1 y_R$ می‌تواند از روی شکل حاصلضربی Q با استفاده از الگوریتم ۵-۴-۲ محاسبه شود.

شمارش اعمال ممیز شناور. با استفاده از متعامدسازی هاوس هولدر $m^2 n - m^3/3$ عمل ممیز شناور (flops) برای پیاده‌سازی الگوریتم فوق مورد نیاز خواهد بود.

خاصیت گرد کردن. نشان داده شده است (Lawson and Hanson (1974, 93)) که بردار محاسبه شده \bar{x} به جواب دقیق با طول می‌نیمم دستگاه اختلال یافته نزدیک است؛ یعنی یک ماتریس E و یک بردار \bar{x} وجود دارند به قسمی که \bar{x} جواب با طول می‌نیمم دستگاه زیر باشد.

$$(A + E)\bar{x} \approx b$$

که در آن

$$\|E\|_F \leq (\epsilon n - 3m + 41)m\mu \|A\|_F + O(\mu^2) \quad \text{و}$$

$$\|\bar{x} - \bar{x}\| \leq (\epsilon n - 3m + 40)m\mu \|\bar{x}\| + O(\mu^2)$$

تذکر MATCOM: الگوریتم ۷-۹-۲ در برنامه MNUDQRH در MATCOM پیاده‌سازی شده است.

مثال ۷-۹-۲

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 6 \\ 9 \end{bmatrix}$$

گام ۱:

$$Q = \begin{bmatrix} -0./2673 & 0./8729 & 0./4082 \\ -0./5345 & 0./2182 & -0./8165 \\ -0./8018 & -0./4364 & -0./4082 \end{bmatrix}$$

$$Q^T A^T = \begin{bmatrix} -3./7417 & -5./3452 \\ 0./\dots & 0./6547 \\ 0./\dots & 0./\dots \end{bmatrix}$$

گام ۲:

$$Q_1 = \begin{bmatrix} -0./2673 & 0./8729 \\ -0./5345 & 0./2182 \\ -0./8018 & -0./4364 \end{bmatrix}$$

$$R_1 = \begin{bmatrix} -3./7417 & -5./3452 \\ 0./\dots & 0./6547 \end{bmatrix}$$

گام ۳:

$$y_R = \begin{bmatrix} -1./6036 \\ 0./6547 \end{bmatrix}$$

گام ۴: جواب با می نیمم نرم عبارت است از

$$x = Q y_R = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

۷-۹-۳ جواب با می نیمم نرم توسط MGS

یک نتیجه جدید (Björck and Paige (1992)) را ملاحظه کنید) این است که فرایند گرام اشمیت اصلاح شده (MGS) می تواند برای محاسبه جواب با می نیمم نرم یک دستگاه فرومعین، به یک طریق عددی پایدار استفاده شود. پایداری پسر و از معادل بودن عددی MGS با تجزیه هاوس هولدر خاص، به

صورت قبلاً ذکر شده، نتیجه می‌شود. برای جزئیات بیشتر مرجع Björck (1994a) و مرجع Björck and Paige (1992) را ملاحظه کنید.

۷-۱۰ تصفیه تکراری

طبیعی است تعجب کنیم اگر یک جواب کم‌ترین توانهای دوم محاسبه شده x بتواند با هزینه کم به یک طریق تکراری نظیر آنچه که در حالت یک دستگاه خطی انجام شد، بهبود یابد. یک مشابه طبیعی از روش تصفیه تکراری برای مسأله دستگاه خطی توصیف شده در بخش ۶-۹ به صورت زیر می‌باشد. این طرح توسط گولوب (Golub (1965) پیشنهاد شده است.

الگوریتم ۷-۱۰-۱ تصفیه تکراری مشابه دستگاه خطی برای جواب کم‌ترین توانهای دوم. فرض کنید $x^{(1)}$ یک جواب تقریبی برای مسأله کم‌ترین توانهای دوم باشد. آنگاه برای $k = 1, 2, \dots$ تا رخ دادن همگرایی انجام دهید:

- گام ۱: $r^{(k)} = b - Ax^{(k)}$ (مانده را با دقت مضاعف محاسبه کنید).
- گام ۲: $c^{(k)}$ را به قسمی تعیین کنید که $\|Ac^{(k)} - r^{(k)}\|_2$ می‌نیم باشد.
- گام ۳: جواب را تصحیح کنید: $x^{(k+1)} = x^{(k)} + c^{(k)}$.

تذکر MATCOM: الگوریتم ۷-۱۰-۱ در برنامه LSITRN1 در MATCOM پیاده‌سازی شده است.

تحلیلی از این روش تصفیه توسط گولوب و ویلکینسون (Golub and Wilkinson (1966) آشکار می‌سازد که روش فقط هنگامی رضایت‌بخش است که بردار مانده واقعی $r = b - Ax$ به اندازه کافی کوچک باشد. یک فرایند موفق که به طور وسیعی در عمل مورد استفاده قرار می‌گیرد به صورت زیر است. روش بر مشاهده جالب زیر که توسط گولوب انجام شده است مبتنی است. جواب x و بردار مانده متناظر در دستگاه خطی زیر صدق می‌کنند.

$$\begin{bmatrix} I & A \\ A^T & \cdot \end{bmatrix} \begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} b \\ \cdot \end{bmatrix}$$

توجه کنید که این دستگاه با دستگاههای

$$Ax + r = b$$

$$A^T r = 0$$

معادل است، که معنای آن این است که x جواب معادله نرمال زیر است

$$A^T A x = A^T b$$

بنابراین می‌توانیم رویه تصفیه تکراری توصیف شده در فصل ۶ را برای این دستگاه خطی افزوده به کار ببریم و طرح زیر را به دست آوریم. این طرح اثر بیورک (Björck (1967; 1968) است.

الگوریتم ۷-۱۰-۲ تصفیه تکراری برای جواب کم‌ترین توانهای دوم.

$$x^{(0)} = 0, r^{(0)} = 0$$

برای $k = 1, 2, \dots$ انجام دهید.

$$\text{گام ۱: بردار} \begin{bmatrix} r_1^{(k)} \\ r_2^{(k)} \\ x^{(k)} \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} - \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} r_1^{(k)} \\ r_2^{(k)} \\ x^{(k)} \end{bmatrix} \text{ را محاسبه کنید (ذخیره‌سازی در}$$

حاصلضرب داخلی را در دقت مضاعف انجام دهید).

$$\text{گام ۲: دستگاه} \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} c_1^{(k)} \\ c_2^{(k)} \end{bmatrix} = \begin{bmatrix} r_1^{(k)} \\ r_2^{(k)} \end{bmatrix} \text{ را حل کنید.}$$

$$\text{گام ۳: جواب و مانده را به هنگام در آورید:} \begin{bmatrix} r^{(k+1)} \\ x^{(k+1)} \end{bmatrix} = \begin{bmatrix} r^{(k)} \\ x^{(k)} \end{bmatrix} + \begin{bmatrix} c_1^{(k)} \\ c_2^{(k)} \end{bmatrix}$$

تبصره: هایام (Higham (1991) نشان داده است که روش تصفیه تکراری با مانده‌های با دقت ساده، هر چند به مؤثری مسائل دستگاه خطی مربعی (الگوریتم ۶-۹-۱) نیست، اما باز هم یک اندازه مؤلفه‌ای از پایداری پسرو را بهبود می‌بخشد. بنابراین حتی اگر گام ۱ الگوریتم ۷-۱۰-۲ با دقت مضاعف انجام نشود، نتایج باز هم ممکن است رضایت‌بخش باشند مرجع Björck (1994b) را نیز ملاحظه کنید.

تسذکر MATCOM: الگوریتم ۷-۱۰-۲ در برنامه LSITRN2 در MATCOM پیاده‌سازی شده است.

پیاده‌سازی گام ۲

چون ماتریس

$$\begin{bmatrix} I & A \\ A^T & \cdot \end{bmatrix}$$

از مرتبه $m + n$ است، طرح پیشین کاملاً گران خواهد بود اگر m بزرگ باشد. خوشبختانه می‌توانیم طرح نسبتاً ارزانی پیاده‌سازی کنیم. مشاهده می‌کنید که اگر

$$Q^T A = \begin{bmatrix} R_1 \\ \cdot \end{bmatrix}$$

تجزیه A باشد، آنگاه دستگاه

$$\begin{bmatrix} I & A \\ A^T & \cdot \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$$

می‌تواند به دستگاه‌های زیر تبدیل شود

$$Q^T c_1 + \begin{bmatrix} R_1 \\ \cdot \end{bmatrix} c_2 = Q^T r_1$$

$$(R_1^T, \cdot) Q^T c_1 = r_2$$

این نتیجه نشان می‌دهد که دستگاه افزوده قبلی می‌تواند با حل دو دستگاه مثلثی و با محاسبه دو ضرب ماتریس - بردار به صورت زیر حل شود:

$$Q^T r_1 = \begin{bmatrix} r'_1 \\ r'_2 \end{bmatrix} \quad \text{۱. تشکیل}$$

$$۲. \text{ حل دستگاه } R_1^T c'_1 = r'_2 \text{ نسبت به } c'_1$$

$$۳. \text{ حل دستگاه } R_1 c_2 = r'_1 - c'_1 \text{ نسبت به } c_2$$

$$۴. \text{ تشکیل } c_1 = Q \begin{bmatrix} c'_1 \\ r'_2 \end{bmatrix}$$

شمارش اعمال ممیز شناور. با فرمول‌بندی قبلی هر تکرار فقط به $4mn - n^2$ عمل ممیز شناور (flops) نیاز دارد، با فرض این که روش هاوس هولدر استفاده شده است و Q به صورت صریح تشکیل نگردیده است. توجه کنید که برای ضربهای ماتریس - بردار در گامهای ۱ و ۴، Q لازم نیست که به طور صریح تشکیل شود؛ این ضربها می‌توانند به دست آیند اگر Q فقط به شکل ضمنی معلوم باشد.

خطای گرد کردن. می توان نشان داد (Björck (1994b)) که جواب x به دست آمده در تکرار s ام الگوریتم ۷-۱۰-۲ در رابطه

$$\frac{\|x^{(s)} - x\|_2}{\|x^{(s-1)} - x\|_2} < c\mu\text{Cond}(A), \quad s = 2, 3, \dots$$

صدق می کند که در آن $c = c(m, n)$ یک ثابت خطا است.

تعبیری از نتیجه و تبصره ها

نتیجه قبل بیان می کند که روش تصفیه تکراری کاملاً رضایت بخش است. حتی برای مسائل کمترین توانهای دوم با مانده های بزرگ رضایت بخش تر است. توجه کنید که برای این مسائل $(\text{Cond}(A))^2$ به عنوان عدد وضعیت به کار برده می شود. از طرف دیگر این نتیجه نشان می دهد که خطا در یک گام تصفیه تکراری به عدد وضعیت A بستگی دارد. روش ممکن است در دقت ساده جوابهایی با دقت کامل ارائه دهد حتی اگر جواب اولیه هیچ رقم بامعنای درستی نداشته باشد (Björck (1994b)). برای یک ماتریس خوش وضع همگرایی می تواند حتی در یک تکرار رخ دهد. بیورک و گولوب (Björck and Golub (1967)) نشان داده اند که با اختیار کردن A به صورت شش ستون آخر معکوس ماتریس هیلبرت 8×8 ، سه رقم دقت در هر گام هم برای جواب و هم برای مانده به دست می آید.

مثال ۷-۱۰-۱

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 5 \\ 9 \end{bmatrix}$$

$$r^{(0)} = 0, \quad x^{(0)} = 0, \quad k = 0.$$

$$\begin{bmatrix} r_1^{(0)} \\ r_2^{(0)} \end{bmatrix} = \begin{bmatrix} b \\ \cdot \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 9 \\ \cdot \\ \cdot \end{bmatrix}$$

گام ۱: حل دستگاه:

$$\begin{bmatrix} I & A \\ A^T & \cdot \end{bmatrix} \begin{bmatrix} c_1^{(0)} \\ c_2^{(0)} \end{bmatrix} = \begin{bmatrix} r_1^{(0)} \\ r_2^{(0)} \end{bmatrix}$$

$$\begin{bmatrix} c_1^{(0)} \\ c_2^{(0)} \end{bmatrix} = \begin{bmatrix} 0/3333 \\ -0/6667 \\ 0/3333 \\ 3/3333 \\ -0/3333 \end{bmatrix}$$

گام ۲: به هنگام در آوردن جواب و مانده:

$$\begin{bmatrix} r^{(1)} \\ x^{(1)} \end{bmatrix} = \begin{bmatrix} r^{(0)} \\ x^{(0)} \end{bmatrix} + \begin{bmatrix} c_1^{(0)} \\ c_2^{(0)} \end{bmatrix} = \begin{bmatrix} 0/3333 \\ -0/6667 \\ 0/3333 \\ 3/3333 \\ -0/3333 \end{bmatrix}$$

محاسبه $c_1^{(0)}$ و $c_2^{(0)}$ به صورت زیر می باشد.

$$r_1^{(0)} = b, \quad r_2^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} r_1' \\ r_2' \end{bmatrix} = Q^T b = \begin{bmatrix} -10/6904 \\ -0/2182 \\ 0/8165 \end{bmatrix}$$

بنابراین

$$r_1' = \begin{bmatrix} -10/6904 \\ 0/2182 \end{bmatrix}$$

$$r_2' = (0/8165)$$

$$c_2' = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$c_2^{(0)} = \begin{bmatrix} 3/3333 \\ -0/3333 \end{bmatrix}$$

$$c_1^{(0)} = \begin{bmatrix} 0/3333 \\ -0/6667 \\ 3/3333 \end{bmatrix}$$

توجه کنید که

$$x^{(1)} = \begin{bmatrix} ۳/۳۳۳۳ \\ -۰/۳۳۳۳ \end{bmatrix}$$

همان جواب کمترین توانهای دوم به دست آمده توسط روشهای QR و معادلات نرمال می باشد.

۷-۱۱ محاسبه ماتریس واریانس و کوواریانس $(A^T A)^{-1}$

در کاربردهای آماری و پردازش تصویر اغلب نیاز به ماتریس واریانس و کوواریانس

$$X = (A^T A)^{-1}$$

داریم. برای دیدن این که چگونه این ماتریس در تحلیل آماری مطرح می شود، مدل رگرسیون خطی کلاسیک

$$b = Ax + \varepsilon$$

را با خاصیتهای زیر ملاحظه می کنیم.

۱. $E(\varepsilon) = 0$ ؛ و

۲. $\text{cov}(\varepsilon) = E(\varepsilon\varepsilon^T) = \sigma^2 I$

در اینجا b بردار پاسخ، A ماتریس طرح، ε جمله خطا هستند. پارامترهای x و σ^2 مجهول هستند.

تحلیل رگرسیون به پیش بینی یک یا چند متغیر پاسخ (وابسته) از روی مقادیر یک گروه متغیرهای پیشگو (مستقل) مربوط می شود. با فرض این که A دارای رتبه کامل است، آنگاه واضح است که برآورد کمترین توانهای دوم x در مدل پیش بینی به صورت زیر ارائه می شود

$$\hat{x} = (A^T A)^{-1} A^T b$$

همچنین

$$\hat{\varepsilon} = b - \hat{b} = (I - A(A^T A)^{-1} A^T) b$$

(توجه کنید که $A^T \hat{\varepsilon} = 0$ و $\hat{\varepsilon}^T \hat{\varepsilon} = 0$). یک برآورد متداول از σ^2 عبارت است از

$$\hat{\sigma}^2 = \frac{\|b - A\hat{x}\|_2}{m - n}$$

که در آن A یک ماتریس $m \times n$ است. برای جزئیات بیشتر مراجع Johnson and Wichern (1988) و Björck (1994b) را ملاحظه کنید.

چون اطلاعات ضروری ممکن است در محاسبه صریح $A^T A$ از بین بروند، ما متعجب نخواهیم شد اگر محاسبه X از تجزیه QR ماتریس A ممکن باشد. ما نشان خواهیم داد که چگونه این عمل انجام خواهد شد. تجزیه QR ماتریس A را در نظر بگیرید:

$$Q^T A = \begin{bmatrix} \tilde{R} \\ \cdot \end{bmatrix}$$

آنگاه

$$A^T A = (\tilde{R})^T \tilde{R}$$

و بنابراین

$$X = (A^T A)^{-1} = (\tilde{R})^{-1} (\tilde{R})^T$$

به سادگی می‌توانیم $(\tilde{R})^{-1}$ را حساب کنیم، زیرا \tilde{R} یک ماتریس بالا مثلثی است. توجه کنید که چون X متقارن است، فقط یک دوم از عناصرش لازم است که محاسبه شوند، و X می‌تواند بر روی R نوشته شود. ولیکن محاسبه $(A^T A)^{-1}$ به این طریق نیاز به ضرب صریح $(\tilde{R})^{-1} (\tilde{R})^T$ دارد.

از این ضرب صریح می‌توان به سادگی صرف‌نظر کرد اگر محاسبات دوباره تنظیم شوند. بنابراین اگر x_1 تا x_n ستونهای متوالی X باشند، آنگاه از

$$X = (\tilde{R})^{-1} (\tilde{R})^{-T}$$

داریم

$$\tilde{R} x_n = (\tilde{R})^{-T}$$

چون ستون آخر \tilde{R}^{-T} دقیقاً بردار $(1/r_{nn})e_n$ می‌باشد، ستون آخر، x_n ، به طور ساده از حل دستگاه بالا مثلثی زیر محاسبه می‌شود.

$$\tilde{R} x_n = \frac{1}{r_{nn}} e_n \quad (1-11-7)$$

با استفاده از تقارن همچنین سطر آخر X را نیز به دست می‌آوریم.

اکنون فرض کنید که تاکنون $x_{ij} = x_{ji}$ ، $j = n, \dots, k+1$ و $i \leq j$ را تعیین کرده باشیم. اکنون x_{kk} ، $i \leq k$ را تعیین می‌کنیم.

برای x_{kk} داریم:

$$x_{kk} r_{kk} + \sum_{j=k+1}^n r_{kj} x_{jk} = \frac{1}{r_{kk}}$$

که از این رابطه به دست می‌آوریم

$$x_{kk} = \frac{1}{r_{kk}} \left[\frac{1}{r_{kk}} - \sum_{j=k+1}^n r_{kj} x_{jk} \right] \quad (2-11-7)$$

(توجه کنید که $x_{kj} = x_{jk}$ ، $j = k+1, \dots, n$ قبلاً محاسبه شده‌اند.) برای $i = k-1, \dots, 1$ داریم:

$$x_{ik} = -\frac{1}{r_{ii}} \left[\sum_{j=i+1}^k r_{ij} x_{jk} + \sum_{j=k+1}^n r_{ij} x_{kj} \right] \quad (3-11-7)$$

بنابراین (۳-۱۱-۷) - (۱-۱۱-۷) همه عناصر X را تعیین می‌کنند.

الگوریتم ۱-۱۱-۷ محاسبه ماتریس واریانس و کوواریانس $X = (A^T A)^{-1}$

گام ۱: آخرین ستون x_{jk} را با حل دستگاه (۱-۱۱-۷) محاسبه کنید.

گام ۲: x_{kk} را از رابطه (۲-۱۱-۷) به ازای $i = k-1, k-2, \dots, 1$ محاسبه کنید.

گام ۳: x_{ik} را از رابطه (۳-۱۱-۷) به ازای $i = k-1, k-2, \dots, 1$ محاسبه کنید.

شمارش اعمال ممیز شناور. محاسبه $X = (A^T A)^{-1}$ با استفاده از این فرمولها حدوداً نیاز به $n^3/3$ عمل ممیز شناور (flops) دارد.

تذکر MATCOM: الگوریتم ۱-۱۱-۷ در برنامه VARCOVAR در MATCOM پیاده‌سازی شده است.

مثال ۷-۱۱-۱

$$A = \begin{bmatrix} 1 & 5 & 7 \\ 2 & 3 & 4 \\ 1 & 1 & 1 \end{bmatrix}$$

$A = QR$ نتیجه می‌دهد

$$Q = \begin{bmatrix} -0/4082 & 0/9045 & 0/1231 \\ -0/8165 & -0/3015 & -0/4924 \\ -0/4082 & -0/3015 & 0/8616 \end{bmatrix}$$

$$R = \begin{bmatrix} -2/4495 & -4/8990 & -6/5320 \\ 0 & 3/3166 & 4/8242 \\ 0 & 0 & -0/2462 \end{bmatrix}$$

آخرین ستون عبارت است از

$$x_3 = \begin{bmatrix} 4 \\ -24 \\ 16/5 \end{bmatrix} \quad \text{(با استفاده از (۷-۱۱-۱))}$$

$$x_{22} = 35 \quad \text{(با استفاده از (۷-۱۱-۲))}$$

$$x_{12} = -6 \quad \text{(با استفاده از (۷-۱۱-۳))}$$

$$x_{11} = 1/5 \quad \text{(با استفاده از (۷-۱۱-۳))}$$

پس

$$X = (A^T A)^{-1} = \begin{bmatrix} 1/5 & -6 & 4 \\ -6 & 35 & -24 \\ 4 & -24 & 16/5 \end{bmatrix}$$

۱۲-۲ مرور و خلاصه

- وجود و یکتایی: جواب کم‌ترین توانهای دوم برای مسأله $Ax = b$ همیشه وجود دارد. در حالت فرامعین یکتا است اگر و فقط اگر A دارای رتبه کامل باشد (قضیه ۷-۳-۱).
- مسائل فرامعین: دو روش را بحث کردیم: روش معادلات نرمال و روش تجزیه QR . روش معادلات نرمال برای پیاده‌سازی ساده است، اما با این روش در برخی از حالتها مشکلات عددی وجود دارند. روشهای تجزیه QR گران‌تر از روش معادلات نرمال هستند اما از لحاظ عددی قابل

اطمینان تر هستند. اگر برای استفاده از معادلات نرمال اصرار می‌ورزیم، استفاده از دقت توسعه یافته توصیه می‌شود. از تبدیلهای هاوس هولدر یا گیونز یا روش گرام اشمیت اصلاح شده استفاده می‌کنیم تا تجزیه QR ماتریس A را که برای حل مسأله کمترین توانهای دوم لازم می‌باشد، به دست آوریم.

تجزیه QR هاوس هولدر کاراترین و از لحاظ عددی قابل اطمینان‌ترین روش می‌باشد؛ روش گرام اشمیت کمی گران‌تر از روش هاوس هولدر است، اما از لحاظ عددی به قابل اطمینانی روش هاوس هولدر و یا گیونز برای تجزیه QR نمی‌باشد. اما با برخی تنظیم‌های دوباره می‌تواند برای محاسبه جواب کمترین توانهای دوم قابل مقایسه با روش هاوس هولدر گردد. در حقیقت محبوبیت استفاده از روش MGS در حل مسأله کمترین توانهای دوم در حال رشد است. حتی برخی تجربه‌های عددی برتری این روش را بر سایر روشهای QR نشان می‌دهند (Björck (1994a, 1994b)). ابتدا رایس^۱ برتری این الگوریتم را با انجام تجربه‌های عددی ذکر کرد.

برای مسائل فرامعین با رتبه ناقص، جوابهای بیشماری وجود دارد. در اینجا ما فقط روش QR با محورگیری ستونی را بحث کرده‌ایم. این روش به طور معقولی دقیق است، اما از لحاظ عددی مطمئن‌ترین روش در ارتباط با رتبه ناقص روش SVD است که در فصل ۱۰ بحث خواهد شد.

۳. دستگاههای فرامعین: برای مسائل فرامعین جوابهای بیشماری در حالت رتبه کامل وجود دارد. ما روشهای معادلات نرمال و تجزیه QR را برای جواب با می‌نیم نرم بحث کردیم. معادلات نرمالی که حل می‌شوند عبارتند از

$$AA^T y = b; \quad x = A^T y$$

در حالی که این معادلات در مسأله فرامعین عبارتند از

$$A^T A x = A^T b$$

مشکلات عددی با این معادلات نرمال نظیر همان مشکلات مسأله فرامعین می‌باشند. تحلیلهای خطای گرد کردن برای مسائل فرامعین و فرامعین با رتبه ناقص به نوعی پیچیده هستند، زیرا جوابها در این حالتها یکتا نیستند مگر شرط داشتن جواب با می‌نیم نرم تحمیل

شود. تحلیل‌های خطای پسر و روش‌های تجزیه QR با استفاده از تبدیلهای هاوس هولدر در این حالتها نشان می‌دهند که جواب با می‌نیمم نرم در هر حالت تقریباً برابر جواب با می‌نیمم نرم مسأله کمی اختلال یافته است. این با نتایج به دست آمده توسط روش تجزیه QR هاوس هولدر برای مسائل فرامعین با رتبه کامل مغایرت دارد. در آنجا نشان داده شده است که جواب محاسبه شده جواب واقعی مسأله کمی اختلال یافته است.

۴. تحلیل اختلال: نتایج تحلیل اختلال برای حالت‌های مختلف اختلالها در داده‌ها متفاوت هستند. اگر فقط b اختلال یابد، آنگاه $\|A^+\| \|A\| = \text{Cond}(A)$ به عنوان عدد وضعیت برای جواب یکتای کم‌ترین توانهای دوم به کار می‌رود (قضیه ۷-۷-۱).

اگر A اختلال یابد، آنگاه در حالت کلی حساسیت جواب یکتای کم‌ترین توانهای دوم به توان دوم عدد وضعیت بستگی دارد (قضیه ۷-۷-۲). در حالت‌های خاصی، نظیر هنگامی که مانده صفر است، حساسیت فقط به عدد وضعیت بستگی دارد.

۵. تصفیه تکراری: همانند مسأله دستگاه خطی، ممکن است دقت یک جواب کم‌ترین توانهای دوم محاسبه شده را به یک طریق تکراری بهبود بخشید. یک الگوریتم که یک مشابه طبیعی برای الگوریتم مربوط به دستگاه خطی می‌باشد (بخش ۶-۹) فقط هنگامی رضایت‌بخش است که بردار مانده $r = b - Ax$ به اندازه کافی کوچک باشد. یک الگوریتم که به طور وسیعی استفاده می‌شود و اثر بیورک می‌باشد در بخش ۷-۱۰ ارائه شد. این الگوریتم نیاز به جواب یک دستگاه افزوده با مرتبه $m + n$ دارد (در اینجا A یک ماتریس $m \times n$ است). نشان داده شد که چگونه می‌توان دستگاه را به یک طریق نسبتاً ارزان با استفاده از تجزیه QR حل کرد. جواب به دست آمده توسط این الگوریتم تصفیه تکراری کاملاً رضایت‌بخش است.

۶. محاسبه ماتریس کوواریانس: یک محاسبه کارا برای ماتریس $(A^T A)^{-1}$ در کاربردهای آماری و پردازش تصویر مهم است. در بخش ۷-۱۱ نشان دادیم که چگونه این ماتریس را می‌توان با استفاده از تجزیه QR ماتریس A و بدون پیدا کردن معکوس صریح محاسبه کرد. جدول ۷-۲ روش‌های این فصل را با هم مقایسه می‌کند.

جدول ۷-۲ مقایسه روشهای کمترین توانهای دوم متفاوت

مسئله	روش	تعداد اعمال ممیز شناور	خاصیتهای عددی
فرامعین با رتبه کامل	معادلات نرمال	$\frac{mn^2}{2} + \frac{n^3}{6}$	۱. مشکلاتی در تشکیل ATA وجود دارد. ۲. در برخی حالتها این روش خطاهایی بیشتر از آنچه که توسط داده‌ها تضمین می‌شوند، در جواب تولید می‌کند.
فرامعین با رتبه کامل	QR هاوس هولدر	$mn^2 - \frac{n^3}{3}$	پایدار؛ جواب محاسبه شده جواب واقعی یک مسئله نزدیک است.
فرامعین با رتبه کامل	$MGS - QR$	mn^2	به همان پایداری روش QR هاوس هولدر است.
فرامعین با رتبه ناقص	QR هاوس هولدر با محورگیری ستونی	$2mnr - r^2(m+n) + \frac{r^2}{3}$ $r = \text{rank}(A)$	نسبتاً پایدار؛ جواب با می‌نیم نرم محاسبه شده نزدیک به جواب یا می‌نیم نرم یک مسئله اختلال یافته است.
فرامعین با رتبه کامل	معادلات نرمال	$\frac{m^2n}{2} + \frac{m^3}{6}$	همان مشکلات حالت مسئله فرامعین وجود دارد.
فرامعین با رتبه کامل	QR هاوس هولدر	$m^2n - \frac{m^3}{3}$	این مسئله دارای همان خاصیتهای مسئله فرامعین با رتبه ناقص است.
فرامعین با رتبه کامل	گرام اشمیت اصلاح شده	m^2n	این روش به همان پایداری روش QR هاوس هولدر است.

تمرینها

بخش ۷-۳

۱. فرض کنید A یک ماتریس $n \times m$ با $m \geq n$ باشد. آنگاه با استفاده از تجزیه QR ماتریس A ثابت کنید که $A^T A$ معین مثبت است اگر و فقط اگر A دارای رتبه کامل n باشد.
۲. نشان دهید که بردار مانده $r = b - Ax$ بر همه بردارهای واقع در $R(A)$ متعامد است.
۳. ثابت کنید که x یک جواب کمترین توانهای دوم است اگر و فقط اگر

$$Ax = b_R \quad \text{و} \quad b - Ax = b_N$$

- که در آن b_R و b_N به ترتیب مؤلفه‌های برد - فضا و ستون - فضای بردار b می‌باشند.
۴. قضیه ۷-۳-۱ را در هر دو حالت رتبه کامل و رتبه ناقص، و مبتنی بر تجزیه QR ماتریس A ثابت کنید.

بخشهای ۷-۵ تا ۷-۷

۵. فرض کنید A دارای ستونهای مستقل خطی باشد. آنگاه ثابت کنید که $(\text{Cond}_p(A))^2 = \text{Cond}_p(A^T A)$.
۶. با استفاده از کمترین توانهای دوم یک خط راست و یک تابع درجه دو به داده‌های زیر برازش دهید.

x	۰	۱	۳	۵	۷	۹	۱۲
y	۱۰	۱۲	۱۸	۱۵	۲۰	۲۵	۳۶

- نتایج خود را با هم مقایسه کنید. عدد وضعیت ماتریس و اندروموند مربوطه را در هر حالت حساب کنید.
۷. معکوس تعمیم یافته و عدد وضعیت هر یک از ماتریسهای زیر را به دست آورید.

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & & & & \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ 0 & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 7 & 6/990 \\ 4 & 4 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 0 & 7 & 8 \end{bmatrix}$$

۸. فرض کنید A و \tilde{A} دارای رتبه کامل باشند. فرض کنید x و \tilde{x} به ترتیب جوابهای یکتای کم‌ترین توانهای دوم مسائل زیر باشند.

$$Ax = b$$

و

$$\tilde{A}\tilde{x} = b$$

که در آن $\tilde{A} = A + E$. آنگاه ثابت کنید که

$$\begin{aligned} \frac{\|\tilde{x} - x\|}{\|\tilde{x}\|} &\leq \text{Cond}(A) \frac{\|E\|}{\|A\|} \left(1 + \frac{\|b\|}{\|A\| \|\tilde{x}\|}\right) \\ &+ (\text{Cond}(A))^2 \frac{\|E\|}{\|A\|} \left(1 + \frac{\|E\|}{\|A\|}\right) \end{aligned}$$

(راهنمایی: تحلیل اختلال دستگاههای خطی با معادلات نرمال را به کار برید).

۹. نامساوی تمرین ۸ را با داده‌های زیر تحقیق کنید.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 7 \\ 11 \end{bmatrix}$$

$$E = 10^{-4}A$$

۱۰. نامساوی قضیه ۷-۷-۴ را در هر یک از حالت‌های زیر تحقیق کنید.

$$E = 10^{-4}A, \quad A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad (\text{الف})$$

$$E = 10^{-4}A, \quad A = \begin{bmatrix} 1 & 1 \\ 10^{-4} & \cdot \\ \cdot & 10^{-4} \end{bmatrix} \quad (\text{ب})$$

$$E = 10^{-3}A, \quad A = \begin{bmatrix} 1 & 1 \\ \cdot & 1 \\ \cdot & \cdot \end{bmatrix} \quad (\text{ج})$$

۱۱. اثباتی برای قضیه ۷-۷-۲ ارائه دهید.

بخشهای ۷-۷ و ۸-۷

۱۲. یک الگوریتم مبتنی بر محورگیری کامل ماتریس دستگاه برای حل معادلات نرمالی تدوین نمایید که ماتریس دستگاه آن نیمه معین مثبت و منفرد باشد. الگوریتم خود را در مورد معادلات نرمال با داده‌های زیر به کار برید.

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \text{ و } b = \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

۱۳. فرض کنید $A = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 0 & 0 \end{bmatrix}$ ، $b = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$

(الف) جواب یکتای کم‌ترین توانهای دوم x را با استفاده از روشهای زیر پیدا کنید.

$x = A^+b$ (i)

(ii) روش معادلات نرمال

(iii) روشهای تجزیه QR هاوس هولدر و گیونز

(iv) روشهای MGS و CGS

(ب) $Cond(A)$ را پیدا کنید.

(ج) نشان دهید که برای این مسأله، حساسیت مسأله کم‌ترین توانهای دوم، هنگامی که فقط A

اختلال می‌یابد، بستگی به $Cond(A)$ دارد.

(د) فرض کنید $E = 10^{-4}A$ و $\tilde{x} = \tilde{A}^+b$ که در آن $\tilde{A} = A + E$. مقدار $\frac{\|\tilde{x} - x\|}{\|x\|}$ را بیابید و

نامساوی قضیه ۷-۷-۲ را برای این مسأله تحقیق کنید.

(ه) $\tilde{\tau}$ و $\tilde{\epsilon}$ را پیدا کنید و نامساوی قضیه ۷-۷-۳ را تحقیق کنید.

۱۴. مثالی بسازید که در آن حساسیت مسأله کم‌ترین توانهای دوم به توان دوم عدد وضعیت ماتریس

بستگی داشته باشد (همه کار خود را نشان دهید).

۱۵. ماتریس بدوضع معروف زیر را در نظر بگیرید

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \epsilon & \cdot & \cdot \\ \cdot & \epsilon & \cdot \\ \cdot & \cdot & \epsilon \end{bmatrix}, \quad |\epsilon| \ll 1$$

(الف) یک ε کوچک به نحوی انتخاب کنید که $\text{rank}(A) = 3$ باشد. سپس $\text{Cond}_p(A)$ را برای بررسی بد وضع بودن A حساب کنید.

(ب) جواب کمترین توانهای دوم را با A داده شده به صورت فوق و

$$b = \begin{bmatrix} 3 \\ \varepsilon \\ \varepsilon \\ \varepsilon \end{bmatrix}$$

با استفاده از روشهای زیر به دست آورید.

(i) روش معادلات نرمال

(ii) روش تجزیه QR با هاوس هولدر، و روشهای MGS و CGS

(ج) b را با

$$\tilde{b} = \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

تغییر دهید. A را بدون تغییر حفظ کنید. یک کران بالا برای خطای نسبی در جواب کمترین توانهای دوم پیدا کنید.

(د) A را به $\tilde{A} = A + E$ تغییر دهید که در آن $E = 10^{-3}A$. b را بدون تغییر حفظ کنید. یک

کران بالا برای خطای نسبی در جواب کمترین توانهای دوم پیدا کنید.

(ه) حداکثر انحراف از متعامد بودن ستونهای محاسبه شده ماتریس Q توسط روشهای CGS و MGS را پیدا کنید.

(و) جواب کمترین توانهای دوم مسأله قسمت (ب) را با استفاده از SVD محاسبه کنید.

۱۶. (چولسکی مستقل از ریشه دوم) با مفروض بودن یک ماتریس معین مثبت متقارن A یک

الگوریتم برای پیدا کردن تجزیه چولسکی A بدون هیچ ریشه دوم تدوین نمایید:

$$A = LDL^T$$

که در آن L یک ماتریس پایین مثلثی واحد و D یک ماتریس قطری با عناصر قطری مثبت است. الگوریتم خود را برای حل مسأله کمترین توانهای دوم با رتبه کامل مبتنی بر حل معادلات نرمال به کار ببرید.

۱۷. (الف) یک گونه ستون گرا از فرایند گرام اشمیت اصلاح شده به دست آورید.
 (ب) نشان دهید که تعداد اعمال ممیز شناور برای هر دو روش گرام اشمیت و گرام اشمیت اصلاح شده برابر mn^2 می باشد، در اینجا A یک ماتریس $m \times n$ ، $(m \geq n)$ ، می باشد.
 (ج) نشان دهید که تعداد اعمال ممیز شناور برای روش هاوس هولدر برابر است با $n^2(m - n/3)$.
۱۸. (الف) یک مثال برای مسأله کم ترین توانهای دوم بسازید که در آن گرام اشمیت یک نتیجه ضعیف ارائه دهد و لیکن گرام اشمیت اصلاح شده نسبتاً خوب عمل کند.
 (ب) روشهای CGS و MGS را برای تجزیه QR ماتریس

$$A = \begin{bmatrix} 8 & 21 \\ 13 & 34 \\ 21 & 35 \\ 34 & 39 \end{bmatrix}$$

- با استفاده از $t = 4$ به کار ببرید. نشان دهید که در هر دو حالت فقدان متعامد بودن بین q_1 و q_2 قابل پذیرش نیست.
۱۹. نشان دهید که تعداد اعمال ممیز شناور برای تجزیه QR با محورگیری ستونی و با استفاده از روش هاوس هولدر برابر $\frac{2}{3}r^3 - r^2(m+n) + 2mnr$ می باشد که در آن $r = \text{rank}(A)$.
۲۰. نشان دهید که برای روش QR با محورگیری ستونی در حالت رتبه ناقص، جواب پایه ای نمی تواند جواب با می نیمم نرم باشد مگر این که R_{12} برابر صفر باشد.
۲۱. (الف) نشان دهید که جواب با می نیمم نرم برای $Ax = b$ با $\text{rank}(A) = r$ ، به دست آمده توسط تجزیه QR کامل با محورگیری ستونی، به صورت زیر ارائه می شود

$$x = PW \begin{bmatrix} T^{-1}c \\ \cdot \end{bmatrix}$$

- که در آن c یک بردار متشکل از r عنصر اول $Q^T b$ می باشد.
 (ب) جواب با می نیمم نرم را برای مسأله کم ترین توانهای دوم با داده های زیر پیدا کنید.

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$

بخش ۷-۹

۲۲. یک الگوریتم مبتنی بر تجزیه QR با محورگیری ستونی برای محاسبه جواب با می نیمم نرم برای دستگاه فرومعیین

$$Ax = b$$

تدوین نمایید که در آن A یک ماتریس $m \times n$ با $m < n$ است. تعداد اعمال ممیز شناور را برای این الگوریتم حساب کنید. الگوریتم خود را در مورد دستگاه زیر به کار ببرید.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 15 \end{bmatrix}$$

۲۳. یک الگوریتم، مبتنی بر تجزیه A با محورگیری کامل، برای حل دستگاه فرومعیین تدوین نمایید، که در آن A یک ماتریس $m \times n$ با $m < n$ است. الگوریتم شما باید یک جواب را محاسبه کند هنگامی که وجود دارد، یا در غیر این صورت اعلام نماید که جوابی وجود ندارد.
(الف) تعداد اعمال ممیز شناور را برای الگوریتم خود حساب کنید.
(ب) الگوریتم خود را با داده‌های مسأله ۲۲ امتحان کنید.

۲۴. با استفاده از قضیه ۷-۹-۱ ثابت کنید که جواب با می نیمم نرم برای یک دستگاه فرومعیین می تواند با تصویر کردن هر جواب دستگاه بر روی $R(A^+)$ ، به دست آید. یعنی، اگر P_A^+ تصویر متعامد بر روی $R(A^+)$ باشد، آنگاه جواب با می نیمم نرم x توسط

$$x = P_A^+ y$$

ارائه می شود که در آن y یک جواب دلخواه است. با استفاده از فرمول بالا جواب با می نیمم نرم را برای دستگاه زیر به دست آورید.

$$\begin{bmatrix} 1 & 10^{-4} & 0 & 0 \\ 1 & 0 & 10^{-4} & 0 \\ 1 & 0 & 0 & 10^{-4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

۲۵. یک تعبیر هندسی برای قضیه ۷-۹-۱ ارائه دهید.

بخش ۷-۱۰

۲۶. (الف) سه تکرار از الگوریتم ۷-۱۰-۱ را برای هر یک از مسائل زیر به کار ببرید.

$$b = \begin{bmatrix} 3 \\ 5 \\ 6 \end{bmatrix}, A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad (i)$$

$$b = \begin{bmatrix} 2 \\ 10^{-4} \\ 10^{-4} \end{bmatrix}, A = \begin{bmatrix} 1 & 1 \\ 10^{-4} & 0 \\ 0 & 10^{-4} \end{bmatrix} \quad (ii)$$

$$b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, A = \begin{bmatrix} 1 & 1 \\ 10^{-4} & 0 \\ 0 & 10^{-4} \end{bmatrix} \quad (iii)$$

(ب) چه رابطه‌ای بین الگوریتم ۷-۱۰-۱ و الگوریتم ۷-۱۰-۲ وجود دارد؟

۲۷. الگوریتم ۷-۱۰-۲ را با و بدون حل صریح دستگاه افزوده در گام ۲ در مورد هر قسمت تمرین ۲۶ به کار ببرید و نتایج را مقایسه کنید.

۲۸. الگوریتم ۷-۱۰-۲ را برای مسأله کم‌ترین توانهای دوم با ماتریس هیلبرت 7×7 و $b = (2/5929, 1/7179, 1/3290, 1/0956, 0/9365, 0/8199, 0/7301)^T$ به کار ببرید. بیان نمایید در هر تکرار چند رقم دقت در جواب و مانده به دست آمده است. توجه کنید که $x = (1, 1, \dots, 1)^T$.

بخش ۷-۱۱

۲۹. در کاربردهای زیادی، فقط عناصر قطری ماتریس وارپانس - کوواریانس $X = (A^T A)^{-1}$ مورد نیاز هستند. نشان دهید که این عناصر قطری دقیقاً برابر نرم دو سطرهای R^{-1} هستند.

برنامه‌های MATLAB و مسائل

تذکر: شما به برنامه‌های `lsfrnme`، `lsfrngs`، `lsfrqrh`، `mdgrsch`، `elgrsch`، `givqr`، `housqr`، `lsitrn2`، `lsitrn1`، `mnudqrh`، `mnudnme` و `varcovar` از MATCOM نیاز خواهید داشت.

۱. مجموعه نقاط داده‌های زیر را در نظر بگیرید.

x	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
y	۱	۲/۹	۶/۸	۱۲	۲۰/۵	۳۰/۹	۴۲/۹	۵۱/۵	۷۳	۹۰/۵

با استفاده از دستور **vander** از **MATLAB** و عمل \setminus ، برازش کمترین توانهای دوم این داده‌ها را به چند جمله‌ای از درجه ۱ تا ۴ حساب کنید. نمودار نقاط داده‌های اصلی و برازشهای کمترین توانهای دوم را با استفاده از دستورهایی **plot** و **polyval** از **MATLAB** رسم کنید و نتایج را مقایسه کنید.

۲. فرض کنید

$$A = \begin{bmatrix} 1 & 1 \\ 10^{-3} & 0 \\ 0 & 10^{-3} \\ 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$E_A = \begin{bmatrix} 0 & -0.0001 \\ 0 & 0.0009 \\ 0 & 0.0003 \\ 0 & 0.0001 \end{bmatrix}, \quad \delta b = \begin{bmatrix} 0.0001 \\ 0.0001 \\ 0.0001 \\ 0.0001 \end{bmatrix}$$

با استفاده از دستورهایی **MATLAB** نظیر **pinv**، **Cond**، **norm**، **orth**، و **null**، نامساویهای قضیه‌های ۱-۷-۷ تا ۴-۷-۷ را در مورد حساسیتهای متفاوت مسائل کمترین توانهای دوم تحقیق کنید.

۳. (پیاپی‌سازی الگوریتم **QR** کمترین توانهای دوم با استفاده از دورانهای گیونز) با استفاده از دستورهایی **givqr** و **backsub** یک برنامه **MATLAB** بنویسید تا الگوریتم **QR** را با استفاده از دورانهای گیونز برای مسائل کمترین توانهای دوم فرامین با رتبه کامل پیاده‌سازی کند:

$$[x] = \text{lsfrqr}(a, b)$$

برنامه خودتان را با $A = \text{hilb}(7)$ امتحان کنید و یک بردار b طوری تولید کنید که جواب کمترین توانهای دوم $x = (1, 1, \dots, 1)^T$ باشد.

داده‌های آزمون برای مسائل ۴ و ۵

برای مسائل ۴ و ۵ از مجموعه داده‌های آزمون زیر استفاده کنید.

۱. یک ماتریس به طور تصادفی تولید شده از مرتبه ۱۰

۲. ماتریس هیلبرت از مرتبه ۱۰

$$3. \begin{bmatrix} 1 & 1 & 1 \\ 10^{-3} & \cdot & \cdot \\ \cdot & 10^{-3} & \cdot \\ \cdot & \cdot & 10^{-3} \end{bmatrix}$$

برای مسأله ۵، b را طوری تولید کنید که جواب کم‌ترین توانهای دوم x در هر حالت دارای همه عناصر برابر ۱ باشد. مثلاً برای ماتریس داده‌ها در داده‌های آزمون ۳ بردار b را به شکل زیر اختیار کنید.

$$b = \begin{bmatrix} 3 \\ 10^{-3} \\ 10^{-3} \\ 10^{-3} \end{bmatrix}$$

۴. (هدف از این تمرین مقایسه دقت، تعداد اعمال ممیز شناور، و زمان برای روشهای مختلف برای تجزیه QR یک ماتریس می‌باشد.)

(الف) تجزیه QR را برای هر ماتریس از مجموعه داده‌ها و با استفاده از روشهای زیر حساب کنید.

(i) $[Q, R] = qr(A)$ از MATLAB یا $[Q, R] = \text{housqr}(A)$ از فصل ۵

(QR هاوس هولدر)

(ii) $[Q, R] = \text{givqr}(A)$ ، (QR گیونز)

(iii) $[Q, R] = \text{clgrsch}(A)$ از MATCOM (گرام اشمیت کلاسیک)

(iv) $[Q, R] = \text{mdgrsch}(A)$ از MATCOM (گرام اشمیت اصلاح شده)

(ب) با استفاده از نتایج (الف) جدولی نظیر جدول ۷-۳ برای هر ماتریس A بسازید. \hat{Q} و \hat{R} را

برای نمایش Q و R محاسبه شده به کار ببرید.

(ج) نتایج خود را بنویسید.

۵. (هدف از این تمرین مقایسه دقت، تعداد اعمال ممیز شناور، و زمان برای روشهای مختلف کم‌ترین

توانهای دوم برای مسائل فرامعین با رتبه کامل است).

(الف) جواب کمترین توانهای دوم \hat{x} را برای هر یک از مجموعه داده‌ها و با استفاده از روشهای زیر حساب کنید.

(i) $[x] = \text{lsfrnms}(A, b)$ ، (کمترین توانهای دوم با استفاده از گرام اشمیت کلاسیک)

(ii) $[x] = \text{lsfrqrh}(A, b)$ ، (کمترین توانهای دوم با استفاده از QR هاوس هولدر)

(iii) $[x] = \text{lsfrqrg}(A, b)$ ، (کمترین توانهای دوم با استفاده از QR گیونز)

(iv) $[x] = \text{lsfrmne}(A, b)$ ، (کمترین توانهای دوم با استفاده از معادلات نرمال)

(v) $[x] = \text{pinv}(A) * b$ ، (کمترین توانهای دوم با استفاده از معکوس تعمیم یافته)

تذکر: lsfrnms ، lsfrqrh ، و lsfrmne همگی در MATCOM موجود هستند. pinv یک دستور MATLAB برای محاسبه معکوس تعمیم یافته یک ماتریس است.

(ب) با استفاده از نتایج (الف) جدولی نظیر جدول ۷-۴ برای هر یک از مجموعه داده‌ها بسازید. (\hat{x} را برای نمایش x محاسبه شده به کار ببرید). توجه کنید که بردار x دارای همه عناصر برابر ۱ است.

(ج) نتایج خود را بنویسید.

جدول ۷-۳ مقایسه روشهای مختلف برای تجزیه QR

روش	$\ (\hat{Q})^T \hat{Q} - I\ _2$	$\ A - \hat{Q}\hat{R}\ _2 / \ A\ _2$	تعداد اعمال ممیز شناور	زمان سپری شده
housqr				
givqr				
clgrsch				
mdgrsch				

جدول ۴-۷ مقایسه روشهای مختلف برای مسأله کمترین توانهای دوم فرامعین با رتبه کامل

روش	$\ x - \hat{x}\ _2 / \ x\ _2$	$\ A\hat{x} - b\ _2$	تعداد اعمال ممیزشناور	زمان سپری شده
Isfrngs				
Isfrqrh				
Isfrqrg				
Isfrnme				
Generalized inverse				

۶. با استفاده از `housqr` از MATCOM یا $[Q, R] = qr(A)$ از MATLAB و `backsub` از فصل ۶، یک برنامه MATLAB به نام `lsrdqrh(A, b)` برای محاسبه جواب کمترین توانهای دوم با می نیم نرم x برای مسأله فرامعین با رتبه ناقص $Ax = b$ و بردار مانده متناظر r ، و مبتنی بر تجزیه QR هاوس هولدر ماتریس A بنویسید:

$$[x, r] = \text{lsrdqrh}(A, b)$$

(این برنامه الگوریتم ۷-۸-۶ از کتاب را باید پیاده سازی کند)

داده های آزمون: A یک ماتریس 20×2 با همه عناصر برابر ۱ است و b یک بردار با همه عناصر برابر ۲ می باشد.

۷. از تابع MATLAB

$$[U, S, V] = \text{svd}(A)$$

برای محاسبه تجزیه مقدار تکین A استفاده کنید، یک برنامه MATLAB به نام `lsrdsvd` برای جواب کمترین توانهای دوم با می نیم نرم x برای دستگاه فرامعین با رتبه ناقص $Ax = b$ بنویسید.

$$[x] = \text{lsrdsvd}(A, b)$$

از همان داده های آزمون مسأله ۶ استفاده کنید و نتایج را از نظر دقت، تعداد اعمال ممیز شناور و زمان سپری شده مقایسه کنید.

۸. برنامه‌های **mnudnme** (جواب با می نیمم نرم برای مسأله فرومعین با رتبه کامل با استفاده از معادلات نرمال) و **mnudqrh** (جواب با می نیمم نرم برای مسأله فرومعین با رتبه کامل با استفاده از تجزیه QR هاوس هولدر) را از MATLAB با مجموعه داده‌های زیر به منظور محاسبه جواب با می نیمم نرم x برای مسأله فرومعین با رتبه کامل $Ax = b$ اجرا کنید و نتایج را از نظر دقت، تعداد اعمال ممیز شناور و زمان سپری شده مقایسه کنید.

$$A = (1 \ 2 \ 3 \ 4 \ 5 \ 6)$$

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{bmatrix}$$

$$A = \begin{bmatrix} 10 & 10 & 10 & 10 & 10 & 10 & 10 & 10 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

b را برای هر A طوری بسازید که جواب با می نیمم نرم x دارای همه عناصر برابر ۱ باشد.

۹. برنامه‌های **lsitrn1** (مبتنی بر الگوریتم ۷-۱۰-۱) و **lsitrn2** (مبتنی بر الگوریتم ۷-۱۰-۲) را از MATCOM با ماتریس هیلبرت 8×8 اجرا کنید، b را به طور تصادفی بسازید. الگوریتمها را بر روی این داده‌ها از نظر تعداد تکرارها، تعداد اعمال ممیز شناور، هم برای جواب و هم برای مانده مقایسه کنید.

۱۰. الف) $(A^T A)^{-1}$ را برای هر یک از ماتریسهای زیر و به روشهای زیر حساب کنید.

(i) $(A^T A)^{-1}$ را با استفاده از MATLAB و دستور **inv** به طور صریح محاسبه کنید.

(ii) برنامه **varcovar** از MATCOM را اجرا کنید.

ب) نتایج را از نظر دقت، تعداد اعمال ممیز شناور و زمان سپری شده مقایسه کنید

داده‌های آزمون

ماتریس هیلبرت 8×8

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 10^{-3} & \cdot & \cdot \\ \cdot & 10^{-3} & \cdot \\ \cdot & \cdot & 10^{-3} \end{bmatrix}$$

ماتریس به طور تصادفی تولید شده 8×8

پیشنهادها برای مطالعه بیشتر

روشهای جوابهای کمترین توانهای دوم در کتابهای درسی جبر خطی عددی و برخی از کتابهای درسی آنالیز عددی ارائه می‌شوند. در بیشتر کتابها تأکید بر مسائل فرامعین است. برای بررسی دقیقی از این موضوع ما خواننده را به کتاب Golub and Van Loan (1989) ارجاع می‌دهیم. کتابهای Stewart (1973)، Watkins (1991) و Gil, Murray, and Wright (1991) نیز حاوی بحثهای مفصل در مورد تحلیلهای اختلال هستند. همچنین مقاله Stewart (1977a) و کتاب Stewart and Sun (1990) را ملاحظه کنید. برای تحلیل اختلال مؤلفه‌ای مرجع Björck (1994b) و Chandrasekaran and Ipsen (1994b) را ملاحظه کنید. برای تحلیل اختلال مسائل فرومعین مرجع Golub and Van Loan (1989, 258-259) را ملاحظه کنید. برای کاربردهای حذفی گاوس در جواب کمترین توانهای دوم مرجع Plemmons (1974) را ملاحظه کنید.

برای کاربردهای مسائل کمترین توانهای دوم برای جوابهای عددی معادلات دیفرانسیل مرجع Compbell (1985) را ملاحظه کنید.

برای مطالعه در مورد مسائل فرومعین مقاله‌های Cline and Plemmons (1976) و Arioli and Laratta (1985) را ملاحظه کنید.

یک کتاب کامل اختصاص یافته به این موضوع کتاب (تقریباً کلاسیک) Lawson and Hanson (1974) است. یک رساله جدید زیبا، ارائه شده توسط بیورک (Björck (1994b))، وجود دارد. این دو کتاب اطلاعات اساسی برای آنهايي هستند که علاقه‌مند به اطلاعات بیشتر درباره این موضوع می‌باشند. بویژه کتاب نوشته شده توسط لاوسون و هانسون اثباتها و تحلیلهای خطای گرد کردن الگوریتمهای مختلف توصیف شده در این کتاب را ارائه می‌دهد.

هر کتاب در مورد تحلیل رگرسیون در آمار شامل کاربردهای مسائل کمترین توانهای دوم در آمار می‌باشد. بویژه ما از کتاب

Applied Linear Regression Models, by John Neter, William Wasserman, and Michael Kunter (1983).

استفاده کرده‌ایم. یک مقاله مروری کلاسیک از گولوب (Golub (1969)) حاوی شرح عالی از روشهای جبرخطی عددی برای مسائل کمترین توانهای دوم و تجزیه مقدار تکین مطرح در آمار و سایر جاهای دیگر می‌باشد. مقاله Stewart (1987) نیز برای خواندن جالب است. کتاب Thisted (1988) و مقاله Efron (1988) را برای مفاهیم محاسباتی رگرسیون آماری ملاحظه کنید.

چندین جنبه از مسائل کمترین توانهای دوم در این کتاب ارائه نگردیده‌اند، آنها شامل مسائل کمترین توانهای دوم تام، مسائل کمترین توانهای دوم مقید، مسائل کمترین توانهای دوم تنک هستند. مرجع Björck (1994b) را برای این زمینه‌ها ملاحظه کنید. کتاب:

The Total Least Squares Problems: Computational Aspects and Analysis by S. Van Huffel and J. Vandewalle, *Frontier in Applied Mathematics* series SIAM, Philadelphia, 1991.

برای مطالعه بیشتر در مورد مسائل کمترین توانهای دوم تام نیز پیشنهاد می‌شود. بعلاوه مقاله جدید Fierro and Bunch (1994) را در مورد هم راستایی و کمترین توانهای دوم تام ملاحظه کنید.

مسائل عددی مقادیر ویژه یک ماتریس

۸-۱ مقدمه

این فصل به مطالعه مسائل عددی مقادیر ویژه یک ماتریس اختصاص داده شده است. این مسائل از مسائل بسیار مهم عملی هستند و در کاربردهای متنوعی، شامل مهندسی، آمار و اقتصاد مطرح می‌شوند.

چون مقادیر ویژه یک ماتریس A صفرهای معادله مشخصه $\det(A - \lambda I)$ هستند، به طور طبیعی می‌توان به محاسبه مقادیر ویژه A توسط پیدا کردن چندجمله‌ای مشخصه آن و سپس محاسبه صفرهای آن توسط یک روش استاندارد اندیشید. متأسفانه این یک راه عملی نیست.

یک الگوریتم استاندارد برای پیدا کردن مقادیر ویژه یک ماتریس روش تکراری QR با انتقال ساده و یا مضاعف است. کاربردهای زیادی نیاز به شناخت تمام طیف ندارند. تعداد کمی مقدار ویژه انتخابی معمولاً تعداد کمی از بزرگ‌ترین و یا کوچک‌ترین مقادیر ویژه کافی خواهند بود. یک روش کلاسیک مبتنی بر به توان رسانیدن ضمنی A ، معروف به روش توانی، برای این منظور مفید است.

مسأله مقدار ویژه متقارن از برخی خاصیت‌های خاص و جالب توجه برخوردار است، یکی از این خاصیتها این است که مقادیر ویژه یک ماتریس متقارن خوش‌وضع هستند. یک روش مبتنی بر بهره‌گیری از برخی خواص خاص یک مسأله مقدار ویژه متقارن، معروف به روش تنصیف دنسباله

استورم، برای پیدا کردن مقادیر ویژه یک ماتریس متقارن در یک بازه مفید است.

متأسفانه روش تکراری QR نمی تواند برای محاسبه مقادیر ویژه و بردارهای ویژه یک ماتریس خیلی بزرگ تنک استفاده شود. تنک بودن از دست می رود و اشغال حافظه گسترش می یابد. یک روش تقریباً کلاسیک، ارائه شده توسط لانزوس، اخیراً توسط جبر خطی عددی دانان از سر گرفته شده است. بویژه روش لانزوس متقارن برای استخراج تعداد کمی مقادیر ویژه اکسترمال و بردارهای ویژه وابسته از یک ماتریس خیلی بزرگ و تنک مفید است.

تنظیم این فصل به صورت زیر است. در بخش ۸-۲ برخی از نتایج نظری اساسی در مورد مقادیر ویژه و بردارهای ویژه بیان می شوند.

بخش ۸-۳ به بحث در مورد این که چگونه مسأله مقدار ویژه در برخی کاربردهای عملی مطرح می شود اختصاص یافته است. این گونه کاربردهای عملی عبارتند از: تحلیل‌های پایداری یک دستگاه معادلات دیفرانسیل و تفاضلی، تحلیل ارتعاش، رفتار عبوری یک مدار الکتریکی، مسائل کمانش، تحلیل مؤلفه‌های اصلی در آمار، و غیره.

در بخش ۸-۴ برخی نتایج کلاسیک در مورد مکان مقادیر ویژه، نظیر قضیه‌های قرص گرشگورین بیان و اثبات می شوند. بخش ۸-۵ روش توانی، روش توانی معکوس، تکرار خارج قسمت ریلی و غیره را برای پیدا کردن یک تعداد مقدار ویژه انتخابی و بردارهای ویژه متناظر توصیف می کند.

در بخش ۸-۶ مشکلات محاسبه مقادیر ویژه یک ماتریس از طریق چند جمله‌ای مشخصه و شکل متعارف جردن مشخص خواهند گردید.

حساسیت مقدار ویژه و بردار ویژه در بخش ۸-۷ و ۸-۸ بحث خواهد شد. مهمترین نتیجه در این بخش قضیه بوئر-فایک^۱ است. بخش ۸-۹ مهمترین بخش در این فصل است. روش تکراری QR با و بدون انتقال و پیاده‌سازی آنها در این بخش توصیف می شوند.

تکرار معکوس هسنبرگی و محاسبه بردارهای ویژه از روی شکل شور حقیقی^۲ در بخش ۸-۱۰ توصیف می شوند. روشهای مبتنی بر خاصیت‌های خاص ماتریسهای متقارن در بخش ۸-۱۱ بحث می شوند. اینها شامل روش تنصیف دنباله استورم و روش تکرار QR متقارن با انتقال ویلکینسون و روش تقسیم و تسخیر می شوند.

این فصل را با یک مقدمه مختصر در مورد الگوریتم لانزوس متقارن برای پیدا کردن مقادیر ویژه اکسترمال یک ماتریس متقارن بزرگ تنک به پایان می رسانیم.

ابزار اصلی زیر و مفاهیم قبلاً ارائه شده برای یادگیری و فهم راحت مطالب این بخش لازم خواهند بود: خاصیت‌های نرم ماتریسها (بخش ۱-۷)، تجزیه QR یک ماتریس دلخواه و یک ماتریس هسبرگی، استفاده از هر دو ماتریس هاوس هولدر و گیونز (بخش ۵-۴-۱ و ۵-۵-۱)، جوابیهای دستگاههای خطی (بخش ۶-۴)، عدد وضعیت و خاصیت‌های آن (بخشهای ۶-۶ و ۶-۷)، و تقلیل به شکلهای هسبرگی و سه‌قطری (بخشهای ۵-۴-۳ و ۵-۵-۳).

۸-۲ برخی نتایج اساسی در مورد مقادیر ویژه و بردارهای ویژه

تا کنون ما با بردارها و ماتریسهای حقیقی سر و کار داشته‌ایم. در بیشتر کاربردها داده‌ها حقیقی هستند. ولیکن یک ماتریس حقیقی می‌تواند مقادیر ویژه و بردارهای ویژه مختلط داشته باشد. در این فصل و دو فصل بعد با مقادیر ویژه و بردارهای ویژه سر و کار خواهیم داشت. بنابراین ما به سرعت مفاهیم بردارهای مختلط و فضای مختلط و برخی تعریفهای وابسته را ارائه خواهیم داد. یک بردار n بعدی مختلط یک آرایش مرتب از n عدد مختلط در یک ستون می‌باشد:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

مجموعه همه بردارهای n بعدی توسط \mathcal{C}^n نمایش داده می‌شود. بیشتر مفاهیم تعریف شده برای \mathbb{R}^n در فصل ۱ خیلی ساده به \mathcal{C}^n تعمیم می‌یابند.

یک ماتریس مختلط $m \times n$ یک آرایه مستطیلی با m سطر و n ستون است. مجموعه همه ماتریسهای $m \times n$ توسط $\mathcal{C}^{m \times n}$ نمایش داده می‌شود. مزدوج مختلط یک ماتریس مختلط توسط \bar{A} نمایش داده می‌شود و ماتریسی است که هر عنصر آن مزدوج مختلط عنصر متناظر از A می‌باشد. ترانهاده مزدوج مختلط A توسط A^* نمایش داده می‌شود؛ یعنی

$$A^* = (\bar{A})^T$$

\bar{x} و x^* یک بردار مختلط x به طور مشابه تعریف می‌شوند. یک ماتریس مختلط مربعی A هرمیتی است اگر و فقط اگر $A^* = A$. یک ماتریس مختلط مربعی A یکانی است اگر $A^*A = AA^* = I$. یک ماتریس هرمیتی A معین مثبت است اگر $x^*Ax > 0$ به ازای همه بردارهای مخالف صفر x . نیمه معین مثبت است اگر $x^*Ax \geq 0$ به ازای همه بردارهای $x \neq 0$.

با این تعریفهای اساسی به بیان (و گاهی اوقات اثبات) برخی نتایج اساسی در مورد مقادیر ویژه و

بردارهای ویژه می‌پردازیم. برخی از این نتایج و تعریفها در فصل ۱ هستند. برخی را در اینجا برای راحتی و ارجاع فوری تکرار خواهیم کرد. با به خاطر داشتن کاربردها، روشها را فقط برای محاسبه مقادیر ویژه و بردارهای ویژه ماتریسهای حقیقی بحث خواهیم کرد.

۱-۲-۸ مقادیر ویژه و بردارهای ویژه

فرض کنید A یک ماتریس $n \times n$ باشد. آنگاه λ یک مقدار ویژه A است اگر یک بردار مخالف صفر x یافت شود به قسمی که

$$Ax = \lambda x$$

یا

$$(A - \lambda I)x = 0$$

بردار x یک بردار ویژه راست (یا فقط یک بردار ویژه) A وابسته به مقدار ویژه λ می‌باشد. (λ, x) یک جفت ویژه A نامیده خواهند شد. بردار y ارائه شده توسط

$$y^T A = \lambda y^T$$

یک بردار ویژه چپ ماتریس A وابسته به مقدار ویژه λ نامیده می‌شود. به طور معمول یک بردار ویژه راست فقط یک بردار ویژه نامیده می‌شود.

دستگاه همگن $(A - \lambda I)x = 0$ دارای یک جواب غیر بدیهی است اگر و فقط اگر

$$\det(A - \lambda I) = 0$$

مشخصه ماتریس A نامیده می‌شود. بنابراین n مقدار ویژه A ، n ریشه چندجمله‌ای مشخصه آن هستند. $P_A(\lambda) = \det(A - \lambda I)$ یک چندجمله‌ای برحسب λ و از درجه n است و چند جمله‌ای

مجموع مقادیر ویژه یک ماتریس A اثر ماتریس A نامیده می‌شود. و توسط $\text{trace}(A)$ یا $\text{Tr}(A)$ نمایش داده می‌شود.

همچنین از فصل ۱ یادآور می‌شویم که اگر $A = (a_{ij})$ ، آنگاه اثر ماتریس A مجموع عناصر قطری A می‌باشد، یعنی $\text{trace}(A) = a_{11} + \dots + a_{nn}$.

معادل بودن این دو تعریف از ملاحظه ضرب λ در چندجمله‌ای مشخصه $\det(A - \lambda I)$ نتیجه می‌شود.

مشاهدات

از تعریفهای مقادیر ویژه و بردارهای ویژه فوراً مطالب زیر را ملاحظه می‌کنیم:

۱. اگر (λ, x) یک جفت ویژه از یک ماتریس نامنفرد A باشد، آنگاه $(\frac{1}{\lambda}, x)$ جفت ویژه متناظر از A^{-1} می‌باشد.

۲. اگر (λ, x) یک جفت ویژه از یک ماتریس A باشد، آنگاه $(\lambda - \sigma, x)$ یک جفت ویژه از $A - \sigma I$ می‌باشد، که در آن σ یک اسکالر است.

تعریف ۸-۲-۱ مجموعه مقادیر ویژه یک ماتریس A طیف A نامیده می‌شود. اکنون برخی نتایج اساسی را در مورد مقادیر ویژه و بردارهای ویژه ارائه می‌دهیم.

قضیه ۸-۲-۱ مقادیر ویژه یک ماتریس مثلثی عناصر قطری آن هستند.

اثبات: اگر $A = (a_{ij})$ مثلثی باشد، آنگاه ماتریس $(A - \lambda I)$ نیز مثلثی است و $a_{11} - \lambda, a_{22} - \lambda, \dots, a_{nn} - \lambda$ عناصر قطری آن می‌باشند. بنابراین

$$\det(A - \lambda I) = \prod_{i=1}^n (a_{ii} - \lambda)$$

این نشان می‌دهد که مقادیر ویژه A عبارتند از $a_{11}, a_{22}, \dots, a_{nn}$. قضیه ۸-۲-۱ به سادگی تعمیم می‌یابد.

قضیه ۸-۲-۲ فرض کنید

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1k} \\ & A_{22} & \dots & A_{2k} \\ & & \ddots & \vdots \\ & & & A_{kk} \end{bmatrix}$$

که در آن A_{ii} یک ماتریس مربعی است. آنگاه طیف A برابر اجتماع طیفهای A_{11}, \dots, A_{kk} می‌باشد.

اثبات: داریم

$$\det(A - \lambda I) = \det(A_{11} - \lambda I) \det(A_{22} - \lambda I) \dots \det(A_{kk} - \lambda I)$$

بنابراین مجموعه صفرهای چندجمله‌ای مشخصه A برابر اجتماع مجموعه‌های صفرهای چندجمله‌ایهای مشخصه A_{11}, \dots, A_{kk} می‌باشد.

قضیه ۸-۲-۳ دو ماتریس متشابه دارای مقادیر ویژه یکسان هستند. یعنی مقادیر ویژه A و TAT^{-1} برای همه ماتریسهای متفرد T یکسان هستند.

اثبات: نشان می‌دهیم که A و TAT^{-1} دارای چندجمله‌ایهای مشخصه یکسان هستند و در نتیجه دارای مقادیر ویژه یکسان هستند.

$$\begin{aligned} \det(TAT^{-1} - \lambda I) &= \det(T(A - \lambda I)T^{-1}) \\ &= \det T \cdot \det(A - \lambda I) \cdot \det(T^{-1}) \\ &= \det T \cdot \det(T^{-1}) \cdot \det(A - \lambda I) \\ &= \det(T \cdot T^{-1}) \cdot \det(A - \lambda I) \\ &= \det(A - \lambda I) \end{aligned}$$

بنابراین TAT^{-1} و A دارای چندجمله‌ایهای مشخصه یکسان هستند و در نتیجه دارای مقادیر ویژه یکسان می‌باشند. \square

تذکر: عکس آن درست نیست. یعنی دو ماتریس دارای مقادیر ویژه یکسان لزوماً متشابه نیستند. در اینجا یک مثال ارائه می‌دهیم. دو ماتریس زیر را در نظر بگیرید

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = I_{2 \times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

A و B دارای مقادیر ویژه یکسان هستند، اما آنها نمی‌توانند متشابه باشند، زیرا $TBT^{-1} = B$ به ازای همه ماتریسهای نامنفرد T .

قضیه ۸-۲-۴ بردارهای ویژه وابسته به مقادیر ویژه متمایز یک ماتریس مستقل خطی هستند.

اثبات: قضیه را با استقرا بر روی m ثابت می‌کنیم. فرض کنید $\lambda_1, \lambda_2, \dots, \lambda_m, \lambda_m$ مقدار ویژه متمایز از A باشند و $x_1, x_2, \dots, x_m, x_m$ بردارهای ویژه متناظر باشند. فرض کنید که قضیه برای $m-1$ برقرار باشد. سپس ثابت می‌کنیم که برای m نیز برقرار است. رابطه

$$c_1 x_1 + c_2 x_2 + \dots + c_m x_m = 0$$

را در نظر بگیرید. که در آن c_i ها، $i = 1, \dots, m$ ثابت هستند. اکنون نشان می‌دهیم که

$$c_1 = c_2 = \dots = c_m = 0.$$

از ضرب رابطه فوق از طرف چپ در $(\lambda_1 I - A) \dots (\lambda_{m-1} I - A)$ ، به دست می‌آوریم

$$(\lambda_1 I - A) \dots (\lambda_{m-1} I - A)(c_1 x_1 + \dots + c_m x_m) = 0. \quad (1-2-8)$$

چون ماتریسهای $(\lambda_i I - A)$ و $(\lambda_j I - A)$ را می‌توان با یکدیگر جابه‌جا کرد. برای $i = 1, \dots, m-1$ داریم

$$(\lambda_1 I - A) \dots (\lambda_{m-1} I - A)c_i x_i = c_i \prod_{\substack{j=1 \\ j \neq i}}^{m-1} (\lambda_j I - A)(\lambda_i I - A)x_i = 0.$$

(زیرا $\lambda_i x_i = Ax_i$)

بنابراین (۱-۲-۸) به صورت زیر در می‌آید

$$(\lambda_1 I - A)(\lambda_2 I - A) \dots (\lambda_{m-1} I - A)c_m x_m = 0.$$

یا

$$c_m(\lambda_1 - \lambda_m)(\lambda_2 - \lambda_m) \dots (\lambda_{m-1} - \lambda_m)x_m = 0.$$

یا $c_m = 0$ ، زیرا $\lambda_i - \lambda_m \neq 0$ ، $i = 1, \dots, m-1$. بر طبق فرض استقرای، رابطه

$$\square \quad c_1 = c_2 = \dots = c_{m-1} = 0. \quad c_1 x_1 + \dots + c_{m-1} x_{m-1} = 0.$$

قضیه ۵-۲-۸ یک ماتریس دلخواه A با یک ماتریس قطری D متشابه است اگر و فقط اگر A دارای یک مجموعه کامل از n بردار ویژه مستقل خطی باشد.

اثبات: فرض کنید x_1, \dots, x_n بردارهای ویژه متناظر با مقادیر ویژه $\lambda_1, \dots, \lambda_n$ باشند. ماتریس

$$X = (x_1, \dots, x_n)$$

را تعریف می‌کنیم. آنگاه

$$AX = A(x_1, \dots, x_n) = (\lambda_1 x_1, \dots, \lambda_n x_n)$$

$$= (x_1, \dots, x_n) \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = XD$$

که در آن $D = \text{diag}(\lambda_1, \dots, \lambda_n)$. چون x_1, \dots, x_n مستقل خطی هستند، ماتریس X نامنفرد است بنابراین

$$X^{-1}AX = D$$

□ عکس قضیه را می توان با عکس کردن استدلالها ثابت کرد.
از قضایای ۸-۲-۴ و ۸-۲-۵ فوراً نتیجه زیر را داریم.

نتیجه ۸-۲-۱ اگر همه مقادیر ویژه A متمایز باشند، آنگاه A با یک ماتریس قطری متشابه است. توجه کنید که عکس این نتیجه برقرار نیست. یک ماتریس واحد یک ماتریس قطری است اما مقادیر ویژه آن متمایز نیستند.

تبصره: یادآوری می کنیم که یک ماتریس $n \times n$ مانند A غیر ناقص نامیده می شود. اگر دارای n بردار ویژه مستقل خطی باشد. بنابراین قضیه ۸-۲-۵ می تواند دوباره به صورت زیر بیان شود. یک ماتریس A غیر ناقص است اگر و فقط اگر با یک ماتریس قطری متشابه باشد.

۸-۲-۲ قضیه مثلثی سازی شور^۱ و کاربردهای آن

قضیه ۸-۲-۶ برای هر ماتریس مختلط یک ماتریس یکانی وجود دارد به قسمی که ماتریس

$$U^*AU = T$$

یک ماتریس بالامثلثی باشد. مقادیر ویژه A عناصر قطری T هستند.

این قضیه معنای نظری مهمی دارد. همان گونه که خواهیم دید چندین نتیجه مهم در مورد مقادیر ویژه و بردارهای ویژه می تواند از این قضیه به دست آید. اثباتی از این قضیه را بعداً در این فصل ارائه خواهیم کرد (قضیه ۸-۹-۱).

قضیه ۸-۲-۷ فرض کنید $\lambda_1, \dots, \lambda_n$ مقادیر ویژه یک ماتریس A باشند. آنگاه مقادیر ویژه ماتریس A^m عبارتند از $\lambda_1^m, \lambda_2^m, \dots, \lambda_n^m$. به طور کلی، اگر $p(x)$ یک چند جمله ای از درجه n باشد

$$p(x) = c_n x^n + c_{n-1} x^{n-1} + \dots + c_0$$

آنگاه مقادیر ویژه $p(A)$ عبارتند از $p(\lambda_1), p(\lambda_2), \dots, p(\lambda_n)$.

اثبات: ما فقط قسمت اول را ثابت می‌کنیم. قسمت دوم به عنوان یک تمرین باقی گذارده می‌شود. از $U^*AU = T$ داریم

$$T^m = (U^*AU)(U^*AU) \dots (U^*AU) = U^*A(U^*U)A(U^*U) \dots (U^*U)AU$$

(m بار)

$$= U^*A^mU$$

اما ماتریس T^m نیز یک ماتریس مثلثی است، و $\lambda_1^m, \lambda_2^m, \dots, \lambda_n^m$ عناصر قطری آن هستند. چون A^m و T^m متشابه هستند، آنها دارای مقادیر ویژه یکسان می‌باشند. \square

قضیه ۸-۲-۸ فرض کنید A یک ماتریس هرمیتی باشد. آنگاه

۱. یک ماتریس یکانی وجود دارد به قسمی که

$$U^*AU = D$$

یک ماتریس قطری باشد.

۲. مقادیر ویژه A حقیقی هستند.

۳. بردارهای ویژه A می‌توانند طوری انتخاب شوند که یکا متعامد باشند.

اثبات ۱: بر طبق قضیه مثلثی سازی شور داریم

$$U^*AU = T$$

که در آن T یک ماتریس مثلثی است. اکنون نشان می‌دهیم که T قطری است. داریم

$$(U^*AU)^* = T^*$$

یا

$$U^*A^*U = T^*$$

یا

$$U^*AU = T^* \quad (A^* = A \text{ چون})$$

یا

$$T = T^*$$

این نشان می‌دهد که T هرمیتی است. چون T مثلثی و هرمیتی است، پس باید یک ماتریس قطری باشد.

اثبات ۲: چون مقادیر ویژه A عناصر قطری ماتریس T هستند و عناصر قطری یک ماتریس هرمیتی

باید حقیقی باشند، نتیجه می شود که مقادیر ویژه A حقیقی هستند.

تبصره: توجه کنید که اگر A حقیقی باشد، آنگاه $A^* = A$ ایجاب می کند که A متقارن باشد ($A^T = A$). مقادیر ویژه یک ماتریس متقارن حقیقی همچنین حقیقی هستند.

اثبات ۳: ستونهای U را توسط u_1 تا u_n نمایش دهید. سپس از

$$U^*AU = D$$

یا

$$AU = UD$$

یا

$$A(u_1, u_2, \dots, u_n) = (u_1, u_2, \dots, u_n) \cdot \text{diag}(d_1, d_2, \dots, d_n)$$

ملاحظه می کنیم که u_i بردار ویژه A نظیر مقدار ویژه d_i است. چون U یک ماتریس یکانی است، بردارهای ویژه u_1, u_2, \dots, u_n یکا متعامد هستند. □

فرض کنید A حقیقی و متقارن باشد، آنگاه چون مقادیر ویژه آن حقیقی هستند و بردارهای ویژه می توانند حقیقی انتخاب شوند، ماتریس یکانی U در قضیه ۸-۲-۸ می تواند متعامد باشد.

نتیجه ۸-۲-۲ اگر A یک ماتریس متقارن حقیقی باشد، آنگاه یک ماتریس متعامد U یافت می شود به طوری که $U^T A U = D$ ، که در آن D قطری است.

قضیه ۸-۲-۹ مقادیر ویژه یک ماتریس معین مثبت هریتی مثبت هستند؛ بالعکس اگر همه مقادیر ویژه یک ماتریس هریتی مثبت باشند، آنگاه آن ماتریس باید معین مثبت باشد.

اثبات: فرض کنید (λ, x) یک جفت ویژه A باشد. آنگاه از رابطه $Ax = \lambda x$ داریم $x^* Ax = \lambda x^* x$ چون A هریتی است، λ حقیقی است. چون A معین مثبت است، $x^* Ax > 0$ همچنین چون $x \neq 0$ ،

$$x^* x = |x_1|^2 + \dots + |x_n|^2 > 0$$

که در آن

$$x = (x_1, x_2, \dots, x_n)^T$$

بنابراین

$$\lambda = \frac{x^* Ax}{x^* x} > 0$$

برای اثبات عکس قضیه، با توجه به هرمیتی بودن ماتریس A ، بر طبق قضیه ۸-۲-۸ داریم:

$$U^T A U = D = \text{diag}(\lambda_1, \dots, \lambda_n)$$

برای هر بردار مخالف صفر x ، بردار $y = U^* x$ را تعریف می‌کنیم. بنابراین $y \neq 0$. اکنون برای هر $x \neq 0$ داریم:

$$x^* A x = x^* U D U^* x = y^* D y = \sum \lambda_i |y_i|^2 > 0.$$

این رابطه ثابت می‌کند که A معین مثبت است.

۸-۲-۳ شکل متعارف جردن

ملاحظه کردیم که یک ماتریس دلخواه از مرتبه $n \times n$ با یک ماتریس قطری متشابه است اگر و فقط اگر دارای یک مجموعه کامل از n بردار ویژه مستقل خطی باشد. بنابراین، بویژه یک ماتریس هرمیتی، یک ماتریس متقارن حقیقی، و یک ماتریس دارای مقادیر ویژه متمایز با یک ماتریس قطری متشابه هستند. یعنی، این ماتریسها غیر ناقص هستند. قضیه بعد نشان می‌دهد که یک ماتریس دلخواه همیشه با یک ماتریس قطری بلوکی متشابه است.

قضیه ۸-۲-۱۰ شکل متعارف جردن. برای هر ماتریس از مرتبه $n \times n$ ، یک ماتریس نامنفرد X وجود دارد به قسمی که

$$X^{-1} A X = \begin{bmatrix} J_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & J_m \end{bmatrix}$$

که در آن

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i \end{bmatrix}$$

که در آن J_i ها مقادیر ویژه A هستند.

اثبات: مرجع (Horn and Johnson (1985, pp. 121-129 را ملاحظه کنید.

ماتریس قطری بلوکی قبلی شکل متعارف جردن A نامیده می‌شود. ماتریسهای J_1, \dots, J_m ماتریسهای جردن هستند. اگر J_i از مرتبه v_i باشد، آنگاه

$$v_1 + \dots + v_m = n$$

تعریف ۸-۲-۲ چندجمله‌ایهای

$$\det(J_i - \lambda I) = (\lambda_i - \lambda)^{v_i}$$

مقسوم علیه‌های ابتدایی A نامیده می‌شوند. اگر $v_1 = v_2 = \dots = v_m = 1$ ، آنگاه مقسوم علیه‌های ابتدایی مقسوم علیه‌های ابتدایی خطی نامیده می‌شوند.

اگر ستون i ام X را به x_i نمایش دهیم آنگاه برای بردارهای x_1, x_2, \dots, x_{v_i} داریم:

$$Ax_1 = \lambda_1 x_1$$

$$Ax_{i+1} = \lambda_i x_{i+1} + x_i, \quad i = 1, 2, \dots, v_i - 1$$

روابط مشابهی برای سایر بردارهای وجود دارد. این بردارها بردارهای اصلی یا بردارهای ویژه تعمیم یافته A نامیده می‌شوند.

تذکرها:

۱. هنگامی که همه مقادیر ویژه A متمایز هستند، هر بلوک جردن دارای مرتبه ۱ است؛ بنابراین مقسوم علیه‌های ابتدایی خطی هستند. در حقیقت اگر m بلوک جردن یافت شود، m بردار ویژه مستقل خطی وجود دارد.
۲. اگر چند مقسوم علیه ابتدایی غیر خطی باشند، آنگاه A دارای مقادیر ویژه‌ای است که مرتبه آنها بزرگ‌تر از تعداد بردارهای ویژه مستقل خطی وابسته به آنها می‌باشند.
۳. برای هر ماتریس متقارن حقیقی یا هرمیتی مختلط، مقسوم علیه‌های ابتدایی خطی هستند.
۴. تعداد بلوکهای جردن متناظر با یک مقدار ویژه مرتبه تکرار هندسی مقدار ویژه نامیده می‌شود. مرتبه تکرار جبری یک مقدار ویژه مجموع مرتبه‌های همه بلوکهای جردن وابسته به آن می‌باشد. مرتبه تکرار هندسی یک مقدار ویژه کوچک‌تر یا مساوی مرتبه تکرار جبری آن است.

۳-۸ مسائل مقدار ویژه مطرح در کاربردهای عملی

مسئله پیدا کردن مقادیر ویژه و بردارهای ویژه در طیف وسیعی از کاربردهای عملی مطرح می‌شود. تقریباً در همه شاخه‌های علوم و مهندسی مطرح می‌شود. همان گونه که قبلاً دیدیم، مدل‌های ریاضی بسیاری از مسائل مهندسی دستگاه‌های خطی معادلات دیفرانسیل و تفاضلی هستند، و جواب این معادلات اغلب بر حسب مقادیر ویژه و بردارهای ویژه ماتریسهای این دستگاهها بیان می‌شوند. علاوه بر بسیاری از مشخصه‌های مهم دستگاههای فیزیکی و مهندسی، نظیر پایداری، اغلب می‌توانند فقط با شناخت طبیعت و مکان مقادیر ویژه تعیین شوند. ما تعداد کمی مثال نمایانگر را در این بخش ارائه می‌دهیم.

۱-۳-۸ مسائل پایداری برای معادلات دیفرانسیل و تفاضلی

یک دستگاه خطی همگن از معادلات دیفرانسیل با ضرایب ثابت به شکل

$$\frac{dx_1(t)}{dt} = a_{11}x_1(t) + a_{12}x_2(t) + \dots + a_{1n}x_n(t)$$

$$\frac{dx_2(t)}{dt} = a_{21}x_1(t) + a_{22}x_2(t) + \dots + a_{2n}x_n(t)$$

⋮

$$\frac{dx_n(t)}{dt} = a_{n1}x_1(t) + a_{n2}x_2(t) + \dots + a_{nn}x_n(t)$$

یا به شکل ماتریسی

$$\dot{x}(t) = Ax(t) \quad (1-3-8)$$

که در آن

$$A = (a_{ij})_{n \times n}$$

و

$$\dot{x}(t) = \frac{d}{dt} \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

در سطح وسیعی از دستگاههای فیزیکی و مهندسی مطرح می‌شود. جواب این دستگاه به مسئله مقدار ویژه ماتریس A به طور نزدیکی وابسته است.

برای دیدن این مطلب، فرض کنید که (۱-۳-۸) دارای یک جواب $x(t) = ve^{t}$ باشد، که در آن v

بستگی به t ندارد. بنابراین از (۱-۳-۸) باید داشته باشیم

$$\lambda v e^{\lambda t} = A v e^{\lambda t} \quad (۲-۳-۸)$$

یعنی

$$A v = \lambda v \quad (۳-۳-۸)$$

که نشان می‌دهد که λ یک مقدار ویژه A و v بردار ویژه متناظر می‌باشد. بنابراین جفت مقدار ویژه (λ, v) ماتریس A می‌تواند برای محاسبه یک جواب معادله (۱-۳-۸) مورد استفاده قرار گیرد. اگر A دارای n بردار ویژه مستقل خطی باشد (همان گونه که ملاحظه نمودیم، اتفاق خواهد افتاد اگر مقادیر ویژه A همگی متمایز باشند)، آنگاه جواب عمومی دستگاه می‌تواند به صورت زیر نوشته شود

$$x(t) = c_1 v_1 e^{\lambda_1 t} + c_2 v_2 e^{\lambda_2 t} + \dots + c_n v_n e^{\lambda_n t} \quad (۴-۳-۸)$$

که در آن $\lambda_1, \lambda_2, \dots, \lambda_n$ مقادیر ویژه A هستند و v_1, v_2, \dots, v_n بردارهای ویژه متناظر می‌باشند. در حالت کلی، جواب عمومی (۱-۳-۸) با $x(0) = x_0$ به صورت

$$x(t) = e^{At} x_0$$

ارائه می‌شود که در آن

$$e^{At} = I + At + \frac{A^2 t^2}{2!} + \dots$$

اکنون مقادیر ویژه λ_i ها و بردارهای ویژه متناظر در محاسبه e^{At} ظاهر می‌شوند.

برای مثال اگر A دارای شکل متعارف جردن $\text{diag}(J_1, J_p, \dots, J_k)$ با $V^{-1}AV = \text{diag}(J_1, J_p, \dots, J_k)$ باشد، آنگاه

$$J_i = \begin{bmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix} \quad (۷-۳-۸)$$

باشد، آنگاه

$$e^{At} = V \text{diag}(e^{J_1 t}, e^{J_p t}, \dots, e^{J_k t}) V^{-1} \quad (۸-۳-۸)$$

اگر $p = v_i - 1$ که v_i مرتبه ماتریس J_i می‌باشد، آنگاه

$$e^{J_i t} = e^{\lambda_i t} \begin{bmatrix} 1 & t & t^2/2 & \dots & t^p/p! \\ 0 & 1 & t & \dots & t^{p-1}/(p-1)! \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & t \\ 0 & \dots & \dots & 0 & 1 \end{bmatrix} \quad (۸-۳-۹)$$

بنابراین، دستگاه معادلات دیفرانسیل (۸-۳-۸) به طور کامل با شناخت مقادیر ویژه و بردارهای ویژه ماتریس A دستگاه حل می شود.

بعلاوه، همان گونه که قبلاً گفتیم، بسیاری از خاصیت‌های جالب و مطلوب دستگاه‌های فیزیکی و مهندسی فقط با شناخت مکان یا طبیعت مقادیر ویژه ماتریس A دستگاه می توانند مطالعه شوند. پایداری یک چنین خاصیتی است. پایداری نسبت به یک جواب تعادلی تعریف می شود.

تعریف ۸-۳-۱ یک جواب تعادلی دستگاه

$$\dot{x}(t) = Ax(t), \quad x(0) = x_e.$$

بردار x_e است که در رابطه زیر صدق می کند

$$Ax_e = 0.$$

واضح است $x_e = 0$ یک جواب تعادلی است و یک جواب منحصر به فرد است اگر و فقط اگر A نامنفرد باشد.

اکنون یک تعریف ریاضی از پایداری را ارائه می دهیم.

تعریف ۸-۳-۲ یک جواب تعادلی x_e پایدار نامیده می شود اگر برای هر $\varepsilon > 0$ یک عدد حقیقی $\delta > 0$ وجود داشته باشد به طوری که $\|x(t) - x_e\| < \varepsilon$ هرگاه $\|x_0 - x_e\| < \delta$. دستگاه (۸-۳-۸) پایدار است اگر جواب تعادلی $x_e = 0$ پایدار باشد.

به عبارت دیگر، پایداری تضمین می کند که دستگاه نزدیک به مبدأ باقی می ماند اگر و فقط اگر وضعیت اولیه نزدیک به آن انتخاب شود.

در بسیاری از وضعیت‌ها پایداری کافی نیست.

تعریف ۳-۳-۸ یک جواب تعادلی x_e به طور مجانبی پایدار است اگر پایدار باشد و یک $\delta > 0$ وجود داشته باشد به قسمی که $\|x(t) - x_e\| \rightarrow 0$ هنگامی که $t \rightarrow \infty$ ، هرگاه $\|x_0 - x_e\| < \delta$.

تعریف ۴-۳-۸ دستگاه (۱-۳-۸) به طور مجانبی پایدار است اگر جواب تعادلی $x_e = 0$ به طور مجانبی پایدار باشد.

به عبارت دیگر، پایداری مجانبی نه تنها پایداری را تضمین می‌کند بلکه همچنین مطمئن می‌سازد که هر جواب $x(t)$ به مبدأ میل می‌کند هرگاه وضعیت اولیه نزدیک به آن انتخاب شود.

چون هر دستگاه به طور مجانبی پایدار لزوماً پایدار است، اما عکس آن درست نیست، قرارداد زیر به طور معمول پذیرفته می‌شود.

تعریف ۵-۳-۸ یک دستگاه به طور حاشیه‌ای پایدار نامیده می‌شود اگر پایدار باشد اما به طور مجانبی پایدار نباشد.

تعریف ۶-۳-۸ یک دستگاه که پایدار نباشد ناپایدار نامیده می‌شود.

معیارهای ریاضی برای پایداری مجانبی

قضیه ۱-۳-۸ قضیه پایداری برای یک دستگاه معادلات دیفرانسیل همگن

۱. یک شرط لازم و کافی برای این که جواب تعادلی $x_e = 0$ دستگاه همگن (۱-۳-۸) به طور

مجانبی پایدار باشد این است که مقادیر ویژه ماتریس A همگی دارای قسمت حقیقی منفی باشند.

۲. یک جواب تعادلی ناپایدار است اگر حداقل یک مقدار ویژه دارای قسمت حقیقی مثبت باشد.

اثبات: کافی است ثابت کنیم که $x(t) \rightarrow 0$ هنگامی که $t \rightarrow \infty$. چون جواب عمومی دستگاه

$\dot{x}(t) = Ax(t)$ توسط $x(t) = e^{At}x(0)$ ارائه می‌شود. اثبات قضیه از روابط (۶-۳-۸) الی (۹-۳-۸)

نتیجه می‌گردد.

توجه کنید که اگر $\lambda_j = \alpha_j + i\beta_j$ ، $j = 1, 2, \dots, n$ ، آنگاه $e^{\lambda_j t} = e^{\alpha_j t} e^{i\beta_j t}$ و $e^{\lambda_j t} \rightarrow 0$ هنگامی که $t \rightarrow \infty$ ، اگر و فقط اگر $\alpha_j < 0$.

□

پایداری یک دستگاه ناهمگن

بسیاری از وضعیتهای عملی منجر به یک مدل ریاضی به شکل

$$\dot{x}(t) = Ax(t) + b \quad (۱۱-۳-۸)$$

می‌شوند که در آن b یک بردار ثابت است. پایداری چنین دستگاهی همچنین توسط مقادیر ویژه A معین می‌شود. در حقیقت پایداری هر جواب (۱۱-۳-۸) با پایداری جواب تعادلی $x_e = 0$ دستگاه همگن (۱۰-۳-۸) معادل است. این را می‌توان به صورت زیر ملاحظه کرد.

فرض کنید $\bar{x}(t)$ یک جواب تعادلی دستگاه همگن (۱۱-۳-۸) باشد. تعریف می‌کنیم

$$z(t) = x(t) - \bar{x}(t)$$

پس

$$\begin{aligned} \dot{z}(t) &= \dot{x}(t) - (\dot{\bar{x}})(t) \\ &= Ax(t) + b - A\bar{x}(t) - b \\ &= A(x(t) - \bar{x}(t)) = Az(t) \end{aligned}$$

بنابراین $\bar{x}(t) \rightarrow x(t)$ اگر و فقط اگر $z(t) \rightarrow 0$. بنابراین قضیه (۲-۳-۸) از قضیه (۱-۳-۸) نتیجه می‌شود.

قضیه ۲-۳-۸ قضیه پایداری برای یک دستگاه معادلات دیفرانسیل ناهمگن

۱. یک جواب تعادلی (۱۱-۳-۸) به طور مجانبی پایدار است اگر و فقط اگر همه مقادیر ویژه ماتریس A دارای قسمت حقیقی منفی باشند.

۲. یک جواب تعادلی ناپایدار است اگر حداقل یک مقدار ویژه دارای قسمت حقیقی مثبت باشد.

تبصره: چون پایداری یک جواب تعادلی بستگی به مقادیر ویژه ماتریس A دستگاه دارد، مرسوم است که به پایداری خود دستگاه یا فقط پایداری ماتریس A اشاره شود.

□

پایداری یک دستگاه معادلات تفاضلی

نظیر دستگاه معادلات دیفرانسیل (۱۱-۳-۸)، بسیاری از دستگاههای عملی توسط یک دستگاه معادلات تفاضلی $x_{k+1} = Ax_k + b$ مدل بندی می‌شوند.

یک معیار ریاضی معروف برای پایداری مجانبی یک چنین دستگاهی در قضیه زیر ارائه می شود. اثبات آن را به خواننده واگذار می کنیم.

قضیه ۸-۳-۳ قضیه پایداری برای دستگاه معادلات تفاضلی. یک جواب تعادلی

$$x_{k+1} = Ax_k + b \quad (8-3-12)$$

به طور مجانبی پایدار است اگر و فقط اگر همه مقادیر ویژه A در داخل دایره واحد قرار گیرند. جواب تعادلی ناپایدار است اگر حداقل یک مقدار ویژه دارای اندازه ای بزرگتر از ۱ باشد. به طور خلاصه همه آنچه که ما لازم است برای تعیین پایداری و پایداری مجانبی یک دستگاه مدل بندی شده توسط یک دستگاه معادلات دیفرانسیل یا تفاضلی مرتبه اول بدانیم این است که آیا همه مقادیر ویژه A به ترتیب در نیم صفحه چپ یا داخل دایره واحد قرار می گیرند یا خیر. شناخت صریح مقادیر ویژه لازم نیست.

مثال ۸-۳-۱ یک مسابقه تسلیحاتی اروپایی. مسابقه تسلیحاتی ۱۹۱۴-۱۹۰۹ بین دو اتحاد اروپایی را در نظر بگیرید.

اتحاد ۱: فرانسه و روسیه

اتحاد ۲: آلمان و اتریش - مجارستان

دو اتحاد بر علیه هم به جنگ پرداختند. اجازه دهید سعی کنیم این حقیقت تاریخی را توسط آگاهی از پایداری بیان کنیم. ابتدا مدل ریاضی خام (اما ساده) زیر را از جنگ بین دو کشور در نظر می گیریم.

$$\frac{dx_1}{dt} = k_1 x_1 - \alpha_1 x_1 + g_1$$

$$\frac{dx_2}{dt} = k_2 x_2 - \alpha_2 x_2 + g_2$$

که در آن

$$x_i(t) = \text{امکان بالقوه جنگی کشور } i, \quad i = 1, 2$$

$$g_i(t) = \text{شکایت هایی که کشور } i \text{ بر علیه کشور دیگر دارد, } \quad i = 1, 2$$

α_i, k_i, g_i و $i = 1, 2$ همگی ثابت های مثبت هستند، هزینه تسلیحاتی کشور i را نمایش می دهد. این مدل ریاضی اثر ریباردسون بوده و به مدل ریباردسون معروف است.

توجه کنید که این مدل ساده از این جهت واقع گرایانه است که نرخ تغییر امکان بالقوه جنگی یک کشور به امکان بالقوه جنگی کشور دیگر، به شکایت هایی که یک کشور بر علیه دشمن خود دارد، و به

هزینه تسلیحاتی که کشور می تواند از عهده آن برآید، بستگی دارد. اگرچه دو عامل اول باعث افزایش نرخ می شوند، ولیکن عامل آخر باعث کاهش سرعت می گردد (بدین علت علامت منفی به این جمله نسبت داده شده است).

در شکل ماتریسی این مدل می تواند به صورت زیر نوشته شود

$$\dot{x}(t) = Ax(t) + g$$

که در آن

$$A = \begin{bmatrix} -\alpha_1 & k_1 \\ k_2 & -\alpha_2 \end{bmatrix}, \quad x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

$$g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$$

مقادیر ویژه A عبارتند از:

$$\lambda = \frac{-(\alpha_1 + \alpha_2) \pm \sqrt{(\alpha_1 - \alpha_2)^2 + 4k_1k_2}}{2}$$

بنابراین جواب تعادلی $x(t)$ پایدار مجانبی است اگر $\alpha_1\alpha_2 - k_1k_2 > 0$ و ناپایدار است اگر $\alpha_1\alpha_2 - k_1k_2 < 0$. این نتایج رخ می دهند زیرا هنگامی که $\alpha_1\alpha_2 - k_1k_2 > 0$ هر دو مقدار ویژه دارای قسمت حقیقی منفی خواهند بود؛ اگر منفی باشد یک مقدار ویژه دارای قسمت حقیقی مثبت خواهد بود.

برای مسابقه تسلیحاتی اروپایی، برآوردهای α_1 ، α_2 و k_1 ، k_2 تحت برخی فرضهای واقع گرایانه انجام شده اند و به صورت زیر پیدا شده اند:

$$\alpha_1 = \alpha_2 = 0.2$$

$$k_1 = k_2 = 0.9$$

(برای جزئیات بیشتر در مورد این که چگونه این برآوردها به دست آمده اند، کتاب M. Braun (1978) را ملاحظه کنید. فرضهای اصلی این است که هر دو اتحاد به طور تقریبی از قدرت یکسانی برخوردار باشند و α_1 و α_2 همانند انگلستان باشند، که معمولاً برابر عکس طول عمر مجلس انگلستان (پنج سال) اختیار می شود.

با این مقادیر $\alpha_1, \alpha_2, k_1, k_2$ داریم

$$\alpha_1 \alpha_2 - k_1 k_2 = \alpha_1^2 - k_1^2 = -0.7700$$

بنابراین جواب تعادلی ناپایدار است. در حقیقت دو مقدار ویژه عبارتند از $-1/1000$ و 0.7000 . برای یک مدل کلی از نظریه ریچاردسون در مورد مسابقه‌های تسلیحاتی و نقش مقادیر ویژه در آن مرجع (Luenberger (1979, 209-14 را ملاحظه کنید.

همگرایی طرحهای تکراری برای دستگاههای خطی

در فصل ۶ (قضیه ۶-۱۰-۱) ملاحظه کردیم که تکرار

$$x_{k+1} = Bx_k + d \quad (13-3-8)$$

برای حل دستگاه خطی

$$Ax = b$$

به جواب x به ازای هر انتخاب تقریب اولیه x_1 همگرا می‌گردد اگر و فقط اگر شعاع طیفی B کوچک‌تر از ۱ باشد، یعنی $\rho(B) < 1$.

بنابراین دوباره ملاحظه می‌کنیم که فقط یک شناخت ضمنی از مقادیر ویژه B لازم است تا تعیین کنیم که آیا یک طرح تکراری مبتنی بر (13-3-8) همگرا است یا خیر.

۲-۳-۸ مسأله ارتعاش، مسأله کمانش، شبیه‌سازی جریان عبوری از یک مدار الکتریکی

مسائل نظیر تحلیل ارتعاش و کمانش سازه‌ها، و شبیه‌سازی جریان عبوری از مدارهای الکتریکی اغلب یک دستگاه معادلات دیفرانسیل از مرتبه دو به شکل

$$By'' + Ay = 0 \quad (14-3-8)$$

را به وجود می‌آورند که در آن

$$\ddot{y} = \begin{bmatrix} \ddot{y}_1(t) \\ \ddot{y}_2(t) \\ \vdots \\ \ddot{y}_n(t) \end{bmatrix} \quad (15-3-8)$$

جواب یک چنین دستگاهی منجر به یک مسأله مقدار ویژه از نوع زیر می‌شود:

$$Ax = \lambda Bx \quad (۱۶-۳-۸)$$

این می‌تواند به صورت زیر ملاحظه شود:

فرض کنید که جوابها به شکل $y = xe^{i\omega t}$ باشند. بنابراین از (۱۶-۳-۸) باید داشته باشیم

$$\omega^2 Bx = Ax$$

با نوشتن $\lambda^2 = \omega^2$ ، این رابطه به صورت زیر در می‌آید

$$Ax = \lambda Bx \quad (۱۷-۳-۸)$$

یک چنین مسأله مقدار ویژه‌ای یک مسأله مقدار ویژه تعمیم یافته نامیده می‌شود.

عدد λ یک مقدار ویژه نامیده می‌شود، و x بردار ویژه متناظر از مسأله مقدار ویژه تعمیم یافته

(۱۷-۳-۸) نامیده می‌شود.

در بسیاری از مواقع در عمل ماتریسهای A و B متقارن و معین مثبت هستند. در این حالت مقادیر

ویژه مسأله مقدار ویژه تعمیم یافته همگی حقیقی و مثبت هستند و بردارهای ویژه می‌توانند مستقل

خطی اختیار شوند. در چنین حالتی یک جواب صریح دستگاه می‌تواند برحسب مقادیر ویژه و

بردارهای ویژه نوشته شود.

جواب مقدار ویژه - بردار ویژه $\ddot{y} + Ay = 0$ با B معین مثبت

فرض کنید x_1, x_2, \dots, x_n بردار ویژه مستقل خطی متناظر با مقادیر ویژه

$\lambda_1, \lambda_2, \dots, \lambda_n$ از (۱۷-۳-۸) با A و B معین مثبت باشند. آنگاه جواب عمومی

(۱۶-۳-۸) به صورت زیر ارائه می‌شود

$$y = \sum_{k=1}^n (c_k e^{i\sqrt{\lambda_k}t} + d_k e^{-i\sqrt{\lambda_k}t}) x_k$$

که در آن c_k و d_k ثابتهای دلخواه هستند (Wilkinson 1965, 34-35).

در مسائل ارتعاش، ماتریسهای A و B به ترتیب ماتریسهای جرم و سختی نامیده می‌شوند و

توسط M و K نمایش داده می‌شوند و مسأله مقدار ویژه تعمیم یافته زیر را به وجود می‌آوردند:

$$Kx = \lambda Mx$$

تعریف ۸-۳-۷ مقادیر $\omega_i = \sqrt{\lambda_i}$ ، $i = 1, \dots, n$ ، فرکانسهای طبیعی و x_1, x_2, \dots, x_n مدهای ارتعاش دستگاه نامیده می‌شوند. فرکانسها می‌توانند برای تعیین دوره‌های T_p برای ارتعاشها مورد استفاده قرار گیرند. بنابراین

$$T_p = \frac{2\pi}{\omega_i}$$

دوره ارتعاش برای i امین مد است.

همان گونه که خواهیم دید رفتار یک دستگاه مرتعش می‌تواند با شناخت فرکانسهای طبیعی و مدها تجزیه و تحلیل شود. یک مثال ساده برای نشان دادن این موضوع ارائه خواهیم داد. یک فصل کامل (فصل ۹) به مسأله مقدار ویژه تعمیم یافته اختصاص داده شده است.

مسائل سازه‌ای دیگر نظیر مسأله کمانش یک تیر مسائل مقدار مرزی برای معادلات دیفرانسیل مرتبه دو را درگیر می‌کنند. جوابهای چنین مسائلی با استفاده از تفاضلهای متناهی نیز به مسائل مقدار ویژه منجر می‌شوند. ما این وضعیت را نیز در این بخش توضیح خواهیم داد.

سپس نشان خواهیم داد که چگونه فراوانیها و مدها می‌توانند برای پیش‌بینی پدیده تشدید در مهندسی ارتعاش مورد استفاده قرار گیرند.

همه ماشینها و سازه‌ها، نظیر پلها، ساختمانها، و هواپیماها، که حاوی جرم و سختی هستند ارتعاش را تا درجه‌ای تحمل می‌کنند، و طرح آنها نیاز به ملاحظه رفتار نوسانی آنها دارد. ارتعاش آزاد هنگامی رخ می‌دهد که یک دستگاه بر اثر نیروهای ذاتی موجود در دستگاه و بدون هیچ نیروی خارجی نوسان کند. تحت ارتعاش آزاد یک چنین دستگاهی در یک یا چند تا از فرکانسهای طبیعی اش ارتعاش می‌کند، که اینها خاصیتهای دستگاه پویا هستند و بستگی به جرم وابسته و توزیع سختی دارند. برای ارتعاش تحمیلی، دستگاهها تحت تحریک نیروهای خارجی نوسان می‌کنند. هنگامی که چنین تحریکی نوسانی است، دستگاه نیز مجبور به ارتعاش در فرکانس تحریک می‌شود.

پدیده تشدید

اگر فرکانس تحریک بر یکی از فرکانسهای طبیعی دستگاه منطبق و یا نزدیک شود، ممکن است به طور خطرناکی موجب نوسانهای بزرگ گردد، و با یک وضعیت تشدید برخورد شود. این از آن وضعیتهایی است که یک مهندس بسیار مایل خواهد بود که از آن پرهیز کند. فروریختن پل دره تاکوما (همچنین معروف به گرتی گالوینگ) بر روی پوگت ساوند در ایالت واشنگتن در ۱۹۴۰ و فروریختن پل معلق برایتن در انگلستان به چنین پدیده‌ای نسبت داده شده‌اند.

این توضیح معمولاً پذیرفته شده برای فروریختن پل تاکوما توسط دانشمندانی که باور دارند که ممکن است چیز بیشتری برای آن وجود داشته باشد، به چالش کشیده شده است. مقاله جالب (Lazer and McKenna, 1990) و مراجع واقع در آن را ملاحظه کنید.

در هر دو حالت یک فرکانس نیروی محرک با یکی از فرکانسهای طبیعی پل در زمان فروریختن برابر شده است. در حالت پل برایتن نیروی محرک بر اثر به طور همساز قدمرو رفتن سربازان بر روی پل وارد شده است. در حالت پل تاکوما نیروی باد موجب شده است. به واسطه آنچه که در برایتن اتفاق افتاد، سربازان دیگر حق ندارند که به طور همساز بر روی پل قدمرو بروند. ساختمانها می‌توانند با وسایل فعال ساخته شوند که نیروی ناشی از باد بتواند کنترل شود.

خاصیت مهم دیگر دستگاههای پویا میرایی است که در همه دستگاهها بر اثر از بین رفتن انرژی توسط اصطکاک و سایر مقاومتها وجود دارد. ولیکن مقادیر کوچک میرایی اثر خیلی کمی بر روی فرکانسهای طبیعی دارند و به طور معمول در برآورد فرکانسهای طبیعی وارد نمی‌شوند. میرایی در محدود کردن دامنه نوسان در تشدید اهمیت دارد؛ اثر مهم میرایی در کاهش دادن دامنه با زمان می‌باشد.

برای یک جسم کثشان پیوسته تعداد مؤلفه‌های مستقل یا درجه‌های آزادی نامتناهی است؛ ولیکن تقریباً تحت همه شرایط دستگاه ممکن است به طور پویایی با یک دستگاه که دارای تعداد متناهی درجه آزادی است، معادل تلقی شود.

به طور خلاصه، رفتار یک دستگاه مرتعش می‌تواند با شناخت فرکانسها و مدها تجزیه و تحلیل شود و مقادیر ویژه و بردارهای ویژه ماتریس مدل ریاضی دستگاه به این کمیتها وابسته

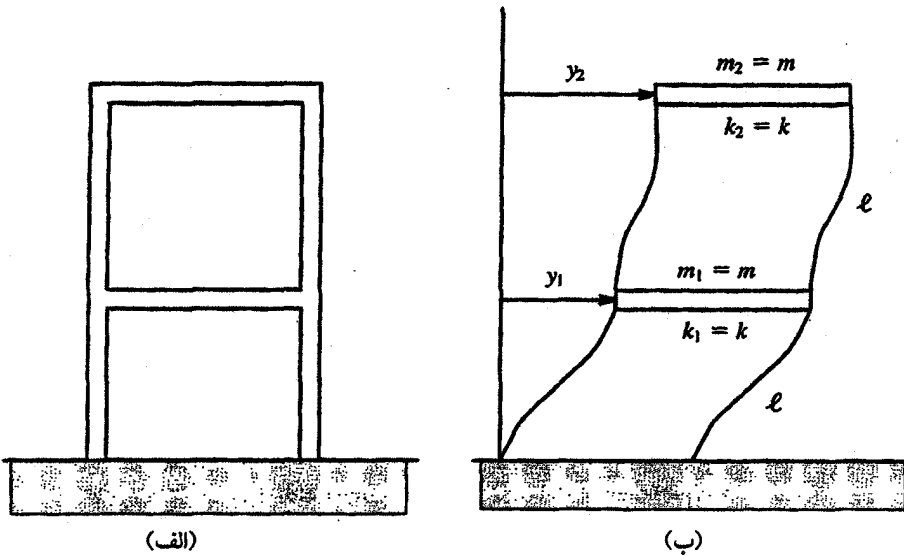
هستند. بویژه فرکانسها برابر ریشه‌های دوم مقادیر ویژه و مدها برابر بردارهای ویژه هستند.

مثال ۸-۳-۲ ارتعاش یک ساختمان. یک ساختمان دو طبقه (شکل ۸-۱ الف)) با یک کف سخت را در نظر بگیرید. فرض می‌شود که توزیع وزن ساختمان می‌تواند به شکل وزن متمرکز شده در هر سطح کف، به صورت نشان داده شده در شکل ۸-۱ ب)، نمایش داده شود، و سختی‌های ستونهای حمال توسط ثابتهای فنر k_i نمایش داده شوند.

معادلات حرکت برای این دستگاه می‌توانند به شکل زیر نوشته شوند

$$m_1 \ddot{y}_1 + (k_1 + k_2)y_1 - k_2 y_2 = 0$$

$$m_2 \ddot{y}_2 - k_2 y_1 + k_2 y_2 = 0$$



شکل ۸-۱

یا

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{bmatrix} + \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \quad (۸-۳-۱)$$

با اختیار کردن $k_1 = k_2 = k$ ، $m_1 = m_2 = m$ داریم

$$\begin{bmatrix} m & \cdot \\ \cdot & m \end{bmatrix} \begin{bmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{bmatrix} + \begin{bmatrix} \gamma k & -k \\ -k & k \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \cdot \quad (۱۹-۳-۸)$$

با تعریف کردن ماتریس جرم

$$M = \begin{bmatrix} m & \cdot \\ \cdot & m \end{bmatrix}$$

و ماتریس سختی

$$K = \begin{bmatrix} \gamma k & -k \\ -k & k \end{bmatrix}$$

معادله (۱۹-۳-۸) به شکل زیر در می آید

$$M\ddot{y} - Ky = \cdot, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad \text{که در آن} \quad (۲۰-۳-۸)$$

با فرض این که یک جواب به شکل $y = xe^{i\omega t}$ است با ضرب معادله (۲۰-۳-۸) در M^{-1} و قراردادن $A = M^{-1}K$ و $\lambda = \omega^2$ ، معادلات حرکت (۲۰-۳-۸) مسأله مقدار اولیه زیر را نتیجه می دهند.

$$\begin{bmatrix} \frac{\gamma k}{m} & -\frac{k}{m} \\ -\frac{k}{m} & \frac{k}{m} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (۲۱-۳-۸)$$

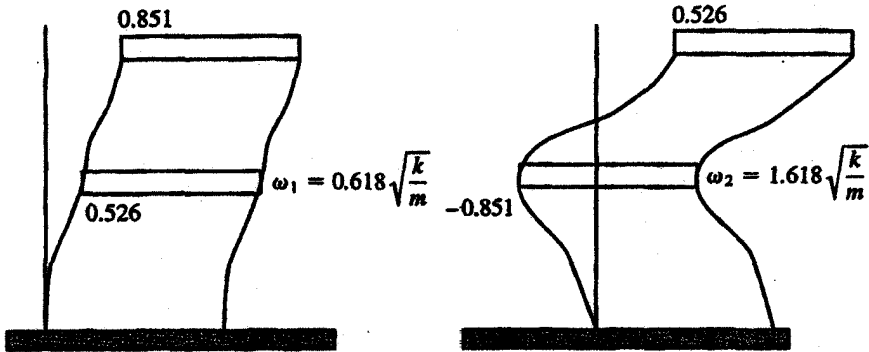
مقادیر ویژه λ_1 و λ_2 و بردارهای ویژه متناظر که نمایش دهنده دو مد نرمال هستند، برای این مسأله 2×2 به سادگی حل می شوند. مقادیر ویژه عبارتند از:

$$\lambda_1 = \frac{k}{m} (۰/۳۸۲۰), \quad \lambda_2 = \frac{k}{m} (۲/۶۱۸۰)$$

بردارهای ویژه عبارتند از:

$$\begin{bmatrix} ۰/۵۲۵۷ \\ ۰/۸۵۰۷ \end{bmatrix}, \quad \begin{bmatrix} -۰/۸۵۰۷ \\ ۰/۵۲۵۷ \end{bmatrix}$$

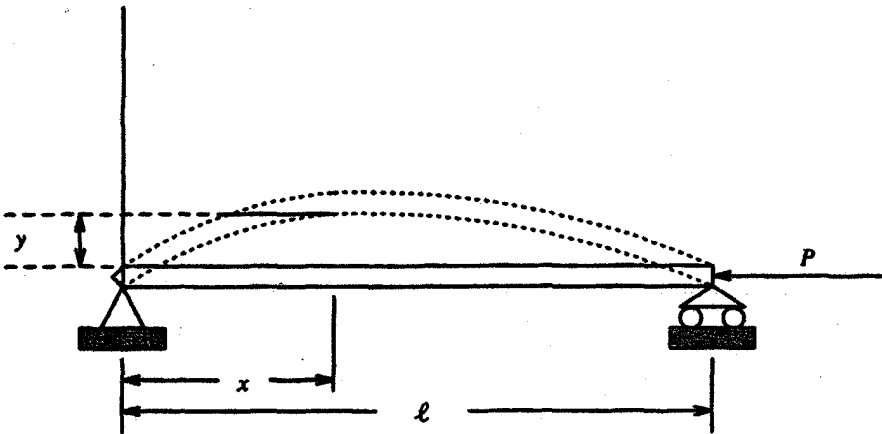
ارتعاش ساختمان در این دو مد در شکل ۸-۲ نشان داده شده است.



شکل ۲-۸

مسأله کمانش (یک مسأله مقدار مرزی)

یک تیر نازک یکنواخت با طول l را در نظر بگیرید. یک بار محوری P در یکی از دو انتها، به صورت شکل ۳-۸، بر تیر اعمال می‌شود.



شکل ۳-۸

به پایداری تیر علاقه‌مندیم - یعنی چگونه و چه موقع تیر خم می‌شود. ما نشان خواهیم داد که چگونه این مسأله به مسأله مقدار ویژه منجر می‌شود و چه نقشی مقادیر ویژه بازی می‌کنند.

فرض کنید y تغییر مکان عمودی یک نقطه از تیر را نمایش دهد که در فاصله x از تکیه‌گاه چپ قرار داشته باشد. فرض کنید که هر دو انتهای تیر به طور ساده نگه داشته شوند. یعنی $y(0) = y(l) = 0$. بر طبق نظریه تیر، رابطه اساسی بین انحنای d^2y/dx^2 و گشتاور داخلی M برای تیر عبارت است از:

$$\frac{d^2y}{dx^2} = \frac{M}{EI}$$

که در آن E ضریب کشسانی و I گشتاور لختی سطح از سطح مقطع ستونی است. با نوشتن گشتاور خمش در هر سطح به صورت $M = Py$ ، این معادله معادله دیفرانسیل حاکم را ارائه می‌دهد، که معادله گشتاور خمش است:

$$EI \frac{d^2y}{dx^2} = -Py \quad (22-3-8)$$

ثابت EI سختی انحنا پذیری تیر نامیده می‌شود.

معادله گشتاور خمش و شرایط مرزی

$$y(0) = y(l) = 0 \quad (23-3-8)$$

یک مسأله مقدار مرزی تشکیل می‌دهند.

این مسأله مقدار مرزی را با تقریب زدن d^2y/dx^2 با یک طرح تفاضل متناهی مناسب، نظیر آنچه در فصل ۶ انجام دادیم، حل می‌کنیم.

فرض کنید بازه $[0, l]$ به n زیربازه با طول مساوی h ، با نقاط x_0, x_1, \dots, x_n ، به عنوان نقاط تقسیم، افزاز شود؛ یعنی

$$0 = x_0 < x_1 < x_2 < \dots < x_j < \dots < x_{n-1} < x_n = l$$

و

$$h = \frac{l}{n}$$

فرض کنید

$$\left. \frac{d^2y}{dx^2} \right|_{x=x_i} \approx \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2}$$

با جایگزین کردن این تقریب d^2y/dx^2 در معادله گشتاور خمش (۲۲-۳-۸) و در نظر گرفتن شرایط مرزی

(۲۳-۳-۸)، مسأله مقدار ویژه ماتریس سه قطری متقارن زیر را به دست می‌آوریم:

$$\begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & -1 \\ 0 & \dots & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \lambda \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (25-3-8)$$

که در آن

$$\lambda = \frac{Ph^2}{EI} \quad (۲۶-۳-۸)$$

هر مقدار λ یک بار

$$P = \frac{\lambda EI}{H^2} \quad (۲۷-۳-۸)$$

را تعیین می‌کند که یک بار بحرانی نامیده می‌شود. این بارهای بحرانی آنهایی هستند که از نظر عملی جالب هستند زیرا آنها کمانش تیر را تعیین می‌کنند.

به طور کلی کوچک‌ترین مقدار P اهمیت ویژه‌ای دارد، زیرا خمش وابسته به مقادیر بزرگ‌تر P ممکن نیست بدون شکستی که تحت اثر کوچک‌ترین مقدار بحرانی P رخ می‌دهد، به دست آیند.

شبیه‌سازی جریان عبوری برای یک مدار الکتریکی (Chapra and Canale 1988)

با مفروض بودن یک مدار الکتریکی متشکل از چندین حلقه، فرض کنید که رفتار جریان عبوری مدار الکتریکی مورد توجه شما باشد. بویژه می‌خواهیم نوسان هر حلقه را نسبت به سایرین بشناسیم. ابتدا مدار حلقه نشان داده شده در شکل ۸-۴ را ملاحظه می‌کنیم. افت ولتاژ در یک خازن برابر است با

$$V_C = \frac{q}{C}; \quad q = \text{بار خازن}, \quad C = \text{ظرفیت}$$

افت ولتاژ در یک القاگر برابر است با

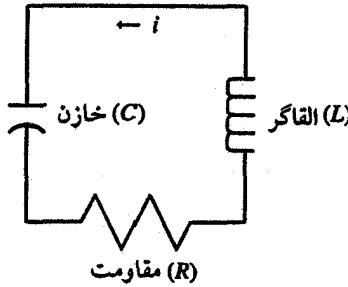
$$V_L = L \frac{di}{dt}; \quad L = \text{ضریب القایی}$$

افت ولتاژ در یک مقاومت برابر است با

$$V_R = iR; \quad R = \text{مقاومت}$$

قانون ولتاژ کیرشهوف^۱ بیان می‌کند که مجموع جبری افت ولتاژ در یک مدار حلقه بسته برابر صفر است. پس برای این مدار داریم:

$$L \frac{di}{dt} + Ri + \frac{q}{C} = 0$$



شکل ۸-۴

یا

$$L \frac{di}{dt} + iR + \frac{1}{C} \int_{-\infty}^{\cdot} idt = 0$$

(چون $V_C = \frac{q}{C} = \frac{1}{C} \int_{-\infty}^t idt$)

اکنون شبکه‌ای را با چهار حلقه به صورت نشان داده شده در شکل ۸-۵ ملاحظه می‌کنیم. به کار بردن قانون ولتاژ کیرشهوف در مورد هر حلقه نتیجه می‌دهد.

حلقه ۱:

$$-L_1 \frac{di_1}{dt} - \frac{1}{C_1} \int_{-\infty}^t (i_1 - i_2) dt = 0 \tag{۲۸-۳-۸}$$

حلقه ۲:

$$-L_2 \frac{di_2}{dt} - \frac{1}{C_2} \int_{-\infty}^t (i_2 - i_3) dt + \frac{1}{C_1} \int_{-\infty}^t (i_1 - i_2) dt = 0 \tag{۲۹-۳-۸}$$

حلقه ۳:

$$-L_3 \frac{di_3}{dt} - \frac{1}{C_3} \int_{-\infty}^t (i_3 - i_4) dt + \frac{1}{C_2} \int_{-\infty}^t (i_2 - i_3) dt = 0 \tag{۳۰-۳-۸}$$

حلقه ۴:

$$-L_4 \frac{di_4}{dt} - \frac{1}{C_4} \int_{-\infty}^t i_4 dt + \frac{1}{C_3} \int_{-\infty}^t (i_3 - i_4) dt = 0 \tag{۳۱-۳-۸}$$

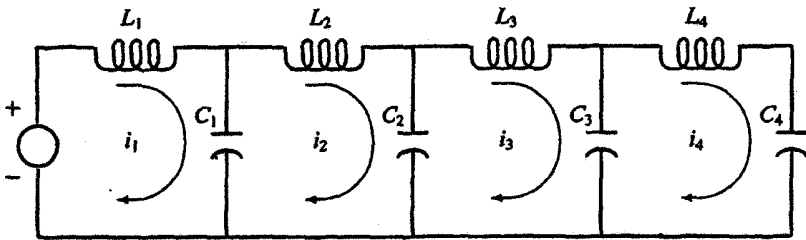
از دستگاه معادلات دیفرانسیل ارائه شده توسط (۲۸-۳-۸) تا (۳۱-۳-۸) می‌توان مشتق گرفت و مرتب نمود تا روابط زیر به دست آید.

$$L_1 \frac{d^2 i_1}{dt^2} + \frac{1}{C_1} (i_1 - i_2) = 0 \quad (32-3-8)$$

$$L_2 \frac{d^2 i_2}{dt^2} + \frac{1}{C_2} (i_2 - i_3) - \frac{1}{C_1} (i_1 - i_2) = 0 \quad (33-3-8)$$

$$L_3 \frac{d^2 i_3}{dt^2} + \frac{1}{C_3} (i_3 - i_4) - \frac{1}{C_2} (i_2 - i_3) = 0 \quad (34-3-8)$$

$$L_4 \frac{d^2 i_4}{dt^2} + \frac{1}{C_4} i_4 - \frac{1}{C_3} (i_3 - i_4) = 0 \quad (35-3-8)$$



شکل ۵-۸

فرض کنید

$$i_j = A_j \sin(\omega t), \quad j = 1, 2, 3, 4 \quad (36-3-8)$$

(یادآور می‌شویم i_j جریان در حلقه j می‌باشد).

از (۳۲-۳-۸) داریم:

$$-L_1 A_1 \omega^2 \sin \omega t + \frac{1}{C_1} A_1 \sin \omega t - \frac{1}{C_1} A_2 \sin \omega t = 0$$

یا

$$\left(\frac{1}{C_1} - L_1 \omega^2 \right) A_1 - \frac{1}{C_1} A_2 = 0 \quad (36-3-8)$$

از (۳۳-۳-۸) داریم:

$$-L_2 A_2 \omega^2 \sin \omega t + \frac{1}{C_2} A_2 \sin \omega t - \frac{1}{C_2} A_1 \sin \omega t - \frac{1}{C_2} A_3 \sin \omega t + \frac{1}{C_1} A_2 \sin \omega t = 0$$

یا

$$-\frac{1}{C_1}A_1 + \left(\frac{1}{C_1} + \frac{1}{C_r} - L_r\omega^2\right)A_r - \frac{1}{C_r}A_r = 0 \quad (38-3-8)$$

از (۳۴-۳-۸) داریم:

$$-L_r A_r \omega^2 \sin \omega t + \frac{1}{C_r} A_r \sin \omega t - \frac{1}{C_r} A_r \sin \omega t - \frac{1}{C_r} A_r \sin \omega t + \frac{1}{C_r} A_r \sin \omega t = 0$$

یا

$$-\frac{1}{C_r}A_r + \left(\frac{1}{C_r} + \frac{1}{C_r} - L_r\omega^2\right)A_r - \frac{1}{C_r}A_r = 0 \quad (39-3-8)$$

از (۳۵-۳-۸) داریم:

$$-L_r A_r \omega^2 \sin \omega t + \frac{1}{C_r} A_r \sin \omega t - \frac{1}{C_r} A_r \sin \omega t - \frac{1}{C_r} A_r \sin \omega t = 0$$

یا

$$-\frac{1}{C_r}A_r + \left(\frac{1}{C_r} + \frac{1}{C_r} - L_r\omega^2\right)A_r = 0 \quad (40-3-8)$$

از گرد هم جمع کردن معادلات (۳۷-۳-۸) تا (۴۰-۳-۸)، داریم:

$$\left(\frac{1}{C_1} - L_1\omega^2\right)A_1 - \frac{1}{C_1}A_r = 0 \quad (41-3-8)$$

$$-\frac{1}{C_1}A_1 + \left(\frac{1}{C_1} + \frac{1}{C_r} - L_r\omega^2\right)A_r - \frac{1}{C_r}A_r = 0 \quad (42-3-8)$$

$$-\frac{1}{C_r}A_r + \left(\frac{1}{C_r} + \frac{1}{C_r} - L_r\omega^2\right)A_r - \frac{1}{C_r}A_r = 0 \quad (43-3-8)$$

$$-\frac{1}{C_r}A_r + \left(\frac{1}{C_r} + \frac{1}{C_r} - L_r\omega^2\right)A_r = 0 \quad (44-3-8)$$

یا

$$\begin{aligned}
 A_1 - A_r &= L_1 C_1 \omega^2 A_1 \\
 -\frac{C_r}{C_1} A_1 + \frac{C_r}{C_1} A_r + A_r - A_r &= L_r C_r \omega^2 A_r \\
 -\frac{C_r}{C_r} A_r + \left(\frac{C_r}{C_r} + 1\right) A_r - A_r &= L_r C_r \omega^2 A_r \\
 -\frac{C_f}{C_r} A_r + \left(\frac{C_f}{C_r} + 1\right) A_r &= L_f C_f \omega^2 A_f
 \end{aligned} \tag{۴۵-۳-۸}$$

دستگاه (۴۵-۳-۸) به شکل ماتریسی به صورت زیر است

$$\begin{bmatrix}
 1 & -1 & \cdot & \cdot \\
 -\frac{C_r}{C_1} & \left(\frac{C_r}{C_1} + 1\right) & -1 & \cdot \\
 \cdot & -\frac{C_r}{C_r} & \left(\frac{C_r}{C_r} + 1\right) & -1 \\
 \cdot & \cdot & -\frac{C_f}{C_r} & \left(\frac{C_f}{C_r} + 1\right)
 \end{bmatrix}
 \begin{bmatrix}
 A_1 \\
 A_r \\
 A_r \\
 A_f
 \end{bmatrix}$$

$$= \omega^2 \begin{bmatrix}
 L_1 C_1 & \cdot & \cdot & \cdot \\
 \cdot & L_r C_r & \cdot & \cdot \\
 \cdot & \cdot & L_r C_r & \cdot \\
 \cdot & \cdot & \cdot & L_f C_f
 \end{bmatrix}
 \begin{bmatrix}
 A_1 \\
 A_r \\
 A_r \\
 A_f
 \end{bmatrix} \tag{۴۶-۳-۸}$$

این مسأله یک مسأله مقدار ویژه است. برای دیدن واضح تر آن، حالت خاص زیر را در نظر بگیرید

$$C_1 = C_r = C_r = C_f = C$$

و

$$L_1 = L_r = L_r = L_f = L$$

و فرض کنید

$$\lambda = LC\omega^2$$

پس مسأله به صورت زیر تقلیل می‌یابد

$$\begin{aligned}(1-\lambda)A_1 - A_2 &= 0 \\ -A_1 + (2-\lambda)A_2 - A_3 &= 0 \\ -A_2 + (2-\lambda)A_3 - A_4 &= 0 \\ -A_3 + (2-\lambda)A_4 &= 0\end{aligned}$$

یا

$$\begin{bmatrix} (1-\lambda) & -1 & 0 & 0 \\ -1 & (2-\lambda) & -1 & 0 \\ 0 & -1 & +(2-\lambda) & -1 \\ 0 & 0 & -1 & +(2-\lambda) \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} = 0 \quad (8-3-47)$$

چون $i_j = A_j \sin \omega t$ ، $j = 1, \dots, 4$ داریم:

$$\begin{bmatrix} (1-\lambda) & -1 & 0 & 0 \\ -1 & (2-\lambda) & -1 & 0 \\ 0 & -1 & (2-\lambda) & -1 \\ 0 & 0 & -1 & (2-\lambda) \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{bmatrix} = 0 \quad (8-3-48)$$

یا

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{bmatrix} = \lambda \begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{bmatrix} \quad (8-3-49)$$

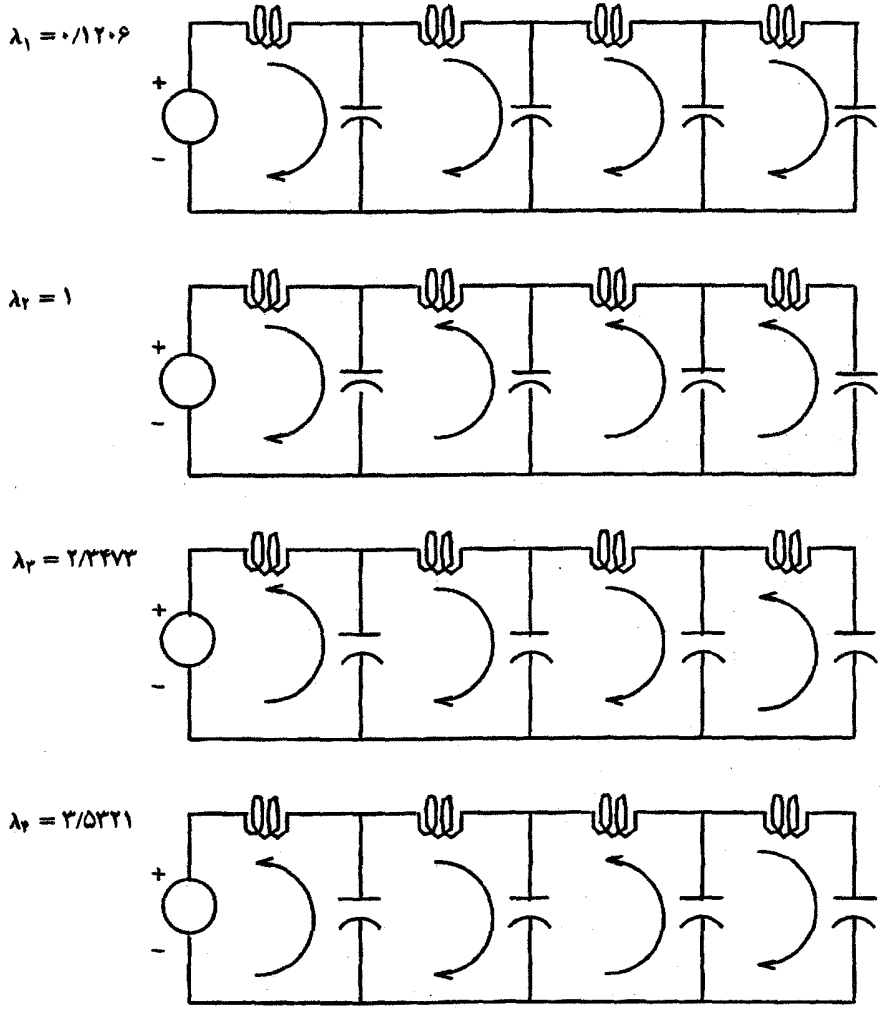
جواب این مسأله مقدار ویژه به ما فرکانسهای طبیعی ($\omega_i^2 = \lambda_i/LC$) را ارائه خواهد داد. بعلاوه شناخت بردارهای ویژه می‌تواند برای مطالعه رفتار فیزیکی مدار، نظیر مدهای طبیعی نوسان، مورد استفاده قرار گیرد.

این مقادیر ویژه و بردارهای ویژه نرمال شده متناظر (با چهار رقم اعشار) عبارتند از

$$\lambda_1 = 0.1206, \quad \lambda_2 = 1, \quad \lambda_3 = 2.3473, \quad \lambda_4 = 3.5321$$

$$\begin{bmatrix} 0/6665 \\ 0/5774 \\ 0/4285 \\ 0/2280 \end{bmatrix}, \begin{bmatrix} 0/5774 \\ -0/0000 \\ -0/5774 \\ -0/5774 \end{bmatrix}, \begin{bmatrix} -0/4285 \\ 0/5774 \\ 0/2280 \\ -0/6565 \end{bmatrix}, \begin{bmatrix} -0/2280 \\ 0/5774 \\ -0/6565 \\ 0/4285 \end{bmatrix}$$

از جهت‌های بردارهای ویژه نتیجه می‌گیریم که برای λ_1 همه حلقه‌ها در یک جهت نوسان می‌کنند. برای λ_3 حلقه‌های دوم و سوم در خلاف جهت حلقه‌های اول و چهارم نوسان می‌کنند، و غیره. این مطالب در شکل ۶-۸ نشان داده شده‌اند.



شکل ۶-۸

۸-۳-۳ یک مثال از مسأله مقدار ویژه و مطرح در آمار: آنالیز مؤلفه‌های اصلی

بسیاری از کاربردهای واقعی حاوی تحلیل آماری (مثلاً معامله اوراق بهادار یا پیش‌بینی هوا) شامل یک مقدار داده می‌باشند. در این حالت حجم و پیچیدگی داده‌ها می‌توانند محاسبات لازم برای تحلیل را عملاً غیر ممکن سازند. در عمل برای راه‌اندازی و تحلیل چنین داده‌های پرحجمی لازم است داده‌ها کاهش یابند. ایده اساسی این است که k مؤلفه از یک مجموعه داده متشکل از n اندازه‌گیری بر روی p ($p > k$) متغیر اصلی به طریقی به طور ماهرانه انتخاب شوند که بیشتر اطلاعات (اگر بیشترین نیست) موجود در p متغیر اصلی در k مؤلفه انتخاب شده موجود باشند.

در آمار چنین k مؤلفه، اولین k مؤلفه اصلی نامیده می‌شوند. شناخت مقادیر ویژه و بردارهای ویژه ماتریس کواریانس برای پیدا کردن این مؤلفه‌های اصلی لازم می‌باشد. بویژه اگر Σ ماتریس کواریانس متناظر با بردارهای تصادفی

$$X = (X_1, X_2, \dots, X_p)$$

باشد، $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ مقادیر ویژه باشند، و x_1 تا x_p بردارهای ویژه ماتریس Σ باشند، آنگاه i امین مؤلفه اصلی به صورت زیر ارائه می‌شود

$$Y_i = x_i^T X, \quad i = 1, 2, \dots, p$$

بعلاوه قسمتی از کل واریانس جامعه که از i امین مؤلفه اصلی ناشی می‌شود توسط نسبت زیر بیان می‌گردد

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} = \frac{\lambda_i}{\text{trace}(\Sigma)}, \quad i = 1, \dots, p \quad (8-3-50)$$

تذکر: ماتریس کواریانس نیمه معین مثبت متقارن است؛ بنابراین مقادیر ویژه آن همگی نامنفی هستند. اگر k نسبت اول بیشترین مقدار کل واریانس جامعه را تشکیل دهند، آنگاه k مؤلفه اصلی می‌توانند در تحلیل آماری مورد استفاده قرار گیرند.

توجه کنید که در محاسبه k امین نسبت ما فقط نیاز به k امین مقدار ویژه ماتریس کواریانس داریم؛ طیف کامل لازم نیست محاسبه شود. برای اتمام این بخش، متذکر می‌شویم که بسیاری از تجربه‌های زندگی واقعی نظیر محاسبه شاخص میانگین صنعتی داو - جونز^۱ می‌توانند بهتر توسط آنالیز مؤلفه‌های اصلی، به صورت توضیح داده شده در مثال زیر، فهمیده و توصیف شوند.

یک مثال از اوراق بهادار بازار (از مرجع Johnson and Wichern (1992) اختیار شده)

فرض کنید که ماتریس کوواریانس برای نرخهای هفتگی برگشت برای اوراق بهادار پنج شرکت بزرگ (Allied Chemical, Dupont, Union Carbide, Exxon, and Texaco) در یک دوره مفروض از زمان به صورت زیر باشد.

$$R = \begin{bmatrix} 1/000 & 0/577 & 0/509 & 0/387 & 0/462 \\ 0/577 & 1/000 & 0/599 & 0/389 & 0/322 \\ 0/509 & 0/599 & 1/000 & 0/436 & 0/426 \\ 0/387 & 0/389 & 0/436 & 1/000 & 0/523 \\ 0/462 & 0/322 & 0/426 & 0/523 & 1/000 \end{bmatrix}$$

دو مقدار ویژه اول R عبارتند از:

$$\lambda_1 = 2/857 \quad (51-3-8)$$

$$\lambda_2 = 0/809 \quad (52-3-8)$$

قسمتی از کل واریانس جامعه که از اولین مؤلفه نتیجه می‌گردد برابر است با

$$\frac{2/857}{5} = 57\% \quad (53-3-8)$$

قسمتی از کل واریانس جامعه که از دومین مؤلفه نتیجه می‌گردد برابر است با

$$\frac{0/809}{5} \cong 16\% \quad (54-3-8)$$

بنابراین دو مؤلفه اصلی اول ۷۳٪ از کل واریانس جامعه را بیان می‌کنند. بردارهای ویژه متناظر با این مؤلفه‌های اصلی عبارتند از

$$x_1^T = (0/464, 0/457, 0/470, 0/421, 0/421) \quad (55-3-8)$$

$$x_2^T = (0/240, 0/509, 0/260, -0/526, -0/582) \quad (56-3-8)$$

این بردارهای ویژه تعبیرهای جالبی دارند. از عبارت x_1 ملاحظه می‌کنیم که اولین مؤلفه (به طور تقریبی) مجموع موزون به طور مساوی پنج اوراق بهادار می‌باشد. این مؤلفه معمولاً مؤلفه بازار نامیده می‌شود. ولیکن عبارت مربوط به x_2 به ما می‌گوید که مؤلفه دوم یک تضاد را بین سهام مواد شیمیایی و سهام صنایع نفتی ظاهر می‌سازد. این مؤلفه معمولاً مؤلفه صنعتی نامیده می‌شود. بنابراین نتیجه می‌گیریم که در حدود ۵۷٪ از کل تغییرات در این برگشتهای سهام از عملکرد بازار و ۱۶٪ از عملکرد صنعت ناشی می‌شود.

مسأله مقدار ویژه همچنین در بسیاری از تحلیلهای آماری مهم دیگر، نظیر محاسبه همبستگیهای کانونی مطرح می‌شوند. خواننده‌های علاقمند برای مطالعه بیشتر به کتاب Johnson and Wichern (1992) مراجعه نمایند.

یک نکته نهایی: بیشترین مسائل مطرح در آمار، نظیر آنالیز مؤلفه‌های اصلی و همبستگیهای کانونی، در حقیقت مسائل تجزیه مقدار تکین هستند و باید با استفاده از تجزیه مقدار تکین توصیف شده در فصل ۱۰ به‌طور محاسباتی مورد عمل قرار گیرند.

۴-۸ محدودسازی مقادیر ویژه

همان‌گونه که در چند مسأله عملی ملاحظه نمودیم، شناخت صریح مقادیر ویژه ممکن است لازم نباشد؛ همه آنچه که لازم است شناخت توزیع مقادیر ویژه در برخی نواحی مفروض از صفحه مختلط یا برآوردهایی از برخی مقادیر ویژه خاص است. راههایی وجود دارند که چنین اطلاعاتی را ممکن است بدون محاسبه واقعی مقادیر ویژه ماتریس فراهم سازند. در این بخش ما برخی از راههای معروف را بحث خواهیم کرد. با نتیجه معروف گرشگورین، ثابت شده در ۱۹۳۱، شروع می‌کنیم.

۱-۴-۸ قضیه‌های قرص گرشگورین

قضیه ۱-۴-۸ قضیه اول گرشگورین. فرض کنید $A = (a_{ij})_{n \times n}$. تعریف می‌کنیم

$$r_i = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, \dots, n$$

آنگاه هر مقدار ویژه λ از A حداقل در یکی از نامساویها زیر صدق می‌کند

$$|\lambda - a_{ii}| \leq r_i, \quad i = 1, 2, \dots, n$$

به بیان دیگر همه مقادیر ویژه A را می‌توان در اجتماع قرصهای $\{z : |z - a_{ii}| \leq r_i, i = 1, \dots, n\}$ پیدا کرد.

اثبات: فرض کنید λ یک مقدار ویژه A و x یک بردار ویژه وابسته به λ باشد. آنگاه از

$$Ax = \lambda x$$

یا از

$$(AI - A)x = 0$$

داریم

$$(\lambda - a_{ii})x_i = \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}x_j, \quad i = 1, \dots, n$$

که در آن x_i مؤلفه i ام بردار x است. فرض کنید بزرگ‌ترین مؤلفه x (از لحاظ قدر مطلق) x_k باشد. آنگاه $|x_j|/|x_k| \leq 1$ به ازای $j \neq k$ ، بنابراین داریم

$$\begin{aligned} |\lambda - a_{kk}| &\leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{kj}| \frac{|x_j|}{|x_k|} \\ &\leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{kj}| \end{aligned}$$

بنابراین λ در دایره $\{\lambda : |\lambda - a_{kk}| \leq r_k\}$ قرار دارد. \square

تعریف ۱-۴-۸ قرصهای $R_i = \{z : |z - a_{ii}| \leq r_i, i = 1, \dots, n\}$ قرصهای گرشگورین در صفحه مختلط نامیده می‌شوند.

مثال ۱-۴-۸

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 9 \\ 1 & 1 & 1 \end{bmatrix}$$

$$r_1 = 5$$

$$r_2 = 12$$

$$r_3 = 2$$

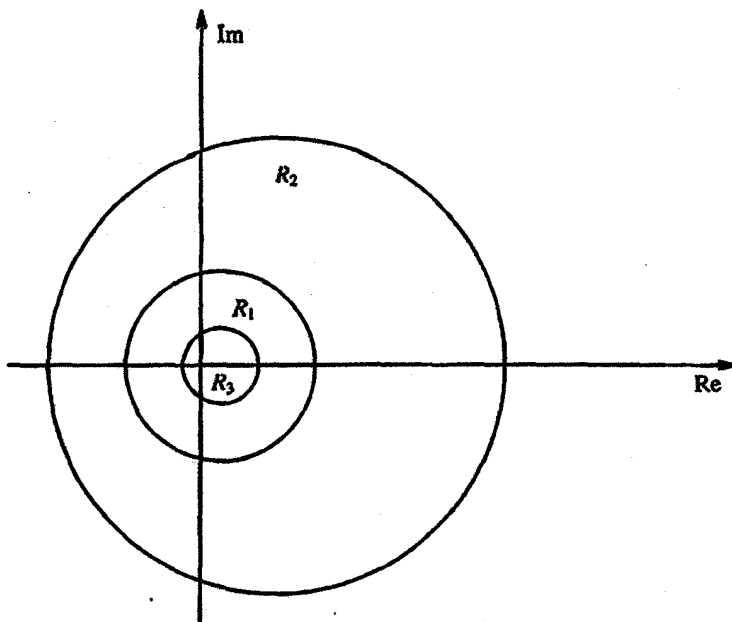
قرصهای گرشگورین عبارتند از:

$$R_1: \{z : |z - 1| \leq 5\}$$

$$R_2: \{z : |z - 4| \leq 12\}$$

$$R_3: \{z : |z - 1| \leq 2\}$$

(مقادیر ویژه عبارتند از $0.7/3.067, 0.34732i \pm 0.65233(-)$) شکل ۷-۸ را ملاحظه کنید.



شکل ۷-۸ - قرصهای گرشگورین: مثال ۱-۴-۸

اگرچه قضیه ۱-۴-۸ فقط بیان می‌کند که مقادیر ویژه A در اجتماع \mathcal{R} قرص گرشگورین قرار دارند، قضیه زیر برخی اطلاعات خاص‌تر را ارائه می‌دهد. ما این قضیه را بدون اثبات بیان می‌کنیم.

قضیه ۲-۴-۸ قضیه دوم گرشگورین. فرض کنید که \mathcal{C} قرص گرشگورین از سایر قرصها مجزا باشند، آنگاه دقیقاً \mathcal{C} مقدار ویژه A در اجتماع این \mathcal{C} قرص قرار دارند.

اثبات: مرجع (Horn and Johnson (1985, 344 - 345) را ملاحظه کنید. □

تذکره: برای تعمیمهایی از قضیه گرشگورین مراجع (Horn and Johnson (1985) و Brualdi and Mallendorf(1994) را ملاحظه کنید.

مثال ۲-۴-۸

$$A = \begin{bmatrix} 1 & 0/1 & 0/2 \\ 0/2 & 4 & 0/3 \\ 0/4 & 0/5 & 8 \end{bmatrix}$$

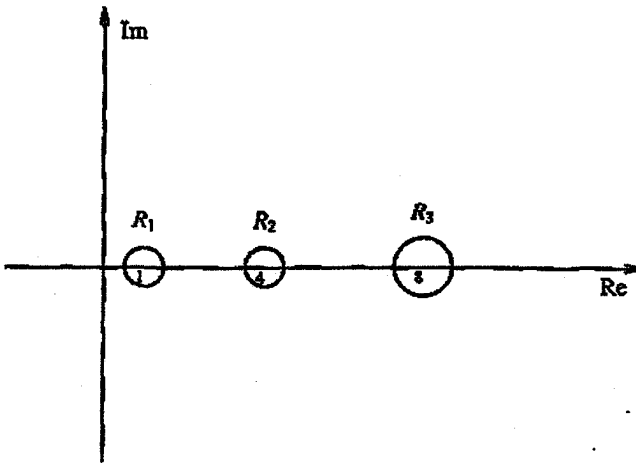
قرصهای گرشگورین عبارتند از:

$$R_1: \{z: |z - 1| \leq 0/3\}$$

$$R_2: \{z: |z - 4| \leq 0/5\}$$

$$R_3: \{z: |z - 8| \leq 0/9\}$$

همه سه قرص مجزا از یکدیگرند، بنابراین بر طبق قضیه ۸-۴-۲ هر قرص باید دقیقاً یک مقدار ویژه در برداشته باشد این در حقیقت درست است. توجه کنید که مقادیر ویژه عبارتند از: $۳/۹۶۷۱$ ، $۰/۹۸۳۴$ و $۸/۰۴۹۵$ را ملاحظه کنید.



شکل ۸-۸ قرصهای گردشگورین: مثال ۸-۴-۲

۸-۴-۲ کرانه‌های مقدار ویژه و نرم‌های ماتریسی

گاهی اوقات بعضی نرم‌های ماتریسی ساده می‌توانند برای به دست آوردن کرانه‌های مفیدی برای مقادیر ویژه به صورتی که قضیه‌های زیر نشان می‌دهند، مورد استفاده قرار گیرند. قضیه ۸-۴-۳ بیان دوباره قضیه ۱-۷-۵ است.

قضیه ۸-۴-۳ برای هر جفت نرم ماتریس - بردار سازگار، داریم

$$|\lambda| \leq \|A\|$$

که در آن λ یک مقدار ویژه A است. بویژه $\rho(A)$ ، شعاع طیفی A ، توسط $\|A\|$ کراندار می‌شود:

$$\rho(A) \leq \|A\|$$

تذکره: در مثال ۸-۴-۱ داریم $\|A\|_\infty = ۱۶$ و $\rho(A) = ۷/۳۰۶۷$

چون مقادیر ویژه A^T همان مقادیر ویژه A هستند، داریم

نتیجه ۸-۴-۱

$$\rho(A) \leq \|A^T\|$$

از ترکیب این دو نتیجه با هم و بویژه اختیار کردن نرم بی نهایت، قضیه ۸-۴-۴ را به دست می آوریم.

قضیه ۸-۴-۴

$$\rho(A) \leq \min\left\{\max_i \sum_{j=1}^n |a_{ij}|, \max_j \sum_{i=1}^n |a_{ij}|\right\}$$

قضیه ۸-۴-۵ فرض کنید $\lambda_1, \lambda_2, \dots, \lambda_n$ مقادیر ویژه A باشند؛ آنگاه

$$\sum_{i=1}^n |\lambda_i|^2 \leq \|A\|_F^2$$

اثبات: قضیه مثلثی سازی شور (قضیه ۸-۲-۶) بیان می کند که یک ماتریس یکانی مانند U وجود دارد به قسمی که

$$U^*AU = T,$$

T یک ماتریس بالا مثلثی

بنابراین

$$TT^* = U^*AA^*U$$

بنابراین AA^* با TT^* به طور یکانی متشابه است. چون ماتریسهای متشابه دارای اثرهای یکسان هستند، پس

$$\text{tr}(T^*T) = \text{tr}(AA^*) = \|A\|_F^2$$

بعلاوه $\text{tr}(T^*T) = \sum_{i=1}^n \sum_{j=1}^n |t_{ij}|^2$. بنابراین

$$\sum_{i=1}^n \sum_{j=1}^n |t_{ij}|^2 = \|A\|_F^2$$

همچنین

$$\sum_{i=1}^n |t_{ii}|^2 = \sum_{i=1}^n |\lambda_i|^2$$

□

$$\sum_{i=1}^n |\lambda_i|^2 = \sum_{i=1}^n |t_{ii}|^2 \leq \sum_{i=1}^n \sum_{j=1}^n |t_{ij}|^2 = \|A\|_F^2$$

بنابراین

۸-۴-۳ اینرسی یک ماتریس

در بخش ۸-۳-۱ ملاحظه نمودیم که پایداری یک دستگاه فیزیکی مدل‌بندی شده توسط یک دستگاه معادلات دیفرانسیل دقیقاً با شناخت این که آیا مقادیر ویژه ماتریس دستگاه A دارای قسمتهای حقیقی منفی هستند، تعیین می‌شود. یک مسأله کلی‌تر از مسأله پایداری مسأله اینرسی است.

اینرسی یک ماتریس مختلط A از مرتبه $n \times n$ نسبت به نیم‌صفحه‌ها به صورت سه‌تایی صحیح $In(A) = (\pi(A), \nu(A), \delta(A))$ تعریف می‌شود، که در آن $\pi(A)$ تعداد مقادیر ویژه A با قسمتهای حقیقی مثبت، $\nu(A)$ تعداد مقادیر ویژه A با قسمتهای حقیقی منفی، و $\delta(A)$ تعداد مقادیر ویژه A با قسمتهای حقیقی صفر با در نظر گرفتن تعداد تکرارها می‌باشند.

بنابراین A یک ماتریس پایدار است اگر $In(A) = (0, n, 0)$. یک نتیجه کلاسیک در مورد اینرسی یک ماتریس هرمیتی A قانون اینرسی سیلوستر^۱ است.

قضیه ۸-۴-۶ قضیه قانون اینرسی سیلوستر. فرض کنید A یک ماتریس هرمیتی و P یک ماتریس نامنفرد باشد. آنگاه

$$In(A) = In(PAP^*)$$

برای اثبات مرجع (Horn and Johnson (1985, 223 - 224) را ملاحظه کنید.

نوشتجات زیادی در مورد اینرسی یک ماتریس ناهرمیتی وجود دارد.

قضیه‌های اینرسی ناهرمیتی سعی بر این دارند که شرایطی را تعیین کنند که تحت آنها اینرسی یک ماتریس ناهرمیتی بتواند برابر با اینرسی یک ماتریس هرمیتی ساخته شود. بحث نظری اینرسی یک ماتریس ناهرمیتی خارج از بحث این کتاب است. برای شرحی از نظریه اینرسی مرجع (Datta(1986) و مقاله اینرسی که در آینده نزدیک از Datta چاپ خواهد شد را ملاحظه کنید.

اهمیت قانون اینرسی سیلوستر در این حقیقت است که به ما می‌گوید که اینرسی یک ماتریس هرمیتی A برابر اینرسی یک ماتریس قطری D است که می‌تواند توسط تجزیه مثلثی A به دست آید (بحث زیر را ملاحظه کنید).

محاسبه اینرسی یک ماتریس متقارن (یا یک ماتریس هرمیتی مختلط)

اگر A متقارن (یا هرمیتی مختلط) باشد، آنگاه قانون سیلوستر اینرسی یک روش مؤثر و ارزان برای محاسبه اینرسی فراهم می‌سازد.

از بخش ۶-۴-۷ یادآور می‌شویم که یک ماتریس متقارن یک تجزیه مثلثی به شکل زیر می‌پذیرد:

$$A = UDU^T$$

که در آن U حاصلضرب ماتریسهای جایگشت و ماتریسهای بالا مثلثی واحد مقدماتی است و D یک ماتریس قطری بلوکی با بلوکهای از مرتبه ۱ و ۲ می‌باشد. این تجزیه به تجزیه با محورگیری قطری معروف است (فصل ۶ و مقاله‌های مربوط به آن راز (Bunch (1971)، Bunch and parlett (1971) و (Bunch and Kaufman (1977) ملاحظه کنید). بنابراین بر طبق قانون سیلوستر اینرسی $In(A) = In(D)$.

فرض کنید D دارای p بلوک از مرتبه ۱ و q بلوک از مرتبه ۲، با $p + 2q = n$ ، باشد. فرض کنید که هیچ یک از بلوکهای 2×2 ماتریس D منفرد نباشند. فرض کنید که از p بلوک مرتبه ۱، p' تا از آنها مثبت باشند، p'' تا از آنها منفی باشند، و p''' تا از آنها صفر باشند (یعنی $p' + p'' + p''' = p$). آنگاه

$$\pi(A) = p' + q$$

$$\nu(A) = p'' + q$$

$$\delta(A) = p'''$$

تجزیه با محورگیری قطری می‌تواند به یک طریق پایدار عددی انجام شود. و این تجزیه نیاز به $n^{3/6}$ عمل ممیز شناور دارد.

محاسبه اینرسی یک ماتریس ناهرمیتی

به طور کلی آنالیز عددی دانان معتقد هستند که اینرسی یک ماتریس ناهرمیتی A بهتر توسط محاسبه صریح مقادیر ویژه A تعیین می‌شود.

با این وجود، اجازه دهید متذکر شویم که یک روش محاسباتی که اثر کارلسون و داتا (Carlson and datta (1977) می‌باشد، برای تعیین اینرسی یک ماتریس ناهرمیتی وجود دارد. این روش نیاز به محاسبه مقادیر ویژه و یا حل یک معادله ماتریسی ندارد. روش بعد از این که ماتریس A به یک ماتریس بالا هسنبرگی تبدیل شد، فقط به $n^{3/3}$ عمل ممیز شناور نیاز دارد.

روش اینرسی کارلسون و داتا مبتنی بر جواب ضمنی یک معادله ماتریسی خاص است: با یک ماتریس ناهرمیتی A شروع می‌شود، روش یک ماتریس هرمیتی X به قسمی می‌سازد که هرگاه X نامنفرد باشد $In(A) = In(X)$ می‌باشد. ماتریس X که باید محاسبه شود جواب یک معادله ماتریسی است، اما لازم نیست یک معادله ماتریسی به طور صریح حل شود. البته به محض آن که ماتریس

هرمیتی X ساخته شود، اینرسی آن می‌تواند با استفاده از قانون سیلوستر اینرسی توصیف شده در بخش قبل حساب شود.

خاصیتهای گرد کردن عددی روش به طور کامل بررسی نشده است. به عنوان یک گام اولیه روش نیاز به تبدیل A به یک ماتریس هسنبرگی H دارد، و به نظر می‌رسد که اگر یک یا بیش از یک عنصر از عناصر زیر قطر اصلی کوچک یا نزدیک به صفر باشند، خطای گرد کردن بزرگی در ساختن X می‌تواند مورد انتظار باشد. روش توسط داتا و داتا (1987) دوباره فرمول‌بندی شده است، در نتیجه هیچ تقسیمی بر عناصر قطری مورد نیاز نیست. این روش در عمل حتی با عناصر زیر قطر اصلی کوچک، خیلی خوب کار می‌کند (به همان خوبی که روش کارلسون - داتا کار می‌کند)، اما این مطلب توسط تحلیل جدی از خطای گرد کردن هر یک از روشها ثابت نشده است.

۸-۵ محاسبه مقادیر ویژه و بردارهای ویژه انتخابی

۸-۵-۱ بحثها در مورد اهمیت بزرگ‌ترین و کوچک‌ترین مقادیر ویژه

قبلاً ملاحظه نمودیم که در کاربردهای متعددی ما فقط به محاسبه یک تعداد کمی از کوچک‌ترین مقادیر ویژه و بزرگ‌ترین مقادیر ویژه نیاز داریم. برای مثال، یادآور می‌شویم که در مسأله کمانش کوچک‌ترین مقدار ویژه است که مهمترین مقدار ویژه می‌باشد.

در تحلیل ارتعاش سازه‌ها، یک عمل معمول در مهندسی محاسبه تعداد کمی از کوچک‌ترین مقادیر ویژه (فرکانسها) و بردارهای ویژه متناظر (مدها) می‌باشد، زیرا در عمل ملاحظه شده است که مقادیر ویژه بزرگ‌تر و بردارهای ویژه آنها سهم خیلی کمی در پاسخ کلی دستگاه دارند. همین اظهارات در حالت مسائل کنترل مدل شده توسط یک دستگاه معادلات دیفرانسیل مرتبه دو و به وجودآمده در عنصر متناهی - مدل تولید کاهش مرتبه ساختارهای فضایی قابل انعطاف بزرگ پا برجا هستند (Inman (1989)).

در کاربردهای آماری نظیر آنالیز مؤلفه‌های اصلی فقط تعداد کمی از بزرگ‌ترین مقادیر ویژه محاسبه می‌شوند. کاربردهای دیگری نیز وجود دارند که در آنها فقط مقادیر ویژه غالب و زیر غالب و بردارهای ویژه متناظر با آنها یک نقش مهم بازی می‌کنند.

۸-۵-۲ نقش مقادیر ویژه غالب و بردارهای ویژه غالب در دستگاههای پویا

اجازه دهید به طور خلاصه در مورد نقش مقادیر ویژه غالب و بردارهای ویژه در زمینه دستگاههای پویا بحث کنیم.

دستگاه باگستگی زمانی همگن زیر را در نظر بگیرید

$$x_{k+1} = Ax_k, \quad k = 0, 1, 2, \dots$$

فرض کنید λ_1 مقدار ویژه غالب A باشد؛ یعنی $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$ ، که در آن $\lambda_1, \lambda_2, \dots, \lambda_n$ مقادیر ویژه A هستند. فرض کنید که A دارای یک مجموعه بردار مستقل خطی v_1, v_2, \dots, v_n باشد. آنگاه حالت x_k در هر زمان $k > 0$ به صورت

$$x_k = \alpha_1 \lambda_1^k v_1 + \alpha_2 \lambda_2^k v_2 + \dots + \alpha_n \lambda_n^k v_n$$

ارائه می‌شود، که در آن $x_0 = \alpha_1 v_1 + \dots + \alpha_n v_n$. چون $|\lambda_1|^k > |\lambda_i|^k$ ، $i = 1, 2, \dots, n$ ، برای مقادیر بزرگ k رابطه

$$\|\alpha_1 \lambda_1^k\| \gg \|\alpha_i \lambda_i^k\|, \quad i = 2, 3, \dots, n$$

نتیجه می‌شود به شرط آن که $\alpha_1 \neq 0$ باشد. این بدین معنی است که برای مقادیر بزرگ k ، بردار حالت x_k به جهت بردار v_1 متناظر با مقدار ویژه غالب λ_1 نزدیک خواهد شد. بعلاوه سرعتی که در آن بردار حالت به v_1 نزدیک می‌شود توسط نسبت دومین مقدار ویژه به اولین مقدار ویژه، $|\lambda_2/\lambda_1|$ تعیین می‌گردد. (یک اثبات بعداً ارائه می‌شود).

در حالت $\alpha_1 = 0$ ، مقدار ویژه غالب دوم λ_2 و بردار متناظر نقش اولین جفت ویژه غالب را بازی می‌کنند. برای دستگاه با زمان پیوسته

$$\dot{x}(t) = Ax(t)$$

نیز به طور مشابه می‌توان نتیجه‌گیری کرد.

برای جزئیات بیشتر مرجع (Luenberger (1979 را ملاحظه کنید.

به طور خلاصه رفتار دراز مدت یک دستگاه پویای همگن را می‌توان اساساً فقط از مقادیر ویژه غالب اول و دوم ماتریس دستگاه و بردارهای ویژه متناظر پیش‌بینی کرد.

بویژه جفت ویژه غالب دوم در حالت $\alpha_1 = 0$ اهمیت دارد. در این حالت رفتار دراز مدت دستگاه توسط این جفت تعیین می‌شود.

یک مثال: مطالعه جمعیت

اجازه دهید حالت یک دستگاه جمعیت را برای توضیح این مطلب در نظر بگیریم. بخوبی می دانیم (Luenberger (1979, 170)) که یک چنین دستگاهی می تواند توسط رابطه

$$P_{k+1} = AP_k, \quad k = 0, 1, 2, \dots$$

فرمول بندی شود که در آن بردار جمعیت می باشد. اگر λ_1 مقدار ویژه غالب ماتریس A کوچک تر از ۱ باشد (یعنی اگر $|\lambda_1| < 1$)، آنگاه از

$$P_k = \alpha_1 \lambda_1^k v_1 + \dots + \alpha_n \lambda_n^k v_n$$

نتیجه می شود که جمعیت به صفر کاهش می یابد هنگامی که k بزرگ می شود. به طور مشابه اگر $|\lambda_1| > 1$ باشد، آنگاه رشد دراز مدت در جمعیت وجود دارد. در حالت اخیر جمعیت اصلی به یک توزیع نهایی میل می کند که توسط بردار ویژه مقدار ویژه غالب تعریف می شود. بعلاوه مقدار ویژه غالب دوم A است که تعیین می کند که با چه سرعتی توزیع جمعیت اصلی به توزیع نهایی میل می کند. بالاخره اگر مقدار ویژه غالب برابر ۱ باشد، آنگاه در دراز مدت در جمعیت نه رشدی و نه زوالی وجود دارد.

۸-۵-۳ روش توانی، تکرار معکوس، تکرار خارج قسمت ریلی

در این بخش به طور خلاصه دو روش کلاسیک معروف را برای پیدا کردن مقادیر ویژه غالب و بردارهای متناظر یک ماتریس شرح می دهیم. روشها بویژه برای ماتریسهای تنک مناسب هستند، زیرا آنها فقط به ضربهای ماتریس بردار متکی هستند.

روش توانی

روش توانی اغلب برای پیدا کردن مقدار ویژه غالب و بردار ویژه متناظر یک ماتریس استفاده می شود. این روش، روش توانی نامیده می شود زیرا مبتنی بر ساخت ضمنی توانهای A می باشد.

فرض کنید مقادیر ویژه $\lambda_1, \lambda_2, \dots, \lambda_n$ ماتریس A به قسمی باشند که

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$$

یعنی λ_1 مقدار ویژه غالب A باشد. فرض کنید v_1 بردار ویژه متناظر باشد. فرض کنید $\max(g)$ عنصر با بزرگترین قدر مطلق بردار g را نمایش دهد. بعلاوه فرض کنید که A قابل قطری شدن باشد.

الگوریتم ۸-۵-۱ روش توانی

Step 1 : Choose x_0

Step 2 : For $k = 1, 2, 3, \dots$ do until convergence

$$\hat{x}_k = Ax_{k-1}$$

$$x_k = \hat{x}_k / \max(\hat{x}_k)$$

قضیه ۸-۵-۱ $\max\{\hat{x}_k\} \rightarrow \lambda_1$ و $\{x_k\} \rightarrow w_1$ ، مضرری از بردار v_1 است، هنگامی که $k \rightarrow \infty$.

اثبات: از قبل داریم

$$x_k = \frac{A^k x_0}{\max(A^k x_0)}$$

چون A قابل قطری شدن است، بردار v_1 تا v_n وابسته به $\lambda_1, \lambda_2, \dots, \lambda_n$ می‌توانند طوری انتخاب شوند که مستقل خطی باشند. پس می‌توانیم بردار x_1 را به صورت $x_0 = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$ ، $\alpha_1 \neq 0$ بنویسیم. در نتیجه

$$A^k x_0 = A^k (\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n)$$

$$= \alpha_1 \lambda_1^k v_1 + \alpha_2 \lambda_2^k v_2 + \dots + \alpha_n \lambda_n^k v_n$$

$$= \lambda_1^k \left[\alpha_1 v_1 + \alpha_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k + \dots + \alpha_n \left(\frac{\lambda_n}{\lambda_1}\right)^k v_n \right]$$

چون λ_1 مقدار ویژه غالب A است، به ازای $i = 2, 3, \dots, n$ داریم

$$\left(\frac{\lambda_i}{\lambda_1}\right)^k \rightarrow 0 \quad k \rightarrow \infty$$

بنابراین

$$x_k = \frac{A^k x_0}{\max(A^k x_0)} \rightarrow c v_1$$

که در آن c یک ثابت است و

□

$$\{\max\{\hat{x}_k\}\} \rightarrow \lambda_1$$

تبصره‌ها: ما روش توانی را تحت دو قید زیر به دست آوردیم

۱. $\alpha_1 \neq 0$

۲. λ_1 تنها مقدار ویژه غالب است.

اولین قید ($\alpha_1 \neq 0$) یک قید عملی جدی نیست، زیرا بعد از یک تعداد کمی تکرار خطاهای گرد کردن تقریباً همیشه باعث رخ دادن آن خواهند شد.

در مورد آنچه که به قید دوم مربوط می‌شود، ملاحظه می‌کنیم که روش باز هم هنگامی که ماتریس A بیش از یک مقدار ویژه غالب دارد، همگرا می‌گردد. برای مثال فرض کنید $|\lambda_1| = |\lambda_2| = \dots = |\lambda_r|$ و $|\lambda_1| > |\lambda_{r+1}| \geq \dots \geq |\lambda_n|$ ، و فرض کنید که بردارهای ویژه وابسته به λ_1 مستقل خطی باشند. آنگاه داریم

$$A^k x_0 = \lambda_1^k \left[\sum_{i=1}^r \alpha_i v_i + \sum_{i=r+1}^n \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k v_i \right]$$

$$\cong \lambda_1^k \sum_{i=1}^r \alpha_i v_i$$

(زیرا $(\lambda_i/\lambda_1)^k$ برای مقادیر بزرگ k کوچک است). این نشان می‌دهد که در این حالت روش روش توانی به یک بردار در زیر فضای پدید آمده توسط v_1, \dots, v_r همگرا می‌گردد.

مثال ۸-۵-۱

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

$$x_0 = (1, 1, 1)^T$$

مقادیر ویژه A عبارتند از $0, 9/6235, -0/6235$ ، و بردار ویژه نرمال شده متناظر با بزرگ‌ترین مقدار ویژه $9/6235$ به صورت زیر است

$$(0/3851, 0/5595, 0/7339)^T$$

$$: k = 1$$

$$\hat{x}_1 = Ax_0 = \begin{bmatrix} 6 \\ 9 \\ 12 \end{bmatrix}$$

$$\max(\hat{x}_1) = 12$$

$$x_1 = \frac{\hat{x}_1}{\max(\hat{x}_1)} = \begin{bmatrix} 1/2 \\ 3/4 \\ 1 \end{bmatrix} = \begin{bmatrix} 0/50 \\ 0/75 \\ 1 \end{bmatrix}$$

$k = 2$

$$\hat{x}_2 = Ax_1 = \begin{bmatrix} 5/0.0 \\ 7/25 \\ 9/5.0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 0.5263 \\ 0.7632 \\ 1/0.000 \end{bmatrix}$$

$\max(\hat{x}_2) = 9/5.0$

$k = 3$

$$\hat{x}_3 = Ax_2 = \begin{bmatrix} 5/0.526 \\ 7/3421 \\ 9/6316 \end{bmatrix}$$

$$x_3 = \frac{\hat{x}_3}{\max(\hat{x}_3)}$$

$$= \begin{bmatrix} 0.5246 \\ 0.7623 \\ 1/0.000 \end{bmatrix}$$

$\max(\hat{x}_3) = 9/6316$

بنابراین $\{\max(\hat{x}_3)\}$ به سمت بزرگ‌ترین مقدار ویژه $9/6235$ میل می‌کند و $\{x_k\}$ به سمت جهت بردار ویژه وابسته به این مقدار ویژه میل می‌کند. (توجه کنید که بردار ویژه غالب نرمال شده زیر مضرری از x_3 است

$$\begin{bmatrix} 0.3851 \\ 0.5595 \\ 0.7339 \end{bmatrix}$$

تذکر MATCOM: الگوریتم ۸-۵-۱ در برنامه POWER در MATCOM پیاده‌سازی شده است.

همگرایی روش توانی

همان‌گونه که به سادگی از مطالب زیر مشاهده می‌شود نرخ همگرایی روش توانی توسط نسبت $|\lambda_2/\lambda_1|$ تعیین می‌شود. از اثبات روش توانی (الگوریتم ۸-۵-۱) داریم

$$\begin{aligned} \|x_k - \alpha_1 v_1\| &= \left\| \alpha_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \dots + \alpha_n \left(\frac{\lambda_n}{\lambda_1}\right)^k v_n \right\| \\ &\leq |\alpha_2| \left| \frac{\lambda_2}{\lambda_1} \right|^k \|v_2\| + \dots + |\alpha_n| \left| \frac{\lambda_n}{\lambda_1} \right|^k \|v_n\| \\ &\leq \left| \frac{\lambda_2}{\lambda_1} \right|^k (|\alpha_2| \|v_2\| + \dots + |\alpha_n| \|v_n\|) \end{aligned}$$

(زیرا $i = 3, 4, \dots, n$ ، $|\lambda_i/\lambda_1| \leq |\lambda_2/\lambda_1|$)

بنابراین داریم

$$\|x_k - \alpha_1 v_1\| \leq \alpha \left(\frac{\lambda_2}{\lambda_1}\right)^k, \quad k = 1, 2, \dots$$

که در آن

$$\alpha = (|\alpha_2| \|v_2\| + \dots + |\alpha_n| \|v_n\|)$$

این نشان می‌دهد که نرخ که در آن x_k به $\alpha_1 v_1$ میل می‌کند بستگی به این دارد که با چه سرعتی $|\lambda_2/\lambda_1|^k$ به صفر میل نماید. قدر مطلق خطا در هر گام با نرخ $|\lambda_2/\lambda_1|$ کاهش می‌یابد، یعنی اگر λ_2 به λ_1 نزدیک باشد، آنگاه همگرایی خیلی آهسته خواهد بود. اگر این نسبت کوچک باشد، همگرایی سریع خواهد بود.

روش توانی با انتقال

در برخی حالتها همگرایی را می‌توان به طور بامعنایی توسط یک انتقال مناسب بهبود بخشید. بنابراین اگر σ یک انتقال مناسب باشد به طوری که $\sigma - \lambda_1$ مقدار ویژه غالب $A - \sigma I$ باشد و اگر روش توانی در مورد ماتریس انتقال یافته $A - \sigma I$ اعمال شود، آنگاه نرخ همگرایی توسط نسبت

$$\left| \frac{\lambda_2}{\lambda_1} \right| \quad \text{به جای} \quad \left| \frac{\lambda_2 - \sigma}{\lambda_1 - \sigma} \right|$$

تعیین خواهد شد. (توجه کنید که با انتقال ماتریس A توسط σ ، مقادیر ویژه توسط σ انتقال می‌یابند، اما بردارهای ویژه بدون انتقال باقی می‌مانند.)

در برخی حالتها با انتخاب σ به طور مناسب، نسبت

$$\left| \frac{\lambda_2 - \sigma}{\lambda_1 - \sigma} \right|$$

می‌تواند به طور بامعنایی کوچک‌تر از $|\lambda_2/\lambda_1|$ باشد، در نتیجه همگرایی سریعتر حاصل می‌شود. یک انتخاب بهینه ((Wilkinson (1965, 572) از σ ، با فرض این که λ_2 ها، همگی حقیقی هستند، $\frac{1}{4}(\lambda_2 + \lambda_{nn})$ است. این انتخاب ساده σ گاهی اوقات همگرایی خیلی سریعی نتیجه می‌دهد، اما مثالهای معمولی زیادی وجود دارند که برای آنها باز هم همگرایی می‌تواند با این انتخاب σ آهسته باشد.

یک ماتریس با مرتبه 20×20 با مقادیر ویژه $19, 20, \dots, 1$ در نظر بگیرید. انتخاب

$$\sigma = (19 + 1)/2 = 10$$

$$\left| \frac{\lambda_2 - \sigma}{\lambda_1 - \sigma} \right| = \frac{9}{10}$$

را نتیجه می‌دهد که باز هم نزدیک به ۱ است. بنابراین نرخ همگرایی باز هم آهسته می‌باشد. بعلاوه در عمل انتخاب فوق برای σ مناسب نیست، زیرا مقادیر ویژه از قبل معلوم نیستند.

روش توانی معکوس / تکرار معکوس

روش تکراری زیر، معروف به تکرار معکوس، یک روش مؤثر برای محاسبه یک بردار ویژه است هنگامی که یک تقریب نسبتاً خوب برای یک مقدار ویژه معلوم می‌باشد.

الگوریتم ۸-۵-۲ تکرار معکوس. فرض کنید σ تقریبی برای یک مقدار ویژه λ_1 باشد به قسمی که $|\lambda_i - \sigma| \ll |\lambda_1 - \sigma|$ ، $(i \neq 1)$ ؛ یعنی σ به λ_1 خیلی نزدیک‌تر از سایر مقادیر ویژه است.

گام ۱: x_0 را انتخاب کنید.

گام ۲: برای $k = 1, 2, 3, \dots$ انجام دهید.

دستگاه $(A - \sigma I)\hat{x}_k = x_{k-1}$ را نسبت به \hat{x}_k با استفاده از روش حذفی گاوس با محورگیری حل کنید.

قرار دهید $x_k = \hat{x}_k / \max(\hat{x}_k)$

متوقف شوید اگر $\|A\| c\mu < \|(A - \sigma I)x_k\|_\infty$ ، که در آن c یک ثابت از مرتبه واحد است.

تبصره: توجه کنید که استفاده از معیار توقف فوق جفت (σ, x_k) را یک جفت ویژه واقعی از یک ماتریس نزدیک خواهد ساخت.

قضیه ۸-۵-۲ دنباله $\{x_k\}$ به جهت بردار ویژه متناظر با λ_1 همگرا می‌گردد.

تبصره: توجه کنید که تکرار معکوس به طور ساده روش توانی است که در مورد $(A - \sigma I)^{-1}$ اعمال می‌شود، و بدین علت به روش توانی معکوس معروف است.

اثبات: مقادیر ویژه $(A - \sigma I)^{-1}$ عبارتند از $(\lambda_1 - \sigma)^{-1}$ ، $(\lambda_2 - \sigma)^{-1}$ ، $(\lambda_3 - \sigma)^{-1}$ ، ... و بردارهای ویژه v_1 ، v_2 ، v_3 ، ... همان بردارهای ویژه A هستند. بنابراین همانند حالت روش توانی

می‌توانیم بنویسیم

$$\begin{aligned} \hat{x}_k &= \frac{c_1}{(\lambda_1 - \sigma)^k} v_1 + \frac{c_2}{(\lambda_2 - \sigma)^k} v_2 + \dots + \frac{c_n}{(\lambda_n - \sigma)^k} v_n \\ &= \frac{1}{(\lambda_1 - \sigma)^k} \left[c_1 v_1 + c_2 \left[\frac{\lambda_1 - \sigma}{\lambda_2 - \sigma} \right]^k v_2 + \dots + c_n \left[\frac{\lambda_1 - \sigma}{\lambda_n - \sigma} \right]^k v_n \right] \end{aligned}$$

چون λ_1 به σ نزدیک تر از هر مقدار ویژه دیگری است، جمله اول در سمت راست، جمله غالب است و بنابراین $\{x_k\}$ به v_1 همگرا می‌گردد. جهت v_1 است که سعی داریم آن را محاسبه نماییم. □

توضیح

اجازه دهید این ایده را با $k = 1$ توضیح دهیم. فرض کنید $x_0 = c_1 v_1 + c_2 v_2 + \dots + c_n v_n$. آنگاه

$$\hat{x}_1 = (A - \sigma I)^{-1} x_0 = (\lambda_1 - \sigma)^{-1} c_1 v_1 + (\lambda_2 - \sigma)^{-1} c_2 v_2 + \dots + (\lambda_n - \sigma)^{-1} c_n v_n$$

چون λ_1 به σ نزدیک تر از هر مقدار ویژه دیگر است، ضریب جمله اول در بسط، $1/(\lambda_1 - \sigma)$ ، جمله غالب است (بزرگ‌ترین است). بنابراین \hat{x}_1 به‌طور تقریبی مضربی از v_1 است، و این چیزی است که می‌خواهیم.

پایداری عددی تکرار معکوس

در نگاه اول تکرار معکوس یک روش خطرناک به نظر می‌رسد، زیرا اگر σ به λ_1 نزدیک باشد ماتریس $(A - \sigma I)$ به طور واضح بدوضع است. در نتیجه این بدوضعی باید بر تقریبهای محاسبه شده بردار ویژه تأثیر بگذارد. خوشبختانه در عمل بدوضعی ماتریس $(A - \sigma I)$ دقیقاً همان چیزی است که ما می‌خواهیم. خطا در هر تکرار به سمت جهت بردار ویژه رشد می‌کند و این جهت بردار ویژه است که مورد نظرمان می‌باشد.

ویلیکینسون متذکر شده است که در عمل \hat{x}_k به طور قابل ملاحظه‌ای نزدیک به جواب معادله

$$(A - \sigma I + F)x_k = x_{k-1}$$

است که در آن F کوچک است. برای جزئیات مرجع (Wilkinson (1965, 620 - 621 را ملاحظه کنید.

مثال ۸-۵-۲

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

$$x_0 = (1, 1, 1)^T, \quad \sigma = 9$$

:k = ۱

$$\hat{x}_1 = (1, 1/5, 2)^T$$

$$x_1 = \frac{\hat{x}_1}{\|\hat{x}_1\|_2} = (0.3714, 0.5571, 0.7428)^T$$

:k = ۲

$$\hat{x}_2 = (0.6190, 0.8975, 1/1761)^T$$

$$x_2 = \frac{\hat{x}_2}{\|\hat{x}_2\|_2} = (0.3860, 0.5597, 0.7334)^T$$

:k = ۳

$$\hat{x}_3 = (0.6176, 0.8974, 1/1772)^T$$

$$x_3 = \frac{\hat{x}_3}{\|\hat{x}_3\|_2} = (0.3850, 0.5595, 0.7340)^T$$

:k = ۴

$$\hat{x}_4 = (0.6176, 0.8974, 1/1772)^T$$

$$x_4 = \frac{\hat{x}_4}{\|\hat{x}_4\|_2} = (0.3850, 0.5595, 0.7340)^T$$

:k = ۵

$$\hat{x}_5 = (0.6177, 0.8974, 1/1772)^T$$

$$x_5 = \frac{\hat{x}_5}{\|\hat{x}_5\|_2} = (0.3851, 0.5595, 0.7339)^T$$

تبصره: در مثال قبل از نرم $\| \cdot \|_2$ به عنوان یک مقیاس بندی برای بردار x استفاده کرده ایم تا تأکید کنیم که مقیاس بندی بی اهمیت است، زیرا ما در جهت بردار ویژه کار می کنیم.

تذکر MATCOM: الگوریتم ۸-۵-۲ در برنامه INVITR در MATCOM پیاده سازی شده است.

انتخاب بردار اولیه x

برای انتخاب بردار اولیه x می توانیم یک تعداد تکرار از روش توانی را اجرا کنیم و سپس به تکرار

معکوس با آخرین بردار تولید شده توسط روش توانی به عنوان بردار اولیه در تکرار معکوس، پردازیم. ویلکینسون (Wilkinson(1965, 627)) بیان کرده است که اگر فرض شود که x_0 به قسمی است که

$$Le = Px_0$$

که در آن P یک ماتریس جایگشت است که در رابطه

$$P(A - \sigma I) = LU$$

صدق می‌کند و e یک بردار با عناصر ۱ می‌باشد، آنگاه دو تکرار معمولاً کافی است، به شرط آن که σ یک تقریب خوب برای مقدار ویژه باشد.

تذکره: اگر x_0 به صورت توصیف شده انتخاب شود، آنگاه محاسبه \bar{x}_1 فقط شامل حل دستگاه مثلی زیر خواهد بود.

$$U\bar{x}_1 = e$$

مثال ۳-۵-۸

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

مقادیر ویژه عبارتند از

$$9/6235 \quad \text{و} \quad -0/6235$$

انتخاب می‌کنیم $\sigma = 9/1$

$$A - \sigma I = \begin{bmatrix} -8/1 & 2/0 & 3/0 \\ 2/0 & -6/1 & 4/0 \\ 3/0 & 4/0 & -4/1 \end{bmatrix}$$

$$L = \begin{bmatrix} 1 & \cdot & \cdot \\ -0/2469 & 1 & \cdot \\ -0/3704 & -0/8456 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} -8/1 & 2 & 3 \\ \cdot & -5/6062 & 4/7407 \\ \cdot & \cdot & 1/0200 \end{bmatrix}$$

$$P = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \cdot & 1 \end{bmatrix}$$

$$x_0 = P^{-1}Le = \begin{bmatrix} 1/000 \\ 0/7531 \\ -0/2160 \end{bmatrix}$$

$$:k = 1$$

$$\hat{x}_1 = (A - \sigma I)^{-1}x_0 = \begin{bmatrix} 0/4003 \\ 0/6507 \\ 0/9804 \end{bmatrix}$$

$$x_1 = \frac{\hat{x}_1}{\max(\hat{x}_1)} = \begin{bmatrix} 0/4083 \\ 0/6637 \\ 1/000 \end{bmatrix}$$

$$:k = 2$$

$$\hat{x}_2 = (A - \sigma I)^{-1}x_1 = \begin{bmatrix} 0/9367 \\ 1/3540 \\ 1/7624 \end{bmatrix}$$

$$x_2 = \begin{bmatrix} 0/5315 \\ 0/7683 \\ 1/000 \end{bmatrix}$$

که تقریباً $1/3$ برابر بردار ویژه نرمال شده است.

بردار ویژه نرمال شده با چهار رقم درست عبارت است از

$$\begin{bmatrix} 0/3851 \\ 0/5595 \\ 0/7339 \end{bmatrix}$$

خارج قسمت ریلی

قضیه ۸-۵-۳ فرض کنید A یک ماتریس متقارن باشد و فرض کنید $x \in \mathbb{R}^n$ یک تقریب خوب

معقول برای یک بردار ویژه باشد. آنگاه خارج قسمت

$$R_q = \sigma = \frac{x^T A x}{x^T x}$$

یک تقریب خوب برای مقدار ویژه λ می خواهد بود که x بردار ویژه متناظرش می باشد.

اثبات: چون A متقارن است، بر طبق قضیه ۸-۲-۸ یک مجموعه بردار ویژه یکا متعامد v_1, v_2, \dots, v_n ،

v_n وجود دارد، در نتیجه می توانیم بنویسیم

$$x = c_1 v_1 + \dots + c_n v_n$$

مجدداً، چون بردارهای v_i ، $i = 1, \dots, n$ نرمال هستند، پس $v_i^T v_i = 1$ بعلاوه چون $Av_i = \lambda v_i$ ،

$i = 1, \dots, n$ ، با توجه به $v_i^T v_j = 0$ ، $i \neq j$ ، داریم:

$$\begin{aligned} \sigma &= \frac{x^T A x}{x^T x} = \frac{(c_1 v_1 + \dots + c_n v_n)^T A (c_1 v_1 + \dots + c_n v_n)}{(c_1 v_1 + \dots + c_n v_n)^T (c_1 v_1 + \dots + c_n v_n)} \\ &= \frac{(c_1 v_1 + \dots + c_n v_n)^T (c_1 \lambda_1 v_1 + \dots + c_n \lambda_n v_n)}{c_1^2 + c_2^2 + \dots + c_n^2} \\ &= \frac{\lambda_1 c_1^2 + \lambda_2 c_2^2 + \dots + \lambda_n c_n^2}{c_1^2 + c_2^2 + \dots + c_n^2} \\ &= \lambda_1 \left[\frac{1 + \left(\frac{\lambda_2}{\lambda_1}\right) \left(\frac{c_2}{c_1}\right)^2 + \dots + \left(\frac{\lambda_n}{\lambda_1}\right) \left(\frac{c_n}{c_1}\right)^2}{1 + \left(\frac{c_2}{c_1}\right)^2 + \dots + \left(\frac{c_n}{c_1}\right)^2} \right] \end{aligned}$$

چون فرض کرده‌ایم که x یک تقریب خوب برای v_1 است، پس c_1 بزرگ‌تر از c_i ، $i = 2, \dots, n$ می‌باشد. بنابراین عبارت داخل [] نزدیک به ۱ است، که بدین معنی است که λ_1 نزدیک به σ است. □

در قضیه ۳-۵-۸ خارج قسمت

$$R = \frac{x^T A x}{x^T x}$$

را تعریف کرده‌ایم که در آن x یک بردار ویژه است و ملاحظه نموده‌ایم که این خارج قسمت یک تقریب خوب برای مقدار ویژه λ می‌است که x یک بردار ویژه حقیقی متناظرش می‌باشد. به طور کلی، اگر x یک بردار ویژه مخالف صفر باشد، آنگاه خارج قسمت

$$\mu = R_q = \frac{x^T A x}{x^T x}$$

مقدار $\| (A - \mu I)x \|_2$ را می‌نیم می‌سازد. توجه کنید که معالات نرمال برای مسأله کمترین توانهای دوم می‌نیم‌سازی

$$\| Ax - \mu x \|_2$$

با مجهول μ توسط رابطه $x^T x \mu = x^T A x$ ارائه می‌شوند.

تعریف ۱-۵-۸ خارج قسمت $R_q = \frac{x^T A x}{x^T x}$ خارج قسمت ریلی بردار x نامیده می‌شود.

مثال ۴-۵-۸

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

فرض کنید

$$x = \begin{bmatrix} 1 \\ -0.5 \end{bmatrix}$$

آنگاه خارج قسمت ریلی

$$\sigma = \frac{x^T A x}{x^T x} = -0.2$$

یک تقریب خوب برای مقدار ویژه -0.2361 است.

تذکر: می توان نشان داد که برای یک ماتریس متقارن A ، $\lambda_n \leq R_q \leq \lambda_1$ ، که در آن λ_1 و λ_n به ترتیب کوچک ترین و بزرگ ترین مقادیر ویژه A هستند.

تکرار خارج قسمت ریلی

ایده تقریب زدن یک مقدار ویژه را می توان با تکرار معکوس ترکیب کرد و تقریبهای متوالی برای یک مقدار ویژه و بردار ویژه متناظرش به یک طریقه تکراری، معروف به تکرار خارج قسمت ریلی که به صورت زیر می باشد، محاسبه کرد. فرض کنید N ماکزیمم تکرارهایی باشد که باید انجام شوند و x_0 یک تقریب اولیه برای یک بردار ویژه باشد.

الگوریتم ۳-۵-۸ تکرار خارج قسمت ریلی

برای $k = 0, 1, 2, \dots$ انجام دهید.گام ۱: $\sigma_k = x_k^T A x_k / x_k^T x_k$ را محاسبه کنیدگام ۲: دستگاه $(A - \sigma_k I) \hat{x}_{k+1} = x_k$ را حل کنیدگام ۳: $x_{k+1} = \hat{x}_{k+1} / \sqrt{\max(\hat{x}_{k+1})}$ را محاسبه کنید.گام ۴: متوقف شوید اگر جفت (σ_k, x_k) جفت مقدار ویژه و بردار ویژه قابل قبول است یا اگر $k > N$.

شمارش اعمال ممیز شناور و همگرایی تکرار خارج قسمت ریلی. در گام ۲ الگوریتم ۳-۵-۸ ما نیاز به حل یک دستگاه جدید برای هر مقدار k داریم. این مطلب باعث می شود که روش در مقایسه با هزینه تکرار معکوس کاملاً گران باشد، مگر البته ماتریس متقارن A سه قطری باشد، که در این حالت (همان گونه که در فصل ۶ دیدیم) هزینه فقط از مرتبه $O(n)$ عمل ممیز شناور در هر تکرار می باشد. اما نرخ همگرایی تکرار خارج قسمت ریلی کاملاً سریع است.

نرخ همگرایی از درجه سه است؛ یعنی تعداد ارقام درست در هر تکرار سه برابر می شوند. برای مثال

مرجع Parlett(1980b) را ملاحظه کنید.

انتخاب x_0 . برای انتخاب یک بردار اولیه x_0 ، شاید بهترین کاری که می‌توان انجام داد، استفاده از خود روش توانی مستقیم چند بار و سپس استفاده از آخرین تقریب به عنوان x_0 باشد. تبصره: تکرار خارج قسمت ریلی همچنین می‌تواند برای حالت نامتقارن تعریف شود. در اینجا هر دو بردار ویژه چپ و راست را در هر گام پیدا می‌کنیم. در اینجا از بحث حالت نامتقارن صرف‌نظر می‌کنیم و خواننده را به مرجعهای Wilkinson(1965, 636) و Parlett(1974) ارجاع می‌دهیم.

مثال ۸-۵-۵

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

اجازه دهید اختیار کنیم

$$x_0 = \begin{bmatrix} 0.5246 \\ 0.7622 \\ 1.000 \end{bmatrix}$$

این بردار بعد از سه تکرار از روش توانی به دست آمده است. سپس نتایج زیر را داریم:

$$:k = 0$$

$$\sigma_0 = \frac{x_0^T A x_0}{x_0^T x_0} = 9.6235$$

$$x_1 = \begin{bmatrix} 0.5247 \\ 0.7623 \\ 1.000 \end{bmatrix}$$

$$:k = 1$$

$$\sigma_1 = \frac{x_1^T A x_1}{x_1^T x_1} = 9.6235$$

$$x_2 = \begin{bmatrix} 1.000 \\ 1.4529 \\ 1.9059 \end{bmatrix}$$

بردار ویژه نرمال شده وابسته به عبارت است از:

$$\begin{bmatrix} 0.2851 \\ 0.5595 \\ 0.7339 \end{bmatrix}$$

توجه کنید که $0/3851$ برابر x_2 همین بردار ویژه با سه رقم اعشار می باشد. بنابراین دو تکرار کافی بوده است. تذکر **MATCOM**: الگوریتم ۸-۵-۳ در برنامه **RAYQOT** در **MATCOM** پیاده سازی شده است.

۸-۵-۴ محاسبه مقادیر ویژه و بردارهای ویژه زیر غالب: تقلیل

به محض این که مقدار ویژه غالب λ_1 و بردار ویژه v_1 متناظر محاسبه شدند، مقدار ویژه غالب بعدی λ_2 می تواند با استفاده از تقلیل محاسبه شود. ایده اساسی در پشت روش تقلیل جایگزین کردن ماتریس اصلی با ماتریس دیگری از همان مرتبه و یا کوچک تر با استفاده از یک جفت ویژه محاسبه شده می باشد به طوری که ماتریس تقلیل یافته دارای همان مقادیر ویژه ماتریس اصلی بجز آن یکی باشد که برای تقلیل دادن استفاده شده است.

تقلیل هاتلینگ^۱

تقلیل هاتلینگ یک فرایند است که به جای ماتریس اصلی $A = A_1$ یک ماتریس A_2 با همان مرتبه جایگزین می کند به طوری که همه مقادیر ویژه A_2 همان مقادیر ویژه A_1 می باشند، بجز آن مقدار ویژه ای که برای ساختن A_2 از A_1 مورد استفاده قرار گرفته است.

تبصره ها: اگرچه تقلیل هاتلینگ عموماً در نوشتجات مهندسی به عنوان یک فرایند عملی برای محاسبه یک تعداد انتخاب شده مقدار ویژه و بردار ویژه توصیه می شود و برای ماتریسهای بزرگ و تنک خوب می باشد، ولیکن روش برای هر دو حالت متقارن و نامتقارن ناپایدار عددی است. (مرجع (Wilkinson (1965, 585).

تقلیل هاوس هولدر

اکنون یک تقلیل با استفاده از یک تبدیل تشابهی بر روی A با ماتریسهای هاوس هولدر خواهیم ساخت. البته روش با هر تبدیل تشابهی دیگری کار خواهد کرد؛ ولیکن ما از ماتریسهای هاوس هولدر به دلایل پایداری عددی استفاده خواهیم کرد. روش بر نتیجه زیر مبتنی است.

قضیه ۸-۵-۴ فرض کنید (λ_1, v_1) یک جفت مقدار ویژه A باشد و H یک ماتریس هاوس هولدر باشد به قسمی که Hv_1 ضربی از e_1 باشد؛ آنگاه

$$A_1 = HAH = \begin{bmatrix} \lambda_1 & * & * & \dots & * \\ \cdot & & & & \\ \vdots & & A_2 & & \\ \cdot & & & & \end{bmatrix} = \left[\begin{array}{c|c} \lambda_1 & b^T \\ \cdot & A_2 \end{array} \right]$$

که در آن A_2 یک ماتریس $(n-1) \times (n-1)$ است و مقادیر ویژه A_2 همان مقادیر ویژه A بجز λ_1 می‌باشند. بویژه اگر $|\lambda_1| > |\lambda_2| > |\lambda_3| \geq \dots \geq |\lambda_n|$ ، آنگاه مقدار ویژه غالب A_2 ، λ_2 است که دومین مقدار ویژه غالب (زیر غالب) ماتریس A می‌باشد.

اثبات: از $Av_1 = \lambda_1 v_1$ داریم:

$$HAHHv_1 = \lambda_1 H v_1 \quad (H^2 = I \text{ چون})$$

یعنی، $H A H (k e_1) = \lambda_1 k e_1$ (چون $H v_1 = k e_1$) یا $H A H e_1 = \lambda_1 e_1$. (این بدین معنی است که ستون اول $H A H$ ، λ_1 برابر ستون اول ماتریس واحد است.) بنابراین $H A H$ باید به شکل زیر باشد

$$H A H = \begin{bmatrix} \lambda_1 & * & * & * & \dots & * \\ \cdot & & & & & \\ \vdots & & & A_2 & & \\ \cdot & & & & & \end{bmatrix}$$

چون $\det(H A H - \lambda I) = \det(\lambda_1 - \lambda) \det(A_2 - \lambda I)$ ، نتیجه می‌شود که مقادیر ویژه A_2 همان مقادیر ویژه A_1 بجز λ_1 می‌باشند. بعلاوه اگر

$$|\lambda_1| > |\lambda_2| > |\lambda_3| \geq \dots \geq |\lambda_n|$$

□ آنگاه مقدار ویژه غالب A_2 ، λ_2 است که مقدار ویژه غالب دوم ماتریس A می‌باشد.

الگوریتم ۸-۵-۴ تقلیل هاوس هولدر برای محاسبه مقدار ویژه زیر غالب

گام ۱: مقدار ویژه غالب λ_1 و بردار ویژه متناظر با آن را با استفاده از روش توانی و روش توانی معکوس محاسبه کنید.

گام ۲: یک ماتریس هاوس هولدر H به قسمی پیدا کنید که

$$H v_1 = \begin{bmatrix} * \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

گام ۳: $H A H$ را محاسبه کنید

گام ۴: سطر اول و ستون اول ماتریس HAH را حذف کنید و مقدار ویژه غالب ماتریس $(n-1) \times (n-1)$ حاصل را پیدا کنید.

مثال ۸-۵-۶

$$A = \begin{bmatrix} 0/2190 & 0/6793 & 0/5194 \\ 0/0470 & 0/9347 & 0/8310 \\ 0/6789 & 0/3835 & 0/0346 \end{bmatrix}$$

مقادیر ویژه عبارتند از $1/4947$ ، $-0/3083$ ، $0/0018$.

$$v_1 = \begin{bmatrix} -0/5552 \\ -0/7039 \\ -0/4430 \end{bmatrix} \quad \lambda_1 = 1/4947 \quad .1$$

$$u = v_1 - \|v_1\|_2 e_1 = \begin{bmatrix} -1/5552 \\ -0/7039 \\ -0/4430 \end{bmatrix} \quad .2$$

$$H = I - \frac{\gamma uu^T}{u^T u} = \begin{bmatrix} -0/5552 & -0/7039 & -0/4430 \\ -0/7039 & 0/6814 & -0/2005 \\ -0/4430 & -0/2005 & -0/8738 \end{bmatrix}$$

$$Hv_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$HAH = \begin{bmatrix} 1/4947 & -0/3223 & -0/3331 \\ 0 & 0/1987 & 0/2672 \\ 0 & 0/3736 & -0/5052 \end{bmatrix} \quad .3$$

$$= \left[\begin{array}{c|cc} 1/4947 & -0/3223 & -0/3331 \\ \hline 0 & A_2 & \\ 0 & & \end{array} \right]$$

۴. مقدار ویژه غالب A_2 برابر $-0/3083$ است که مقدار ویژه زیر غالب A می باشد.

محاسبه بردار ویژه زیر غالب

به محض این که مقدار ویژه زیر غالب λ_2 با استفاده از روش فوق محاسبه شود، بردار ویژه متناظر v_2 می‌تواند با استفاده از تکرار معکوس به دست آید.

محاسبه سایر بزرگ‌ترین مقادیر ویژه و بردارهای ویژه

به محض این که جفت (λ_3, v_3) محاسبه شود، ماتریس A_3 می‌تواند با استفاده از این جفت تقلیل یابد و سپس جفت مقدار ویژه و بردار ویژه (λ_3, v_3) محاسبه گردد. پس از آن با استفاده از جفت (λ_3, v_3) جفت (λ_4, v_4) می‌تواند محاسبه شود و غیره. بنابراین با به کار بردن این فرایند به طور تکراری می‌توانیم به طور متوالی همه n مقدار ویژه و بردار ویژه ماتریس A را محاسبه کنیم.

تبصره: اگر بیش از تعداد کمی مقدار ویژه و بردار ویژه مورد نیاز باشد، روش تکراری QR که کمی بعد توصیف می‌شود، باید استفاده گردد، زیرا در این حالت تکرار QR از نظر هزینه مؤثر تر است و یک روش پایدار می‌باشد.

محاسبه کوچک‌ترین مقدار ویژه

به سادگی ملاحظه می‌شود که به کار بردن روش توانی در مورد A^{-1} کوچک‌ترین مقدار ویژه از نظر اندازه (مقدار ویژه کمترین غالب) را ارائه می‌دهد.

فرض کنید A نامنفرد باشد و فرض کنید مقادیر ویژه A به قسمی مرتب شده باشند که

$$|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_{n-1}| > |\lambda_n| > 0$$

آنگاه مقادیر ویژه A^{-1} (که معکوس مقادیر ویژه A هستند) به صورت زیر مرتب می‌شوند

$$\left| \frac{1}{\lambda_n} \right| > \left| \frac{1}{\lambda_{n-1}} \right| \geq \left| \frac{1}{\lambda_{n-2}} \right| \geq \dots \geq \frac{1}{|\lambda_1|} > 0$$

یعنی $1/\lambda_n$ مقدار ویژه غالب A^{-1} می‌باشد. از این حقیقت نتیجه می‌گیریم که معکوس کوچک‌ترین مقدار ویژه را می‌توان با اعمال روش توانی بر A^{-1} محاسبه نمود.

الگوریتم ۸-۵-۵

گام ۱: روش توانی را در مورد A^{-1} به کار ببرید.

گام ۲: معکوس مقدار ویژه به دست آمده در گام ۱ را اختیار کنید.

تذکره: چون روش توانی فقط توسط ضربهای ماتریس - بردار پیاده‌سازی می‌شود، معکوس A نباید به طور صریح محاسبه شود. این مطلب درست است زیرا محاسبه $A^{-1}x$ ، که در آن x یک بردار می‌باشد معادل با حل دستگاه $Ay = x$ است.

برای محاسبه کوچک‌ترین مقدار ویژه بعدی از نظر اندازه، کوچک‌ترین مقدار ویژه و بردار ویژه نظیرش را محاسبه می‌کنیم و سپس تقلیل را اعمال می‌کنیم. با به کار بردن روش توانی معکوس در مورد ماتریس $(n-1) \times (n-1)$ واقع در گوشه راست پایینی معکوس دومین کوچک‌ترین مقدار ویژه به دست می‌آید. به محض این که مقادیر ویژه لازم محاسبه شدند، می‌توانیم همواره از روش توانی معکوس برای محاسبه بردارهای ویژه متناظر استفاده کنیم و یا از تقلیل به صورت نشان داده در قبل استفاده کنیم.

تبصره: برای سرعت بخشیدن به همگرایی، یک انتقال مناسب باید انجام شود.

مثال ۸-۵-۷

$$A = \begin{bmatrix} 1 & 4 & 5 \\ 2 & 3 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

با به کار بردن روش توانی (بدون انتقال) در مورد A^{-1} با بردار آغازین $x_0 = (1, -1, 1)^T$ مقدار $\lambda = 9/5145$ را بعد از ۵ تکرار به دست می‌آوریم. بنابراین کوچک‌ترین مقدار ویژه A برابر است با

$$\frac{1}{\lambda} = 0.1051$$

(توجه کنید که مقادیر ویژه A عبارتند از $6/3850$ ، $1/4901$ ، و 0.1051).

خلاصه فرایند پیدا کردن مقادیر ویژه و بردارهای ویژه انتخاب شده

برای محاسبه یک تعداد انتخاب شده از مقادیر ویژه (بسیار تعداد کمی از بزرگترین و کوچکترین مقادیر ویژه و بردارهای متناظر)، ترکیبی از روش توانی، تکرار معکوس، و تقلیل می تواند به صورت زیر استفاده شود.

۱. از روش توانی (از اعمال روش توانی بر A^{-1}) استفاده کنید و تقریب خوب و معقولی از بزرگترین (کوچکترین) مقدار ویژه از نظر اندازه و بردارهای متناظر را حساب کنید.

۲. از تکرار معکوس با مقدار ویژه تقریبی (با ثابت نگهداشتن آن) و بردار ویژه تقریبی به دست آمده در گام ۱ به عنوان بردار آغازین استفاده کنید.

۳. برای محاسبه مجموعه جفت ویژه بعدی تقلیل را به کار ببرید.

۸-۶ تبدیلیهای تشابهی و محاسبات مقدار ویژه

به خاطر دارید (قضیه ۸-۲-۳) که دو ماتریس متشابه A و B دارای مقادیر ویژه یکسانند؛ یعنی اگر X یک ماتریس نامنفرد باشد به قسمی که

$$X^{-1}AX = B$$

آنگاه مقادیر ویژه A همان مقادیر ویژه B هستند.

بنابراین یک راه واضح و روشن برای محاسبه مقادیر ویژه A کاهش A به یک شکل «ساده تر» و مناسب B با تشابه می باشد به قسمی که مقادیر ویژه بتوانند به طور ساده تری محاسبه شوند. ولیکن احتیاطهای لازم باید انجام پذیرد. می توان نشان داد (Golub and Van Loan (1989, 340)) که

$$\text{Cond}_p(X^{-1}AX) = \text{Cond}_p(A) + E \quad \text{قضیه ۸-۶-۱} \quad \text{که در آن}$$

$$\|E\|_p \approx \mu \|X\|_p \|X^{-1}\|_p \|A\|_p$$

تبصره: چون ماتریس خطا E به طور واضح بستگی به $\text{Cond}_p(X)$ دارد، قضیه ۸-۶-۱ بیان می کند که در محاسبه مقادیر ویژه و بردارهای ویژه باید از ماتریسهای بدوضع برای تبدیل A به شکل ساده تر پرهیز کرد. عمدتاً به این خاطر مقادیر ویژه یک ماتریس با استفاده از چند جمله ای مشخصه A یا تبدیل A به شکل متعارف کردن محاسبه نمی شوند. در این مورد به طور مفصل در بخش زیر بحث خواهیم کرد.

۸-۶-۱ مشکلات در استفاده از چندجمله‌ای مشخصه

مسائل در استفاده از چندجمله‌ای مشخصه

چون مقادیر ویژه یک ماتریس صفرهای چندجمله‌ای مشخصه هستند، طبیعی است که به محاسبه مقادیر ویژه A توسط پیدا کردن صفرهای چندجمله‌ای مشخصه‌اش فکر کنیم. ولیکن این راه به طور عددی مؤثر نیست.

• اولاً فرایند محاسبه صریح ضرایب چندجمله‌ای مشخصه ممکن است به طور عددی ناپایدار باشد.

• ثانیاً صفرهای چندجمله‌ای مشخصه ممکن است در برابر اختلالاتی ضرایب چندجمله‌ای مشخصه خیلی حساس باشند. بنابراین اگر ضرایب چندجمله‌ای مشخصه به طور دقیق محاسبه نشوند، خطاهایی در مقادیر ویژه محاسبه شده وجود خواهند داشت.

در فصل ۳ حساسیت مسأله پیدا کردن ریشه‌ها را توسط چندجمله‌ای ویلکینسون و سایر مثالها توضیح دادیم. اکنون مشکل محاسبه چندجمله‌ای مشخصه را به طور مفصل بحث خواهیم کرد.

محاسبه چندجمله‌ای مشخصه به طور صریح به قیمت تبدیل ماتریس به یک مجموع مستقیم از ماتریسهای همراه تمام می‌شود. هر ماتریس A می‌تواند به یک ماتریس متشابه به شکل

$$C = \begin{bmatrix} C_1 & & \cdot \\ & \ddots & \\ \cdot & & C_k \end{bmatrix}$$

کاهش یابد که در آن هر C_i یک ماتریس همراه است. گفته می‌شود که ماتریس C به شکل فرینوس می‌باشد. اگر $k = 1$ ، ماتریس A ناموهن است.

فرض کنید که A ناموهن باشد؛ اجازه دهید ببینیم که چگونه A را می‌توان توسط تبدیل تشابهی به یک ماتریس همراه کاهش داد. این کاهش می‌تواند در دو مرحله انجام شود.

کاهش یک ماتریس به یک ماتریس همراه

مرحله ۱: ماتریس A توسط تبدیل تشابهی متعامد با استفاده از روش هاوس هولدر یا گیونز

به یک ماتریس هستبرگی کاهش ناپذیر H تبدیل می‌شود.

مرحله ۲: ماتریس هستبرگی کاهش ناپذیر تبدیل یافته H توسط تبدیل تشابهی به یک ماتریس

همراه کاهش بیشتری می‌یابد.

در فصل ۵ ملاحظه نمودیم که چگونه یک ماتریس A را می توان توسط تبدیل تشابهی متعامد به یک طریق عددی پایدار با استفاده از روش هاوس هولدر و یا گونز به یک ماتریس بالا هسنبرگی تبدیل کرد. اکنون مرحله ۲ را ملاحظه می کنیم، یعنی تبدیل ماتریس هسنبرگی کاهش یافته H به یک ماتریس همراه C .

فرض کنید X ماتریس تبدیل کننده نامنفرد باشد، یعنی $HX = XC$ ، که در آن

$$C = \begin{pmatrix} \cdot & \cdot & \cdot & \dots & \cdot & c_1 \\ 1 & \cdot & \cdot & \dots & \cdot & c_2 \\ \cdot & 1 & \cdot & \dots & \cdot & c_3 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \cdot & \cdot & \cdot & \dots & 1 & c_n \end{pmatrix}$$

اگر x_1, x_2, \dots, x_n بردار ستونی متوالی از X باشند، آنگاه از

$$HX = XC$$

نتیجه می شود که اگر x_1 انتخاب شود، x_2 تا x_n می توانند توسط رابطه بازگشتی زیر تعیین شوند.

$$x_{i+1} = Hx_i, \quad i = 1, \dots, n-1$$

بنابراین نتیجه می شود که اگر x_1 به قسمی انتخاب شود که

$$X = (x_1, Hx_1, \dots, Hx_{n-1})$$

نامنفرد باشد، آنگاه X ماتریس H را توسط تبدیل تشابهی به C تبدیل خواهد کرد.

با انتخاب $x_1 = (1, 0, \dots, 0)^T$ ؛ آنگاه ماتریس بالا مثلثی

$$X = (x_1, Hx_1, \dots, Hx_{n-1}) = \begin{pmatrix} 1 & * & * & \dots & * \\ \cdot & h_{21} & * & \dots & * \\ \cdot & \cdot & h_{21}h_{22} & \dots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdot & \dots & \cdot & h_{21}h_{22} \dots h_{n,n-1} \end{pmatrix}$$

نامنفرد است، زیرا $h_{i+1,i} \neq 0$ ، $i = 1, \dots, n-1$.

بنابراین اگر یک یا چند عنصر زیر قطر اصلی $h_i, i+1$ ماتریس H به طور بامعنایی کوچک باشند، آنگاه معکوس X عناصر خیلی بزرگی خواهد داشت؛ در نتیجه X بدوضع خواهد بود. بنابراین در یک چنین حالتی تبدیل یک ماتریس هسنبرگی کاهش ناپذیر H به یک ماتریس همراه ناپایدار خواهد بود. اکنون این مطالب را خلاصه می‌کنیم.

اولین مرحله، که در آن A با استفاده از روش هاوس هولدر یا گونز به H تبدیل می‌شود، به طور عددی پایدار است، در حالی که دومین مرحله، که در آن H دوباره به C کاهش می‌یابد، می‌تواند از ناپایداری بالایی برخوردار باشد.

مثال ۸-۶-۱

$$H = \begin{bmatrix} 1 & 2 & 3 \\ 0/0001 & 1 & 1 \\ 0 & 2 & 3 \end{bmatrix}$$

$$x_1 = (1, 0, 0)^T, \quad x_2 = Hx_1 = (1, 0/0001, 0)^T$$

$$x_3 = Hx_2 = (1/00002, 0/00002, 0/00002)^T$$

$$X = \begin{bmatrix} 1 & 1 & 1/0002 \\ 0 & 0/0001 & 0/0002 \\ 0 & 0 & 0/0002 \end{bmatrix}$$

$$X^{-1}HX = C = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -4/9998 \\ 0 & 1 & 5 \end{bmatrix}$$

$$\text{Cond}_p(X) = 3/1329 \times 10^4$$

(توجه کنید که وجود یک عنصر زیر قطر اصلی کوچک از H ، مثلاً h_{21} ، ماتریس تبدیل X را بدوضع می‌سازد).

روشهای معادل دیگری نیز برای پیدا کردن چند جمله‌ای مشخصه، نظیر روش لووریر^۱، وجود دارند

که ناپایدارند. برای جزئیات بیشتر مرجع (Wilkinson(1965, 434 را ملاحظه کنید. جهت تأکید بر خطر استفاده از شکل فروبنیوس در محاسبه مقادیر ویژه یک ماتریس، اجازه دهید برخی تذکرات ویلکینسون را در مورد شکلهای فروبنیوس ماتریسهای مطرح در برخی کاربردها نظیر دستگاههای مکانیکی و الکترونیکی ذکر نماییم.

اگرچه روشن ساخته‌ایم که استفاده از شکل فروبنیوس - که در آن به طور فاجعه‌انگیزی بدو وضعی بدتری از ماتریس اصلی کاملاً امکان‌پذیر است، خطرناک می‌باشد، دریافته‌ایم که برنامه‌های مبتنی بر استفاده از این شکل در حالت کلی برای ماتریسهای مطرح در دستگاههای مکانیکی یا الکتریکی میرا به طور شگفت‌برانگیزی رضایت‌بخش هستند و معمولاً چندجمله‌ای مشخصه متناظر خوش‌وضع می‌باشد. هنگامی که چنین وضعیتی برقرار است روشهای مبتنی بر استفاده از چندجمله‌ایهای مشخصه صریح هم سریع و هم دقیق هستند (Wilkinson 1965, 482).

تبصره: تذکرات قبلی ویلکینسون به طور واضح از سنت قدیمی استفاده از چندجمله‌ایهای مشخصه برای محاسبه مقادیر ویژه توسط مهندسان حمایت می‌کند.

۸-۶-۲ محاسبه مقادیر ویژه از طریق شکل متعارف جردن

اجازه دهید اکنون به بحث استفاده از برخی شکلهای متعارف در محاسبه مقادیر ویژه بپردازیم. بی‌گمان در این رابطه اولین شکلی که به خاطر می‌آید شکل متعارف جردن است. به خاطر دارید (قضیه ۸-۲-۱۰) که با مفروض بودن یک ماتریس A از مرتبه $n \times n$ ، یک ماتریس نامنفرد وجود دارد به طوری که

$$X^{-1}AX = \text{diag}(J_1, \dots, J_k)$$

که در آن:

$$J_i = \begin{bmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix}$$

اگر J_i از مرتبه v_i باشد، آنگاه

$$v_1 + v_2 + \dots + v_k = n$$

ماتریس واقع در طرف راست شکل متعارف جردن (JCF) ماتریس A است، و λ ها مقادیر ویژه هستند. بنابراین مقادیر ویژه A ظاهر خواهند شد به محض آن که JCF محاسبه شود.

متأسفانه این محاسبه نیز بسیار ناپایدار است.

اگر A نزدیک به یک ماتریس غیر قابل قطری شدن باشد، ماتریس تبدیل X می تواند خیلی بدوضع باشد (Golub and Van Loan (1989, 339-40)). سپس از قضیه ۸-۶-۱ نتیجه می شود که در این حالت JCF محاسبه شده نادقیق خواهد بود.

۸-۷ حساسیت مقدار ویژه

در دو بخش قبل در مورد خطر محاسبه مقادیر ویژه از طریق شکل متعارف جردن یا شکل فروبنیوس یک ماتریس هشدار دادیم. خطر به طور عمده امکان بدوضع شدن ماتریس تبدیل X بود. در این بخش ملاحظه می کنیم که چه نقش خاصی عدد وضعیت ماتریس تبدیل، $\text{Cond}(X) = \|X\| \|X^{-1}\|$ در حساسیت مقدار ویژه بازی می کند. در فصل ۳ به کمک چند مثال پدیده حساسیت مقدار ویژه را توضیح دادیم. در اینجا یک نتیجه خاص تر، معروف به قضیه بوئر-فایک^۱، را در مورد حساسیت مقدار ویژه یک ماتریس قابل قطری شدن ارائه می دهیم.

۸-۷-۱ قضیه بوئر-فایک

قضیه ۸-۷-۱ فرض کنید A قابل قطری شدن باشد؛ یعنی یک ماتریس نامنفرد X وجود داشته باشد به طوری که $X^{-1}AX$ یک ماتریس قطری D باشد. سپس برای یک مقدار ویژه λ از $A + E$ داریم

$$\min |\lambda_i - \lambda| \leq \|X\| \|X^{-1}\| \|E\|$$

که در آن $\| \cdot \|$ یک نرم ماتریسی طبیعی و $\lambda_1, \lambda_2, \dots, \lambda_n$ مقادیر ویژه A هستند.

اثبات: دو حالت در نظر می گیریم

حالت ۱: $\lambda = \lambda_i$ برای مقداری از i . قضیه به طور بدیهی درست است.

حالت ۲: $\lambda \neq \lambda_i$ به ازای هر i . آنگاه عناصر قطری ماتریس $\lambda I - D$ مخالف صفر هستند. چون دترمینان یک ماتریس برابر حاصلضرب مقادیر ویژه آن می باشد، ماتریس $(\lambda I - D)$ نامنفرد است. از

$$(A + E)x = \lambda x$$

داریم

$$\begin{aligned} Ex &= (\lambda I - D)x \\ &= (\lambda I - XDX^{-1})x \quad (A = XDX^{-1} \text{ توجه کنید که}) \\ &= X(\lambda I - D)X^{-1}x \end{aligned}$$

قرار می‌دهیم $X^{-1}x = y$. سپس از آخرین معادله، با ضرب معادله از دست چپ در X^{-1} خواهیم داشت:

$$(\lambda I - D)y = X^{-1}Ex$$

یا

$$y = (\lambda I - D)^{-1}X^{-1}Ex \quad (x = Xy \text{ توجه کنید که})$$

با گرفتن نرم وابسته از هر دو طرف، داریم

$$\begin{aligned} \|y\| &= \|(\lambda I - D)^{-1}X^{-1}Ex\| \\ &\leq \|(\lambda I - D)^{-1}\| \|X^{-1}Ex\| \|y\| \end{aligned}$$

با تقسیم هر دو طرف بر $\|y\|$ ، به دست می‌آوریم

$$1 \leq \|(\lambda I - D)^{-1}\| \|X^{-1}\| \|E\| \|X\|$$

برای نرم طبیعی داریم:

$$\begin{aligned} \|(\lambda I - D)^{-1}\| &= \max_i \left| \frac{1}{\lambda - \lambda_i} \right| \\ &= \frac{1}{\min_i |\lambda - \lambda_i|} \end{aligned}$$

بنابراین

$$1 \leq \frac{1}{\min_i |\lambda - \lambda_i|} \|X^{-1}\| \|X\| \|E\|$$

یا

$$\min_i |\lambda - \lambda_i| \leq \|X^{-1}\| \|X\| \|E\|$$

پیامدهای قضیه

قضیه ۸-۷-۱ بیان می‌کند که اگر

$$\text{Cond}(X) = \|X\| \|X^{-1}\|$$

بزرگ باشد، آنگاه یک مقدار ویژه λ از ماتریس اختلال یافته $A + E$ می‌تواند به طور بامعنایی متفاوت از یک مقدار ویژه λ از A باشد. در حالت کلی، هر چه ماتریس بردار ویژه X بدوضع تر باشد مسأله ویژه برای A بدوضع تر خواهد بود.

تبصره: اگر A قابل قطری شدن نباشد، یک نتیجه مشابه نیز برقرار خواهد بود (برای جزئیات بیشتر مرجع Golub and Van Loan (1989, 342-343) ملاحظه شود).

مثال ۸-۷-۱

$$A = \begin{pmatrix} 1 & 2 & 3 \\ \cdot & \cdot/9990 & 1 \\ \cdot & \cdot & 2 \end{pmatrix}$$

مقادیر ویژه عبارتند از ۱، $\cdot/9990$ ، و ۲.

$$X = \begin{pmatrix} 1 & -1 & \cdot/9623 \\ \cdot & \cdot/0005 & \cdot/1923 \\ \cdot & \cdot & \cdot/1925 \end{pmatrix}$$

$$X^{-1}AX = \text{diag}(1, \cdot/9990, 2)$$

فرض کنید

$$E = \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$$

مقادیر ویژه $A + E$ عبارتند از: $\cdot/0044i$ ، $\cdot/9995 + \cdot/0044i$ ، $\cdot/0044i - \cdot/9995$ ، و ۲.

یک تغییر 10^{-5} در عنصر (۳، ۱) ماتریس A ، کاملاً دو مقدار ویژه اول را تغییر داد. این به علت این است که $\text{Cond}(X) = 6/8708 \times 10^2$ (بزرگ!)، توجه کنید که $\text{Cond}(X) \times \|E\| = \cdot/0687$. بنابراین بر طبق قضیه بوئر - فایک یک چنین تغییری انتظار می‌رفت. ولیکن سؤالی که مطرح می‌شود

این است که چرا دو مقدار ویژه اول A تغییر می‌کنند و سومی تغییر نمی‌نماید. ما این را در مثال ۸-۷-۲ توضیح خواهیم داد.

۸-۷-۲ حساسیت مقادیر ویژه فردی

عدد وضعیت $\|X^{-1}\| \|X\|$ یک ارزیابی کلی از تغییرات در مقادیر ویژه را نسبت به تغییرات در ماتریس ضرایب ارائه می‌دهد. ولیکن همان‌گونه که در مثالهای فصل ۳ ملاحظه نمودیم ممکن است که برخی از مقادیر ویژه A از بقیه حساس تر باشند. در حقیقت برخی ممکن است خیلی خوش‌وضع باشند، در حالی که بقیه بدوضع هستند. همچنین برخی بردارهای ویژه ممکن است خوش‌وضع باشند، در حالی که بقیه چنین نباشند.

بنابراین بهتر است در مورد وضعیت مقادیر ویژه به طور فردی صحبت شود تا در مورد وضعیت مسأله ویژه بحث گردد. به خاطر دارید که در فصل ۳ تحلیلی از بدوضعی مقادیر ویژه فردی ماتریس ویلکینسون کمی اختلال یافته بر حسب اعداد s_i ارائه شد. به طور کلی این کار می‌تواند برای هر ماتریس قابل قطری شدن انجام شود.

فرض کنید $X^{-1}AX = \text{diag}(\lambda_1, \dots, \lambda_n)$. آنگاه بردارهای ویژه راست و چپ نرمال شده به صورت زیر ارائه می‌شوند

$$x_i = \frac{Xe_i}{\|Xe_i\|_2}, \quad y_i = \frac{(X^{-1})^T e_i}{\|(X^{-1})^T e_i\|_2}$$

تعریف ۸-۷-۱ عدد s_i ، که در آن s_i به صورت

$$s_i = |y_i^T x_i|$$

تعریف می‌شود عدد وضعیت مقدار ویژه λ_i نامیده می‌شود.

اگر این عدد بزرگ باشد، آنگاه y_i یک مقدار ویژه بدوضع است. این معمولاً زمانی اتفاق می‌افتد که A نزدیک به یک ماتریس با چند مقسوم‌علیه ابتدایی غیرخطی باشد.

تذکرها:

۱. n عدد وضعیت متناظر با n مقدار ویژه وجود دارد.
۲. اگر x_i و y_i حقیقی باشند، آنگاه s_i کسینوس زاویه بین x_i و y_i خواهد بود.

مثال ۸-۷-۲

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0.999 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

$$x_1 = (1, 0, 0)^T$$

$$x_2 = (1, -0.00005, 0)^T$$

$$x_3 = (0.9623, 0.1923, 0.1925)^T$$

$$y_1 = (0.0004, 0.7066, -0.7076)^T$$

$$y_2 = (0, -0.7065, -0.7068)^T$$

$$y_3 = (0, 0, 1)^T$$

$$s_1 = |y_1^T x_1| = 3/5329 \times 10^{-4}$$

$$s_2 = |y_2^T x_2| = 3/5373 \times 10^{-4}$$

$$s_3 = |y_3^T x_3| = 0.1925$$

بنابراین این محاسبات به طور واضح نشان می‌دهند که $\lambda_1 = 1$ و $\lambda_2 = 0.999$ بدو وضع هستند، درحالی که $\lambda_3 = 2$ خوش وضع است.

درحقیقت، هنگامی که $a(3, 1)$ به 0.000001 تغییر می‌یابد و مقادیر ویژه ماتریس اختلال یافته محاسبه می‌شوند، اولین دو مقدار ویژه ماتریس اختلال یافته (آنهايي که متناظر با ۱ و 0.999 ماتریس اصلی هستند) مختلط می‌گردند. مقادیر ویژه محاسبه شده ماتریس اختلال یافته با سه رقم اعشار عبارتند از:

$$0.999 + 0.001i, \quad 0.999 - 0.001i, \quad 2$$

برای دیدن یک مثال غیر بدیهی دیگر تمرین ۲۳ را ملاحظه کنید.

یک رابطه بین s_i و $\text{Cond}_p(X)$

به آسانی ملاحظه می‌شود که اعداد s_i و $\text{Cond}_p(X)$ نسبت به نرم دو وابسته هستند.

$$\begin{aligned} s_i &= |y_i^T x_i| = \frac{|e_i^T X^{-1} X e_i|}{\|X e_i\|_p \|(X^{-1})^T e_i\|_p} \\ &= \frac{1}{\|X e_i\|_p \|(X^{-1})^T e_i\|_p} \end{aligned}$$

اکنون

$$\|Xe_i\|_r \leq \|X\|_r \|e_i\|_r = \|X\|_r$$

و

$$\begin{aligned} \|(X^{-1})^T e_i\|_r &\leq \|(X^{-1})^T\|_r \|e_i\|_r \\ &= \|(X^{-1})^T\|_r = \|X^{-1}\|_r \end{aligned}$$

بنابراین

$$\frac{1}{s_i} \leq \|X\|_r \|X^{-1}\|_r = \text{Cond}_r(X)$$

بنابراین برای هر i

$$\frac{1}{s_i} \leq \text{Cond}_r(X)$$

به طور مشابه می‌توان نشان داد که یک ماتریس X از بردارهای ویژه وجود دارد به طوری که

$$\text{Cond}_r(X) \leq \sum_{i=1}^n |s_i^{-1}|$$

مرجع (Wilkinson (1965, 1989) را ملاحظه کنید.

مثال ۸-۷-۳

$$A = \begin{bmatrix} 1 & 2 & 3 \\ . & .999 & 1 \\ . & . & 2 \end{bmatrix}$$

$$\frac{1}{s_1} = 2/8305 \times 10^3$$

$$\frac{1}{s_2} = 2/8270 \times 10^3$$

$$\frac{1}{s_3} = 5/1940$$

قرار دهید

$$X = \begin{bmatrix} 53/20 & -53/17 & 2/193 \\ . & .0266 & .4383 \\ . & . & .4383 \end{bmatrix}$$

$$\text{Cond}_r(X) = 5/6577 \times 10^3$$

بنابراین

$$\frac{1}{s_i} < \text{Cond}_p(X), \quad i = 1, 2, 3$$

$$\text{و } \text{Cond}_p(X) < s_1^{-1} + s_2^{-1} + s_3^{-1}$$

عددهای وضعیت و وابستگی خطی بردارهای ویژه

برای یک ماتریس قابل قطری شدن، ستونهای ماتریس X بردارهای ویژه A هستند، بنابراین $\text{Cond}_p(X)$ نشانه‌ای از چگونگی وابسته بودن بردارهای ویژه را به ما ارائه می‌دهد. اگر $\text{Cond}_p(X)$ بزرگ باشد، معنی آن این است که بردارهای ویژه تقریباً وابسته وجود دارند.

برای ماتریس A مثال قبل، یک ماتریس X بردارهای ویژه عبارت است از:

$$X = \begin{bmatrix} 1 & 1 & 0.9623 \\ 0 & 0.0005 & 0.1923 \\ 0 & 0 & 0.1925 \end{bmatrix}$$

به وابستگی خطی اولین دو مقدار ویژه توجه کنید:

$$\text{Cond}_p(X) = 6/8708 \times 10^3$$

حساسیت مقادیر ویژه یک ماتریس نرمال

یک ماتریس A نرمال نامیده می‌شود اگر $AA^* = A^*A$ که در آن $A^* = (\bar{A})^T$. با استفاده از قضیه مثلثی سازی شور (قضیه ۸-۲-۶) می‌توان به آسانی نشان داد (تمرین ۱۹) که اگر A یک ماتریس نرمال باشد آنگاه یک ماتریس یکانی یافت می‌شود به طوری که U^*AU قطری است. بنابراین یک نتیجه فوری از قضیه بوئر - فایک قضیه زیر است.

نتیجه ۸-۷-۱ نتیجه‌ای از قضیه بوئر - فایک. فرض کنید A یک ماتریس نرمال باشد و $\lambda_1, \lambda_2, \dots, \lambda_n$ مقادیر ویژه A باشند. آنگاه برای یک مقدار ویژه λ از $A + E$ داریم.

$$\min |\lambda_i - \lambda| \leq \|E\|_p$$

حساسیت مقادیر ویژه یک ماتریس نرمال

به عبارت دیگر، مقادیر ویژه یک ماتریس نرمال کاملاً خوش وضع هستند.

تصوره: بیشتر ماتریسهای مطرح در کاربردهای عملی متقارن (یا هرمیتی، در صورت مختلط بودن) هستند. بنابراین بر طبق نتیجه قبل مقادیر ویژه یک ماتریس متقارن (یا هرمیتی) خوش وضع هستند. ما در مورد مسأله مقادیر ویژه متقارن در بخش ۸-۱۱ بیشتر بحث خواهیم کرد.

تذکر MATCOM: برنامه SENSEIG از MATCOM یک بردار حاوی معکوسهای اعداد وضعیت مقادیر ویژه یک ماتریس را حساب می‌کند.

۸-۸ حساسیت بردار ویژه

ما در این بحث زیاد به حساسیت بردارهای ویژه نخواهیم پرداخت. ما فقط یک قضیه را (به شکل خام و تا حدی مختصر) بیان خواهیم کرد که تفاوت اصلی بین حساسیتهای مقادیر ویژه و بردارهای ویژه را روشن خواهد ساخت. برای بیانی دقیق و اثبات مرجع (Watkins (1991, 332-333) را ملاحظه کنید.

قضیه ۸-۸-۱ فرض کنید ΔA یک اختلال خیلی کوچک از A باشد و فرض کنید مقدار ویژه λ_k از A توسط $\delta\lambda_k$ اختلال یابد؛ یعنی $\lambda_k + \delta\lambda_k$ یک مقدار ویژه $A + \Delta A$ باشد. فرض کنید $x_k + \delta x_k$ بردار ویژه متناظر با $\lambda_k + \delta\lambda_k$ باشد. آنگاه با فرض این که مقادیر ویژه و بردارهای ویژه A متمایز هستند داریم

$$x_k + \delta x_k = x_k + \sum_{j \neq k} \frac{a_{jk}}{(\lambda_k - \lambda_j)} x_j + O(\|\Delta A\|^2)$$

که در آن

$$a_{jk} = y_j^T (\Delta A) x_k$$

و x_k و y_k بردارهای ویژه نرمال شده A وابسته به λ_k می‌باشند.

پیامدهای قضیه

قضیه ۸-۸-۱ بیان می‌کند که اگر A توسط مقدار کوچکی اختلال یابد، آنگاه مقدار اختلال در یک بردار ویژه x_k توسط مقادیر زیر تعیین می‌شود.

۱. اعداد وضعیت همه مقادیر ویژه بجز λ_k ، و

۲. فاصله λ_k از مقادیر ویژه دیگر.

یک نتیجه فوری از این قضیه این است که اگر یک مقدار ویژه چندگانه وجود داشته باشد یا یک مقدار ویژه نزدیک به مقدار ویژه دیگر وجود داشته باشد، آنگاه چند بردار ویژه بدوضع وجود دارد. این

بویژه برای یک ماتریس متقارن هرمیتی بامعنی است، زیرا همان طور که می دانیم مقادیر ویژه یک ماتریس هرمیتی خوش وضع هستند، اما بردارهای ویژه می توانند بدوضع باشند. اگر مقادیر ویژه به خوبی متمایز و خوش وضع باشند، آنگاه بردارهای ویژه خوش وضع هستند.

مثال ۸-۸-۱

$$A = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & 0/99 & \cdot \\ \cdot & \cdot & 2 \end{bmatrix}$$

$$A' = (A + \Delta A) = \begin{bmatrix} 1 & \cdot/0001 & \cdot \\ \cdot/0001 & 0/99 & \cdot \\ \cdot & \cdot & 2 \end{bmatrix}$$

مقادیر ویژه $A + \Delta A$ عبارتند از ۱، ۰/۹۹، و ۲. (تغییر کوچکی وجود دارد؛ چون A متقارن است، مقادیر ویژه خوش وضع هستند.) ولیکن بردار ویژه A' عبارتند از

$$\begin{bmatrix} -1 \\ -0/01 \\ \cdot \end{bmatrix}, \quad \begin{bmatrix} \cdot/01 \\ -1 \\ \cdot \end{bmatrix}, \quad \begin{bmatrix} \cdot \\ \cdot \\ 1 \end{bmatrix}$$

در حالی که بردارهای ویژه A عبارتند از:

$$\begin{bmatrix} 1 \\ \cdot \\ \cdot \end{bmatrix}, \quad \begin{bmatrix} \cdot \\ 1 \\ \cdot \end{bmatrix}, \quad \begin{bmatrix} \cdot \\ \cdot \\ 1 \end{bmatrix}$$

توجه کنید که بردار ویژه متناظر با $\lambda = 2$ تغییر نکرده است، در حالی که دو بردار ویژه دیگر به واسطه نزدیکی مقادیر ویژه ۱ و ۰/۹۹ و ساختار اختلال تغییر کرده اند.

۹-۸ شکل شور حقیقی و تکرارهای QR

در بحثهای قبل ملاحظه کردیم که محاسبه مقادیر ویژه A از طریق کاهش A به شکل فروبنیوس یا شکل متعارف جردن از لحاظ عددی مؤثر نیست. اگر ماتریس تبدیل بدوضع باشد، آنگاه ممکن است خطاهای بزرگ در شکل متعارف محاسبه شده وجود داشته باشد؛ این باعث وارد شدن خطاها در مقادیر ویژه می شود.

بنابراین یک سؤال به این ترتیب مطرح می شود که آیا می توان با استفاده از یک ماتریس تبدیل خوش وضع یک کاهش تشابهی از A به شکل متعارف مناسب به دست آورد.

یک ماتریس کاملاً خوش وضع برای مثال یک ماتریس متعامد (یا یکانی، در صورت مختلط بودن) می باشد، عدد وضعیت (نسبت به نرم دو و F -نرم) یک چنین ماتریسی برابر ۱ است.

در حقیقت اگر یک ماتریس A با استفاده از یک تبدیل تشابهی یکانی به یک ماتریس B تبدیل شود، آنگاه یک اختلال در A یک اختلال در B به همان اندازه نتیجه خواهد داد؛ یعنی اگر

$$B = U^*AU$$

و

$$U^*(A + \Delta A)U = B + \Delta B$$

آنگاه

$$\|\Delta B\|_F = \|\Delta A\|_F$$

مثال ۸-۹-۱

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}, \quad U = \begin{bmatrix} -0.5774 & -0.5774 & -0.5774 \\ -0.5774 & 0.7887 & -0.2113 \\ -0.5774 & -0.2113 & 0.7887 \end{bmatrix}$$

$$B = \begin{bmatrix} 13 & -0.6340 & -2.3660 \\ -0.9019 & \cdot & \cdot \\ -6.981 & \cdot & \cdot \end{bmatrix}$$

$$\Delta A = 10^{-5} \times I_{3 \times 3}$$

$$A_1 = A + \Delta A = \begin{bmatrix} 1/0.0001 & 2 & 3 \\ 3 & 4/0.0001 & 5 \\ 6 & 7 & 8/0.0001 \end{bmatrix}$$

$$B_1 = U^*(A + \Delta A)U = \begin{bmatrix} 13/0.0001 & -0.633994 & -2.3660 \\ -0.9019 & 0.0001 & \cdot \\ -6.981 & \cdot & 8/0.0001 \end{bmatrix}$$

$$\Delta B = B_1 - B = 10^{-5} \times I_{3 \times 3}$$

$$\|\Delta A\| = \|\Delta B\| = 10^{-5}$$

یک شکل متعارف کامل که مقادیر ویژه را آشکار می‌سازد یک شکل مثلثی است (عناصر قطری مقادیر ویژه هستند). در این زمینه اکنون یک نتیجه کلاسیک را که اثر شور (قضیه ۸-۲-۶) می‌باشد، یادآوری می‌کنیم. ما این نتیجه مهم را اکنون بیان می‌کنیم و یک اثبات برای آن ارائه می‌دهیم.

قضیه ۸-۹-۱ قضیه مثلثی‌سازی شور (بیان دوباره قضیه ۸-۲-۶) اگر A یک ماتریس $n \times n$ باشد، آنگاه یک ماتریس یکانی U وجود دارد به قسمی که

$$U^*AU = T$$

که در آن T یک ماتریس مثلثی با مقادیر ویژه $\lambda_1, \lambda_2, \dots, \lambda_n$ به صورت عناصر قطری می‌باشد.

اثبات: قضیه را با استفاده از استقرا بر روی n ثابت می‌کنیم.

اگر $n = 1$ باشد قضیه به صورت بدیهی درست است. حال فرض می‌کنیم که قضیه برای $n - 1$ درست باشد؛ سپس نشان خواهیم داد که قضیه برای n نیز درست می‌باشد.

فرض کنید u یک بردار ویژه A با $\|u\|_2 = 1$ و وابسته به مقدار ویژه λ_1 باشد. سپس همواره می‌توانیم یک ماتریس $(n - 1) \times n$ بعدی V را به قسمی انتخاب کنیم که

$$U_1 = (u, V)$$

یکانی باشد. چون $AU_1 = A(u, V) = (\lambda_1 u, AV)$ داریم

$$A_1 = U_1^*AU_1 = \begin{bmatrix} u^* \\ V^* \end{bmatrix} (\lambda_1 u, AV) = \left[\begin{array}{c|ccc} \lambda_1 & * & * & * \\ \hline \cdot & & & \\ \vdots & & \hat{A} & \\ \cdot & & & \end{array} \right]$$

که در آن $\hat{A} = V^*AV$ از بعد $(n - 1) \times (n - 1)$ می‌باشد. بجز λ_1 مقادیر ویژه A و \hat{A} یکسان هستند. بر طبق فرضمان آنگاه ماتریس یکانی V_1 و از مرتبه $(n - 1)$ به قسمی وجود دارد که ماتریس

$$\hat{T} = V_1^*(\hat{A})V_1$$

مثلثی باشد. سپس با تعریف

$$U_2 = \left[\begin{array}{c|ccc} 1 & \cdot & \dots & \cdot \\ \hline \cdot & & & \\ \vdots & & V_1 & \\ \cdot & & & \end{array} \right]$$

ملاحظه می‌کنیم که U_2 یکانی است (زیرا V_1 یکانی است) و

$$\begin{aligned} U_2^* A_1 U_2 &= U_2^* U_1^* A U_1 U_2 \\ &= U^* A U \end{aligned}$$

که در آن $U = U_1 U_2$. در نتیجه

$$U^* A U = \left[\begin{array}{c|ccc} \lambda_1 & * & \dots & * \\ \hline \cdot & & & \\ \vdots & & & \\ \cdot & & & \end{array} \right]$$

چون \hat{T} مثلثی است، در نتیجه $U^* A U$ نیز مثلثی است. چون مقادیر ویژه یک ماتریس مثلثی روی قطر ظاهر می‌شوند، قضیه را ثابت کرده‌ایم. \square

چون یک ماتریس حقیقی می‌تواند مقادیر ویژه مختلط داشته باشد (که به صورت جفت مزدوج مختلط ظاهر می‌شوند)، حتی برای یک ماتریس حقیقی A ماتریسهای U و T در قضیه شور می‌توانند مختلط باشند. ولیکن اگر به جای T یک ماتریس شبه‌مثلثی R جایگزین کنیم آنگاه می‌توانیم U را طوری انتخاب کنیم که متعامد باشد. این شکل به شکل شور حقیقی A (RSF) معروف است.

قضیه ۸-۹-۲ قضیه مثلثی‌سازی شور حقیقی. فرض کنید A یک ماتریس $n \times n$ حقیقی باشد، آنگاه یک ماتریس $n \times n$ متعامد Q وجود دارد به قسمی که

$$Q^T A Q = R = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1k} \\ \cdot & R_{22} & \dots & R_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \dots & R_{kk} \end{bmatrix}$$

که در آن R_{ii} یک اسکالر یا یک ماتریس 2×2 می‌باشد. عناصر قطری اسکالر متناظر با مقادیر ویژه حقیقی هستند و ماتریسهای 2×2 واقع بر روی قطر متناظر با مقادیر ویژه مزدوج مختلط می‌باشند.

اثبات: یک شرح مختصر از اثبات را در اینجا بیان خواهیم کرد. جزئیات آن را می‌توانید در مرجع Golub and Van Loan (1989, 362) پیدا کنید.

قضیه به استقرا ثابت می‌شود. فرض کنید که s مقدار ویژه مختلط برای ماتریس A وجود داشته

باشد. برای $\lambda = 0$ قضیه از قضیه ۸-۹-۱ نتیجه می‌شود. (توجه کنید که چون همه مقادیر ویژه حقیقی هستند، U در قضیه ۸-۹-۱ متعامد حقیقی است).

اکنون $\lambda \geq 1$ را در نظر می‌گیریم. چون مقادیر ویژه مختلط به صورت جفت ظاهر می‌شوند، به سادگی ملاحظه می‌شود که یک ماتریس $n \times n$ متعامد حقیقی U_1 وجود دارد به قسمی که

$$U_1^T A U_1 = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

که در آن A_{11} یک ماتریس 2×2 است و یک جفت مقدار ویژه مزدوج مختلط دارد. به طور مشابه یک ماتریس U_2 متعامد از مرتبه $(n - 2)$ وجود دارد به قسمی که $U_2^T A_{22} U_2$ دارای ساختار فوق باشد.

اکنون قرار دهید

$$Q = U_1 \begin{bmatrix} I_2 & 0 \\ 0 & U_2 \end{bmatrix}$$

تعریف ۸-۹-۱ ماتریس R قضیه ۸-۹-۲ به شکل شور حقیقی (RSF) ماتریس A معروف است. تذکرها:

۱. معمولاً ماتریسهای 2×2 واقع بر روی قطر «دست‌انداز» نامیده می‌شوند.

۲. RSF و شکل شور (T قضیه ۸-۹-۱) منحصر به فرد نیستند.

تبصره: چون اثباتهای هر دو قضیه بر شناخت مقادیر ویژه و بردارهای ویژه ماتریس مبتنی هستند، اثبات‌ها نمی‌توانند به صورت سازنده مورد ملاحظه قرار گیرند. آنها به ما در محاسبه مقادیر ویژه و بردارهای ویژه کمکی نمی‌کنند.

در اینجا ما یک روش معروف به روش تکرار QR را برای محاسبه شکل شور حقیقی A ارائه می‌دهیم. امروزه یک پیاده‌سازی خاص از روی QR به طور وسیع برای محاسبه مقادیر ویژه یک ماتریس دلخواه مورد استفاده قرار می‌گیرد. همان‌گونه که نام روش اظهار می‌دارد، روش بر تجزیه QR مبتنی است و طبیعتی تکراری دارد. روش تکرار QR توسط جی. جی فرانسیس^۱ (۱۹۶۱) به شکل الگوریتمی پیشنهاد شده است، اگرچه ریشه آن را می‌توان به کار اچ. روتیس‌هازر^۲ (۱۹۵۸) ارتباط داد. همچنین روش به‌طور مستقل توسط ریاضیدان روسی وی. ان. کوبلاتوفسکایا^۳ (۱۹۶۱) ارائه شده است.

تذکر: چون مقادیر ویژه A ، n صفر چندجمله‌ای مشخصه هستند و به خوبی می‌دانیم (توسط آبل و گالوا بیش از یک قرن قبل ثابت شده است) که ریشه‌های یک معادله چندجمله‌ای دلخواه از درجه بالاتر از چهار نمی‌توانند در تعداد متناهی گام به دست آیند، هر روش مقدار ویژه عددی برای یک ماتریس دلخواه باید دارای طبیعتی تکراری باشد.

۸-۹-۱ تکرار QR پایه‌ای

ابتدا روش تکرار QR پایه‌ای را معرفی می‌کنیم.

$$A = A_0 \text{ قرار می‌دهیم}$$

حال یک دنباله از ماتریسها (A_k ها) را که به صورت زیر تعریف می‌شوند، محاسبه می‌کنیم

$$A_0 = Q_0 R_0$$

$$A_1 = R_0 Q_0 = R_1 Q_1$$

$$A_2 = R_1 Q_1 = Q_2 R_2$$

و غیره. به طور کلی

$$A_k = Q_k R_k = R_{k-1} Q_{k-1} \quad k = 1, 2,$$

ماتریسها در دنباله $\{A_k\}$ یک خاصیت خیلی جالب دارند. هر ماتریس در دنباله به طور متعامد متشابه با ماتریس قبلی خود است، در نتیجه به طور متعامد متشابه با ماتریس اصلی است. این مطلب به آسانی ملاحظه می‌شود. برای مثال

$$A_1 = R_0 Q_0 = Q_0^T A_0 Q_0 \quad (\text{زیرا } Q_0^T A_0 = R_0)$$

$$A_2 = R_1 Q_1 = Q_1^T A_1 Q_1$$

بنابراین A_1 با A به طور متعامد متشابه است، و A_2 به طور متعامد با A_1 متشابه است. بنابراین همان‌گونه که محاسبه زیر نشان می‌دهد ماتریس A_2 با A به طور متعامد متشابه است.

$$A_2 = Q_1^T A_1 Q_1 = Q_1^T (Q_0^T A_0 Q_0) Q_1 = (Q_0 Q_1)^T A (Q_0 Q_1)$$

چون هر ماتریس با ماتریس اصلی A به طور متعامد متشابه است، پس دارای همان مقادیر ویژه A خواهد بود، در نتیجه اگر دنباله $\{A_k\}$ به یک ماتریس مثلثی یا شبه‌مثلثی همگرا گردد، مقادیر ویژه را خواهیم داشت. نتیجه زیر نشان می‌دهد که تحت شرایط خاصی همگرایی رخ خواهد داد (Wilkinson (1965, 518-519)).

شرطی برای همگرایی

قضیه ۸-۹-۳ یک قضیه همگرایی برای تکرار QR پایه‌ای. فرض کنید که مقادیر ویژه $\lambda_1, \lambda_2, \dots, \lambda_n$ به قسمی باشند که $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$ ، و فرض کنید که ماتریس بردارهای ویژه Y از بردارهای ویژه چپ (یعنی $Y = (X^T)^{-1}$) به قسمی باشد که کهادهای اصلی پیشرو آن مخالف صفر باشند، آنگاه $\{A_k\}$ به یک ماتریس بالامثلثی یا به شکل شور حقیقی میل خواهد کرد.

در حقیقت می‌توان نشان داد که تحت شرایط قبلی، اولین ستون A_k به مضربی از e_1 میل خواهد کرد. بنابراین برای k به اندازه کافی بزرگ به دست می‌آوریم

$$A_k = \begin{bmatrix} \lambda_1 & u \\ \cdot & \hat{A}_k \end{bmatrix}$$

می‌توانیم تکرار QR را دوباره بر \hat{A}_k اعمال کنیم و فرایند می‌تواند برای دیدن این که دنباله به یک ماتریس بالامثلثی همگرا می‌گردد، ادامه یابد.

مثال ۸-۹-۲

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

مقادیر ویژه A عبارتند از $5/3723$ و $-0/3723$ ؛ $|\lambda_1| > |\lambda_2|$.

$k = 0$

$$A_0 = A = Q_0 R_0$$

$$Q_0 = \begin{bmatrix} -0/3162 & -0/9487 \\ -0/9487 & 0/3162 \end{bmatrix}$$

$$R_0 = \begin{bmatrix} -3/1623 & -4/4222 \\ \cdot & -0/6325 \end{bmatrix}$$

$k = 1$

$$A_1 = R_0 Q_0 = \begin{bmatrix} 5/2 & 1/6 \\ 0/6 & -0/2 \end{bmatrix} = Q_1 R_1$$

$$Q_1 = \begin{bmatrix} -0/9934 & -0/1146 \\ -0/1146 & -0/9934 \end{bmatrix}$$

$$R_1 = \begin{bmatrix} -5/2345 & -1/5665 \\ \cdot & -0/3821 \end{bmatrix}$$

$k = 2$

$$A_2 = R_1 Q_1 = \begin{bmatrix} 5/3796 & -0/9562 \\ 0/0438 & -0/3796 \end{bmatrix} = Q_2 R_2$$

(توجه کنید که تاکنون در جهت به دست آوردن مقادیر ویژه مقادیر ویژه مقادیر پیشرفت کرده‌ایم.)

$$Q_2 = \begin{bmatrix} -1 & -0/0082 \\ -0/0081 & 1 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} -5/3797 & 0/9593 \\ . & -0/3718 \end{bmatrix}$$

$k = 3$

$$A_3 = R_2 Q_2 = \begin{bmatrix} 5/3718 & 1/0030 \\ 0/0030 & -0/3718 \end{bmatrix} = Q_3 R_3$$

$$Q_3 = \begin{bmatrix} -1 & -0/0006 \\ -0/0006 & 1 \end{bmatrix}$$

$$R_3 = \begin{bmatrix} -5/3718 & -1/0028 \\ . & -0/3723 \end{bmatrix}$$

$k = 4$

$$A_4 = R_3 Q_3 = \begin{bmatrix} 5/3723 & -0/9998 \\ 0/0002 & -0/3723 \end{bmatrix}$$

تذکر MATCOM: تکرار QR پایه‌ای به صورتی که در بالا توصیف شد در برنامه QRITRB در MATCOM پیاده‌سازی شده است.

۸-۹-۲ تکرار QR هسنبرگی

روش تکراری QR توصیف شده در فوق کارا نیست اگر ماتریس A پر و متراکم باشد. قبلاً دیده‌ایم که تجزیه QR یک چنین ماتریس A بی نیاز به $O(n^3)$ عمل ممیز شناور (flops) دارد؛ بنابراین n تکرار روش QR به $O(n^4)$ عمل ممیز شناور (flops) نیاز خواهد شد و این مطلب روش را غیر عملی می‌سازد.

خوشبختانه عمل ساده‌ای می‌تواند انجام شود.

ماتریس A را قبل از شروع تکرارهای QR توسط تبدیلات تشابهی متعامد به یک ماتریس هسنبِرگی کاهش می‌دهیم. یک پیامد عملی جالب از این عمل این است که اگر $A = A_0$ در ابتدا به یک ماتریس بالا هسنبِرگی کاهش یابد و فرض شود که کاهش ناپذیر باشد، آنگاه هر عنصر از دنباله $\{A_k\}$ نیز بالا هسنبِرگی می‌باشد.

این نتیجه می‌تواند به صورت زیر ملاحظه شود. فرض کنید A_k یک ماتریس بالا هسنبِرگی کاهش ناپذیر باشد و دورانهای گیونز برای تجزیه A_k به صورت $Q_k R_k$ استفاده شوند. آنگاه

$$Q_k = (J(n-1, n, \theta) \dots J(2, 3, \theta) J(1, 2, \theta))^T$$

نیز یک ماتریس بالا هسنبِرگی است زیرا $J(i, i+1, \theta)$ فقط عناصر غیر قطری در مکانهای $(i+1, i)$ و $(i, i+1)$ دارد.

چون R_k بالا مثلثی است، $A_{k+1} = R_k Q_k$ نیز بالا هسنبِرگی است.

چون تجزیه QR یک ماتریس بالا هسنبِرگی فقط نیاز به $O(n^2)$ عمل ممیز شناور (flops) دارد روش تکرار QR با کاهش اولیه A به هسنبِرگی توسط $O(n^3)$ کراندار خواهد شد و به روشی کارا تبدیل خواهد گردید.

مثال ۸-۹-۳

$$A = A_0 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \\ 0 & 1 & 1 \end{bmatrix}$$

$$Q_1 = \begin{bmatrix} -0.4472 & 0.3651 & -0.8165 \\ -0.8944 & -0.1826 & 0.4082 \\ 0 & 0.9129 & 0.4082 \end{bmatrix}$$

$$R_1 = \begin{bmatrix} -2/2361 & -3/5777 & -5/8138 \\ 0 & 1/0.954 & 1/0.954 \\ 0 & 0 & -0.0000 \end{bmatrix}$$

$k = 1$

$$A_1 = R_1 Q_1 = \begin{bmatrix} 4/2000 & -5/4705 & -2/0083 \\ -0/9798 & 0/8000 & 0/8944 \\ . & -0/0000 & -0/0000 \end{bmatrix}$$

$$Q_1 = \begin{bmatrix} -0/9739 & -0/2272 & -0/0000 \\ 0/2272 & -0/9739 & -0/0000 \\ . & -0/0000 & 1/0000 \end{bmatrix}$$

$$R_1 = \begin{bmatrix} -4/3128 & 5/5092 & 2/1590 \\ . & 0/4637 & -0/4148 \\ . & -0/0000 & -0/0000 \end{bmatrix}$$

$k = 2$

$$A_2 = R_1 Q_1 = \begin{bmatrix} 5/4516 & -4/3854 & 2/1590 \\ 0/1054 & -0/4516 & -0/4148 \\ . & . & . \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} -0/9998 & 0/0193 & -0/0000 \\ -0/0193 & -0/9998 & 0/0000 \\ . & . & 1/0000 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} -5/4526 & 4/3933 & -2/1506 \\ . & 0/3668 & 0/4564 \\ . & . & -0/0000 \end{bmatrix}$$

توجه کنید که هر Q_i و هر A_i بالا هسنبرگی هستند. همچنین مقادیر ویژه A_2 عبارتند از $5/3723$ ، 0 و $-0/3722$ و مقادیر ویژه A عبارتند از $5/3723$ ، 0 ، $-0/3723$.

تذکر MATCOM: تکرار QR هسنبرگی در برنامه QRITRH از MATCOM پیاده سازی شده است.

۸-۹-۳ همگرایی تکرارهای QR و انتقال از مبدأ

به محض این که دانستیم که روش تکرار QR می تواند با استفاده از کاهش اولیه A به یک ماتریس هسنبرگی، کارا گردد؛ طبیعی است در مورد نرخ همگرایی تکرارها متعجب شویم. منظورمان از نرخ همگرایی این است که با چه سرعتی عناصر زیر قطر ماتریس هسنبرگی تبدیل یافته به سمت صفر میل

می‌کنند.

فرض کنید A ابتدا به یک ماتریس بالا هسنبرگی H کاهش یابد؛ تکرار QR را بر H اعمال می‌کنیم تا دنباله H_k به دست آید. می‌توان نشان داد (Golub and Van Loan (1989, 374)) که عناصر زیر قطر اصلی $h_{i,i-1}^{(k)}$ از ماتریس H_k با یک نرخ همگرایی که توسط نسبت

$$\left| \frac{\lambda_i}{\lambda_i - 1} \right|^k$$

تعیین می‌شود به صفر میل می‌کند. بنابراین نرخ همگرایی خیلی آهسته خواهد بود اگر قدر مطلق دو مقدار ویژه λ_i و λ_{i-1} خیلی به یکدیگر نزدیک باشند. خوشبختانه نرخ همگرایی می‌تواند به طور بامعنایی با انتقال مبدأ بهبود یابد.

فرض کنید $\tilde{\lambda}_i$ یک تقریب از یک مقدار ویژه λ_i ماتریس H باشد. فرض کنید تکرار QR بر ماتریس

$$\tilde{H} = H - \tilde{\lambda}_i I$$

اعمال شود. مقادیر ویژه \tilde{H} عبارتند از $\lambda_1 - \tilde{\lambda}_i, \lambda_2 - \tilde{\lambda}_i, \dots, \lambda_n - \tilde{\lambda}_i$. فرض کنید این مقادیر ویژه طوری مرتب شوند که

$$|\lambda_1 - \tilde{\lambda}_i| > |\lambda_2 - \tilde{\lambda}_i| > \dots > |\lambda_n - \tilde{\lambda}_i|$$

آنگاه در این حالت $(i-1)$ امین عنصر زیر قطر \tilde{H}_k با یک نرخ همگرایی که توسط نسبت

$$\left| \frac{\lambda_i - \tilde{\lambda}_i}{\lambda_{i-1} - \tilde{\lambda}_i} \right|^k$$

به جای $|\lambda_i / \lambda_{i-1}|^k$ تعیین خواهد شد به صفر میل خواهد کرد. معمولاً نسبت اولی کوچک‌تر از دومی است.

مثال قانع‌کننده زیر را از مرجع Ortega and Poole (1981, 227) در نظر بگیرید.

فرض کنید $\lambda_2 = 0/99, \lambda_1 = 1/1, \lambda_i = 1, \tilde{\lambda}_i = 1$. آنگاه

$$\left| \frac{\lambda_i - \tilde{\lambda}_i}{\lambda_{i-1} - \tilde{\lambda}_i} \right| = 0/1$$

$$\left| \frac{\lambda_i}{\lambda_{i-1}} \right| = 0/99$$

این مشاهده بیان می‌کند که اگر ما تکرار QR را در مورد ماتریس انتقال یافته \hat{H} به جای ماتریس اصلی H به کار ببریم، آنگاه نرخ همگرایی سریعتر خواهد شد. البته به محض یافتن یک مقدار ویژه \hat{H} مقدار ویژه متناظر H را می‌توان با اضافه کردن مقدار انتقال محاسبه کرد.

۴-۹-۸ تکرار QR تک انتقالی

برای پیاده‌سازی تکرار QR تک انتقالی، لازم است که یک تقریب \hat{H} از مقدار ویژه λ داشته باشیم. چون \hat{H} در عمل موجود نیست، $h_{nn}^{(0)}$ ، n امین عنصر $H = H_0$ به عنوان یک تقریب اولیه اختیار می‌شود و سپس تکرار با استفاده از عنصر (n, n) ام ماتریس جاری به عنوان یک انتقال جدید ادامه می‌یابد. از بحث قبل روش تکرار QR اصلاح شده معروف به روش تکرار QR تک انتقالی زیر نتیجه می‌شود.

گام ۱: A را به یک ماتریس بالا هستبرگی H تبدیل کنید.

گام ۲: قرار دهید $H_0 = H$

برای $k = 0, 1, 2, \dots$ تا این که $h_{n,n-1}^{(k)}$ از لحاظ محاسباتی صفر شود، انجام دهید.

$$H_k - h_{nn}^{(k)} I = Q_k R_k$$

$$H_{k+1} = R_k Q_k + h_{nn}^{(k)} I$$

در این الگوریتم $h_{ij}^{(k)}$ عنصر (i, j) ام ماتریس H_k را نشان می‌دهد. البته هر یک از ماتریسهای $\{H_k\}$ می‌توانند بر روی H نوشته شوند.

همگرایی. نرخ همگرایی روش تکراری تک انتقالی QR سرانجام از درجه دو می‌باشد (مرجع Stewart (1973, 366-7) را ملاحظه کنید).

تبصره: یک گونه ضمنی از تکرار تک انتقالی معروف به تکرار QR ضمنی می‌توان به دست آورد که در آن انتقالها را کم نمی‌کنیم اما به طور ضمنی ماتریس H_{k+1} را می‌سازیم. این روش در تمرین ۲۹ توضیح داده شده است.

مثال ۴-۹-۸ تکرار QR تک انتقالی

$$H = H_0 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix}$$

:k = ۰

$$H_0 - h_{33}^{(0)} I = \begin{bmatrix} \cdot & 1 & 1 \\ 1 & 1 & 3 \\ \cdot & 1 & \cdot \end{bmatrix} = Q_0 R_0$$

$$Q_0 = \begin{bmatrix} \cdot & \cdot/\sqrt{0.71} & -\cdot/\sqrt{0.71} \\ -1 & \cdot & \cdot \\ \cdot & \cdot/\sqrt{0.71} & \cdot/\sqrt{0.71} \end{bmatrix}$$

$$R_0 = \begin{bmatrix} -1/\cdot\cdot\cdot\cdot & -1/\cdot\cdot\cdot\cdot & -3/\cdot\cdot\cdot\cdot \\ \cdot & 1/4142 & \cdot/\sqrt{0.71} \\ \cdot & \cdot & -\cdot/\sqrt{0.71} \end{bmatrix}$$

$$H_1 = R_0 Q_0 + h_{33}^{(0)} I = \begin{bmatrix} 2/\cdot\cdot\cdot & -2/8284 & -1/4142 \\ -1/4142 & 1/5\cdot\cdot\cdot & \cdot/5\cdot\cdot\cdot \\ \cdot & -\cdot/5\cdot\cdot\cdot & \cdot/5\cdot\cdot\cdot \end{bmatrix}$$

:k = ۱

$$H_1 - h_{33}^{(1)} I = Q_1 R_1$$

$$Q_1 = \begin{bmatrix} -\cdot/\sqrt{276} & -\cdot/6342 & -\cdot/2615 \\ \cdot/686\cdot & -\cdot/6727 & -\cdot/2774 \\ \cdot & -\cdot/3812 & \cdot/9245 \end{bmatrix}$$

$$R_1 = \begin{bmatrix} -2/\cdot616 & 2/744\cdot & 1/372\cdot \\ \cdot & 1/3117 & \cdot/56\cdot6 \\ \cdot & \cdot & \cdot/2311 \end{bmatrix}$$

$$H_2 = R_1 Q_1 + h_{33}^{(1)} I = \begin{bmatrix} 3/8824 & -1/\cdot613 & 1/\cdot464 \\ \cdot/8998 & -\cdot/596\cdot & \cdot/1544 \\ \cdot & -\cdot/\cdot881 & \cdot/7137 \end{bmatrix}$$

$k = 2$

$$H_2 - h_{22}^{(2)} I = Q_2 R_2$$

$$Q_2 = \begin{bmatrix} -0./962. & 0./2721 & 0./247 \\ -0./2732 & -0./958. & -0./087. \\ . & -0./905 & 0./9959 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} -2./294. & 1./3787 & -1./488 \\ . & 0./974. & 0./1367 \\ . & . & 0./124 \end{bmatrix}$$

$$H_2 = R_2 Q_2 + h_{22}^{(2)} I = \begin{bmatrix} 2./5057 & -2./1221 & -1./2459 \\ -0./2661 & -0./2318 & 0./0514 \\ . & -0./0011 & 0./726. \end{bmatrix}$$

$k = 3$

$$H_3 - h_{33}^{(3)} I = Q_3 R_3$$

$$Q_3 = \begin{bmatrix} -0./9955 & -0./953 & -0./001 \\ 0./953 & -0./9955 & -0./0010 \\ . & -0./0010 & 1./000 \end{bmatrix}$$

$$R_3 = \begin{bmatrix} -2./7924 & 2./212 & 1./2451 \\ . & 1./1556 & 0./675 \\ . & . & 0./0001 \end{bmatrix}$$

$$H_3 = R_3 Q_3 + h_{33}^{(3)} I = \begin{bmatrix} 2./6983 & -1./7472 & 1./2434 \\ 0./1101 & -0./2244 & 0./664 \\ . & 0./0000 & 0./7261 \end{bmatrix}$$

واضح است که تکرار به سمت مقدار ویژه $0./7261$ میل می‌کند. (مقادیر ویژه H با چهار رقم اعشار عبارتند از $0./7261$ ، $3./6511$ و $-0./3772$).

تذکر MATCOM: تکرار QR تک انتقالی صریح در برنامه QRITRSE در MATCOM پیاده‌سازی شده است.

۵-۹-۸ تکرار QR با انتقال مضاعف

اگر مقادیر ویژه حقیقی باشند تکرار تک انتقالی قبلی خوب کار می‌کند. ولیکن یک ماتریس حقیقی می‌تواند دارای مقادیر ویژه مختلط باشد. در یک چنین حالتی،

۱. در مرحله‌ای از تکرار ممکن است به یک ماتریس 2×2 پسر و بر روی گوشه راست انتهایی برخورد نماییم که متناظر با یک جفت مقدار ویژه مزدوج باشد اگر چنین مقادیر ویژه‌ای وجود داشته باشند.
۲. عنصر (n, n) ام ماتریس 2×2 پسر و که حقیقی است، یک تقریب خوب نخواهد بود.
۳. طبیعی است که از مقادیر ویژه ماتریس 2×2 به عنوان پارامترهای انتقال استفاده نماییم. این کار تکرار QR با انتقال مضاعف را نتیجه می‌دهد.

یک گام تکراری از QR با انتقال مضاعف (مختلط)

فرض کنید که مقادیر ویژه ماتریس 2×2 گوشه راست انتهایی ماتریس هسنبرگی H_s مقادیر k_1 و $k_2 = \bar{k}_1$ باشند. آنگاه یک گام تکراری از تکرار QR با انتقال مضاعف به صورت زیر خواهد بود.

$$\begin{aligned} H_s - k_1 I &= Q_s R_s, & H_{s+1} &= R_s Q_s + k_1 I \\ H_{s+1} - k_2 I &= Q_{s+1} R_{s+1}, & H_{s+2} &= R_{s+1} Q_{s+1} + k_2 I \end{aligned}$$

مثال ۵-۹-۸

$$H_s = H = \begin{bmatrix} 1 & 2 & 2 \\ \cdot & \cdot & 1 \\ \cdot & -1 & \cdot \end{bmatrix}$$

$$k_1 = i, \quad k_2 = -i$$

$$: H_s - k_1 I = Q_s R_s$$

$$Q_s = \begin{bmatrix} -1 & \cdot & \cdot \\ \cdot & -\sqrt{0.71} & -\sqrt{0.71}i \\ \cdot & \sqrt{0.71}i & \sqrt{0.71} \end{bmatrix}$$

$$R_s = \begin{bmatrix} -1+i & -2 & -2 \\ \cdot & \sqrt{4142}i & -\sqrt{4142} \\ \cdot & \cdot & -1 \end{bmatrix}$$

$$H_1 = R_s Q_s + k_1 I = \begin{bmatrix} 1 & \sqrt{4142} - \sqrt{4142}i & -\sqrt{4142} + \sqrt{4142}i \\ \cdot & -i & \cdot \\ \cdot & \cdot & i \end{bmatrix}$$

$$:H_1 - k_p I = Q_1 R_1$$

$$Q_1 = \begin{bmatrix} -1 & \cdot & \cdot \\ \cdot & -\cdot/۸۱۶۵ & -\cdot/۵۷۴۴i \\ \cdot & \cdot/۵۷۴۴i & \cdot/۸۱۶۵ \end{bmatrix}$$

$$R_1 = \begin{bmatrix} -1-i & -1/۴۱۴۲ + 1/۴۱۴۲i & 1/۴۱۴۲ - 1/۴۱۴۲i \\ \cdot & \cdot & 1/۱۵۴۷ \\ \cdot & \cdot & 1/۶۳۳\cdot i \end{bmatrix}$$

$$H_p = R_1 Q_1 + k_p I = \begin{bmatrix} 1 & 1/۹۷۱۲ - \cdot/۳۳۸۲i & 1/۹۷۱۲ - \cdot/۳۳۸۲i \\ \cdot & -\cdot/۳۳۳۳i & \cdot/۹۴۲۸ \\ \cdot & -\cdot/۹۴۲۸ & \cdot/۳۳۳۳i \end{bmatrix}$$

توجه کنید که مقادیر ویژه ماتریس ۲×۲ گوشه راست انتهایی

$$\begin{bmatrix} -\cdot/۳۳۳۳i & \cdot/۹۴۲۸ \\ -\cdot/۹۴۲۸ & \cdot/۳۳۳۳i \end{bmatrix}$$

عبارتند از i و $-i$.

بنابراین مقادیر ویژه H_p عبارتند از ۱ ، i و $-i$.

پرهیز از حساب مختلط در تکرار QR با انتقال مضاعف

چون k_1 و k_p می توانند مختلط باشند، گام تکراری QR با انتقال مضاعف قبلی ممکن است برای پیاده سازی نیاز به حساب مختلط داشته باشد، حتی اگر ماتریس اولیه H_s حقیقی باشد. ولیکن با کمی انجام اعمال حساب مختلط می توان از آن پرهیز کرد. این وضع را اکنون توضیح خواهیم داد.

نشان می دهیم که ماتریس H_{s+p} از طریق یک ماتریس تبدیل حقیقی با H_s به طور متعامد متشابه می باشد، و می تواند به طور مستقیم از روی H_s بدون محاسبه H_{s+1} محاسبه شود. ماتریس زیر را در نظر بگیرید

$$N = (H_s - k_p I)(H_s - k_1 I) = H_s^2 - (k_1 + k_p)H_s + k_1 k_p I$$

چون $k_p = \bar{k}_1$ ، ماتریس N حقیقی است. اکنون نشان می دهیم که $(Q_s Q_{s+1})(R_{s+1} R_s)$ تجزیه QR ماتریس N هستند.

$$\begin{aligned}
 N &= (H_s - k_\gamma I)(H_s - k_1 I) \\
 &= (H_s - k_\gamma I)Q_s R_s \\
 &= Q_s Q_s^* (H_s - k_\gamma I)Q_s R_s \\
 &= Q_s (H_{s+1} - k_\gamma I)R_s = Q_s Q_{s+1} R_{s+1} R_s
 \end{aligned}$$

چون N حقیقی است و $(Q_s Q_{s+1})(R_{s+1} R_s)$ تجزیه QR ماتریس N است، ماتریس $Q_s Q_{s+1}$ می تواند طوری انتخاب شود که حقیقی و متعامد باشد.

بالاخره نشان می دهیم که H_{s+2} از طریق این ماتریس تبدیل حقیقی $Q_s Q_{s+1}$ با ماتریس H_s به طور متعامد متشابه است.

$$\begin{aligned}
 H_{s+2} &= R_{s+1} Q_{s+1} + k_\gamma I \\
 &= Q_{s+1}^* (H_{s+1} - k_\gamma I) Q_{s+1} + k_\gamma I \\
 &= Q_{s+1}^* (R_s Q_s + (k_1 - k_\gamma) I) Q_{s+1} + k_\gamma I \\
 &= Q_{s+1}^* [Q_s^* (H_s - k_1 I) Q_s + (k_1 - k_\gamma) I] Q_{s+1} + k_\gamma I \\
 &= Q_{s+1}^* Q_s^* H_s Q_s Q_{s+1} \\
 &= (Q_s Q_{s+1})^T H_s (Q_s Q_{s+1})
 \end{aligned}$$

چون $Q_s Q_{s+1}$ و H_s حقیقی هستند H_{s+2} نیز حقیقی است.

از بحثهای قبل نتیجه می گیریم که می توان H_{s+2} را مستقیماً از روی H_s و به کمک تبدیلهای متعامد حقیقی به دست آورد.

مقادیر ویژه k_1 و k_γ لازم نیست به طور صریح محاسبه شوند.

اگرچه محاسبه مقادیر ویژه ماتریس 2×2 تقریباً یک عمل بدیهی است ولیکن توجه می کنیم

که k_1 و k_γ لازم نیست به طور صریح محاسبه شوند. برای تشکیل ماتریس

$$N = (H_s - k_1 I)(H_s - k_\gamma I) = H_s^2 - (k_1 + k_\gamma)H_s + k_1 k_\gamma I$$

همه آنچه را که ما برای محاسبه نیاز داریم اثر و دترمینان ماتریس 2×2 می باشد.

فرض کنید

$$\begin{bmatrix} h_{n-1, n-1} & h_{n-1, n} \\ h_{n, n-1} & h_{nn} \end{bmatrix}$$

ماتریس 2×2 گوشه راست ماتریس جاری H_s باشد. آنگاه

$$t = k_1 + k_2 = \text{مجموع مقادیر ویژه}$$

$$= h_{n-1, n-1} + h_{nn}, \text{ حقیقی است}$$

$$d = k_1 k_2 = \text{حاصلضرب مقادیر ویژه}$$

$$= h_{n-1, n-1} h_{nn} - h_{n, n-1} h_{n-1, n}, \text{ حقیقی است}$$

این به ما اجازه می‌دهد که یک گام از تکرار QR با انتقال مضاعف رادر حساب حقیقی به صورت زیر

بنویسیم

یک گام از تکرار QR با انتقال مضاعف (حساب حقیقی)

$$1. \text{ تشکیل ماتریس } N = H_s^T - tH_s + dI$$

$$2. \text{ پیدا کردن تجزیه } QR \text{ ماتریس } N: N = QR$$

$$3. \text{ تشکیل ماتریس } H_{s+2} = Q^T H_s Q$$

ما این محاسبات را به علتی که در بخش بعد ذکر خواهیم کرد، تکرار QR با انتقال مضاعف صریح

می‌نامیم.

مثال ۸-۹-۶

$$H = H_0 = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & -2 & 2 \end{bmatrix}$$

$$t = 2, \quad d = 2$$

$$N = H^T - tH + dI = \begin{bmatrix} 3 & -8 & 5 \\ -1 & 2 & 3 \\ -2 & 0 & 0 \end{bmatrix}$$

ابتدا تجزیه QR ماتریس N را پیدا می‌کنیم

$$Q = \begin{bmatrix} -0./8081 & -0./5470 & -0./2408 \\ 0./2673 & 0./3222 & -0./9631 \\ 0./5345 & -0./8365 & 0./1204 \end{bmatrix}$$

$$H_7 = Q^T H \cdot Q = \begin{bmatrix} -0./8571 & 1/1007 & 2/5740 \\ -1/1867 & 3/0455 & -0./8289 \\ 0./0000 & 1/8437 & 0./8116 \end{bmatrix}$$

تذکر MATCOM: تکرار QR با انتقال مضاعف صریح در برنامه QRITRDSE در MATCOM پیاده‌سازی شده است.

۸-۹-۶ تکرار QR ضمنی

بعد از همه اینها، با ناامیدی تمام متذکر می‌شویم که تکرار QR با انتقال مضاعف قبلی (صریح) عملی نیست. زیرا تشکیل خود ماتریس N در گام ۱ نیاز به $O(n^3)$ عمل ممیز شناور دارد. خوشبختانه یک عمل زیرکانه به ما اجازه می‌دهد که این گام را با $O(n^2)$ عمل ممیز شناور پیاده‌سازی کنیم.

یک تکرار از QR ضمنی با انتقال مضاعف

۱. محاسبه اولین ستون n_1 از ماتریس N .

۲. پیدا کردن یک ماتریس هاوس هولدر P_0 به قسمی که $P_0 n_1$ مضربی از e_1 باشد.

۳. تشکیل ماتریس $H'_S = P_0^T H_S P_0$.

۴. تبدیل H'_S به یک ماتریس بالا هسنبرگی با استفاده از تبدیلات تشابهی متعامد با استفاده

از ماتریسهای هاوس هولدر

$$(P_{n-2}^T \cdots P_2^T P_1^T) H'_S (P_1 P_2 \cdots P_{n-2}) = H'_{S+2}$$

با استفاده از قضیه Q ضمنی (قضیه ۵-۵-۳) می‌توان نشان داد که ماتریس H'_{S+2} حاصل از روش

QR صریح و H'_{S+2} حاصل از QR ضمنی هر دو ماتریسهای بالا هسنبرگی کاهش‌ناپذیر هستند و

اساساً یک ماتریس می‌باشند.

چهار گام قبلی یک تکرار از روش QR ضمنی با انتقال مضاعف را تشکیل می‌دهد. برای فهمیدن این که چگونه این چهار گام می‌توانند در $O(n^2)$ عمل ممیز شناور انجام شوند، ما باید به ساختارهای این محاسبات نگاه کنیم.

اولاً چون H_s یک ماتریس بالا هسنبرگی است، محاسبه ستون اول $N = (H_s - k_1 I)(H_s - k_2 I)$ تقریباً بدیهی است. این ستون فقط سه عنصر مخالف صفر دارد. آنها می‌توانند به صورت صریح نوشته شوند. فرض کنید

$$Ne_1 = n_1 = \begin{bmatrix} n_{11} \\ n_{21} \\ n_{31} \\ \vdots \\ \cdot \end{bmatrix}$$

پس

$$n_{11} = h_{11}^2 - th_{11} + d + h_{12}h_{21}$$

$$n_{21} = h_{21}(h_{11} + h_{22} - t)$$

$$n_{31} = h_{21}h_{32}$$

در اینجا h_{ij} به عنصر (i, j) ماتریس H_s اطلاق می‌شود. ثانیاً چون فقط سه عنصر از n_1 مخالف صفر هستند ماتریس هاوس هولدر P_* دارای شکل زیر است

$$P_* = \begin{bmatrix} \hat{P}_* & \cdot \\ \cdot & I_{n-3} \end{bmatrix}$$

که در آن \hat{P}_* یک ماتریس هاوس هولدر 3×3 است. به واسطه این شکل P_* و همچنین هسنبرگی بودن ماتریس H_s ، ماتریس $H_s' = P_* H_s P_*$ یک ماتریس پر نیست. یک ماتریس هسنبرگی با یک پرشدگی است. مثلاً برای $n = 6$ داریم

$$H_s' = P_* H_s P_* = \begin{bmatrix} * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ \cdot & \cdot & \cdot & * & * & * \\ \cdot & \cdot & \cdot & \cdot & * & * \end{bmatrix}$$

یک پرشدگی در هر گام کاهش H_k' به شکل هسنبرگی ایجاد خواهد شد و ساختن ماتریسهای هاوس هولدر P_1 تا P_{n-2} جهت از بین بردن این پرشدگیها به طور منظم انجام خواهند شد. هر P_k ، $k = 1, 2, \dots, n-3$ دارای شکل زیر است.

$$P_k = \begin{bmatrix} I_k & & \\ & \hat{P}_k & \\ & & I_{n-k-2} \end{bmatrix}$$

که در آن یک ماتریس هاوس هولدر 3×3 است. آخرین ماتریس هاوس هولدر، P_{n-2} ، دارای شکل زیر است.

$$P_{n-2} = \begin{bmatrix} I_{n-2} & \circ \\ \circ & \hat{P}_{n-2} \end{bmatrix}$$

با در نظر گرفتن ساختارهای محاسبات قبلی، می توان نشان داد که یک گام از تکرار QR ضمنی نیاز به $O(n^2)$ عمل ممیز شناور دارد.

برای جزئیات بیشتر این محاسبات از مرتبه $O(n^2)$ یک تکرار QR ضمنی با انتقال مضاعف کتاب Stewart (1973, 375-8) و کتاب جدید Watkins (1991, 277-8) را ملاحظه کنید.

الگوریتم ۸-۹-۱ یک گام تکراری از QR ضمنی با انتقال مضاعف. فرض کنید H یک ماتریس بالا هسنبرگی کاهش ناپذیر $n \times n$ باشد. آنگاه الگوریتم زیر یک گام از تکرار QR ضمنی با انتقال مضاعف را انجام می دهد، ماتریسهای متعامد P_1, P_2, \dots, P_{n-2} را به قسمی تولید می کند که ماتریس $Q^T H Q$ ، که در آن $Q = P_1 \dots P_{n-2}$ ، یک ماتریس بالا هسنبرگی باشد. این الگوریتم بر روی ماتریس H را می نویسد.

۱: انتقالها را محاسبه کنید:

$$t = h_{n-1, n-1} + h_{nn}$$

$$d = h_{n-1, n-1} h_{nn} - h_{n, n-1} h_{n-1, n}$$

۲: سه عنصر اول مخالف صفر ماتریس $N = H^2 - tH + dI$ را محاسبه کنید:

$$x = n_{11} = h_{11}^2 - t h_{11} + d + h_{12} h_{21}$$

$$y = n_{21} = h_{21} (h_{11} + h_{22} - t)$$

$$z = n_{31} = h_{21} h_{32}$$

۳: ماتریسهای هاوس هولدر P_0, P_1, \dots, P_{n-2} را به قسمی محاسبه کنید که ماتریس نهایی بالا هسنبرگی باشد.

۳(الف) برای $k = 1, 2, \dots, n-3$ اعمال زیر را انجام دهید.
 یک ماتریس 3×3 هاوس هولدر \hat{P}_k به قسمی پیدا کنید که

$$\hat{P}_k \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} * \\ \cdot \\ \cdot \end{bmatrix}$$

ماتریس $P_k H P_k^T$ را روی H بنویسید، که در آن

$$P_k = \begin{bmatrix} I_k & & \cdot \\ & \hat{P}_k & \\ \cdot & & I_{n-k-3} \end{bmatrix}$$

x, y و z را به هنگام در آوردن

$$x \equiv h_{k+2, k+1}$$

$$y \equiv h_{k+3, k+1}$$

$$z \equiv h_{k+2, k+1} \quad (k < n-3)$$

۳(ب) یک ماتریس هاوس هولدر \hat{P}_{n-2} از مرتبه ۲ به قسمی پیدا کنید که

$$\hat{P}_{n-2} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} * \\ \cdot \end{bmatrix}$$

ماتریس P_{n-2} را به شکل زیر تشکیل دهید

$$P_{n-2} = \begin{bmatrix} I_{n-2} & \cdot \\ \cdot & \hat{P}_{n-2} \end{bmatrix}$$

شمارش اعمال ممیز شناور. یک گام از تکرار QR ضمنی با انتقال مضاعف حدود $6n^2$ عمل ممیز شناور نیاز دارد. اگر ماتریس تبدیل Q مورد نیاز باشد و اندوخته گردد، آنگاه $6n^2$ عمل ممیز شناور (flops) دیگر نیز لازم خواهد بود.

مثال ۸-۹-۷

$$H = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & -2 & 2 \end{bmatrix} \quad , d = 2 \quad , t = 2.1$$

$$z = n_{r1} = -2, y = n_{r1} = -1, x = n_{11} = 3.2 \\ k = 0.3$$

$$P_0 = I - \frac{\gamma uu^T}{u^T u}$$

$$u = \begin{bmatrix} 6/\sqrt{417} \\ -1/\dots \\ -2/\dots \end{bmatrix} \quad \text{که در آن}$$

$$P_0 = \begin{bmatrix} -.18018 & .2673 & .5345 \\ .2673 & .9604 & -.0793 \\ .5345 & -.0793 & .8414 \end{bmatrix}$$

$$H \equiv P_0^T H P_0 = \begin{bmatrix} -.18571 & -2/6248 & -.9733 \\ .0581 & .1666 & 1/9505 \\ 1/1852 & -.7221 & 2/9906 \end{bmatrix}$$

x و y را به هنگام در می آوریم: $x = h_{r1} = 0.0581$, $y = h_{r1} = 1/1852$
ماتریس P_1 را پیدا می کنیم

$$\hat{P}_1 = \begin{bmatrix} .0490 & -.9988 \\ -.9988 & .0490 \end{bmatrix}, \quad \hat{P}_1 \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1/1866 \\ . \end{bmatrix}$$

$$P_1 = \begin{bmatrix} 1 & . & . \\ . & .0490 & -.9988 \\ . & -.9988 & .0490 \end{bmatrix}$$

$$H \equiv P_1^T H P_1 = \begin{bmatrix} -.18571 & 1/1008 & 2/5740 \\ -1/1867 & 3/0455 & -.1289 \\ . & 1/8437 & .1116 \end{bmatrix}$$

توجه کنید که ماتریس H به دست آمده از QR ضمنی همان ماتریس H_2 به دست آمده قبلی در بخش ۸-۹-۵ و توسط روش QR صریح می‌باشد (مثال ۸-۹-۶).

تذکر $MATCOM$: الگوریتم ۸-۹-۱ در برنامه QRITRDSI در $MATCOM$ پیاده سازی شده است.

۸-۹-۷ به دست آوردن شکل شعور حقیقی

۱. ماتریس A را به شکل هسنبرگی تبدیل می‌کنیم (الگوریتم ۵-۴-۴ یا الگوریتم ۵-۵-۵).

۲. با روش QR ضمنی با انتقال مضاعف تکرارها را انجام می‌دهیم (الگوریتم ۸-۹-۱).

معمولاً بعد از دو یا سه گام از روش تکرار QR ضمنی با انتقال مضاعف یک یا دو عنصر (بعضی اوقات تعداد بیشتری) از عناصر زیر قطر ماتریس هسنبرگی به سمت صفر میل می‌کنند. این پدیده یک مقدار ویژه حقیقی و یا یک جفت مقدار ویژه مزدوج مختلط ارائه خواهد داد. به محض این که یک مقدار ویژه حقیقی و یا یک جفت مقدار ویژه مزدوج مختلط محاسبه شدند، آخرین سطر و آخرین ستون در حالت اول و دو سطر آخر و دو ستون آخر در حالت دوم می‌توانند حذف شوند و محاسبه سایر مقادیر ویژه می‌تواند با زیر ماتریس ادامه یابد. این فرایند همچنین معروف به تقلیل است.

توجه کنید که مقادیر ویژه زیر ماتریس تقلیل یافته همچنین مقادیر ویژه ماتریس اصلی هستند. فرض کنید که بلافاصله بعد از تقلیل ماتریس دارای شکل

$$H_k = \begin{bmatrix} A' & C' \\ \cdot & B' \end{bmatrix}$$

باشد که در آن B' یک زیر ماتریس 2×2 و یا یک ماتریس 1×1 است. آنگاه معادله مشخصه H_k به صورت زیر است

$$\det(\lambda I - H_k) = \det(\lambda I - A') \det(\lambda I - B')$$

بنابراین مقادیر ویژه H_k مقادیر ویژه A' همراه با مقادیر ویژه B' هستند. چون H_k به طور متعامد مشابه با ماتریس اصلی A است بنابراین دارای همان مقادیر ویژه A می‌باشد. برای یک بیان الگوریتمی دقیق از پیدا کردن RSF مرجع (Golub and Van Loan (1989, 380) را ملاحظه کنید.

چه موقع یک عنصر زیر قطر به عنوان صفر پذیرفته می‌شود

یک تصمیم بزرگ که باید در طی روش تکراری بگیریم این است که چه موقع یک عنصر زیر قطر را به عنوان صفر بپذیریم تا ماتریس را بتوان تقلیل داد. به نظر می‌رسد که هیچ قرارداد صریح و روشنی وجود ندارد؛ ما در اینجا معیار زیر را که معمولاً استفاده می‌شود، ارائه می‌دهیم. برای بحث بیشتر در این

مورد کتاب *Symmetric Eigenvalue Problem* (Parlett 1980b) را ملاحظه کنید.

یک عنصر زیر قطر را برابر صفر می‌پذیریم اگر داشته باشیم

$$|h_{i, i-1}| \leq \varepsilon (|h_{ii}| + |h_{i-1, i-1}|)$$

در اینجا ε تولرانس می‌باشد.

مثال ۸-۹-۸ شکل شور حقیقی (RSF) ماتریس زیر را پیدا کنید

$$H = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & -2 & 2 \end{bmatrix}$$

h_{r1}	تکرار
-۱/۱۸۶۷	۱
۰/۳۵۴۳	۲
۰/۰۱۲۹	۳
۰/۰۰۰۰	۴

RSF محاسبه شده به صورت زیر است

$$\begin{bmatrix} -۱/۱۶۶۳ & -۱/۳۳۲۶ & -۲/۰۵۳۱ \\ \cdot & ۱/۲۳۸۴ & ۱/۶۶۵۹ \\ \cdot & -۱/۹۴۰۹ & ۲/۹۲۷۹ \end{bmatrix}$$

مقادیر ویژه زیر ماتریس 2×2 گوشه راست پایینی عبارتند از $۱/۵۸۷۴i \pm ۲/۰۸۳۲$.

مثال ۸-۹-۹ شکل شور حقیقی ماتریس زیر را پیدا کنید

$$H = \begin{bmatrix} ۰/۲۱۹۰ & -۰/۰۷۵۶ & ۰/۶۷۸۷ & -۰/۶۳۹۱ \\ -۰/۹۶۱۵ & ۰/۹۰۳۲ & -۰/۴۵۷۱ & ۰/۸۸۰۴ \\ \cdot & -۰/۳۸۲۲ & ۰/۴۵۲۶ & -۰/۰۶۴۱ \\ \cdot & \cdot & -۰/۱۰۶۹ & -۰/۰۲۵۲ \end{bmatrix}$$

h_{33}	h_{32}	h_{21}	تکرار
-۰/۰۰۰۶۴	-۰/۵۰۸۴	۰/۳۸۶۰	۱
۰/۰۰۰۰۱	-۰/۳۷۷۳	-۰/۰۶۷۲	۲
۰	-۰/۳۶۷۳	۰/۰۰۰۸۹	۳
۰	-۰/۳۵۹۰	-۰/۰۰۰۱۱	۴
۰	-۰/۳۹۰۵	۰/۰۰۰۰۱	۵

RSF محاسبه شده به صورت زیر است

$$H = \begin{bmatrix} 1/4.95 & 0/7632 & -0/1996 & 0/8394 \\ 0/0001 & 0/1922 & 0/5792 & 0/494 \\ \cdot & -0/39.5 & 0/0243 & -0/4.89 \\ \cdot & \cdot & \cdot & -0/0763 \end{bmatrix}$$

مقادیر ویژه ماتریس عبارتند از $0/4681i \pm 0/1082$ $\begin{bmatrix} 0/1922 & 0/5792 \\ -0/39.5 & 0/0243 \end{bmatrix}$

متعادل کردن

همانند فرآیند حل مسایل دستگاه خطی، بهتر است قبل از شروع تکرارهای QR عناصر ماتریس اصلی چنانچه با هم تفاوت زیادی داشته باشند، متعادل شوند.

متعادل کردن، بدین معنی است که بر ماتریس A پیش فرآیندی اعمال شود تا مسأله مقدار ویژه ساده تر گردد. دو نوع پیش فرآیند وجود دارد. اولی جایگشت می باشد. بدین معنی که یک ماتریس P به قسمی پیدا کنیم که ماتریس $A' = P A P^T$ تقریباً بالا مثلثی یا نزدیک به شکل شور حقیقی باشد. دومی مقیاس کردن است. این معادل با پیدا کردن یک ماتریس قطری D است به قسمی که $D A' D^{-1}$ دارای سطرها و ستونهای با نرم کم و بیش مساوی باشد. این دو عمل متعادل کردن نامیده می شود. زیر روال SGEBAL از LAPACK متعادل کردن را قبل از محاسبه مقادیر ویژه انجام می دهد.

به طور کلی از قبل متعادل کردن ماتریس دقت روش تکراری QR را بهبود می بخشد. توجه کنید که هیچ خطای گرد کردنی در این محاسبه وارد نمی شود، و فقط به $O(n^2)$ عمل ممیز شناور (flops) نیاز دارد.

شمارش اعمال ممیز شناور. چون روش QR یک روش تکراری است، ارائه تعداد صحیح اعمال ممیز شناور کار سختی است ولیکن مشاهده‌های تجربی نشان داده‌اند که در حدود دو تکرار QR برای هر مقدار ویژه لازم می‌باشد. بنابراین با در نظر گرفتن تقلیل در حدود $6n^3$ عمل ممیز شناور لازم است تا همه مقادیر ویژه محاسبه شوند. اگر ماتریس تبدیل Q و ماتریس شبه مثلثی نهایی T لازم باشند، آنگاه هزینه در حدود $15n^3$ عمل ممیز شناور خواهد بود. این تعداد اعمال خیلی تقریبی هستند.

خاصیت گرد کردن. تکرار QR کاملاً پایدار است. یک تحلیل از خاصیت گرد کردن الگوریتم نشان می‌دهد که RSF محاسبه شده به طور متعادل متشابه با یک ماتریس نزدیک $A + E$ است که در آن

$$\|E\|_F \leq \phi(n)\mu \|A\|_F$$

$\phi(n)$ یک تابع رشد آهسته از n است. ماتریس محاسبه شده Q همچنین تقریباً متعادل است.

۸-۹-۸ شکل شور حقیقی و زیر فضاهای پایا

تعریف ۸-۹-۸ فرض کنید S یک زیر فضا از صفحه مختلط \mathbb{C}^n باشد. آنگاه S یک زیر فضای پایا (نسبت به پیش ضرب در A) نامیده می‌شود اگر $x \in S$ رابطه $Ax \in S$ را ابراجاب کند. چون برای هر λ داریم

$$Ax = \lambda x$$

بنابراین هر بردار ویژه یک زیر فضای پایای از بعد ۱ وابسته به مقدار ویژه متناظر می‌باشد. شکل شور حقیقی A اطلاعات مربوط به زیر فضاهای پایا را آشکار می‌سازد.

پایه یک زیر فضای پایا از RSF

فرض کنید

$$Q^T A Q = R = \begin{bmatrix} R_{11} & R_{12} \\ \cdot & R_{22} \end{bmatrix}$$

فرض کنید که R_{11} و R_{22} مقادیر ویژه مشترک نداشته باشند. آنگاه اولین p ستون Q ، که p مرتبه R_{11} است، تشکیل یک پایه برای زیر فضای پایای وابسته به مقادیر ویژه R_{11} را می‌دهند.

در بسیاری از کاربردها، مثلاً در حل معادلات ریکاتی جبری (Laub (1979))، ما نیاز به

محاسبه پایه یکا متعامد یک زیر فضای پایا وابسته به تعدادی مقدار ویژه انتخابی داریم. متأسفانه شکل شور حقیقی تبدیل یافته و به دست آمده از تکرار QR ممکن است برخی مقادیر ویژه را به ترتیب دلخواه ارائه ندهد. بنابراین اگر مقادیر ویژه به ترتیب دلخواه نباشند، باید مقداری کار اضافی انجام داد تا آنها را به ترتیب مورد نظر درآورد. آنچه را که باید در حقیقت انجام داد می‌توان در بحث ساده زیر ملاحظه کرد. فرض کنید A یک ماتریس 2×2 باشد. فرض کنید

$$Q_1^T A Q_1 = \begin{bmatrix} \lambda_1 & r_{12} \\ 0 & \lambda_2 \end{bmatrix}, \quad \lambda_1 \neq \lambda_2$$

اگر λ_1 و λ_2 به ترتیب درست قرار نگیرند، همه آنچه را که ما برای عکس کردن ترتیب نیاز داریم تشکیل یک دوران گونز $J(1, 2, \theta)$ به قسمی می‌باشد که

$$J(1, 2, \theta) \begin{bmatrix} r_{12} \\ \lambda_2 - \lambda_1 \end{bmatrix} = \begin{bmatrix} * \\ 0 \end{bmatrix}$$

آنگاه $Q = Q_1 J(1, 2, \theta)^T$ به قسمی است که

$$Q^T A Q = \begin{bmatrix} \lambda_2 & \pm r_{12} \\ 0 & \lambda_1 \end{bmatrix}$$

مثال ۸-۹-۱۰

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

$$Q_1 = \begin{bmatrix} 0.8507 & 0.5257 \\ -0.5257 & 0.8507 \end{bmatrix}$$

$$Q_1^T A Q_1 = \begin{bmatrix} -0.2361 & 0.0000 \\ 0.0000 & 4.2361 \end{bmatrix}$$

$$J(1, 2, \theta) = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

$$J(1, 2, \theta) \begin{bmatrix} 0 \\ -4/4.722 \end{bmatrix} = \begin{bmatrix} 4/4.722 \\ 0 \end{bmatrix}$$

$$Q = Q_1 J(1, 2, \theta)^T = \begin{bmatrix} -0.5257 & -0.8507 \\ -0.8507 & 0.5257 \end{bmatrix}$$

$$Q^T A Q = \begin{bmatrix} 4/2.361 & 0.00 \\ 0.00 & -0.2361 \end{bmatrix}$$

این فرایند را می توان به سادگی توسعه داد و هر ترتیب دلخواه از مقادیر ویژه را در شکل شور حقیقی به دست آورد. برای جزئیات بیشتر مرجع Golub and Van Loan (1989, 386) را ملاحظه کنید. فرایند نسبتاً ارزان تمام می شود. فقط نیاز به $k(N)$ عمل ممیز شناور (flops) دارد، که در آن k تعداد تعویضهای لازم برای رسیدن به ترتیب دلخواه می باشد. استوارت ((Stewart (1976b)) برنامه های فرتون مفیدی را برای چنین مرتب سازی از مقادیر ویژه فراهم کرده است.

۸-۱۰ محاسبه بردارهای ویژه

۸-۱۰-۱ تکرار معکوس هسنبیگی

به محض آن که یک مقدار ویژه λ توسط تکرار QR محاسبه شود، می توان از تکرار معکوس (الگوریتم ۸-۵-۲) برای محاسبه بردار ویژه متناظر استفاده کرد. ولیکن چون A در ابتدا برای تکرار QR به یک ماتریس هسنبیگی H کاهش یافته است، طبیعی است که از مزیت ساختار ماتریس هسنبیگی H در حل دستگاه خطی که لازم است در فرایند تکرار معکوس حل شود، بهره برداری شود. بنابراین تکرار معکوس هسنبیگی می تواند به صورت زیر بیان شود.

الگوریتم ۸-۱۰-۱ تکرار معکوس هسنبیگی

گام ۱: ماتریس A را به یک ماتریس بالا هسنبیگی کاهش دهید.

$$P^T A P = H$$

گام ۲: یک مقدار ویژه λ را، که بردار ویژه x آن نیز مورد جستجو است، با استفاده از الگوریتم QR ضمنی محاسبه کنید.

گام ۳: تکرار معکوس را اعمال کنید.

برای $k = 1, 2, \dots$ تا همگرایی انجام دهید.

دستگاه $(H - \lambda I)z^{(k)} = y^{(k-1)}$ را حل کنید.

قرار دهید $y^{(k)} = z^{(k)} / \max(z^{(k)})$

متوقف شوید اگر $\|(H - \lambda I)y^{(k)}\| \leq c\mu \|A\|_\infty$ ، که در آن c یک ثابت از مرتبه واحد است.

گام ۴: بردار x را بازیابی کنید:

$$x = P y^{(k)}$$

در اینجا $y^{(k)}$ تقریبی از بردار ویژه λ به دست آمده از گام ۳ است.

۸-۱۰-۲ محاسبه بردارهای ویژه از شکل شور حقیقی

بردارهای ویژه را همچنین می توان به طور مستقیم از شکل شور حقیقی بدون استفاده از تکرار معکوس حساب کرد. این فرایند به صورت زیر توصیف می شود.

فرض کنید A توسط تکرار QR ضمنی به شکل شور حقیقی T تبدیل شود

$$Q^T A Q = T$$

پس $Ax = \lambda x$ را می توان به صورت زیر نوشت

$$Q^T A Q Q^T x = \lambda Q^T x$$

یعنی با نوشتن $Q^T x = y$ ، خواهیم داشت

$$T y = \lambda y$$

بنابراین بعد از این که A به شکل شور حقیقی T تبدیل شد، یک بردار ویژه x نظیر یک مقدار ویژه λ را می توان به صورت زیر محاسبه کرد:

۱. دستگاه مثلثی یا شبه مثلثی همگن $T y = \lambda y$ را حل کنید.

۲. $x = Q y$ را محاسبه کنید.

اکنون نشان می دهیم که چگونه حل $T y = \lambda y$ را می توانیم با فرض این که T مثلثی است و همه مقادیر ویژه A (یعنی عناصر قطری T) متمایز هستند، ساده کنیم.

فرض کنید $t_{kk} = \lambda$ ؛ یعنی سعی داریم یک بردار y به قسمی پیدا کنیم که

$$(T - t_{kk} I) y = 0$$

می توان نوشت

$$T - t_{kk} I = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}$$

که در آن T_{11} یک ماتریس $k \times k$ است. y را به طور مناسب افراز می کنیم:

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

در اینجا y_1 دارای k عنصر است. سپس

$$(T - t_{kk}I)y = 0$$

نتیجه می دهد

$$\begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0$$

یعنی

$$T_{11}y_1 + T_{12}y_2 = 0, \quad T_{22}y_2 = 0$$

اکنون T_{22} نامنفرد است، زیرا عناصر قطری آن $t_{jj} - t_{kk}$ ، $n, \dots, k+1, j$ می باشند که مخالف صفر هستند. بنابراین دستگاه همگن

$$T_{22}y_2 = 0$$

فقط دارای جواب بدیهی $y_2 = 0$ است. بنابراین خواهیم داشت

$$T_{11}y_1 = 0$$

چون k امین عنصر قطری T_{11} صفر است، T_{11} منفرد است؛ بنابراین $y_1 \neq 0$. دوباره توجه کنید که T_{11} دارای شکل زیر است

$$T = \left[\begin{array}{c|c} \hat{T}_{11} & s \\ \hline \cdot & \cdot \end{array} \right]$$

که در آن \hat{T}_{11} یک ماتریس $(k-1) \times (k-1)$ و نامنفرد است. بنابراین $T_{11}y_1 = 0$ به

$$\hat{T}_{11}\hat{y}_1 + sz = 0$$

کاهش می یابد که در آن

$$y_1 = \begin{bmatrix} \hat{y}_1 \\ z \end{bmatrix}$$

z را می توان هر عدد مخالف صفری اختیار کرد. چون \hat{T}_{11} بالا مثلثی است، \hat{y}_1 را می توان با جایگزینی پسرو به دست آورد.

الگوریتم ۸-۱۰-۲ محاسبه مستقیم یک بردار ویژه از روی RSF

گام ۱: A را به RSF با استفاده از تکرار ضمنی QR تبدیل کنید.

$$Q^T A Q = T$$

(فرض کنید T مثلثی و عناصر قطری T متفاوت باشند).

گام ۲: مقدار ویژه $\lambda = t_{kk}$ را که بردار ویژه x آن باید محاسبه شود، انتخاب کنید.

گام ۳: ماتریس $(T - t_{kk}I)$ را به صورت زیر افراز کنید.

$$(T - t_{kk}I) = \begin{bmatrix} T_{11} & T_{12} \\ \cdot & T_{22} \end{bmatrix}$$

گام ۴: ماتریس T_{11} را به صورت زیر افراز کنید.

$$T_{11} = \begin{bmatrix} \hat{T}_{11} & s \\ \cdot & \cdot \end{bmatrix}$$

گام ۵: دستگاه زیر را با جایگزینی پسرو حل کنید:

$$\hat{T}_{11} \hat{y}_1 = -sz$$

z را یک عدد مخالف صفر انتخاب کنید.

گام ۶: $y_1 = \begin{bmatrix} \hat{y}_1 \\ z \end{bmatrix}$ را تشکیل دهید.

گام ۷: $y = \begin{bmatrix} y_1 \\ \cdot \end{bmatrix}$ را تشکیل دهید.

گام ۸: $x = Qy$ را محاسبه کنید.

مثال ۸-۱۰-۱

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 4 & 6 & 7 \end{bmatrix}$$

$$T = \begin{bmatrix} 1.0/18998 & 2/5145 & 2/7563 \\ \cdot & 0/3571 & 0/2829 \\ \cdot & \cdot & -0/2569 \end{bmatrix}$$

$$Q = \begin{bmatrix} -0/1358 & -0/9606 & -0/2423 \\ -0/4747 & 0/2778 & -0/1352 \\ -0/1696 & -0/0016 & 0/4937 \end{bmatrix}$$

فرض کنید که می‌خواهیم بردار ویژه متناظر با $\lambda = \lambda_{22} = 0/3571$ را حساب کنیم. پس

$$T_{11} = \begin{bmatrix} 10/5428 & 2/5145 \\ \cdot & \cdot \end{bmatrix}$$

$$\hat{T}_{11} = 10/5428$$

$$s = 2/5145$$

با انتخاب $z = 1$ داریم

$$\hat{y}_1 = -(\hat{T}_{11})^{-1}sz = -0/2385$$

$$y_1 = \begin{bmatrix} \hat{y}_1 \\ z \end{bmatrix} = \begin{bmatrix} -0/2385 \\ 1 \end{bmatrix}$$

$y_2 = 0$ را انتخاب می‌کنیم. آنگاه

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -0/2385 \\ 1 \\ \cdot \end{bmatrix}$$

$$x = Qy = \begin{bmatrix} -0/9283 \\ 0/3910 \\ 0/2058 \end{bmatrix}$$

به سادگی می‌توان تحقیق کرد که

$$Ax = \begin{bmatrix} -0/2315 \\ 0/1396 \\ 0/0735 \end{bmatrix}$$

و

$$\lambda_1 x = \begin{bmatrix} -0/2315 \\ 0/1396 \\ 0/0735 \end{bmatrix}$$

۸-۱۱ مسئله مقدار ویژه متقارن

البته می‌توان تکرار QR را برای پیدا کردن مقادیر ویژه یک ماتریس متقارن به کار برد. در حقیقت در حالت متقارن روش تا حدود زیادی ساده می‌شود. ما تکرار QR متقارن را به طور مختصر توضیح

می‌دهیم. ولیکن چون مقادیر ویژه و بردارهای ویژه یک ماتریس متقارن خاصیت‌های قابل توجه و خاصی نسبت به مقادیر ویژه حالت نامتقارن دارند، برخی از روشهای خاص با بهره‌برداری از این خاصیتها می‌توانند برای مسائل متقارن توسعه یابند. یکی از این روشها بر خاصیت دنباله استورم^۱ چندجمله‌ای مشخصه ماتریس مبتنی است.

برخی خاصیت‌های خاص

مقادیر ویژه یک ماتریس متقارن حقیقی. مقادیر ویژه یک ماتریس متقارن حقیقی، حقیقی هستند. بردارهای ویژه می‌توانند طوری انتخاب شوند که متعامد باشند (قضیه ۸-۲-۸).

قضیه ۸-۱۱-۱ شکل شور حقیقی یک ماتریس متقارن حقیقی. شکل شور حقیقی یک ماتریس متقارن حقیقی یک ماتریس قطری است.

اثبات: از قضیه مثلثی سازی شور حقیقی (قضیه ۸-۹-۲) داریم:

$$Q^T A Q = R$$

که در آن R به شکل شور حقیقی است؛ یعنی R یک ماتریس بالا مثلثی است که هر عنصر قطری آن یا به صورت یک اسکالر یا به صورت یک ماتریس 2×2 می‌باشد. هر ماتریس 2×2 بر روی قطر متناظر با یک جفت مقدار ویژه مختلط می‌باشد. چون یک ماتریس متقارن حقیقی نمی‌تواند یک مقدار ویژه مختلط داشته باشد، نتیجه می‌شود که R نمی‌تواند دارای یک ماتریس 2×2 بر روی قطر باشد، یعنی R یک ماتریس قطری است.

قضیه ۸-۱۱-۲ خاصیت اختلال کلی مسأله مقدار ویژه متقارن. فرض کنید A یک ماتریس متقارن حقیقی $n \times n$ باشد. فرض کنید $A' = A + E$ ، که در آن E یک اختلال متقارن حقیقی از ماتریس A است، و فرض کنید $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ و $\lambda'_1 \geq \lambda'_2 \geq \dots \geq \lambda'_n$ به ترتیب مقادیر ویژه A و A' باشند، آنگاه

$$\lambda_i - \|E\|_2 \leq \lambda'_i \leq \lambda_i + \|E\|_2, \quad i = 1, 2, \dots, n$$

با استفاده از قضیه معروف زیر اثباتی برای قضیه ۸-۱۱-۲ ارائه خواهیم داد.

قضیه ۸-۱۱-۳ قضیه کمینه بیشینه کورانت - فیشر^۲. فرض کنید $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ مقادیر ویژه یک ماتریس متقارن A باشند. آنگاه

$$\lambda_i = \min_S \max_{x \neq 0} \left(\frac{x^T A x}{x^T x} \right)$$

که در آن می‌نیمم بر روی همه زیر فضاهای S از بعد $(n - i + 1)$ اختیار شده است و ماکزیمم بر روی همه بردارهای مخالف صفر x در زیر فضای S اختیار شده است.

اثبات: مرجع (Saad (1992, 30-31) را ملاحظه کنید. □

مشخص سازی مقادیر ویژه اکسترمم

جالب است توجه داشته باشیم که دو حالت خاص قضیه ۸-۱۱-۳ به مشخص سازی مقادیر ویژه اکسترمم λ_1 و λ_n ماتریس A منجر می‌گردند.

$i = 1$. در این حالت می‌نیمم فقط بر روی زیر فضاهای S از بعد n ، که خود R^n می‌باشد، گرفته

می‌شود. بنابراین

$$\lambda_1 = \lambda_{\max} = \max_{x \neq 0} \frac{x^T A x}{x^T x}$$

$i = n$. در این حالت می‌نیمم فقط بر روی زیر فضاهای S از بعد ۱ و متشکل از فقط بردار x و

مضربهای آن ax گرفته می‌شود. بنابراین

$$\lambda_n = \lambda_{\min} = \min_{x \neq 0} \frac{x^T A x}{x^T x}$$

اثبات قضیه ۸-۱۱-۲: با به کار بردن قضیه ۸-۱۱-۳ در مورد A' ، داریم:

$$\begin{aligned} \lambda'_i &= \min \max_x \frac{x^T A' x}{x^T x} = \min \left(\max \left(\frac{x^T A x}{x^T x} + \frac{x^T E x}{x^T x} \right) \right) \\ &\leq \min \left(\max \frac{x^T A x}{x^T x} + \max \frac{x^T E x}{x^T x} \right) \\ &= \min \max \left(\frac{x^T A x}{x^T x} \right) + \min \max \left(\frac{x^T E x}{x^T x} \right) \\ &\leq \min \max \frac{x^T A x}{x^T x} + \|E\|_F \\ &= \lambda_i + \|E\|_F \end{aligned}$$

توجه کنید که در اثبات قبلی ما از نتیجه

$$\frac{x^T E x}{x^T x} \leq E \text{ بزرگ‌ترین مقدار ویژه}$$

و

$$\|E\|_2 \leq \text{قدر مطلق بزرگ‌ترین مقدار ویژه } E$$

(قضیه ۸-۴-۳) استفاده کرده‌ایم.

به طور مشابه می‌توان ثابت کرد که

$$\lambda_i' - \|E\|_2 \leq \lambda_i'$$

□

نتیجه اختلال بر روی مقادیر ویژه یک ماتریس متقارن

مقادیر ویژه یک ماتریس متقارن حقیقی خوش‌وضع هستند؛ یعنی تغییرهای کوچک در عناصر A فقط موجب تغییرهای کوچک در مقادیر ویژه A می‌گردند.

مثال ۸-۱۱-۱

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 6 \end{bmatrix}, \quad E = 10^{-4} \times I_{3 \times 3}$$

مقادیر ویژه A عبارتند از $10/1867$ ، $0/2336$ ، $-0/4203$ و مقادیر ویژه $A + E$ عبارتند از $10/1868$ و $0/2337$ ، $-0/4202$. $\|E\|_2 = 10^{-4}$ توجه کنید که

خاصیت اختلال از رتبه ۱. یک قضیه بیان می‌کنیم که نشان می‌دهد که چگونه مقادیر ویژه اختلال می‌یابند اگر E یک اختلال از رتبه ۱ باشد. نتیجه آن یک نقش مهم در الگوریتم تقسیم و تسخیر^۱ (Dongarra and Sorensen 1987, and Cuppen (1981)) برای مسأله مقدار ویژه متقارن دارد.

قضیه ۸-۱۱-۴ مقادیر ویژه یک ماتریس اختلال یافته رتبه ۱. فرض کنید $B = A + abb^T$ ، که در آن A یک ماتریس متقارن $n \times n$ است، α یک اسکالر، و b یک بردار n بعدی است. فرض کنید $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ مقادیر ویژه A و $\lambda_1' \geq \lambda_2' \geq \dots \geq \lambda_n'$ مقادیر ویژه B باشند. آنگاه

$$\begin{aligned} \lambda'_i &\in [\lambda_i, \lambda_{i-1}], & i = 2, \dots, n, & \alpha \geq 0 \text{ اگر} \\ \lambda'_i &\in [\lambda_{i+1}, \lambda_i], & i = 1, \dots, n-1, & \alpha < 0 \text{ اگر} \end{aligned}$$

□ اثبات: مرجع (Wilkinson (1965, 97-98) را ملاحظه کنید.

مثال ۸-۱۱-۲

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}, \quad \alpha = -1, b = (1, 2, 3)^T$$

مقادیر ویژه B عبارتند از $\lambda'_3 = -3/3028$ ، $\lambda'_2 = 0$ ، $\lambda'_1 = 0/3028$. مقادیر ویژه A عبارتند از $\lambda_3 = -0/5157$ ، $\lambda_2 = 0/1709$ ، $\lambda_1 = 11/3448$. به آسانی می توان تحقیق کرد که $\lambda_3 < \lambda'_3 < \lambda_2$ و $\lambda_2 < \lambda'_2 < \lambda_1$.

قضیه هافمن-ویلندت^۱. در قسمت قبل فرض کردیم که A و $A + E$ هر دو متقارن هستند. قضیه بعد یک حالت کلی تر را، که در آن A و $A + E$ هر دو نرمال هستند، در نظر می گیرد.

قضیه ۸-۱۱-۵ قضیه هافمن-ویلندت. فرض کنید A و $A + E$ هر دو نرمال باشند. فرض کنید مقادیر ویژه $\lambda_1, \lambda_2, \dots, \lambda_n$ ماتریس A به یک ترتیب مفروضی ارائه شده باشند. همچنین فرض کنید $\lambda'_1, \lambda'_2, \dots, \lambda'_n$ مقادیر ویژه ماتریس $A + E$ نیز دارای ترتیبی باشند، آنگاه یک جایگشت $\sigma(i)$ از n اولین عدد طبیعی $1, 2, \dots, n$ وجود دارد به قسمی که

$$\left[\sum_{i=1}^n |\lambda'_{\sigma(i)} - \lambda_i|^2 \right]^{1/2} \leq \|E\|_F$$

برای اثبات و چند نتیجه جالب از این قضیه مرجع (Horn and Johnson (1985, 368-371) را ملاحظه کنید.

۸-۱۱-۱ دنباله استورم و روش تنصیف

در این بخش یک روش برای پیدا کردن مقادیر ویژه یک ماتریس متقارن شرح می دهیم. بویژه این روش مفید است اگر مقادیر ویژه در یک بازه مورد نیاز باشند. اساس این روش می تواند برای پیدا کردن همه مقادیر ویژه مورد استفاده قرار گیرد. ولیکن در عمل اگر همه مقادیر ویژه به طور صریح مورد نیاز

قضیه ۸-۱۱-۶ خاصیت یک در میانی. فرض کنید T یک ماتریس سه قطری متقارن کاهش ناپذیر باشد و $\lambda_i^{(k)}$ کوچک ترین مقدار ویژه i ام زیر ماتریس اصلی $k \times k$ آن باشد. آنگاه

$$\lambda_1^{(k+1)} < \lambda_1^{(k)} < \lambda_2^{(k+1)} < \dots < \lambda_i^{(k+1)} < \lambda_i^{(k)} < \lambda_{i+1}^{(k+1)} < \dots < \lambda_k^{(k+1)} < \lambda_k^{(k)} < \lambda_{k+1}^{(k+1)}$$

(مرجع Golub and Van Loan (1989, 438) ملاحظه شود.)

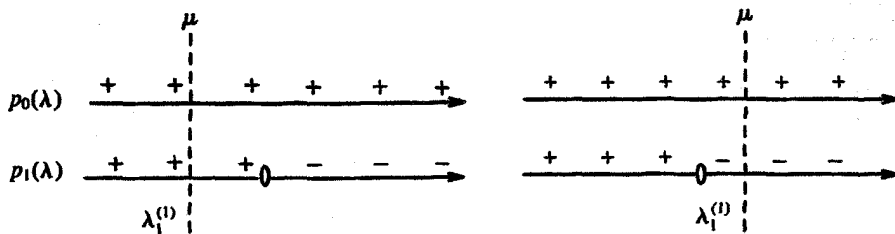
نشان خواهیم داد که این خاصیت یک در میانی به یک نتیجه جالب در مورد تعداد مقادیر ویژه واقع در یک بازه منجر می شود.

واضح است که $p_k(\lambda) = (-1)^k \lambda^k + \dots$ بنابراین اگر λ منفی بوده و از لحاظ اندازه بزرگ باشد $P_k(\lambda)$ مثبت است. صفرهای $\lambda_1^{(k)} < \lambda_2^{(k)} < \dots < \lambda_k^{(k)}$ چند جمله ای $p_k(\lambda)$ خط حقیقی را به $k + 1$ بازه زیر تقسیم می کنند:

$$(-\infty, \lambda_1^{(k)}), (\lambda_1^{(k)}, \lambda_2^{(k)}), \dots, (\lambda_{k-1}^{(k)}, \lambda_k^{(k)}), (\lambda_k^{(k)}, \infty)$$

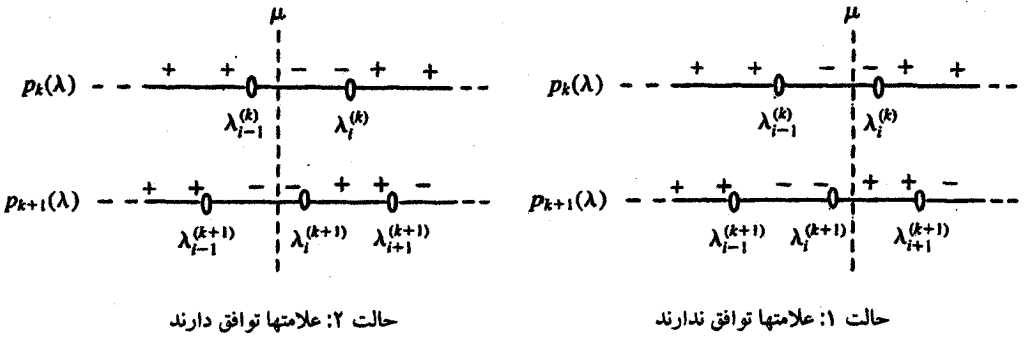
$p_k(\lambda)$ در بازه اول مثبت است و علامتهای متناوب در بازه های متوالی اختیار می کند. خاصیت یک در میانی و تغییر علامتهای $p_k(\lambda)$ در شکل ۸-۹ نشان داده شده اند.

فرض کنید که μ یک عدد حقیقی باشد. دو حالت برای علامتهای $P_0(\mu)$ و $P_1(\mu)$ وجود دارد (شکل ۸-۱۰).



حالت ۲: علامتها توافق دارند

حالت ۱: علامتها توافق ندارند



شکل ۸-۱۱

واضح است که یک توافق علامتی بین $P_0(\mu)$ و $P_1(\mu)$ ، $\lambda_1^{(1)}$ را در طرف راست μ قرار می دهد. بر طبق خاصیت یک در میانی این مسأله ایجاب می کند که

$$\mu < \lambda_1^{(1)} < \lambda_1^{(2)} < \dots < \lambda_n^{(n)}$$

بنابراین یک مقدار ویژه $\lambda_n^{(n)}$ از A (یا T)، همچنین یک مقدار ویژه از هر زیر ماتریس $k \times k$ را در طرف راست μ قرار می دهد.

فرض کنید که تعداد توافقه‌های علامتی بین دو جمله متوالی از $P_0(\mu)$ ، $P_1(\mu)$ ، \dots ، $P_k(\mu)$ را شمارش کرده ایم. اجازه دهید علامتهای $P_k(\mu)$ و $P_{k+1}(\mu)$ و خاصیت یک در میانی را ملاحظه کنیم (شکل ۸-۱۱).

فرض کنید μ بین $\lambda_{i-1}^{(k)}$ و $\lambda_i^{(k)}$ باشد. از شکل ۸-۱۱ واضح است که یک توافق علامتی بین $P_k(\mu)$ و $P_{k+1}(\mu)$ ، $\lambda_i^{(k+1)}$ را در طرف راست μ قرار می دهد، که معنای آن این است که P_{k+1} یک صفر بیشتر از P_k در بازه (μ, ∞) دارد.

هیچ صفری از P_0 در (μ, ∞) وجود ندارد. تعداد صفرهای P_1 را در بازه (μ, ∞) از روی موافق بودن یا ناموافق بودن علامتهای $P_0(\mu)$ و $P_1(\mu)$ می شناسیم. به طور کلی وقتی ما تعداد صفرهای P_k را در (μ, ∞) بشناسیم، تعداد صفرهای P_{k+1} را در بازه (μ, ∞) با بررسی توافق علامتی بین $P_k(\mu)$ و $P_{k+1}(\mu)$ خواهیم شناخت. این بازگشت می تواند ادامه یابد تا تعداد صفرهای P_n را بشناسیم، یعنی تعداد مقادیر ویژه T را در بازه (μ, ∞) بشناسیم. این قاعده روشن است.

قضیه ۸-۱۱-۷ تعداد توافقه‌های علامتی بین جمله‌های متوالی $P_0(\mu)$ ، $P_1(\mu)$ ، \dots ، $P_n(\mu)$ برابر تعداد مقادیر ویژه T بزرگتر از μ است.

تذکره: واضح است که همچنین $p_{k+1}(\mu) = 0$ باید یک توافق علامتی با $p_k(\mu)$ در نظر گرفته شود، زیرا یک صفر بیشتر در بازه $[\mu, \infty)$ قرار می‌گیرد.

مثال ۸-۱۱-۳ فرض کنید

$$T = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

$$p_0(\lambda) = 1, \quad p_1(\lambda) = 2 - \lambda$$

$$p_2(\lambda) = (2 - \lambda)^2 - 1$$

$$\begin{aligned} p_3(\lambda) &= (2 - \lambda)p_2(\lambda) - p_1(\lambda) \\ &= (2 - \lambda)[(2 - \lambda)^2 - 1] - (2 - \lambda) \\ &= (2 - \lambda)^3 - 2(2 - \lambda) \end{aligned}$$

فرض کنید $\mu = 0$.

پس دنباله $p_0, p_1(\mu), p_2(\mu), p_3(\mu)$ عبارات است از ۱، ۲، ۳، ۴. سه توافق در علامت وجود دارد. بنابراین همه مقادیر ویژه T بزرگ‌تر یا مساوی صفر هستند. در حقیقت چون $0 \neq p_3(\mu) = 4$ نتیجه گرفته می‌شود همه مقادیر ویژه T مثبت هستند. به آسانی مشاهده می‌شود که همه مقادیر ویژه T عبارتند از ۲، $2 + \sqrt{2}$ و $2 - \sqrt{2}$.

فرض کنید $\mu = 2$. پس دنباله $p_0, p_1(\mu), p_2(\mu), p_3(\mu)$ عبارات است از ۱، ۰، -۱، ۰. تعداد توافقات را در علامتهای

$$+ \quad + \quad - \quad -$$

می‌شماریم. دو توافق وجود دارد؛ تأیید می‌شود که T دارای دو مقدار ویژه بزرگ‌تر یا مساوی ۲ است. اکنون می‌توان روش دنباله استورم - تنصیف را برای محاسبه یک صفر مشخص از $p_n(\lambda)$ به صورت زیر بیان کرد:

الگوریتم ۸-۱۱-۱ الگوریتم دنباله استورم - تنصیف. فرض کنید $\lambda_1 < \lambda_2 < \dots < \lambda_n$ مقادیر ویژه T باشند؛ یعنی صفرهای $p_n(\lambda)$ باشند. فرض کنید به ازای یک m ($m \leq n$) مفروض مقدار ویژه مورد نظر λ_{n-m-1} باشد. آنگاه

گام ۱: یک بازه $[s_1, s_2]$ شامل λ_{n-m-1} پیدا کنید. چون در ابتدا $\lambda_i \leq \|T\|$ می‌توان اختیار کرد

$$s_1 = -\|T\|_\infty, \quad s_2 = \|T\|_\infty$$

گام ۲: $s_3 = \frac{s_1 + s_2}{2}$ را محاسبه کنید.

گام ۳: $N(s_3)$ تعداد توافقتها در علامتهای ۱، $p_1(s_3)$ ، $p_2(s_3)$ ، ...، $p_n(s_3)$ را محاسبه کنید. اگر $N(s_3) < m$ ، قرار دهید $s_2 = s_3$ ؛ در غیر این صورت قرار دهید $s_1 = s_3$.

فرض کنید $\varepsilon > 0$ یک عدد مثبت کوچک از قبل مشخص شده باشد. آزمون کنید که آیا $|s_2 - s_1| < \varepsilon$ است. اگر چنین است، $s_3 = (s_1 + s_2)/2$ را به عنوان یک مقدار تقریبی از λ_{n-m-1} بپذیرید. در غیر این صورت به گام ۲ بروید.

تذکره: بعد از k گام، صفر مورد نظر در یک بازه با طول $\frac{(s_2 - s_1)}{2^k}$ قرار می‌گیرد.

مثال ۸-۱۱-۳ فرض کنید

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

فرض کنید می‌خواهیم $\lambda_1 = 2 - \sqrt{2}$ را تقریب بزنیم:

$$m = 3$$

تکرار ۰: در آغاز اختیار می‌کنیم

$$s_1 = -4, \quad s_2 = 4$$

$$s_3 = 0$$

$$N(s_3) = N(0) = 3$$

قرار می‌دهیم $s_3 = s_1$

تکرار ۱:

$$s_1 = 0, \quad s_2 = 4$$

$$s_3 = \frac{0 + 4}{2} = 2$$

$$N(s_3) = N(2) = 2 < 3$$

قرار می‌دهیم $s_3 = s_2$

تکرار ۲:

$$s_1 = 0, \quad s_2 = 2, \quad s_3 = \frac{0 + 2}{2} = 1$$

$N(s_3) =$ تعداد توافقتها در علامتهای دنباله

$$1, -1, 0, -1$$

$$(+ \quad - \quad - \quad -)$$

$$N(s_3) = 2 < 3$$

قرار می‌دهیم $s_3 = s_2$

تکرار ۳:

$$s_1 = 0, \quad s_2 = 1, \quad s_3 = 0/5$$

$$N(s_3) = 3$$

$$s_1 = s_3 \text{ قرار می دهیم}$$

λ_1 به طور واضح در بازه $[0/5, 1]$ قرار دارد. که در حقیقت چنین می باشد. تکرارها را تا به دست آوردن طول بازه $\varepsilon < |s_2 - s_1|$ ادامه می دهیم.

۸-۱۱-۲ روش تکرار QR متقارن

هنگامی که A متقارن است، ماتریس هسبرگی تبدیل یافته $PAPT = H$ نیز متقارن است؛ بنابراین سه قطری متقارن است. در نتیجه در این حالت ما باید به مسأله مقدار ویژه یک ماتریس سه قطری متقارن به جای ماتریس هسبرگی پردازیم.

برای اعمال روش تکرار QR بر یک ماتریس سه قطری متقارن، توجه می کنیم که اگر ماتریس آغازین یک ماتریس سه قطری متقارن T باشد، آنگاه هر ماتریس T_k در دنباله

$$T_k - \mu_k I = Q_k R_k$$

نیز چنین خواهد بود، و بعلاوه ما فقط به $O(n)$ عمل ممیز شناور (flops) برای تولید هر T_k نیاز داریم (توجه کنید که تجزیه QR یک ماتریس سه قطری متقارن فقط به $O(n)$ عمل ممیز شناور (flops) نیاز دارد).

از طرف دیگر چون مقادیر ویژه یک ماتریس متقارن همگی حقیقی هستند، استراتژی انتقال مضاعف بحث شده برای مسأله مقدار ویژه کلی در این حالت لازم نیست. ولیکن در این حالت یک انتقال متداول، معروف به انتقال ویلکینسون، معمولاً استفاده می شود.

به جای اختیار کردن عنصر (n, n) در هر تکرار به عنوان انتقال، معمولاً آن مقدار ویژه از ماتریس 2×2 پسر و که به عنصر (n, n) نزدیک تر است به عنوان انتقال انتخاب می شود. این معروف به انتقال ویلکینسون است. بنابراین اگر یک زیر ماتریس 2×2 پسر و T_k به صورت

$$\begin{bmatrix} t_{n-1, n-1}^{(k)} & t_{n, n-1}^{(k)} \\ t_{n, n-1}^{(k)} & t_{nn}^{(k)} \end{bmatrix}$$

ارائه شود، آنگاه انتقال ویلکینسون به صورت زیر می باشد

$$w = t_{nn}^{(k)} + r - \text{sign}(r) \sqrt{r^2 + (t_{n,n-1}^{(k)})^2}$$

که در آن

$$r = \frac{(t_{n-1,n-1}^{(k)} - t_{nn}^{(k)})}{2}$$

یک راه بهتر برای محاسبه w در تمرین ۳۸ ارائه شده است.

الگوریتم ۸-۱۱-۲ تکرار QR متقارن

گام ۱: ماتریس A را با استفاده از تبدیلهای تشابهی متعامد به یک ماتریس سه قطری متقارن تبدیل کنید:

$$PAP^T = T$$

گام ۲: تکرار QR تک انتقالی را در مورد T با انتقال ویلکینسون w اعمال کنید:

قرار دهید $T = T_1$ ، w را به صورت نشان داده شده در فوق محاسبه کنید.

برای $k = 1, 2, \dots$ تا برقراری همگرایی انجام دهید.

$$T_k - wI = Q_k R_k$$

$$T_{k+1} = R_k Q_k + wI$$

یک انتقال جدید w را انتخاب کنید.

تذکره: ممکن است T_{k+1} را از روی T_k بدون تشکیل صریح ماتریس $T_k - w_k I$ حساب کرد. این معروف به QR متقارن ضمنی است. تمرین ۲۹ را ملاحظه کنید.

همگرایی تکرار QR متقارن

می توان نشان داد (مراجع Lawson and Hanson (1974, 109) و Parlett (1980, chapter 8)

ملاحظه کنید). که همگرایی $t_{n,n-1}^{(k)}$ به صفر از درجه دو است؛ یعنی $e > 0$ و وابسته به A وجود دارد به قسمی که به ازای همه مقادیر k رابطه زیر برقرار می باشد

$$\left| t_{n,n-1}^{(k+1)} \right| \leq e \left| t_{n,n-1}^{(k)} \right|^2$$

تبصره: در عمل مشاهده شده است که همگرایی تقریباً همیشه از درجه سه است؛ اما تنها همگرایی درجه دو ثابت شده است.

شمارش اعمال ممیز شناور. چون تجزیه QR یک ماتریس سه قطری نیاز به $O(n)$ عمل ممیز شناور (flops) دارد، برای محاسبه همه مقادیر ویژه یک ماتریس سه قطری متقارن با استفاده از گام ۲ الگوریتم ۸-۱۱-۲ نیاز به $O(n^2)$ عمل ممیز شناور (flops) خواهد بود. کاهش ماتریس متقارن A به ماتریس سه قطری متقارن T نیز حدوداً به $\frac{2}{3}n^3$ عمل ممیز شناور (flops) نیاز خواهد داشت.

خاصیت گرد کردن. همانند حالت کلی نامتقارن، QR متقارن با انتقال ضمنی پایدار است. می توان نشان داد که، با مفروض بودن یک ماتریس متقارن A ، الگوریتم QR متقارن با انتقال ضمنی یک ماتریس Q متعامد و یک ماتریس قطری D تولید می کند به قسمی که

$$Q^T A Q = D + E$$

که در آن

$$\|E\|_F \approx \mu \phi(n) \|A\|_F$$

و $\phi(n)$ یک تابع رشد آهسته از n است.

بعلاوه، هر مقدار محاسبه شده $\hat{\lambda}_i$ در نامساوی زیر صدق می کند

$$|\lambda_i - \hat{\lambda}_i| \leq \phi(n) \mu \|A\|_2$$

بنابراین خطای مطلق در هر مقدار ویژه محاسبه شده کوچک است، اما خطای نسبی در مقادیر ویژه بسیار کوچک می تواند بزرگ باشد.

۸-۱۱-۳ روش تقسیم و تسخیر

همان گونه که از عنوان این مطلب بر می آید، این روش مبتنی بر اصل «تقسیم و تسخیر» است. الگوریتم ابتدا یک مسأله مقدار ویژه سه قطری متقارن مفروض را به دو مسأله کوچک تر تقسیم می کند، و سپس از جوابهای زیر مسائل برای تسخیر جواب مسأله اصلی استفاده می کند. روش ابتدا توسط کاپن^۱ (۱۹۸۱) پیشنهاد شده است.

روش می تواند برای محاسبه همه مقادیر ویژه و بردارهای متناظر یک ماتریس متقارن استفاده شود، و سریع تر از روش تکراری QR شرح داده شده در بخش قبل است. در اینجا یک شمای خیلی خلاصه از

این روش را ارائه می‌دهیم.

فرض کنید که ماتریس متقارن A توسط تبدیلات تشابهی متعامد به یک ماتریس سه قطری متقارن T تبدیل شده باشد. فرض کنید

$$T = \begin{bmatrix} a_1 & b_1 & & \cdot \\ b_1 & \ddots & \ddots & \\ & \ddots & \ddots & b_{n-1} \\ \cdot & & b_{n-1} & a_n \end{bmatrix}$$

تعریف می‌کنیم

$$T_1 = \begin{bmatrix} a_1 & b_1 & & \cdot \\ b_1 & \ddots & \ddots & \\ & \ddots & a_{k-1} & b_{k-1} \\ \cdot & & b_{k-1} & a_k - b_k \end{bmatrix}$$

$$T_2 = \begin{bmatrix} a_{k+1} - b_k & b_{k+1} & & \cdot \\ b_{k+1} & \ddots & \ddots & \\ & \ddots & \ddots & b_{n-1} \\ \cdot & & b_{n-1} & a_n \end{bmatrix}$$

آنگاه

$$T = \begin{bmatrix} T_1 & \cdot \\ \cdot & T_2 \end{bmatrix} + b_k \nu \nu^T$$

که در آن

$$\nu = (\cdot, \cdot, \dots, 1, 1, \cdot, \dots, \cdot)^T$$

چون T_1 و T_2 سه قطری متقارن هستند، می‌توانیم ماتریسهای متعامد Q_1 و Q_2 را به قسمی پیدا

کنیم که

$$T_1 = Q_1 D_1 Q_1^T, \quad T_2 = Q_2 D_2 Q_2^T$$

آنگاه

$$T = \begin{bmatrix} Q_1 & \cdot \\ \cdot & Q_2 \end{bmatrix} \left[\begin{bmatrix} D_1 & \cdot \\ \cdot & D_2 \end{bmatrix} + b_k u u^T \right] \begin{bmatrix} Q_1^T \\ \cdot \\ Q_2^T \end{bmatrix}$$

که در آن

$$u = \begin{bmatrix} Q_1^T & \cdot \\ \cdot & Q_2^T \end{bmatrix} v$$

بنابراین مقادیر ویژه T همان مقادیر ویژه ماتریس

$$\hat{D} = D + b_k u u^T = D + \rho u u^T$$

خواهند بود که در آن D یک ماتریس قطری حقیقی $n \times n$ ارائه شده به صورت

$$D = \begin{bmatrix} D_1 & \cdot \\ \cdot & D_2 \end{bmatrix}$$

می‌باشد و $\rho = b_k$. بنابراین اکنون توجه خود را به چگونه به دست آوردن مقادیر ویژه و بردارهای ویژه یک ماتریس قطری اختلال یافته از رتبه یک $\hat{D} = D + \rho u u^T$ معطوف می‌داریم.

بدون از دست دادن کلیت مسأله، فرض کنید که $\|u\|_2 = 1$ و $\rho = b_k \neq 0$. همچنین فرض کنید که $D = \text{diag}(d_1, d_2, \dots, d_n)$ و $d_1 < d_2 < d_3 < \dots < d_n$ و هیچ یک از مؤلفه‌های بردار u صفر نباشند.

در حقیقت یک مؤلفه صفر u یک توفیق اجباری است. می‌توان نشان داد (تمرین ۴۱) که در این حالت یک زوج آزاد مقدار ویژه و بردار ویژه به دست می‌آوریم. بعلاوه این فرصتی را برای تقلیل ارائه می‌دهد. همچنین اگر k مقدار ویژه D مساوی باشند، آنگاه مسأله می‌تواند با حذف $(k-1)$ سطر و ستون آخر تقلیل یابد (تمرین ۴۲).

فرض کنید (λ, q) یک جفت ویژه از \hat{D} باشد. آنگاه نشان می‌دهیم که

۱. λ یک ریشه معادله زیر است

$$1 + \rho u^T (D - \lambda I)^{-1} u = 0$$

۲. $q = (D - \lambda I)^{-1} u$ یک بردار ویژه $(D + \rho u u^T)$ متناظر با λ است.

برای نشان دان (۱)، توجه می‌کنیم که چون (λ, q) یک جفت ویژه از \hat{D} است ما باید داشته باشیم

$$(D + \rho u u^T) q = \lambda q$$

یعنی

$$(D - \lambda I)q = -\rho(u^T q)u$$

اکنون فرضهای ما $\rho \neq 0$ ، $d_1 < d_2 < d_3 < \dots < d_n$ و این که هیچ مؤلفه‌ای از u صفر نیست ایجاب می‌کنند که اولاً $(D - \lambda I)$ نامنفرد باشد و ثانیاً $u^T q$ مخالف صفر باشد (تمرین ۴۴).

از ضرب آخرین معادله در $(D - \lambda I)^{-1}$ خواهیم داشت

$$q = -\rho(u^T q)(D - \lambda I)^{-1}u$$

از ضرب هر دو طرف در u^T و تقسیم بر اسکالر $u^T q$ ، به دست می‌آوریم

$$1 + \rho u^T (D - \lambda I)^{-1}u = 0$$

برای نشان دادن (۲)، توجه می‌کنیم که

$$\begin{aligned} (D + \rho uu^T)(D - \lambda I)^{-1}u &= (D - \lambda I + \lambda I + \rho uu^T)(D - \lambda I)^{-1}u \\ &= u + \lambda(D - \lambda I)^{-1}u + \rho u u^T (D - \lambda I)^{-1}u \\ &= u + \lambda(D - \lambda I)^{-1}u + u(-1) \\ &= u + \lambda(D - \lambda I)^{-1}u - u \\ &= \lambda(D - \lambda I)^{-1}u \end{aligned}$$

معادله

$$1 + \rho u^T (D - \lambda I)^{-1}u = 0$$

می‌تواند بر حسب مؤلفه‌های u_i بردار u به صورت زیر نوشته شود

$$f(\lambda) = 1 + \rho u^T (D - \lambda I)^{-1}u = 1 + \rho \sum_{j=1}^n \frac{u_j^2}{d_j - \lambda} = 0$$

این معادله معمولاً معروف به معادله ویژه است.

دوباره چون d_i ها همگی متمایز هستند و هیچ یک از مؤلفه‌های u صفر نیستند، می‌توانیم نشان دهیم (تمرین ۴۳ و ۴۵) که $f(\lambda) = 0$ دقیقاً دارای n ریشه است، یک ریشه در هر بازه (d_j, d_{j+1}) ، $j = 1, 2, \dots, n-1$ و یک ریشه در طرف راست d_n اگر $\rho > 0$ یا یک ریشه در طرف چپ d_1 اگر $\rho < 0$.

سپس می‌توان روش تنصیف (الگوریتم ۸-۱۱-۱) یا روش نیوتن را برای پیدا کردن این ریشه‌ها در هر یک از این بازه‌ها به کار برد. به محض پیدا شدن یک ریشه، بردار ویژه متناظر را می‌توان از (۲)

به دست آورد.

توجه کنید که اگر $q = (D - \lambda I)^{-1}u$ ، آنگاه مؤلفه های q توسط رابطه زیر ارائه می شوند

$$q_i = \frac{u_i}{d_i - \lambda}, \quad i = 1, \dots, n$$

بحث فوق به الگوریتم زیر منجر می شود.

الگوریتم ۸-۱۱-۳ روش تقسیم و تسخیر. با مفروض بودن یک ماتریس سه قطری متقارن T ،

الگوریتم مقادیر ویژه و بردارهای ویژه T را پیدا می کند.

گام ۱: ماتریس زیر را تشکیل دهید.

$$T = \begin{bmatrix} T_1 & \cdot \\ \cdot & T_2 \end{bmatrix} + b_k v v^T$$

گام ۲: ماتریسهای متعامد Q_1 و Q_2 را به قسمی پیدا کنید که $Q_1^T T_1 Q_1 = D_1$ و

$$Q_2^T T_2 Q_2 = D_2$$

گام ۳: ماتریس $D = \text{diag}(D_1, D_2) = \text{diag}(d_1, \dots, d_n)$ و بردار $u = \text{diag}(Q_1^T, Q_2^T)v$

تشکیل دهید.

گام ۴: مقادیر ویژه T را از حل معادله ویژه

$$f(\lambda) = 1 + \rho \sum_{j=1}^n \frac{u_j^2}{d_j - \lambda} = 0$$

پیدا کنید و بردارهای ویژه $D + \rho u u^T$ را از گام ۲ به دست آورید.

گام ۵: بردارهای ویژه T را بازبازی کنید: اگر Q' ماتریس بردارهای ویژه $D + \rho u u^T$ باشد، آنگاه

ماتریس بردارهای ویژه T توسط رابطه زیر ارائه می شود

$$\begin{bmatrix} Q_1 & \cdot \\ \cdot & Q_2 \end{bmatrix} Q'$$

شمارش اعمال ممیز شناور. با فرض این که یک گام تکراری نیوتن در حدود $O(n)$ عمل ممیز شناور

(flops) هزینه دارد، الگوریتم فقط $O(n^2)$ عمل ممیز شناور (flops) برای همه n مقدار ویژه نیاز

خواهد داشت. هزینه محاسبه هر بردار ویژه از یک مقدار ویژه محاسبه شده نیز برابر $O(n)$ عمل ممیز

شناور (flops) می باشد.

تبصره‌ها: بانچ، نیلسن، و سورنسن (۱۹۷۸)^۱ یک روش برای حل $f(\lambda) = 0$ با استفاده از تقریبهای توابع گویا پیشنهاد کرده‌اند. روش آنها به طور درجه دو همگرا می‌گردد. برای جزئیات بیشتر در مورد این روش و پیاده‌سازی آن مرجع Bunch, Nielsen, and Sorensen (1978) و مقاله Dongarra and Sorensen (1987) را ملاحظه کنید. روش تقسیم و تسخیر دارای طبیعت موازی است.

توجه کنید که چون T_1 و T_2 هر دو سه‌قطری متقارن هستند، مسأله مقدار ویژه برای هر یک از این ماتریسها می‌تواند باز هم به دو زیر مسأله تجزیه شود و در نتیجه مسأله اصلی به چهار زیر مسأله تجزیه می‌شود. این چهار مسأله دوباره می‌توانند به هشت مسأله کوچک‌تر تجزیه شوند، و فرایند می‌تواند تا جایی که مورد نظر باشد (احتمالاً تا آن جا که اندازه مسأله‌ها ۱ در ۱ و یا ۲ در ۲ شوند) ادامه یابد. چون هر یک از مسأله‌ها مستقل هستند، مسأله اصلی به تعداد زیادی زیر مسأله تقسیم می‌شود. مرجع Dongarra and Sorensen (1987) را از نظر پیاده‌سازی موازی این روش ملاحظه کنید. در حقیقت هدف اصلی از الگوریتم ارائه یک طرح موازی بوده است، اما سریعتر از الگوریتم QR متقارن از کار درمی‌آید حتی اگر به صورت معمول پیاده‌سازی شود. مقاله جدید Gu and Eisenstat (1994a, 1994b) را ملاحظه کنید.

۸-۱۲ الگوریتم لانزوس برای ماتریسهای متقارن

زمینه‌های کاربردی نظیر دستگاههای نیرو، علوم فضایی، شیمی و فیزیک کوانتومی، و مطالعات هسته‌ای وجود دارند که در آنها معمولاً مسائل ویژه برای ماتریسهای از مرتبه خیلی بزرگ پیدا می‌شوند. بیشتر مسائل بزرگی که در کاربردها مطرح می‌شوند، تنک هستند. تحقیق در این زمینه خیلی فعال است. مسأله متقارن بهتر از مسأله نامتقارن فهمیده می‌شود. اکنون روشهای تثبیت شده خوبی برای طیف ماتریسهای متقارن خیلی بزرگ و تنک وجود دارند. یک روش که ابتدا توسط لانزوس (۱۹۵۰) اختراع گردید به طور قابل ملاحظه‌ای در این زمینه مورد توجه قرار گرفته است. در این بخش این روش و کاربردهای آن را برای محاسبه مقادیر ویژه مورد بحث قرار خواهیم داد. روش QR توصیف شده در بخش قبل برای مسائل مقدار ویژه بزرگ و تنک عملی نیست. تنک بودن از بین می‌رود و ذخیره‌سازی یک مسأله خواهد شد.

الگوریتم لانزوس متقارن

با مفروض بودن یک ماتریس متقارن $n \times n$ مانند A و یک بردار واحد v_1 الگوریتم لانزوس

به طور همزمان یک ماتریس سه قطری متقارن T و یک ماتریس یکا متعامد V را به قسمی می سازد که

$$T = V^T A V \quad (۸-۱۲-۱)$$

الگوریتم می تواند به سادگی به صورت زیر استنتاج شود.
فرض کنید

$$T = \begin{bmatrix} \alpha_1 & \beta_1 & \dots & 0 \\ \beta_1 & \alpha_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \beta_{n-1} \\ 0 & \dots & \beta_{n-1} & \alpha_n \end{bmatrix}$$

و $V = (v_1, v_2, \dots, v_n)$ آنگاه از معادله

$$V^T A V = T$$

یا

$$A V = V T$$

یا

$$A(v_1, v_2, \dots, v_n) = (v_1, v_2, \dots, v_n) \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_1 & \beta_2 & & 0 \\ & \beta_2 & \ddots & \ddots & \\ 0 & & \ddots & \ddots & \ddots \\ & & & \beta_{n-1} & \alpha_n \end{bmatrix}$$

خواهیم داشت :

$$A v_j = \alpha_j v_j + \beta_{j-1} v_{j-1} + \beta_j v_{j+1} \quad j = 1, 2, \dots, n-1 \quad (۸-۱۲-۲)$$

(در اینجا فرض کرده ایم که $\beta_0 = 0$). از ضرب طرفین این معادله از طرف چپ در v_j^T و ملاحظه این که شرط یکا متعامد بودن روابط

$$v_j^T v_j = 1$$

$$v_j^T v_k = 0, \quad j \neq k$$

را ارائه می دهد، به دست می آوریم

$$\alpha_j = v_j^T A v_j, \quad j = 1, 2, \dots, n \quad (3-12-8)$$

همچنین اگر بنویسیم

$$r_j = A v_j - \alpha_j v_j - \beta_{j-1} v_{j-1} \quad (4-12-8)$$

آنگاه از (۲-۱۲-۸) و (۴-۱۲-۸) به دست می آوریم

$$v_{j+1} = \frac{r_j}{\beta_j},$$

به شرط آن که

$$\beta_j \neq 0$$

مخالف صفر بودن β_j تضمین می شود اگر اختیار کنیم $\beta_j = \|r_j\|_2$.

این بحث الگوریتم لانزوس پایه‌ای زیر را که به نام ریاضیدان و فیزیکدان مجارستانی کورنلیوس لانزوس (۱۸۹۳-۱۹۷۴) نام گذاری شده است، به ما ارائه می دهد.

الگوریتم ۱-۱۲-۸ لانزوس متقارن پایه‌ای. با مفروض بودن یک ماتریس متقارن $n \times n$ مانند A و یک بردار واحد v_1 ، الگوریتم زیر همزمان عناصر یک ماتریس سه قطری متقارن T و یک ماتریس یکا متعامد $V = (v_1, v_2, \dots, v_n)$ را به قسمی می سازد که

$$V^T A V = T$$

Set $v_0 = 0, \beta_0 = 1, r_0 = v_1$

For $j = 1, 2, \dots, n$ do

$$v_j = r_{j-1} / \beta_{j-1}$$

$$\alpha_j = v_j^T A v_j$$

$$r_j = (A - \alpha_j I) v_j - \beta_{j-1} v_{j-1}$$

$$\beta_j = \|r_j\|_2$$

تذکرها:

۱. بردارهای v_1, v_2, \dots, v_n بردارهای لانزوس نامیده می شوند.

۲. هر بردار لانزوس v_{j+1} بر همه بردارهای قبلی عمود است به شرط آن که $\beta_j \neq 0$ باشد (در حساب دقیق).

۳. همه الگوریتم می تواند فقط با استفاده از یک زیر روال که بتواند ضرب ماتریس - بردار را انجام دهد، پیاده سازی گردد، و بنابراین تنک بودن ماتریس اصلی A می تواند حفظ شود. در حقیقت، برعکس روش هاوس هولدر و یا روش گیونز، ماتریس A هرگز در تمام طول روش تغییر نمی کند. یک چنین طرحی الگوریتم لانزوس را برای محاسبه های تنک جذاب می سازد.
۴. بردارهای $\{v_1, \dots, v_j\}$ یک پایه یکا متعامد برای زیر فضای کرلیف $K_j(A, v_1)$ می سازد.
۵. روش آرنولدی در حالت متقارن، ارائه شده در فصل ۶، همان روش لانزوس متقارن توصیف شده در اینجا است.

اکنون یک نسخه گرام - اشمیت اصلاح شده از الگوریتم ۸-۱۲-۱ را که از لحاظ محاسباتی مؤثرتر است ارائه می دهیم. فرض کنید $m \leq n$.

الگوریتم ۸-۱۲-۲ لانزوس دوباره فرمول بندی شده

$$\text{Set } v_0 = 0, \beta_0 = 1, r_0 = v_1$$

$$\text{For } j = 1, 2, \dots, m \text{ do}$$

$$\text{Step 1: } v_j = r_{j-1} / \beta_{j-1}$$

$$\text{Step 2: } u_j \equiv Av_j$$

$$\text{Step 3: } r_j \equiv u_j - \beta_{j-1}v_{j-1}$$

$$\text{Step 4: } \alpha_j = v_j^T u_j$$

$$\text{Step 5: } r_j = r_j - \alpha_j v_j$$

$$\text{Step 6: } \beta_j = \|r_j\|_2$$

مثال ۸-۱۲-۱

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}, \quad r_0 = v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \beta_0 = 1$$

$j = 1$

$$\alpha_1 = 1, \quad \beta_1 = 3/6.056, \quad v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$j = 2$:

$$\alpha_2 = 1/0.769, \quad \beta_2 = 0/6154, \quad v_2 = \begin{bmatrix} \cdot \\ 0/5547 \\ 0/8321 \end{bmatrix}$$

$$T_2 = \begin{bmatrix} \alpha_1 & \beta_1 \\ \beta_1 & \alpha_2 \end{bmatrix} = \begin{bmatrix} 1 & 3/6.56 \\ 3/6.56 & 1/0.769 \end{bmatrix}$$

$j = 3$:

$$\alpha_3 = -0/0.769, \quad \beta_3 = 1/5466 \times 10^{-14}, \quad v_3 = \begin{bmatrix} \cdot \\ 0/8321 \\ -0/5547 \end{bmatrix}$$

$$T = T_3 = \begin{bmatrix} 1 & 3/6.56 & \cdot \\ 3/6.56 & 1/0.769 & 0/6154 \\ \cdot & 0/6154 & -0/0.769 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & 0/5547 & 0/8321 \\ \cdot & 0/8321 & -0/5547 \end{bmatrix}$$

توجه کنید که $V^TAV = T$.

خاصیتهای گرد کردن الگوریتم لانزوس متقارن

از (۸-۱۲-۲) واضح است که اگر الگوریتم لانزوس متقارن از $j = 1$ تا $j = k$ ($k < n$) اجرا شود، آنگاه داریم:

$$AV_j = V_jT_j + r_j e_j^T$$

که در آن $V_j = (v_1, v_2, \dots, v_j)$ و T_j زیر ماتریس اصلی $j \times j$ ماتریس T می باشد. اگر \hat{V}_j ، \hat{T}_j ، و \hat{r}_j به ترتیب مقادیر محاسبه شده متناظر را نمایش دهند، آنگاه می توان نشان داد (Paige (1980)) که

$$A\hat{V}_j = \hat{V}_j\hat{T}_j + \hat{r}_j e_j^T + E_j$$

که در آن $\|E_j\|_2 \approx \mu \|A\|_2$ ، و نشان می دهد که تا آنجا که به معادله $AV_j = V_jT_j + r_j e_j^T$ مربوط

می شود الگوریتم لانزوس دارای خاصیت‌های عددی خیلی مساعدی می باشد. ولیکن همان گونه که در ذیل توضیح داده می شود از دست دادن تعامد بین بردارهای لانزوس امری واقعی است.

از دست دادن تعامد

به طور واضح الگوریتم لانزوس با شکست مواجه می گردد هرگاه هر یک از β_j ها صفر شوند. ولیکن همان گونه که بعد ملاحظه خواهیم کرد این امر از لحاظ محاسباتی یک توفیق اجباری است. بلافاصله ما یک زیر فضای پایا به دست می آوریم. ولیکن این امر در عمل بندرت اتفاق می افتد. مشکل اصلی این است که بر اثر حذفهایی که در طی محاسبه های r_j رخ می دهند $\hat{\beta}_j$ محاسبه شده می تواند خیلی کوچک باشد. و یک $\hat{\beta}_j$ کوچک می تواند موجب از دست رفتن جدی تعامد بین بردارهای محاسبه شده \hat{v}_j گردد. این مطلب را می توان در نتیجه زیر که می تواند به سادگی با استفاده از قوانین پایه ای حساب ممیز شناور به دست آید، مشاهده کرد.

$$\left| \hat{v}_{j+1}^T \hat{v}_j \right| \approx \frac{\left| \hat{r}_j^T \hat{v}_j \right| + \mu \|A\|_2}{\left| \hat{\beta}_j \right|}$$

بنابراین اگر $\hat{\beta}_j$ کوچک باشد، آنگاه \hat{v}_j و \hat{v}_{j+1} از متعامد بودن دور خواهند بود.

چون الگوریتم لانزوس یک نقش مهم در محاسبات ماتریسی تنک بازی می کند، میزان مهمی از تحقیقات به ساختن یک الگوریتم عملی اختصاص داده شده است. اکنون روشهایی وجود دارند. نظیر لانزوس با متعامدسازی کامل که بردارهای لانزوسی را تولید می کند که با دقتی که کار می شود بر هم عمود باشند، یا لانزوس با متعامدسازی انتخابی، که برای تحمیل کردن تعامد فقط بر بردارهای انتخابی استفاده می شود، هرگاه مورد نیاز باشد. برای جزئیات بیشتر خواننده را به کتابهای معروف در این زمینه نظیر Parlett (1980b) و Cullum and Willoughby (1985) ارجاع می دهیم. چندین مقاله از پیچ (Paige 1970, 1976, 1980) نیز که کار پیشگام او در سالهای اول ۱۹۷۰ توجه پژوهشگران را در این زمینه از نو جلب کرد، برای مطالعه بسیار مفید هستند. برنامه هایی برای متعامدسازی کامل یا انتخابی در کتاب Golub and Van Loan (1989, 487-9) با توضیحاتی ارائه شده اند. برای متعامدسازی انتخابی مرجع Parlett and Scott (1979) را ملاحظه کنید. همچنین کتاب Saad (1992) را مطالعه کنید.

محاسبه مقادیر ویژه A

تنها هدف الگوریتم لانزوس در اینجا نشان دادن این است که ماتریسهای لانزوس T_j می توانند برای

محاسبه برخی مقادیر ویژه ماتریس A استفاده شوند.

قبلاً ملاحظه نمودیم که وقتی یک $\hat{\beta}_j$ دقیقاً برابر صفر می‌شود، یک زیرفضای پایا داریم. در حقیقت این نشانه خوبی است. متأسفانه این خیلی بندرت در عمل رخ می‌دهد. در عمل برای j به قدر کافی بزرگ، مقادیر ویژه T_j تقریبهای خوبی برای مقادیر ویژه اکسترم A فراهم می‌سازند. سؤالی که مطرح می‌شود این است: آیا می‌توان کرانه‌های پسینی ارائه داد؟ برای اتمام این مطلب تعریفهای زیر را ارائه می‌دهیم.

تعریف ۸-۱۲-۱ فرض کنید (θ_j, s_j) یک جفت ویژه از T_j باشند. آنگاه جفت (θ_j, y_j) که در آن y_j توسط $y_j = V_j s_j$ تعریف می‌شود، جفت ریتز نامیده می‌شود، θ_j ها معروف به مقادیر ریتز هستند و y_j ها بردارهای ریتز نامیده می‌شوند.

اکنون به این سؤال برمی‌گردیم که تا چه اندازه خوب یک جفت ریتز یک جفت ویژه را تقریب می‌زند. نتیجه زیر را که می‌تواند به آسانی ثابت شود (تمرین ۴۶)، بیان می‌کنیم.

قضیه ۸-۱۲-۱ فرض کنید $r_i = Ay_i - y_i \theta_i$ ، $i = 1, \dots, j$ ، آنگاه در هر بازه $[r_i, \theta_i + \|r_i\|_2, \theta_i - \|r_i\|_2]$ یک مقدار ویژه از A وجود دارد.

بنابراین از قضیه ۸-۱۲-۱ نتیجه می‌شود که $\|r_i\|_2$ یک اندازه خوب از میزان دقت جفت ریتز (θ_i, y_i) ارائه می‌دهد.

محاسبه $\|r_i\|_2$

خوشبختانه می‌توانیم در هرگام $\|r_i\|_2$ را از T_j بدون محاسبه $y_i = V_j s_j$ حساب کنیم. تجزیه شور T_j را در نظر بگیرید:

$$S_j^T T_j S_j = \text{diag}(\theta_1, \theta_2, \dots, \theta_j)$$

و فرض کنید

$$Y_j = (y_1, y_2, \dots, y_j) = V_j S_j = V_j (s_1, s_2, \dots, s_j)$$

آنگاه با استفاده از نرم ۲ داریم

$$\begin{aligned} \|r_i\| &= \|Ay_i - y_i \theta_i\| = \|AV_j s_i - V_j s_i \theta_i\| \\ &= \|(AV_j - V_j T_j) s_i\| && \text{(زیرا } T_j s_i = s_i \theta_i \text{)} \\ &= \|(\beta_j v_{j+1} e_j^T) s_i\| && \text{(توجه کنید که } AV_j - V_j T_j = \beta_j v_{j+1} e_j^T \text{)} \end{aligned}$$

$$\begin{aligned} &= |\beta_j| \|e_j^T s_i\| && (\|v_{j+1}\|_2 = 1 \text{ زیرا}) \\ &= |\beta_j| |s_{ji}| \end{aligned}$$

که s_{ji} عنصر (j, i) ماتریس S_j است. این بحث می‌تواند در قضیه زیر خلاصه شود.

قضیه ۸-۱۲-۲ قضیه مانده برای یک جفت ریتز. فرض کنید T_j ماتریس سه قطری متقارن $j \times j$ را بعد از j گام الگوریتم لانزوس نمایش دهد. فرض کنید $S_j^T T_j S_j$ شکل شور حقیقی T_j را نمایش دهد.

$$S_j^T T_j S_j = \text{diag}(\theta_1, \theta_2, \dots, \theta_j)$$

یعنی $\theta_1, \dots, \theta_j$ ، مقادیر ویژه T_j هستند. فرض کنید

$$Y_j = (y_1, y_2, \dots, y_j) = V_j S_j$$

آنگاه برای هر i از ۱ تا j داریم

$$\|r_i\| = \|Ay_i - \theta_i y_i\| = |\beta_j| |s_{ji}|$$

که در آن $s_{ji} = e_j^T s_i$ ستون i ام ماتریس S_j است

مثال ۸-۱۲-۲

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

مقادیر ویژه A عبارتند از $0, -0.6235, 9.6235$.

$j = 2$:

$$T_2 = \begin{bmatrix} 1 & 3/6.056 \\ 3/6.056 & 1/0.769 \end{bmatrix}, \quad \beta_2 = -0.6154$$

$$S_2 = \begin{bmatrix} -0.9221 & 0.3870 \\ -0.3870 & -0.9221 \end{bmatrix}$$

$$\theta_1 = -0.5133, \quad \theta_2 = 9.5903$$

(توجه کنید که $\theta_2 = 9/5903$ یک تقریب خوب و معقول از بزرگ‌ترین مقدار ویژه $9/6235$ ماتریس A است).

$$\|r_1\|_2 = |\beta_2| |s_{21}| = 0/6154 \times 0/3970 = 0/2382$$

$$\|r_2\|_2 = |\beta_2| |s_{22}| = 0/5674$$

$$:j = 3$$

$$T_3 = \begin{bmatrix} 1 & 3/6065 & 0 \\ 3/6065 & 1/0769 & 0/6154 \\ 0 & 0/6154 & -0/0769 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 0/8277 & -0/4082 & 0/3851 \\ -0/3727 & 0/1132 & 0/9210 \\ 0/4196 & 0/9058 & 0/0584 \end{bmatrix}$$

$$\|r_1\| = |\beta_3| |s_{31}| = 6/487 \times 10^{-15}$$

$$\|r_2\| = |\beta_3| |s_{32}| = 1/4009 \times 10^{-14}$$

$$\|r_3\| = |\beta_3| |s_{33}| = 9/0367 \times 10^{-16}$$

(در این حالت به کوچک بودن r_3 ، r_2 ، r_1 و r_3 توجه کنید.)

۱۳-۸ مرور و خلاصه

این فصل به مطالعه مسأله مقدار ویژه، مسأله محاسبه مقادیر ویژه و بردارهای ویژه یک ماتریس A اختصاص داده شده است.

۱. کاربردهای مقادیر ویژه و بردارهای ویژه: مسأله مقادیر ویژه در کاربردهای عملی با تنوع وسیعی مطرح می‌شود. مدل‌های ریاضی بسیاری از مسائل که در کاربردهای مهندسی مطرح می‌شوند دستگاه‌های معادلات دیفرانسیل و معادلات تفاضلی هستند، و اساساً مسأله مقدار ویژه در تحلیل پایداری این معادلات مطرح می‌شود. حفظ پایداری یک دستگاه یک اصل مهم برای مهندسان است. برای مثال در مطالعه ارتعاشهای سازه‌ها، مقادیر ویژه و بردارهای ویژه به فرکانسهای طبیعی و دامنه جرمها بستگی دارند، و اگر هر یک از فرکانسهای طبیعی برابر یا نزدیک به یک فرکانس از نیروی دوره‌ای تحمیل شده بر سازه شوند، تشدید رخ می‌دهد. این حالتی است که مهندسين ترجیح می‌دهند از آن پرهیز کنند. ما در این فصل مثالهایی را در مورد

مسابقه تسلیحاتی اروپایی، کمانش یک تیر، شبیه‌سازی جریان عبوری از یک مدار الکتریکی، ارتعاش یک ساختمان، و آنالیز مؤلفه‌های اصلی در آمار با ارجاع به یک تحلیل بورس اوراق بهادار ارائه دادیم. ما سعی کردیم نشان دهیم که تا چه حد مسأله مقدار ویژه در کاربردهای عملی اهمیت دارد.

۲. محدود سازی مقادیر ویژه: در کاربردهای زیادی شناخت صریح مقادیر ویژه لازم نیست؛ همه آن‌چه که مورد نیاز است، شناخت توزیع مقادیر ویژه در یک ناحیه از صفحه مختلط یا برآوردهایی از برخی مقادیر ویژه خاص است.

(الف) از قضیه‌های قرص گرشگورین (قضیه‌های $1-4-8$ و $2-4-8$) می‌توان برای به دست آوردن یک ناحیه از صفحه مختلط حاوی همه مقادیر ویژه یا در برخی حالتها تعدادی از مقادیر ویژه در یک ناحیه استفاده کرد. ولیکن برآوردها می‌توانند خیلی دقیق نباشند.

(ب) همچنین $\|A\| \leq |\lambda|$ (قضیه $3-4-8$). این قضیه بیان می‌کند که کران بالای هر مقدار ویژه A را می‌توان با محاسبه نرم آن به دست آورد. یادآور می‌شویم که این نتیجه در تحلیل همگرایی روشهای تکراری برای حل دستگاهها اهمیت دارد.

۳. روش توانی و تکرار معکوس: کاربردهایی وجود دارند نظیر تحلیل دستگاههای پویا، تحلیل ارتعاش سازه‌ها، کمانش یک تیر و آنالیز مؤلفه‌های اصلی در آمار که فقط بزرگ‌ترین و کوچک‌ترین (از لحاظ اندازه) مقدار ویژه و یا تنها چند مقدار ویژه اول یا آخر و بردارهای ویژه متناظر آنها مورد نیاز می‌باشند.

از روش توانی و روش توانی معکوس مبتنی بر ساختن ضمنی توانهایی از A می‌توان برای محاسبه این مقادیر ویژه و بردارهای ویژه استفاده کرد. روش توانی خیلی از جهت پیاده‌سازی ساده است و برای ماتریسهای بزرگ و تنک مناسب می‌باشد، اما برخی محدودیتهای عددی وجود دارند.

در عمل روش توانی باید با یک انتقال مناسب استفاده شود. روش توانی معکوس به طور ساده روش توانی اعمال شده بر $(A - \sigma I)^{-1}$ می‌باشد، که در آن σ یک انتقال مناسب است. این روش برای محاسبه یک بردار ویژه استفاده می‌شود هنگامی که یک تقریب انصافاً خوب برای یک مقدار ویژه معلوم باشد.

۴. تکرار خارج قسمت ریلی: خارج قسمت

$$R_q = \frac{x^T A x}{x^T x}$$

به خارج قسمت ریلی معروف است، این خارج قسمت یک برآورد از یک مقدار ویژه λ را وقتی ارائه می‌دهد که یک برآورد از بردار ویژه x معلوم باشد. این ایده، هنگامی که با روش تکراری معکوس ترکیب می‌شود، می‌تواند برای محاسبه یک تقریب برای یک مقدار ویژه و بردار ویژه نظیرش استفاده شود. فرایند به تکرار خارج قسمت ریلی معروف است (الگوریتم ۸-۵-۳).

۵. حساسیت مقادیر ویژه و بردارهای ویژه:

(الف) قضیه بوئر - فایک (قضیه ۸-۷-۱) بیان می‌کند که اگر A یک ماتریس قابل قطری شدن باشد، آنگاه عدد وضعیت ماتریس تبدیل X ، $\|X\| \|X^{-1}\|$ ، $\text{Cond}(X)$ ، نقش عدد وضعیت مسأله مقدار ویژه را بازی می‌کند. اگر این عدد بزرگ باشد، آنگاه یک تغییر کوچک در A می‌تواند موجب تغییرات بامعنی در مقادیر ویژه گردد. چون یک ماتریس متقارن A می‌تواند با تبدیلات تشابهی متعامد به یک ماتریس قطری تبدیل شود و عدد وضعیت یک ماتریس متعامد (نسبت به نرم ۲) برابر ۱ است، بنابراین از قضیه بوئر - فایک نتیجه می‌شود که مقادیر ویژه یک ماتریس متقارن نسبت به اختلالهای کوچک حساس نیستند. قضیه ۸-۱۱-۲ را نیز ملاحظه کنید.

(ب) اگر یک مسأله مقدار ویژه بدووضع باشد، آنگاه ممکن است برخی از مقادیر ویژه از سایر مقادیر ویژه حساس تر باشند. بنابراین مهم است که حساسیت مقادیر ویژه فردی را بشناسیم. متأسفانه برای اندازه‌گیری حساسیت یک مقدار ویژه فردی نیاز به شناخت هر دو بردار ویژه راست و چپ متناظر با آن مقدار ویژه داریم. عدد وضعیت مقدار ویژه λ معکوس عدد $|y_i^T x_i|$ است که در آن x_i و y_i به ترتیب بردارهای ویژه راست و چپ متناظر با λ هستند.

(ج) حساسیت یک بردار ویژه λ متناظر با یک مقدار λ اولاً به عدد وضعیت همه مقادیر ویژه بجز λ و ثانیاً به فاصله λ از سایر مقادیر ویژه بستگی دارد (قضیه ۸-۸-۱). بنابراین اگر مقادیر ویژه از یکدیگر خوب مجزای باشند و خوش‌وضع باشند، آنگاه بردارهای ویژه خوش‌وضع هستند. از طرف دیگر اگر یک مقدار ویژه چندگانه یافت شود یا یک مقدار ویژه نزدیک به مقدار ویژه دیگر یافت شود، آنگاه چند بردار ویژه بدووضع یافت می‌شود. بویژه این برای یک ماتریس متقارن معنی‌دار است. مقادیر ویژه یک ماتریس متقارن خوش‌وضع هستند، اما بردارهای ویژه می‌توانند کاملاً بدووضع باشند.

۶. محاسبه مقدار ویژه از طریق چندجمله‌ای مشخصه و شکل متعارف جردن: یک تبدیل تشابهی مقادیر ویژه را حفظ می‌کند و می‌دانیم که یک ماتریس A را می‌توان با تبدیل تشابهی به

شکل متعارف جردن (JCF) و به شکل فربنیوس (یا یک شکل همراه اگر A تخفیف‌ناپذیر باشد) تبدیل کرد. مقادیر ویژه این شکل‌های متراکم به طور ساده‌تر محاسبه می‌شوند. JCF مقادیر ویژه را به طور صریح نمایش می‌دهد و با شکل فربنیوس یا همراه چندجمله‌ای مشخصه A به طور بدیهی محاسبه می‌شود؛ سپس برای به دست آوردن مقادیر ویژه که صفرهای چندجمله‌ای مشخصه هستند یک روش پیدا کردن ریشه می‌تواند بر چندجمله‌ای مشخصه اعمال شود.

ولیکن در عمل محاسبه مقادیر ویژه از طریق چندجمله‌ای مشخصه یا JCF توصیه نمی‌شود. به دست آوردن این شکلها ممکن است به یک ماتریس تبدیل بدوضع نیاز داشته باشد و حساسیت مسأله مقدار ویژه به عدد وضعیت این ماتریس تبدیل بستگی دارد. به طور کلی در محاسبه مقدار ویژه از تبدیل تشابهی بدوضع باید پرهیز شود.

استفاده از ماتریسهای تبدیل خوش‌وضع نظیر ماتریسهای متعامد مطلوب است.

۷. الگوریتم تکرار QR : مورد استفاده‌ترین الگوریتم برای پیدا کردن مقادیر ویژه یک ماتریس الگوریتم تکرار QR است. برای یک ماتریس حقیقی A ، الگوریتم اساساً به طور تکراری شکل شور حقیقی (RSF) ماتریس A را توسط تبدیل تشابهی متعامد می‌سازد. چون الگوریتم بر تکرار تجزیه‌های QR مبتنی است و هر تجزیه QR یک ماتریس $n \times n$ به $O(n^3)$ عمل ممیز شناور (flops) نیاز دارد، n گام از الگوریتم تکراری QR ، اگر به طور ساده پیاده‌سازی شود (که آن را تکرار QR پایه‌ای می‌نامیم) نیاز به $O(n^4)$ عمل ممیز شناور خواهد داشت که الگوریتم را غیر عملی می‌سازد.

(الف) بنابراین ابتدا A قبل از شروع تکرار QR توسط تبدیل تشابهی به یک ماتریس هسنبِرگی H کاهش می‌یابد. در اینجا مشاهدات کلیدی این هستند که کاهش A به H فقط یک بار انجام می‌شود و شکل هسنبِرگی در هر تکرار حفظ می‌گردد.

(ب) ولیکن همگرایی الگوریتم تکراری QR هسنبِرگی می‌تواند در صورت وجود یک مقدار ویژه نزدیک - یا مقدار ویژه چندگانه کاملاً آهسته باشد. همگرایی می‌تواند با استفاده از انتقال مناسب سرعت بخشیده شود. در عمل انتقالهای مضاعف استفاده می‌شوند. در هر تکرار انتقالها مقادیر ویژه زیر ماتریس 2×2 واقع در گوشه راست پایینی هستند. چون مقادیر ویژه یک ماتریس حقیقی می‌توانند مختلط باشند، معمولاً حساب مختلط مورد نیاز است. ولیکن محاسبات می‌توانند طوری تنظیم شوند که از حساب مختلط بتوان پرهیز

کرد. همچنین مقادیر ویژه ماتریس گوشه راست پایینی در هر تکرار لازم نیست به طور صریح محاسبه شوند. فرایند به تکرار QR ضمنی با انتقال مضاعف معروف است (بخش ۸-۹-۶).

با انتقالهای مضاعف دو مقدار ویژه در یک زمان با هم محاسبه می‌شوند. به محض این که دو مقدار ویژه محاسبه شدند ماتریس تقلیل می‌یابد و فرایند بر ماتریس تقلیل یافته تکمیل می‌گردد. به نظر می‌رسد که تکرار QR ضمنی با انتقال مضاعف عملی‌ترین الگوریتم برای محاسبه مقادیر ویژه یک ماتریس متراکم با اندازه نسبتاً کوچک می‌باشد.

۸. مرتب کردن مقادیر ویژه: مقادیر ویژه ظاهر شده در RSF حاصل از الگوریتم تکراری QR ممکن است به ترتیب دلخواه ظاهر نشوند، هر چند بعضی از کاربردها وجود دارند که به آن نیاز دارند. ولیکن با کمی کار اضافی مقادیر ویژه می‌توانند به هر ترتیب دلخواهی ارائه شوند. یک برنامه فرترن عالی، طراحی شده توسط استوارت وجود دارد، که این کار را انجام می‌دهد. (بخش ۸-۹-۸).

۹. محاسبه بردارهای ویژه: به محض این که تقریبی برای یک مقدار ویژه توسط تکرار QR به دست آید تکرار معکوس می‌تواند برای محاسبه بردار ویژه نظیر آن به کار برده شود. چون در ابتدا برای پیاده‌سازی عملی الگوریتم تکراری QR ماتریس A به یک ماتریس هسنبرگی کاهش می‌یابد، می‌توان از مزیت ساختار یک ماتریس هسنبرگی در محاسبه یک بردار ویژه با استفاده از تکرار معکوس استفاده کرد. متناوباً می‌توانیم از RSF ماتریس A مستقیماً برای محاسبه بردارهای ویژه استفاده کنیم (بخش ۸-۱۰-۲).

۱۰. مسأله مقدار ویژه:

(الف) البته از الگوریتم تکراری QR (الگوریتم ۸-۱۱-۲) می‌توان برای محاسبه مقادیر ویژه یک ماتریس متقارن A استفاده کرد. یک انتقال که انتقال ویلکینسون نامیده می‌شود، معمولاً در اینجا استفاده می‌شود. ثابت شده است که همگرایی الگوریتم تکراری QR متقارن با انتقال ویلکینسون از درجه دو می‌باشد؛ ولیکن در عمل اغلب از درجه سه می‌باشد.

(ب) مقادیر ویژه و بردارهای ویژه ماتریس متقارن A از برخی خاصیت‌های خاص و جالب توجه بهره‌مند می‌باشند و روشهایی برای حل مسأله متقارن وجود دارند که از این خاصیتها

بهره‌برداری می‌کنند. یکی از این روشها روش تنصیف است. ابتدا ماتریس متقارن مفروض A با استفاده از تبدیل تشابهی متعامد به یک ماتریس متقارن سه‌قطری T تبدیل می‌شود، سپس الگوریتم معروف تنصیف برای مسأله ریشه‌یابی بر چندجمله‌ای مشخصه T ، به دست آمده از یک فرمول بازگشتی ساده، اعمال می‌شود (بخش ۸-۱۱-۱). این فرمول بازگشتی نه تنها چندجمله‌ای مشخصه T را ارائه می‌دهد، بلکه چندجمله‌ایهای مشخصه همه زیر ماتریسهای اصلی پیشرو را نیز فراهم می‌سازد. یک حقیقت قابل توجه این است که این چندجمله‌ایها دارای خاصیت دنباله استورم هستند، که از آن برای پیاده‌سازی روش تنصیف استفاده می‌شود. بویژه روش تنصیف برای پیدا کردن مقادیر ویژه یک ماتریس متقارن در یک بازه مفروض از خط حقیقی مفید است.

(ج) روشهای دیگری نیز نظیر روش تقسیم و تسخیر و روش ژاکوبی برای مسأله مقدار ویژه وجود دارند. در این فصل ما فقط روش اول (الگوریتم ۸-۱۱-۳) را به طور خلاصه معرفی کردیم. این روشها بویژه برای محاسبات موازی ارزشمند هستند.

۱۱. مسأله مقدار ویژه بزرگ و تنک: مسأله مقدار ویژه برای ماتریسهای بزرگ و تنک یک زمینه فعال تحقیق است. پیشرفته‌ترین روشهایی که از روشهای از نوع لانزوس و آرنولد با برخی گونه‌های دوباره متعامدسازی و پیش شرط سازی خاص استفاده می‌کنند می‌توانند فقط تعدادی از مقادیر ویژه اکسترم را محاسبه کنند. روشها برای مسائل ویژه متقارن تنک و بزرگ بهتر از مسائل نامتقارن ساخته شده‌اند و بهتر نیز قابل فهم می‌باشند.

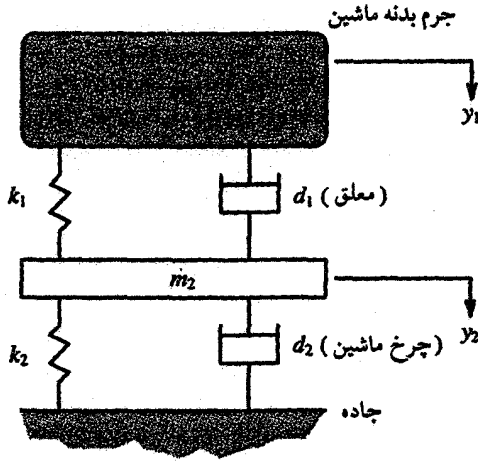
به خاطر کامل شدن مطالب و ارائه یک ایده از این که چگونه روشهای از نوع لانزوس در محاسبات مقدار ویژه استفاده می‌شوند، یک توصیف خلاصه از روش لانزوس متقارن را بیان کردیم (بخش ۸-۱۲).

تمرینها

تذکر: هرگاه لازم است از MATLAB استفاده کنید.

بخش ۳-۸

۱. مدل زیر را برای ارتعاش عمودی یک موتور اتومبیل ملاحظه کنید (Muller and Schienlen, (1985):



(الف) معادلات زیر و مربوط به حرکت اتومبیل را فرمول‌بندی کنید، از ثابتهای میرایی d_1 جذب‌کننده شوک و d_2 چرخ اتومبیل صرف‌نظر کنید:

$$M\ddot{y} + Ky = 0$$

که در آن $M = \text{diag}(m_1, m_2)$

$$K = \begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_1 + k_2 \end{bmatrix}$$

و

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

پایداری حرکت را تعیین کنید هنگامی که

$$m_1 = m_2 = 1200 \text{ kg}$$

$$k_1 = k_2 = 300 \frac{\text{N}}{\text{cm}}$$

(ب) نشان دهید هنگامی که فقط از میرایی d_2 چرخ اتومبیل صرف‌نظر می‌شود معادله حرکت به صورت زیر ارائه می‌شود

$$M\ddot{y} + Dy + Ky = 0$$

M و K همانند قسمت (الف) می‌باشند و

$$D = \begin{bmatrix} d_1 & -d_1 \\ -d_1 & d_1 \end{bmatrix}$$

در این حالت پایداری حرکت را بررسی کنید هنگامی که

$$d_1 = 4500 \frac{\text{Ns}}{\text{cm}}$$

راهنمایی: نشان دهید که دستگاه

$$M\ddot{y} + D\dot{y} + Ky = 0$$

با دستگاه مرتبه اول زیر معادل است

$$\dot{x}(t) = Ax(t)$$

در اینجا

$$A = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}D \end{bmatrix}$$

و

$$x(t) = \begin{bmatrix} y(t) \\ \dot{y}(t) \end{bmatrix}$$

۲. جواب معادله $M\ddot{y} + Ky = 0$ را با M و K ارائه شده در مسأله ۱ (الف) و استفاده از شرایط اولیه

$$y(0) = 0 \quad \text{و} \quad \dot{y}(0) = (1, 1, \dots, 1)^T \quad \text{بنویسید.}$$

۳. یک مسأله مقدار ویژه را برای یک شبکه LC مشابه با مورد ارائه شده در بخش ۸-۲-۲، اما فقط با

سه حلقه تدوین نمایید.

نشان دهید که در این حالت فرکانسهای طبیعی به صورت زیر ارائه می شوند

$$w = \begin{cases} \frac{0.4451}{\sqrt{LC}} \\ \frac{1.2470}{\sqrt{LC}} \\ \frac{1.8019}{\sqrt{LC}} \end{cases}$$

مدها را پیدا کنید و شرح دهید که چگونه در این مدها جریانها نوسان می کنند.

بخش ۸-۴

۴. برای ماتریسهای زیر قضیه‌های قرص گرشگورین را برای به دست آوردن کرانه‌های مقادیر ویژه به کار ببرید.

$$A = \begin{bmatrix} 1.0 & 1 & 1 \\ 2 & 1.0 & 1 \\ 2 & 2 & 1.0 \end{bmatrix} \quad (\text{الف})$$

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 1 & 1 & 6 \end{bmatrix} \quad (\text{ب})$$

$$A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \quad (\text{ج})$$

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \quad (\text{د})$$

$$A = \begin{bmatrix} 1/0.00 & 0/577 & 0/509 & 0/387 & 0/462 \\ 0/577 & 1/0.00 & 0/599 & 0/389 & 0/322 \\ 0/509 & 0/599 & 1/0.00 & 0/436 & 0/426 \\ 0/387 & 0/389 & 0/436 & 1/0.00 & 0/523 \\ 0/452 & 0/322 & 0/426 & 0/523 & 1/0.00 \end{bmatrix} \quad (\text{ه})$$

$$A = \begin{bmatrix} 1-i & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1-i & 1+i \end{bmatrix} \quad (\text{و})$$

۵. با استفاده از قضیه قرص گرشگورین ثابت کنید که ماتریس غالب قطری نامنفرد است.

۶. فرض کنید x یک بردار ویژه متناظر با یک مقدار ویژه λ تنها در قرص گرشگورین R_k باشد. ثابت کنید که $|x_k| > |x_i|$ به ازای $i \neq k$

۷. فرض کنید $A = (a_{ij})$ یک ماتریس متقارن $n \times n$ باشد. سپس با استفاده از قضیه قرص

گرشگورین ثابت کنید که هر مقدار ویژه A در یکی از بازه‌های $[a_{ij} - r_i, a_{ij} + r_i]$ قرار خواهد گرفت. بازه‌ای پیدا کنید که همه مقادیر ویژه A در آن قرار گیرند.

بخش ۵-۸

۸. با به کار بردن روش توانی و روش توانی معکوس، مقدار ویژه غالب و بردار ویژه متناظر را برای هر یک از ماتریسهای تمرین ۴ پیدا کنید.

۹. ثابت کنید که اگر $\lambda_1, \lambda_2, \dots, \lambda_n$ مقادیر ویژه A باشند، و v_1, v_2, \dots, v_n بردارهای ویژه متناظر باشند، آنگاه $\sigma - \lambda_1, \dots, \sigma - \lambda_n$ مقادیر ویژه $A - \sigma I$ هستند و بردارهای ویژه متناظر v_1, \dots, v_n می‌باشند.

۱۰. سرعت کند همگرایی روش توانی را با ماتریسهای زیر توضیح دهید.

$$A = \begin{bmatrix} 3 & 3 & 2 \\ 0 & 1 & 0 \\ 0 & 1 & 2/9 \end{bmatrix} \quad (\text{الف})$$

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 10 & 0 \\ 1 & 1 & 9/8 \end{bmatrix} \quad (\text{ب})$$

یک انتقال مناسب σ انتخاب کنید و سپس روش توانی انتقال یافته را در مورد هر یک از ماتریسها به کار برید و بهبود سرعت همگرایی را ملاحظه کنید.

۱۱. تکرار متعامد: روش تکراری زیر روش توانی را تعمیم می‌دهد و معروف به فرایند تکرار متعامد است. از این فرایند می‌توان برای محاسبه p ($p > 1$) مقدار ویژه بزرگ‌تر و بردارهای ویژه متناظر استفاده کرد. فرض کنید Q_1 یک ماتریس یکا متعامد $n \times p$ باشد. آنگاه

For $k = 2, 3, \dots$, do

Step 1: Compute $B_k = AQ_{k-1}$

Step 2: Factorize into $QR: B_k = Q_k R_k$

یک الگوریتم کامل مبتنی بر این روش برای محاسبه p مقدار ویژه اول و بردارهای ویژه متناظر بنویسید و سپس الگوریتم خودتان را با هر یک از ماتریسهای تمرین ۱۰ آزمایش کنید، اختیار کنید $p = 2$.

۱۲. تکرار متعامد معکوس: تکرار زیر که تکرار متعامد معکوس نامیده می‌شود روش توانی معکوس

را تعمیم می‌دهد و می‌تواند برای محاسبه p مقدار ویژه کوچک‌تر یک ماتریس نامنفرد A استفاده شود. فرض کنید Q_1 یک ماتریس یکا متعامد $n \times p$ باشد

For $k = 2, 3, \dots$

Step 1: Solve for $B_k: AB_k = Q_{k-1}$

Step 2: Factorize into $QR: B_k = Q_k R_k$

تکرار متعامد معکوس را برای محاسبه دو مقدار ویژه کوچک‌تر هر یک از ماتریسهای تمرین ۴ به کار برید.

۱۳. فرض کنید T یک ماتریس متقارن سه قطری باشد. فرض کنید الگوریتم تکراری خارج قسمت ریلی (الگوریتم ۸-۵-۳) با استفاده از نرم ۲ در گام ۳ بر T با $e_n = x_0$ اعمال شود؛ سپس ثابت کنید که $x_1 = q_n$ که در آن q_n ستون آخر Q در $(T - \sigma_1 I) = QR$ می‌باشد.

۱۴. مقدار ویژه زیر غالب هر یک از ماتریسهای تمرین ۴ را با استفاده از تقلیل هاوس هولدر محاسبه کنید. سپس بردار ویژه متناظر را بدون استفاده از تکرار معکوس محاسبه کنید.

۱۵. کوچک‌ترین مقدار ویژه هر یک از ماتریسهای تمرین ۴ را با اعمال روش توانی بر A^{-1} بدون محاسبه صریح A^{-1} محاسبه کنید.

۱۶. تقلیل با استفاده از یک زیر فضای پایا: فرض کنید که یک ماتریس $n \times m$ مانند X با ستونهای مستقل و یک ماتریس $m \times m$ مانند M به قسمی داشته باشیم که

$$AX = XM$$

تجزیه QR ماتریس X را در نظر بگیرید:

$$Q^T X = \begin{bmatrix} R \\ \cdot \end{bmatrix}$$

سپس مطالب زیر را ثابت کنید

$$Q^T A Q = \begin{bmatrix} A_1 & A_2 \\ \cdot & A_3 \end{bmatrix} \quad (\text{الف})$$

(ب) مقادیر ویژه A مقادیر ویژه A_1 و A_3 هستند.

(ج) مقادیر ویژه A مقادیر ویژه M هستند.

بخشهای ۸-۶ تا ۸-۸

۱۷. یک مثال ساده بسازید که نشان دهد که یک تبدیل تشابهی دلخواه می تواند وضعیت مقادیر ویژه ماتریس تبدیل یافته را بدتر کند.

۱۸. اگر $B = U^*AU$ ، که در آن U یکانی است، و $U^*(A + \Delta A)U = B + \Delta B$ ، آنگاه نشان دهید که $\|\Delta B\|_2 = \|\Delta A\|_2$.

۱۹. (الف) ثابت کنید که A نرمال است اگر یک ماتریس یکانی U وجود داشته باشد به قسمی که $U^*AU = D$ ، که در آن D قطری است.

(ب) ثابت کنید که یک ماتریس A یک مجموعه از n بردار ویژه یکا متعامد دارد اگر و فقط اگر A نرمال باشد.

(ج) ثابت کنید که یک ماتریس نرمال A یکانی است اگر و فقط اگر مقادیر ویژه آن روی دایره واحد باشند؛ یعنی برای هر مقدار ویژه λ از A داشته باشیم $|\lambda| = 1$.

(د) شکل شور حقیقی یک ماتریس نرمال چگونه به نظر می رسد؟

(ه) با استفاده از قضیه حساسیت بردار ویژه (قضیه ۸-۸-۱) نشان دهید که اگر A نرمال باشد، آنگاه بردار ویژه x_k متناظر با مقدار ویژه λ_k خوش وضع است اگر λ_k به خوبی از سایر مقدارهای ویژه جدا باشد.

۲۰. هم به طور نظری و هم به طور تجربی نشان دهید که دو بردار ویژه ماتریس

$$A = \text{diag}(1 + \varepsilon, 1 - \varepsilon, 2),$$

که در آن ε یک عدد خیلی کوچک مثبت است، بدوضع هستند. ولیکن مقادیر ویژه خوش وضع هستند.

۲۱. نشان دهید که تبدیل تشابهی یکانی عدد وضعیت در نرم دو یک مقدار ویژه را حفظ می کند.

۲۲. قضیه بوئر - فایک (قضیه ۸-۷-۱) را با استفاده از قضیه اول گرشگورین ثابت کنید.

۲۳. (الف) با مفروض بودن ε ، یک عدد مثبت کوچک، و

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 + \varepsilon \end{bmatrix}$$

ماتریس بردار ویژه X را به قسمی پیدا کنید که $X^{-1}AX$ قطری باشد؛ سپس نشان دهید که مقدارهای ویژه A بدوضع هستند. بدوضع A را به طور محاسباتی با ایجاد یک اختلال کوچک در A و پیدا کردن مقادیر ویژه ماتریس اختلال یافته، تحقیق کنید

(ب) ماتریس زیر را در نظر بگیرید

$$A = \begin{pmatrix} 12 & 11 & 10 & \dots & \dots & 2 & 1 \\ 11 & 11 & 10 & \dots & \dots & 2 & 1 \\ \cdot & 10 & 10 & \dots & \dots & 2 & 1 \\ \vdots & \ddots & \ddots & \ddots & & \vdots & \vdots \\ \vdots & & & \ddots & \ddots & \vdots & \vdots \\ \vdots & & & & \ddots & \vdots & \vdots \\ \cdot & \dots & \dots & \dots & \cdot & 1 & 1 \end{pmatrix}$$

با استفاده از MATLAB و محاسبه مقادیر ویژه A نشان دهید که بزرگ‌ترین مقادیر ویژه A خوش وضع هستند، در حالی که کوچک‌ترین مقادیر ویژه خیلی بد وضع هستند.

بخشهای ۸-۹ و ۸-۱۰

۲۴. ثابت کنید که اگر R بالا مثلثی و Q بالا هسنبرگی باشد، آنگاه RQ بالا هسنبرگی است.

۲۵. یک مثال بسازید و نشان دهید که برای هر ماتریس بالا هسنبرگی دلخواه H لازم نیست که ماتریس Q در

$$H = QR$$

یک ماتریس بالا هسنبرگی باشد. ولیکن اگر R عناصر قطری مخالف صفر داشته باشد، آنگاه Q

در $H = QR$ همیشه بالا هسنبرگی است.

۲۶. ثابت کنید که

(الف) تجزیه QR یک ماتریس هسنبرگی $n \times n$ به $O(n^2)$ عمل ممیز شناور (flops) نیاز دارد.

(ب) تجزیه QR یک ماتریس سه قطری فقط به $O(n)$ عمل ممیز شناور (flops) نیاز دارد.

۲۷. سه تکرار از تکرار QR تک انتقالی را بر ماتریسهای زیر اعمال کنید و همگرایی و یا ناهمگرایی عناصر زیر قطر را ملاحظه کنید

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 3 & 4 \\ 0 & 4 & 1 \end{pmatrix} \quad \text{(الف)}$$

$$A = \begin{pmatrix} 4 & 5 & 6 \\ 1 & 0 & 1 \\ 0 & -2 & 2 \end{pmatrix} \quad \text{(ب)}$$

۲۸. یک گام از تکرار QR تک انتقالی را با ماتریس زیر انجام دهید

$$H = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 10 & 6 \\ 0 & 0.1 & 1 \end{bmatrix}$$

۲۹. QR تک انتقالی ضمنی: یک گام از تکرار QR تک انتقالی را ملاحظه کنید

$$H_k - h_{nn}^{(k)} I = Q_k R_k, \quad H_{k+1} = R_k Q_k + h_{nn}^{(k)} I$$

یا (به طور ساده)

$$H - \lambda I = QR, \quad \bar{H} = RQ + \lambda I \quad (\lambda \text{ حقیقی است})$$

(الف) ثابت کنید که ستون اول Q مضرری از ستون اول $H - \lambda I$ است و در نتیجه حاوی فقط دو عنصر مخالف صفر است.

(ب) اولین ستون $H - \lambda I$ را به $h_1 = (h_{11} - \lambda, h_{21}, 0, \dots, 0)^T$ نشان دهید. یک دوران گیونز P_0 به قسمی پیدا کنید که $P_0 h_1$ یک مضرری از e_1 باشد. نشان دهید که اولین ستون P_0 ، بجز احتمالاً در علامت، همان ستون اول Q است.

(ج) ماتریس $H^{\sim} = P_0 H P_0^T$ را تشکیل دهید. با استفاده از دورانهای گیونز $J(2, 3, \theta)$ ، $J(3, 4, \theta)$ ، ...، $J(n-1, n, \theta)$ ماتریس H^{\sim} را به یک ماتریس بالاهسنبرگی H_1^{\sim} تبدیل کنید. نشان دهید که ماتریس

$$\tilde{Q} = P_0^T J^T(2, 3, \theta) \dots J^T(n-1, n, \theta)$$

دارای همان ستون اول Q است. بالاخره از قضیه ضمنی Q نتیجه بگیرید که ماتریس هسنبرگی H_1^{\sim} اساساً همان \bar{H} است.

گامهای (الف) تا (ج) یک گام از تکرار QR تک انتقالی ضمنی را تشکیل می دهند. یک گام از تکرار QR تک انتقالی ضمنی را بر ماتریس سه قطری متقارن زیر اعمال کنید.

$$\begin{bmatrix} 2 & -1 & \dots & 0 \\ -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ 0 & \dots & -1 & 2 \end{bmatrix}$$

۳۰. یک گام از تکرار QR تک انتقالی صریح (حساب حقیقی) و یک گام از تکرار QR ضمنی با انتقال

مضاعف را برای ماتریس

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 6 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -2 & 2 \end{bmatrix}$$

انجام دهید و نشان دهید که ماتریسهای هسنبرگی نهایی (اساساً) یکسان هستند.

۳۱. تکرار LR : مشابه الگوریتم تکرار QR پایه‌ای (بخش ۸-۹-۱)، یک الگوریتم تکرار LR مبتنی بر تجزیه LU ماتریس A همراه با فرضهای لازم، تدوین نمایید. چرا این الگوریتم بر الگوریتم تکرار QR ترجیح داده نمی‌شود.

۳۲. با در نظر گرفتن ساختار ماتریسهای P_i ، $i = 0, 1, \dots, n-2$ در گام تکرار QR ضمنی با انتقال مضاعف (تمرین ۲۹)، نشان دهید که پیاده‌سازی این گام فقط به $4n^2$ عمل ممیز شناور (flops) نیاز دارد.

۳۳. نشان دهید که ماتریسهای H_0 و H_{s+2} در تکرار QR با انتقال مضاعف دارای مقادیر ویژه یکسان هستند.

۳۴. فرض کنید که $H = H_0$. دنباله $\{H_k\}$ را تولید کنید:

$$H_k - \mu_k I = Q_k R_k, \quad H_{k+1} = R_k Q_k + \mu_k I$$

نشان دهید که

$$\prod_{i=1}^n (H - \mu_i I) = (Q_1 \dots Q_n)(R_n \dots R_1)$$

بخشهای ۸-۱۱ و ۸-۱۲

۳۵. (الف) فرض کنید

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

بدون محاسبه مقادیر ویژه نشان دهید که برای هر مقدار ویژه A داریم $|\lambda| < 4$. نشان دهید که دقیقاً دو مقدار ویژه بزرگ‌تر از ۲ و دو مقدار ویژه کوچک‌تر از ۲ وجود دارد. الگوریتم دنباله استورم - تنصیف (الگوریتم ۸-۱۱-۱) را برای محاسبه مقدار ویژه نزدیک به ۲ به کار ببرید.

(ب) الگوریتم تکرار هسبرگی معکوس را برای A و برای محاسبه بردار ویژه وابسته به مقدار ویژه نزدیک به ۲ به کار ببرید.

(ج) با به کار بردن تکرار QR متقارن با انتقال و یلکینسون مقادیر ویژه A را محاسبه کنید.

۳۶. (الف) ثابت کنید که مقادیر ویژه یک ماتریس سه قطری متقارن حقیقی کاهش ناپذیر حقیقی و متمایز هستند.

(ب) ثابت کنید که اگر k یک مقدار ویژه با مرتبه تکرار k از یک ماتریس سه قطری متقارن حقیقی کاهش ناپذیر T باشد، آنگاه حداقل $(k - 1)$ عنصر زیر قطر T باید صفر باشد.

۳۷. (الف) یک الگوریتم از نوع QR برای محاسبه مقادیر ویژه یک ماتریس معین مثبت متقارن A ، مبتنی بر تجزیه چولسکی (بخش ۶-۴-۶) تدوین نمایید.

(ب) الگوریتم خود را با ماتریس A تمرین ۳۵ آزمایش کنید.

۳۸. فرض کنید

$$A = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}$$

ثابت کنید که مقدار ویژه δ از A که نزدیک ترین به γ می باشد، به صورت زیر است

$$w = \gamma - \frac{\text{sign}(r)\beta^2}{|r| + \sqrt{r^2 + \beta^2}}$$

که در آن $r = (\alpha - \gamma)/2$. توضیح دهید که چرا این فرمول بهتر از فرمول ارائه شده در بخش ۸-۱۱-۲ می باشد.

۳۹. فرض کنید $A = A_1 + iA_2$ یک ماتریس هرمیتی باشد. آنگاه ثابت کنید که

$$B = \begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix}$$

متقارن است. چگونه مقادیر ویژه و بردارهای ویژه A به مقادیر ویژه و بردارهای ویژه B وابسته هستند؟

۴۰. ماتریس سه قطری متقارن T با ابعاد 50×50 را در نظر بگیرید که عناصر قطری آن برابر ۲ بوده و

عناصر قطر زیری و قطر بالایی آن برابر ۱- باشند. الگوریتم ۸-۱۲-۲ را با $20, \dots, 2, 1 = z$ بر

T اعمال کنید. تقریبهایی برای تعداد کمی از مقادیر ویژه اکسترمم ماتریس T با استفاده از

قضیه های ۸-۱۲-۱ و ۸-۱۲-۲ پیدا کنید.

تمرینهای ۴۱ تا ۴۵ مربوط به روش تقسیم و تسخیر هستند (بخش ۸-۱۱-۳).

۴۱. ثابت کنید که اگر مؤلفه u_i بردار u برابر صفر باشد، آنگاه

(الف) i امین ستون \hat{D} برابر $d_i e_i$ است.

(ب) i امین سطر \hat{D} برابر $d_i e_i^T$ است.

(ج) d_i یک مقدار ویژه \hat{D} با بردار ویژه وابسته e_i می باشد.

۴۲. ثابت کنید که اگر k مقدار ویژه D برابر باشند، آنگاه مسأله مقدار ویژه \hat{D} می تواند با حذف

(۱ - k) سطر و ستون تقلیل یابد.

۴۳. فرض کنید که

(الف) $\rho \neq 0$,

(ب) مقادیر ویژه D به صورت $d_1 < d_2 < \dots < d_n$ مرتب شده باشند.

(ج) هیچ کدام از مؤلفه های u صفر نباشند.

آنگاه ثابت کنید که هر بازه (d_i, d_{i+1}) دقیقاً یک مقدار ویژه از $\hat{D} = D + \rho uu^T$ را دربردارد.

۴۴. ثابت کنید که اگر فرضهای تمرین ۴۳ برقرار باشند، آنگاه $D - \lambda I$ نامنفرد است و $u^T q \neq 0$.

۴۵. با رسم نمودارهای معادله ویژه یا طور دیگر نشان دهید که

(الف) $\rho > 0$ ، اگر $d_1 < \lambda_1 < d_2 < \lambda_2 < \dots < \lambda_n < d_n + \rho$

(ب) $\rho < 0$ ، اگر $d_1 + \rho < \lambda_1 < d_2 < \lambda_2 < \dots < d_{n-1} < \lambda_n < d_n$

۴۶. قضیه ۸-۱۲-۱ را ثابت کنید.

برنامه های MATLAB و مسائل

۱. یک برنامه MATLAB برای محاسبه مقدار ویژه غالب یک ماتریس با استفاده از روش توانی

بنویسید.

$$[\text{lambda1}] = \text{power}(A, x0, N)$$

که در آن N تعداد تکرارهایی است که باید انجام شود. نتیجه هر تکرار را چاپ کنید.

(الف) برنامه `power` را برای درج یک انتقال `sigma` تغییر دهید.

$$[\text{lambda1}] = \text{powershift}(A, x0, \text{sigma}, N)$$

(ب) برنامه های `power` و `powershift` را برای ماتریسهای زیر به کار ببرید و سرعت همگرایی را

مقایسه کنید.

داده‌های آزمون

$$A = \begin{bmatrix} 3 & 2 & 3 \\ 0 & 0.99 & 1 \\ 0 & 0 & 2/9 \end{bmatrix}$$

یک ماتریس تولید شده تصادفی از مرتبه ۱۵

ماتریس دو قطری ویلکینسون از مرتبه ۲۰

۲. (الف) با استفاده از `linsyspp` از فصل ۶ یا دستور `\` از MATLAB یک برنامه از MATLAB به

نام `invitr` برای پیاده‌سازی الگوریتم تکرار معکوس بنویسید.

$$[x] = \text{invitr}(A, x_0, \text{sigma}, N)$$

N تعداد تکرارهایی است که باید انجام شود.

(ب) با استفاده از `linsyspp` از فصل ۶ یا دستور `\` از MATLAB یک برنامه از MATLAB به

نام `powersmall` برای محاسبه کوچک‌ترین مقدار ویژه (از لحاظ قدر مطلق) ماتریس A بنویسید.

$$[\text{lambdan}] = \text{powersmall}(A, x_0, N)$$

N تعداد تکرارهایی است که باید انجام شود.

داده‌های آزمون و آزمایشها برای ۲ (الف) و ۲ (ب)

۱. ماتریس سه قطری متقارن 20×20 ظاهر شده در مسأله کمانش بخش ۸-۳-۲ را در نظر بگیرید.

برنامه `power` را با انتخاب دلخواه x_0 برای مقدار ویژه غالب `lambda1` به کار ببرید.

۲. اکنون کوچک‌ترین مقدار ویژه از لحاظ قدر مطلق، `lambdan`، را با استفاده از برنامه‌های

زیر حساب کنید.

(الف) برنامه `powershift` با `sigma = lambda1` و x_0 مانند ۱.

(ب) برنامه `powersmall` با همان x_0 استفاده شده در پیاده‌سازی `power`.

۳. سرعت‌های (الف) و (ب) در ۲ را مقایسه کنید.

۴. کوچک‌ترین بار بحرانی را پیدا کنید که تیر را خم خواهد کرد.

۵. با اختیار کردن `sigma = lambdan` بردار ویژه متناظر با کوچک‌ترین مقدار ویژه A را

توسط `invitr` پیدا کنید.

۳. با استفاده از برنامه‌های `power`، `invtr`، `houszero` (از فصل ۵)، و `phousmul` (از فصل ۴) یک برنامه `MATLAB` بنویسید که الگوریتم تقلیل هاوس هولدر را پیاده‌سازی نماید و مقدار ویژه زیر غالب یک ماتریس را حساب کند.

$$[\text{lambda2}] = \text{housdeflt}(a)$$

داده‌های آزمون

یک تک - جنس، مدل جمعیت گروهی می‌تواند توسط دستگاه معادلات تفاضلی

$$p_0(k+1) = \alpha_0 p_0(k) + \alpha_1 p_1(k) + \dots + \alpha_n p_n(k)$$

$$p_{i+1}(k+1) = \beta_i p_i(k), \quad i = 0, 1, \dots, n-1$$

یا به شکل ماتریسی زیر نمایش داده شود

$$p_{k+1} = A p_k$$

در اینجا α_i نرخ تولد i امین گروه سنی و β_i نرخ بقاء آن گروه می‌باشند (Luenberger (1979, 170)). فرض کنید $n = 50$.

با اختیار کردن $\alpha_0 = 0$ ، $\alpha_1 = \alpha_2 = \dots = \alpha_n = 1$ ، و $\beta_i = 1$ ، $i = 0, 1, \dots, n-1$ و با استفاده از برنامه‌های `power`، `invtr`، `housflt` تعیین کنید که آیا یک رشد با مدت طولانی در جمعیت وجود دارد؛ اگر چنین است توزیع جمعیت نهایی چیست و با چه سرعتی توزیع اصلی به توزیع نهایی نزدیک می‌شود؟ مثال مربوط به مطالعه جمعیت در بخش ۸-۵-۲ را ملاحظه کنید.

۴. قرصهای گرشگورین هر یک از ماتریسهای A تمرین ۴ بخش ۸-۴ را رسم کنید.

۵. (هدف از این تمرین مطالعه چگونه تحت تأثیر قرار گرفتن مقادیر ویژه یک ماتریس توسط وضعیت ماتریس تبدیل می‌باشد).

(الف) برای هر یک از ماتریسهای زیر یک ماتریس X از مرتبه مناسب بسازید که بالا مثلثی با همه عناصر برابر ۱ بوده، به استثنای تعداد کمی از عناصر قطری که خیلی کوچک می‌باشند. سپس با استفاده از دستورهای `eig` و `inv` از `MATLAB` مقادیر ویژه A و مقادیر ویژه AX^{-1} را محاسبه کنید از قالب "long" در `MATLAB` استفاده کنید.

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.9 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & -5 \\ 0 & 1 & 4 \end{bmatrix}$$

ماتریس دو قطری ویلکینسون از مرتبه ۲۰

(ب) قسمت (الف) را با اختیار کردن X به صورت یک ماتریس هاوس هولدر از مرتبه مناسب تکرار کنید.

(ج) نتایج (الف) و (ب) را مقایسه کنید.

۶. (الف) مقادیر ویژه ماتریسهای زیر را با استفاده از دستورهایی زیر حساب کنید.

(i) دستورهایی `poly` و `roots` از MATLAB.

(ii) دستور `eig` از MATLAB.

$$A = \begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & -5 \\ 0 & 1 & 4 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0.19 & 1 \\ 0 & 0 & 0.9999 \end{bmatrix}$$

ماتریس دو قطری ویلکینسون از مرتبه ۲۰

ماتریس تصادفی از مرتبه ۵۰

(ب) نتایج حاصل از (i) و (ii) را برای هر ماتریس مقایسه کنید.

۷. (هدف از این تمرین مطالعه حساسیتهای مقادیر ویژه برخی ماتریسهای معروف با مقادیر ویژه بدو وضع است).

(الف) با استفاده از دستور $[V, D] = \text{eig}(A)$ از MATLAB مقادیر ویژه و ماتریس بردارهای ویژه راست را پیدا کنید. سپس ماتریس بردارهای چپ W را به صورت نشان داده شده در بخش ۸-۷-۲ پیدا کنید.

(ب) مقادیر $s_i = w_i^T v_i$ ، $i = 1, \dots, n$ را حساب کنید که در آن v_i و w_i ستونهای V و W هستند.

(ج) مقادیر $c_i = 1/s_i$ عدد وضعیت i امین مقدار ویژه، $i = 1, 2, \dots, n$ را حساب کنید.

(د) در عنصر $(n, 1)$ ماتریس A توسط 10^{-10} ، 10^{-7} ، 10^{-5} اختلال ایجاد کنید. سپس مقادیر ویژه $\hat{\lambda}_i$ ، $i = 1, \dots, n$ ، ماتریس اختلال یافته را با استفاده از دستور `eig` از

MATLAB حساب کنید.

(ه) یک جدول به شکل زیر برای هر ماتریس تشکیل دهید. همچنین $|\lambda_i|$ ، $|\lambda_i - \hat{\lambda}_i|$ ، $i = 1, 2, \dots, n$ را برای هر اختلال رسم کنید.

λ_i	$\hat{\lambda}_i$	$ \lambda_i - \hat{\lambda}_i $	Cond(V)	c_i

(و) نتایج خود را بنویسید.

دادهای آزمون

۱. $A =$ ماتریس دوقطری و یلکینسون از مرتبه ۲۰

۲. $A =$ ترانواده ماتریس دوقطری و یلکینسون از مرتبه ۲۰

$$A = \begin{bmatrix} 12 & 11 & 10 & \dots & \dots & 2 & 1 \\ 11 & 11 & 10 & \dots & \dots & 2 & 1 \\ \cdot & 10 & 10 & \dots & \dots & 2 & 1 \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \cdot & \dots & \dots & \dots & \cdot & 1 & 1 \end{bmatrix} \quad .3$$

۸. حساسیتهای بردارهای ویژه ماتریسهای زیر را با محاسبه بردارهای ویژه ماتریسهای اصلی و اختلال یافته مطالعه کنید. برای این منظور از دستورهای $[V, D] = \text{eig}(A)$ ، $[V, D] = \text{eig}(\hat{A})$ از MATLAB استفاده کنید که در آن \hat{A} ماتریس حاصل از A با ایجاد اختلال در عنصر $(n, 1)$ توسط 10^{-10} ، 10^{-7} ، 10^{-5} می باشد.

$$A = \begin{bmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & 0/9999 & \cdot & \cdot \\ \cdot & \cdot & \cdot & 2 \\ \cdot & \cdot & \cdot & \cdot/0005 \end{bmatrix} \quad (\text{الف})$$

$$A = \text{diag}(1, 0/9999, 1, 0/9999, 1) \quad (\text{ب})$$

(ج) ماتریسهای تولید شده تصادفی از مرتبه ۱۰، ۱۵، ۲۰، ۲۵، ۳۰، ۵۰.

۹. یک برنامه MATLAB به نام **qritrb** برای پیاده‌سازی تکرار QR پایه‌ای با استفاده از **givqr** از فصل ۵ بنویسید:

$$[A] = \text{qritrb}(A, N) \quad ; \quad N \text{ تعداد تکرارها است.} \quad (\text{الف})$$

(ب) برنامه را برای پیاده‌سازی تکرار QR هستبرگی تغییر دهید. $[H] = \text{qritrh}(H, N)$; تعداد تکرارها است.

(ج) سرعت برنامه‌ها را با محاسبه تعداد اعمال ممیز شناور، و زمان سپری شده برای ماتریسهای تصادفی از مرتبه ۱۰، ۱۵، ۲۰، ۲۵، ۳۰ مقایسه کنید.

۱۰. (الف) یک برنامه MATLAB برای محاسبه یک گام از تکرار QR با انتقال مضاعف صریح بنویسید: $[A] = \text{qritrdse}(A, N)$; N تعداد تکرارهایی است که باید انجام شود.

(ب) یک برنامه MATLAB برای محاسبه یک گام از تکرار QR با انتقال مضاعف ضمنی بنویسید: $[A] = \text{qritrdsi}(A, N)$; N تعداد تکرارهایی است که باید انجام شود.

(ج) نتایج حاصل از (الف) و (ب) را مقایسه کنید و نتیجه بگیرید که آنها اساساً یکی هستند. برنامه‌ها را بر روی ماتریسهای تصادفی از مرتبه ۱۰، ۱۵، ۲۰، ۲۵، ۳۰، ۵۰ آزمایش کنید.

۱۱. یک برنامه MATLAB برای تقلیل دادن k سطر و k ستون آخر یک ماتریس هستبرگی بنویسید.

$$[hprime] = \text{deflat}(H, k)$$

این برنامه را با ماتریسهای تصادفی مرتبه ۲۰ با مقادیر مختلف k آزمایش کنید. توجه کنید که برای $k = 1$ ، $hprime$ از مرتبه $n - 1$ خواهد بود، و برای $k = 2$ ، $hprime$ از مرتبه $n - 2$ خواهد بود، و غیره.

۱۲. با استفاده از **qritrdsi** و **deflat** یک برنامه MATLAB برای تعیین شکل شور حقیقی یک ماتریس هستبرگی A بنویسید.

$$[h] = \text{rsf}(h, eps)$$

که در آن eps تولرانس است.

آزمون: یک ماتریس هسنبرگی 20×20 تصادفی تولید کنید و جدول زیر را با استفاده از rsf تشکیل دهید.

تکرار	h_{21}	h_{32}	h_{43}	h_{54}	$h_{20,19}$

۱۳. (الف) یک برنامه MATLAB به نام **polysymtri** بنویسید که چند جمله‌ای مشخصه p_i ، i امین کهاد اصلی پیشرو ($i = 1, 2, \dots, n$) یک ماتریس سه قطری متقارن کاهش ناپذیر T از مرتبه n را بر اساس فرمول بازگشتی بخش ۸-۱۱-۱ محاسبه کند.

$$[\text{valpoly}] = \text{polysymtri}(T, \text{lambda}, i)$$

(ب) با استفاده از **polysymtri** یک برنامه MATLAB به نام **signagree** بنویسید که تعداد مقادیر ویژه T بزرگ‌تر از یک عدد حقیقی مفروض μ را بر اساس قضیه ۸-۱۱-۷ پیدا کند.

$$[\text{number}] = \text{signagree}(T, \text{meu})$$

(ج) با استفاده از **polysymtri** و **signagree** الگوریتم تصنیف ۸-۱۱-۱ را پیاده‌سازی کنید.

$$[\text{lambda}] = \text{bisection}(T, m, n)$$

$n-m+1$ بار را به ازای $m = 1, 2, 3, \dots$ با استفاده از برنامه **bisection** محاسبه کنید و سپس نتایج حاصل را با نتایج به دست آمده از برنامه **eig(T)** مقایسه کنید.

داده‌های آزمون

$T =$ ماتریس سه قطری متقارن مطرح شده در مسأله کمانش در بخش ۸-۳-۲ با $n = 20$

۱۴. یک برنامه MATLAB به نام **lanczossym** بنویسید که الگوریتم لانزوس متقارن دوباره فرمول بندی شده بخش ۸-۱۱-۳ را پیاده سازی کند. با استفاده از **lanczossym** ۱۰ مقدار ویژه اول یک ماتریس متقارن تولید شده تصادفی از مرتبه ۱۰۰ را پیدا کنید. (برای تولید یک ماتریس متقارن B ، ابتدا ماتریس A را تولید کنید و سپس اختیار کنید $B = A + A^T$). مقادیر ویژه واقعی و مقادیر ویژه محاسبه شده را رسم کنید.

تذکر: برنامه های **power**، **invtr**، **qritrb**، **qritrh**، **qritrdsi** و غیره در **MATCOM** هستند، اما بهتر است که خودتان این برنامه ها را بنویسید.

۱۵. (الف) یک برنامه MATLAB برای پیاده سازی روش تقسیم و تسخیر بنویسید (الگوریتم ۸-۱۱-۳):

$$[D, V] = dc(T)$$

که در آن D یک ماتریس قطری حاوی مقادیر ویژه ماتریس متقارن T و V ماتریس بردارهای ویژه هستند. این برنامه را با ماتریس متقارن تولید شده تصادفی T از مرتبه ۱۰۰ اجرا کنید.

(ب) مقادیر ویژه و بردارهای ویژه T را با استفاده از دستور **eig** از **MATLAB** محاسبه کنید.

$$[V, D] = eig(T)$$

(ج) با محاسبه زمان اجرای هر کدام از این برنامه ها سرعت این برنامه ها را مقایسه کنید.

تمرینهای ۱۶ و ۱۷ در مورد کاربرد شکل شور حقیقی (RSF) هستند.

۱۶. جواب معادلات ماتریسی سیلوستر و لیاپانوف: مسأله حل معادله ماتریسی سیلوستر در نظریه کنترل را ملاحظه کنید:

$$AX - XB = C$$

که در آن A ، B و C مفروض هستند و X باید پیدا شود. فرض کنید A از مرتبه $m \times m$ و B از مرتبه $n \times n$ ($m \geq n$) باشند.

(معادله ماتریسی لیاپانوف $AX - XA^T = C$ یک حالت خاص است.) این مسأله

می تواند با تبدیل A به یک ماتریس بالا هسنبرگی و B به شکل شور حقیقی به صورت زیر حل شود.

گام ۱: مسأله ابتدا به مسأله معادل زیر کاهش می یابد

$$A'Y - YB' = C'$$

که در آن A' یک ماتریس بالا هسنبرگی است و B' به شکل شور حقیقی است.
 گام ۲: سپس Y در مسأله کاهش یافته با حل یک سری از دستگاههای خطی پیدا می شود.
 گام ۳: بالاخره جواب X مسأله اصلی از جواب Y مسأله کاهش یافته توسط ضربهای ماتریسی متعامد (یا یکانی) به دست می آید.

ابتدا جزئیات گامهای ۱ تا ۳ را روشن سازید و سپس یک برنامه MATLAB بنویسید. این برنامه را با یک ماتریس تولید شده تصادفی از مرتبه ۲۰ با استفاده از دستور **lyap** از MATLAB آزمایش کنید. (روش مبتنی بر مقاله های (Bartels and Stewart (1972) و Golub, Nash, and Van Loan (1979) می باشد).

۱۷. قطری سازی بلوکی: فرض کنید $Q^T A Q = T$ شکل شور حقیقی با T_{11}, \dots, T_{pp} به عنوان بلوکهای قطری با طیفهای مجزا باشند. یک الگوریتم با استفاده از معادله ماتریسی سیلوستر تدوین نمایید که Y را به قسمی پیدا کند که $Y^{-1} T Y = \text{diag}(T_{11}, \dots, T_{pp})$.
 یک برنامه MATLAB بنویسید و الگوریتم خودتان را با یک ماتریس از مرتبه ۱۰۰ آزمایش کنید.

پیشنهادها برای مطالعه بیشتر

اغلب کتابهای مربوط به ارتعاش در مورد مسائل مقدار ویژه مطرح در ارتعاش سازهها بحث می کنند. ولیکن اغلب مسائل مقدار ویژه در اینجا مسائل مقدار ویژه تعمیم یافته هستند، در حقیقت مسائل معین متقارن هستند.

برای مرجعها و کتابهای معروف در ارتعاش بخش ۹-۱۰ را ملاحظه کنید. بویژه کتابهای Inman (1994) و Thomson (1993) کتابهای خیلی مفید و مهمی در این زمینه هستند.

برای یادگیری بیشتر در مورد این که چگونه مسأله مقدار ویژه در سایر زمینه های مهندسی مطرح می شود کتابهای عددی در مهندسی نظیر Chapra and Canale (1988) و Peter O'Neil (1991) را ملاحظه کنید. کتابهای مهندسی دیگر (که برای فهرست کردن در اینجا بی شمار هستند) بویژه در زمینه های الکتریکی، مکانیکی، راه و ساختمان، و مهندسی شیمی، حاوی بحثهایی در مورد مسائل مقدار ویژه در مهندسی هستند. شکل شور حقیقی (RSF) و شکلهای هسنبرگی یک ماتریس وسیله های مهمی برای حل مؤثر عددی بسیاری از مسائل مهم کنترل هستند،
 نظیر:

(i) حل معادله ماتریسی لیاپانوف (Bartels and Stewart, (1972))

(ii) حل معادله ماتریسی سیلوستر (Golub, Nash, and Van Loan, (1979))

(iii) حل معادله ریکاتی جبری که در مراجع زیر ارائه شده است.

Laub (1979), Van Dooren (1981a), Byers (1986), Ammar and Mehrmann (1991),

Mehrmann (1988), and Bunse-Gerstner and Mehrmann (1986).

برای جزئیات بیشتر مراجع زیر را ملاحظه کنید.

The book *Computational Methods for Linear Control Systems* by Petkov Christov, and Konstantinov (1991)

The IEEE reprint by Patel et al (1994)

The book *Numerical Methods in Control Theory*, by B. N. Datta

The survey paper of Datta (1994)

برخی مرجعها در مورد این که چگونه مسائل مقدار ویژه (بویژه مسائل مقدار ویژه بزرگ و تنک) در سایر زمینه‌های علوم و مهندسی مطرح می‌شود نظیر دستگاه‌های توانی، فیزیک، شیمی، مکانیکهای ساختاری و اقیانوس‌شناسی در کتاب زیر ارائه شده‌اند.

Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Vol .1, by Jane Cullum and Ralph Willoughby (1985).

همچنین کتاب جدید (Saad (1992) را ملاحظه کنید.

برای برخی تعمیم‌های قضیه‌های قرص گرشگورین، مقاله جدید Brualdi and Mallendorf (1994) را ملاحظه کنید. این مقاله حاوی نتایجی است که یک ناحیه از صفحه مختلط را برای هر مقدار ویژه ارائه می‌دهد؛ برای یک توصیف کامل از قضیه‌های قرص گرشگورین و کاربردها کتاب زیر را ملاحظه کنید.

Matrix Analysis by Roger Horn and Charles Johnson (1985, Chapter 6)

همچنین کتاب (Varga (1962) و Johnston (1971) را ملاحظه کنید.

در کتاب زیر توصیف زیبایی از نظریه پایداری در دستگاه‌های پویا ارائه شده است.

Introduction to Dynamic Systems: Theory, Models, and Applications by David Luenberger (1979).

برای نتایجی در مورد کرانه‌های مقدار ویژه مقاله‌های (Kahan, Parlett, and Varah (1968b) و

Jiang (1982) و کتاب (Stewart and Sun (1990) را ملاحظه کنید.

برای محاسبه شکل متعارف جردن، مقاله‌های زیر را ملاحظه کنید.

Golub and Wilkinson (1976), Kagstrom and Ruhe (1980a and 1980b), Demmel (1983)

توصیف روشهای متداول برای محاسبه مقدار ویژه و بردار ویژه نظیر روش توانی، روش معکوس توانی، روش خارج قسمت ریلی، و روش تکرار QR را می‌توان در همه کتابهای جبر خطی عددی پیدا کرد:

Golub and Van Loan (1983 and 1989), Stewart(1973), Hager(1988), Watkins (1991), Wilkinson (1965)

البته کتاب Wilkinson (1965) معتبرترین کتاب در این زمینه است.

مقاله‌های Varah (1968a, 1970) و Peter and Wilkinson (1979) در زمینه تکرار معکوس مهم‌ترین هستند.

یک برنامه فرتون برای مرتب کردن مقادیر ویژه یک ماتریس بالا هسنبرگی حقیقی در مرجع Stewart(1976b) ارائه شده است. همچنین مرجعهای زیر را ملاحظه کنید.

Dongarra, Hammarling, and Wilkinson (1992)

Bai and Demmel (1993a)

یک کتاب مهم در زمینه مسأله مقدار ویژه متقارن کتاب زیر است.

Symmetric Eigenvalue Problem by Beresford Parlett (1980b)

برای اثباتی از همگرایی جامع تکرار QR متقارن با انتقالهای ویلکینسون مرجعهای Lawson and Honson (1974, 240-7) و Parlett (1980b) را ملاحظه کنید.

برای رابطه بین روش توانی و تکرارهای QR مقاله‌های Watkins (1982, 1993) را ملاحظه کنید. این دو مقاله همراه با مقاله Nanda (1985) مقداری دید نظری در زمینه همگرایی الگوریتم تکرار QR فراهم می‌سازند.

چندین مقاله مهم در مورد همگرایی و پایداری الگوریتم تکرار QR عبارتند از:

Parlett and Le (1993) و Parlett (1965, 1966, 1968)

اخیراً یک راه جدید در نظریه انتقال ضمنی در الگوریتم تکرار QR اعمال شده بر یک ماتریس هسنبرگی در مرجع Miminis and Paige (1991) پیشنهاد شده است. این راه با راه بحث شده در کتاب، یعنی راه فرانسویس متفاوت است و به نظر می‌رسد که می‌تواند بر ماتریسهای هسنبرگی بلوکی با هر شکلی اعمال شود. بعلاوه این راه با تعداد زیادی از الگوریتمها که با الگوریتم QR تفاوت دارند، اما از برخی جهات شبیه آن می‌باشند، تفاوت دارد. برای جزئیات بیشتر مرجع Miminis and Paige (1991) را

ملاحظه کنید.

برای محاسبه دستگاههای ویژه دقیق ماتریسهای غالب قطری مقیاس شده مرجع Barlow and Demmel (1990) را ملاحظه کنید. برای نتایج جدیدی از یک در میانی مرجع Hill and Parlett (1993) را ملاحظه کنید.

برای یک توصیف مفصل از روش تقسیم و تسخیر مرجع Dongarra and Sorensen (1987) و مقاله اصلی Cuppen (1981) را ملاحظه کنید. برای یک پیاده‌سازی کارا از روش تقسیم و تسخیر مرجع Gu and Eisenstat (1994a, 1994b) را ملاحظه کنید. برای روش تقسیم و تسخیر برای مسائل مقدار ویژه نامتقارن مرجع Adams and Arbenz (1994) را ملاحظه کنید. برای تحلیل خطای روشهای به هنگام برای مسائل مقدار ویژه متقارن مرجع Barlow (1993) را ملاحظه کنید. برای یک روش به نام هوموتوپی برای مسأله مقدار ویژه متقارن که برای پیاده‌سازی موازی مناسب می‌باشد، مرجع Zeng (1991) را ملاحظه کنید. همچنین مراجع Li and Zeng (1992) و Li, Zeng, and Cong (1992) را ملاحظه کنید. کتابهای Golub and VanLoan (1989) ، Parlett (1980) ، Watkins (1991) ، و Wilkinson (1965) حاوی توصیفهایی از روش ژاکوبی هستند. مقاله‌های زیادی در مورد روش ژاکوبی نوشته شده است، از قدیمی‌ترین مقاله‌های در دسترس مقاله Eberlein (1962) بوده و مقاله اخیر عبارت است از Demmel and Veselic (1992). کتاب Golub and Van Loan (1989) حاوی یک فهرست کامل از مرجعها است.

ما روشی را برای برآورد عددی وضعیت مقادیر ویژه و بردارهای ویژه مورد بحث قرار نداده‌ایم. مقاله‌های زیر را ملاحظه کنید.

Bai, Demmel, and Mckenney (1993), Boley (1990), Byers (1984), Van Loan (1987) برای روشهای تصفیه مقدار ویژه مراجع Dongarra, Moler and Wilkinson (1983) و Demmel (1987a) را ملاحظه کنید.

یک بحث اساسی در مورد روشهای لانزوس با و بدون دوباره متعامد سازی و کاربردهای آن برای حل دستگاههای خطی معین مثبت و محاسبات مقدار ویژه در کتاب Golub and Van Loan (1989, chapter 9) ارائه شده است. دو کتاب معتبر در این زمینه عبارتند از:

Lanczos Algorithms for Large Symmetric Eigenvalue Computations, volumes I and II, by J. K. Cullum and R. A. Willoughby (1985).

The Symmetric Eigenvalue Problem by B. N Parlett (1980b).

کتاب مهم و جدید دیگر در این زمینه کتاب زیر می‌باشد.

Numerical Methods for Large Eigenvalue Problems by Y. Saad (1992)

از سال ۱۹۷۰ که کار ابتکاری پیچ ارائه شد، مقاله‌های زیادی در این زمینه منتشر شده‌اند و این مطلب یک زمینه فعال تحقیق است. مراجع زیر را ملاحظه کنید.

Parlett and Scott (1979), Parlett, Simon, and Stringer (1982), Parlett (1989), Saad (1981), Ruche (1994), Simon (1984), Calvetti, Reichel, and Sorensen (1994).

رساله دکترای پیچ ((Chris Paige (1971)) و تعداد زیادی از مقاله‌های او پس از آن ((Paige (1976, 1980, etc)) همچنین خیلی باارزش و خواندنی هستند. این مقاله‌ها به عنوان «مقاله‌های بذر» برای تحقیق‌های جدید و جدی دیگر در این زمینه در نظر گرفته می‌شوند.

علاوه بر الگوریتم لانزوس روش تکرار همزمان روش مفید دیگری برای محاسبه تعداد کمی از مقادیر ویژه و بردارهای ویژه یک ماتریس تنک و بزرگ می‌باشد. مقاله‌های Jennings and Stewart (1975) و Stewart (1976c) را ملاحظه کنید.

مسأله مقدار ویژه تعمیم یافته

۱-۹ مقدمه

در این فصل مسأله مقدار ویژه زیر را که به مسأله مقدار ویژه تعمیم یافته معروف است، ملاحظه می‌کنیم.

بیان مسأله مقدار ویژه تعمیم یافته

با مفروض بودن ماتریسهای $n \times n$ بعدی A و B ، اسکالرهای λ و بردارهای مخالف صفر x را به قسمی پیدا کنید که

$$Ax = \lambda Bx$$

توجه کنید که مسأله مقدار ویژه استاندارد برای ماتریس A یک حالت خاص از این مسأله است (اختیار کنید $B = I$).

تعریف ۹-۱-۱ در مسأله $Ax = \lambda Bx$ ، λ یک مقدار ویژه نامیده می‌شود، و بردار x یک بردار ویژه وابسته به λ برای مسأله تعمیم یافته $Ax = \lambda Bx$ می‌باشد.

به آسانی ملاحظه می‌کنیم که λ یک ریشه معادله مشخصه زیر است

$$\det(A - \lambda B) = 0$$

تعریف ۹-۱-۲ ماتریس $(A - \lambda B)$ ، یک مداد ماتریسی نامیده می‌شود، و برای راحتی توسط (A, B) نمایش داده می‌شود. جفت (A, B) منظم نامیده می‌شود اگر $\det(A - \lambda B)$ متحد با صفر نباشد. در غیر این صورت منفرد نامیده می‌شود.

یک تذکر در مورد استفاده از کلمه مداد

استفاده نسبتاً عجیب از کلمه مداد از نورشناسی و هندسه می‌آید. یک مجموعه از شعاعهای (نوری) همگرا به یک نقطه اشاره بر نوک تیز یک مداد دارند، و با یک تعمیم طبیعی، این کلمه برای هر دسته از منحنیهای یک پارامتری، فضاها، و یا شیءهای ریاضی دیگر به کار برده می‌شود (Parlett (1980, 302).

مقادیر ویژه یک مداد منظم

اکنون برخی حقیقتهای مهم اما به آسانی قابل اثبات را در مورد مقادیر ویژه یک مداد منظم بیان می‌کنیم. فرض کنید (A, B) یک مداد منظم باشد. آنگاه حالت ۱. اگر B نامنفرد باشد، آنگاه همه مقادیر ویژه جفت (A, B) متناهی هستند و برابر مقادیر ویژه AB^{-1} یا $B^{-1}A$ می‌باشند.

اثبات: چون B نامنفرد است، از $Ax = \lambda Bx$ داریم $B^{-1}Ax = \lambda x$.

بنابراین λ یک مقدار ویژه $B^{-1}A$ است. همچنین AB^{-1} و $B^{-1}A$ متشابه هستند، زیرا، $AB^{-1} = BB^{-1}AB^{-1}$ ؛ بنابراین آنها مقادیر ویژه یکسان دارند. □

حالت ۲. اگر B منفرد باشد، آنگاه درجه $p(\lambda) = \det(A - \lambda B)$ کمتر از n است. فرض کنید r باشد. آنگاه r صفر $p(\lambda)$ مقادیر ویژه جفت (A, B) هستند. رسم این است که $(n - r)$ مقدار ویژه باقیمانده برابر ∞ قرار داده شوند.

مثال ۹-۱-۱

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & \\ 0 & 0 \end{bmatrix}$$

$\det(A - \lambda B) = 2 - \lambda$. مقادیر ویژه (A, B) برابر ۲ و ∞ هستند.

مثال ۹-۱-۲

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

هر دو مقدار ویژه (A, B) برابر ∞ هستند.

مقادیر ویژه یک مداد منفرد

اگر (A, B) یک مداد منفرد باشد، آنگاه چون $\det(A - \lambda B)$ متحد با صفر است، هر عدد λ می تواند یک مقدار ویژه (A, B) باشد. ما در اینجا این حالت خاص را بررسی نخواهیم کرد. بیشتر مسائل مقدار ویژه تعمیم یافته ای که در عمل مطرح می شوند منظم هستند.

بسیاری از کاربردهای مهندسی منجر به مسائل مقدار ویژه تعمیم یافته می شوند. اکثر مسائل مقدار ویژه ای که در ارتعاش مکانیکی مطرح می شوند، مسائل مقدار ویژه تعمیم یافته هستند. برای مثال مسائل ویژه برای ارتعاشهای سازه ها نظیر بناها و پلها هستند که همچنین مسائل مقدار ویژه تعمیم یافته معین متقارن برای ماتریسهای جرم و سختی $Kx = \lambda Mx$ نامیده می شوند. در اینجا M ماتریس جرم و K ماتریس سختی است. M و K معمولاً متقارن حقیقی هستند و بعلاوه M معین مثبت متقارن است. تعریف ۹-۱-۳ اگر A و B ماتریسهای متقارن حقیقی باشند، و بعلاوه اگر A معین مثبت باشد، آنگاه مسأله مقدار ویژه تعمیم یافته

$$Ax = \lambda Bx$$

مسأله مقدار ویژه تعمیم یافته معین متقارن نامیده می شود.

این فصل به مطالعه حل عددی مسأله مقدار ویژه تعمیم یافته اختصاص داده شده است و بر ویژه به مسأله معین متقارن توجه خاصی خواهیم نمود (بخشهای ۹-۵ و ۹-۶). کاربردهای مسأله مقدار ویژه معین متقارن را برای حل معادلات دیفرانسیل مرتبه دو در ارتعاش مکانیکی تا حد عمیق مطالعه خواهیم کرد (بخشهای ۹-۶ و ۹-۷). بعلاوه در بخش ۹-۲ در مورد تجزیه شور تعمیم یافته بحث خواهیم کرد، و در بخش ۹-۳ الگوریتم QR را مورد مطالعه قرار خواهیم داد. در بخش ۹-۴ تکرار معکوس را برای بردارهای ویژه مداد توصیف خواهیم کرد.

بیشتر مطالب زیر و مفاهیم قبلاً توسعه یافته برای یادگیری و فهم روان مطالب این فصل لازم

خواهند بود: روشهای هاوس هولدر و گیونز برای ایجاد صفر در یک بردار و الگوریتم تجزیه QR متناظر (بخشهای ۴-۵، ۱-۴-۵، ۲-۴-۵، ۵-۵، ۱-۵-۵)، الگوریتم چولسکی برای حل دستگاههای معین مثبت متقارن (بخش ۶-۴-۷)، و الگوریتم تکرار معکوس و الگوریتم تکرار خارج قسمت ریلی (بخش ۸-۵).

۲-۹ تجزیه شور تعمیم یافته

به عنوان تعمیمی از روش حل مسائل مقدار ویژه استاندارد، طبیعی است که به حل مسأله مقدار ویژه تعمیم یافته توسط کاهش همزمان ماتریسهای A و B به شکلهای مناسبی فکر کرد که مقادیر ویژه تعمیم یافته بتوانند به سادگی استخراج شوند. قضیه ۹-۳-۱ وجود چنین شکل مناسبی را نشان می دهد و قضیه ۹-۲-۱ استخراج مقادیر ویژه را توضیح می دهد به محض این که ماتریسهای A و B به چنین شکلهایی کاهش یافته باشند.

قضیه ۹-۲-۱

فرض کنید A و B ماتریسهای $n \times n$ و جفت (A, B) منظم باشند. فرض کنید U_1 و U_2 ماتریسهای یکانی باشند به قسمی که

$$U_2 A U_2^T = T_1 = \begin{bmatrix} t_{11} & & & \\ & \ddots & & * \\ & & \ddots & \\ & & & t_{nn} \end{bmatrix}$$

$$U_1 B U_1^T = T_2 = \begin{bmatrix} t'_{11} & & & \\ & \ddots & & * \\ & & \ddots & \\ & & & t'_{nn} \end{bmatrix}$$

آنگاه مقادیر ویژه تعمیم یافته λ_i ، $i = 1, \dots, n$ ، مداد منظم (A, B) به صورت زیر ارائه می شوند

$$\lambda_i = t_{ii} / t'_{ii}, \quad \text{اگر } t'_{ii} \neq 0$$

$$\lambda_i = \infty, \quad \text{اگر } t'_{ii} = 0 \text{ و } t_{ii} \neq 0$$

اثبات: از $Ax = \lambda Bx$ داریم

$$U_1 A U_1^* x = \lambda U_1 B U_1^* x$$

سپس با تعریف $y = U_1^* x$ خواهیم داشت

$$U_1 A U_1 y = \lambda U_1 B U_1 y$$

یعنی $T_1 y = \lambda T_1 y$. بنابراین λ یک مقدار ویژه (A, B) است اگر و فقط اگر یک مقدار ویژه جفت (T_1, T_2) باشد. چون T_1 و T_2 هر دو مثلثی هستند، $\det(T_1 - \lambda T_2) = \prod_{i=1}^n (t_{ii} - \lambda t'_{ii})$ ، و صفر می شود هنگامی که $\lambda_i = t_{ii}/t'_{ii}$. همچنین هنگامی که $t_{ii} \neq 0$ اما $t'_{ii} = 0$ مقدار ویژه λ_i می تواند برابر ∞ قرار داده شود.

□

تبصره ها:

۱. توجه کنید که $x = U_1 y$ یک بردار ویژه $(A - \lambda B)$ وابسته به λ می باشد.
۲. اگر A و B ماتریسهای حقیقی باشند، آنگاه U_1 و U_2 می توانند به قسمی اختیار شوند که $U_1 A U_1 = T_1$ به شکل شور حقیقی باشد.

۹-۳ الگوریتم QZ

یک روش مشابه با روش تکرار QR و معروف به الگوریتم تکرار QZ برای محاسبه مقادیر ویژه توصیف خواهیم کرد. این الگوریتم توسط مولر و استوارت^۱ تدوین شده است. فرض کنید که B نامنفرد باشد. آنگاه ایده اصلی به کار بردن روش تکرار QR در مورد ماتریس $C = B^{-1}A$ (یا $A^{-1}B$) بدون تشکیل صریح ماتریس C می باشد، زیرا اگر B نزدیک به منفرد باشد، آنگاه تشکیل B^{-1} مطلوب نیست. در این حالت عناصر C خیلی بزرگتر از عناصر A و B خواهند بود و مقادیر ویژه C به طور نادقیقی محاسبه خواهند شد.

اگر محاسبه صریح $A^{-1}B$ یا $B^{-1}A$ مورد نظر نباشند، آنگاه البته راه حل بهتر بعدی این است که A و B به شکلهای کاهش یافته ای تبدیل شوند و سپس مقادیر ویژه تعمیم یافته λ_i از روی این شکلهای کاهش یافته استخراج شوند. کاهش همزمان A و B به شکلهای مثلثی به وسیله هم ارزی توسط قضیه زیر تضمین می شود.

قضیه ۹-۳-۱ تجزیه شور حقیقی تعمیم یافته. با مفروض بودن دو ماتریس حقیقی $n \times n$ بعدی

A و B ، ماتریسهای متعامد Q و Z وجود دارند به قسمی که $Q^T A Z = A'$ یک ماتریس بالا شور حقیقی بوده و $Q^T B Z = B'$ یک ماتریس بالا مثلثی باشد.

$$\begin{aligned} A' &= Q^T A Z: & A' &= \text{یک ماتریس بالا شور حقیقی} \\ B' &= Q^T B Z: & B' &= \text{یک ماتریس بالا مثلثی} \end{aligned}$$

گفته می شود که جفت (A', B') به شکل شور حقیقی تعمیم یافته هستند. جفت (A', B') همان مقادیر ویژه جفت (A, B) را دارا می باشد. اکنون یک اثبات ساختنی از این قضیه ارائه خواهیم داد. کاهش به شکل شور حقیقی تعمیم یافته در دو مرحله انجام می شود.

مرحله ۱. A و B به ترتیب به یک ماتریس بالا هسنبرگی و یک ماتریس بالا مثلثی، توسط هم ارزی متعامد همزمان، کاهش می یابند:

$$\begin{aligned} A &\equiv Q^T A Z & \text{یک ماتریس بالا هسنبرگی} \\ B &\equiv Q^T B Z & \text{یک ماتریس بالا مثلثی} \end{aligned}$$

مرحله ۲. جفت هسنبرگی - مثلثی (A, B) باز هم بابه کاربردن تکرار QR ضمنی در مورد AB^{-1} به شکل شور حقیقی تعمیم یافته کاهش می یابد.

این فرایند معروف به الگوریتم QZ است.

اکنون به طور مختصر این دو مرحله را شرح می دهیم. برای جزئیات بیشتر مرجع Moler and Stewart (1973) را ملاحظه کنید. در این فصل برای راحتی نمادگذاری از این به بعد یک دوران گیونز $J(i, j, \theta)$ را به J_{ij} نشان می دهیم.

۹-۳-۱ کاهش به شکل هسنبرگی - مثلثی

فرض کنید A و B دو ماتریس $n \times n$ باشند. آنگاه

۱. یک ماتریس متعامد U به قسمی پیدا می کنیم که

$$B \equiv U^T B : \quad \text{بالا مثلثی}$$

ماتریس زیر را تشکیل می دهیم

$$A \equiv U^T A$$

(در حالت کلی A پر خواهد بود)

۲. A را به شکل هسنبرگی کاهش می‌دهیم در حالی که ساختار مثلثی B را حفظ می‌کنیم. این گام به صورت زیر انجام می‌شود. در آغاز داریم

$$A \equiv U^T A = \begin{bmatrix} * & * & \dots & * \\ * & * & \dots & * \\ \vdots & \vdots & & \vdots \\ * & * & \dots & * \\ * & * & \dots & * \end{bmatrix}$$

$$B \equiv U^T B = \begin{bmatrix} * & * & \dots & \dots & * \\ \cdot & * & \dots & \dots & * \\ \cdot & \cdot & \ddots & \dots & * \\ \vdots & \vdots & & \ddots & \vdots \\ \cdot & \cdot & \cdot & \dots & * \end{bmatrix}$$

ابتدا عنصر $(n, 1)$ ماتریس A را با اعمال یک دوران $Q_{n-1, n}$ در صفحه $(n-1, n)$ به صفر تبدیل می‌کنیم.

$$A \equiv Q_{n-1, n} A = \begin{bmatrix} * & * & \dots & * \\ * & * & \dots & * \\ \vdots & \vdots & & \vdots \\ * & * & \dots & * \\ \cdot & * & \dots & * \end{bmatrix}$$

هنگامی که این تبدیل از طرف چپ بر B اعمال می‌شود یک عنصر در وضعیت $(n-1, n)$ را پر خواهد کرد:

$$B \equiv Q_{n-1, n} B = \begin{bmatrix} * & * & \dots & \dots & * \\ \cdot & * & \dots & \dots & * \\ \cdot & \cdot & * & \dots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdot & * & * \end{bmatrix}$$

اکنون دوران $Z_{n-1, n}$ را از طرف راست بر B اعمال می‌کنیم تا عنصر $(n, n-1)$ ماتریس B به صفر تبدیل شود. خوشبختانه وقتی این دوران از طرف راست بر A اعمال می‌شود صفر قبلاً تولید شده را از بین نمی‌برد. به طور شمایی داریم

$$B \equiv BZ_{n-1, n} = \begin{bmatrix} * & * & * & \dots & * \\ \cdot & * & * & \dots & * \\ \cdot & \cdot & * & \dots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \cdot & \cdot & \dots & \cdot & * \end{bmatrix}$$

$$A \equiv AZ_{n-1, n} = \begin{bmatrix} * & * & * & \dots & * \\ * & * & * & \dots & * \\ * & * & * & \dots & * \\ \vdots & \vdots & \vdots & & \vdots \\ * & * & * & \dots & * \\ \cdot & * & * & \dots & * \end{bmatrix}$$

عناصر $(1, 1), (1, 2), \dots, (1, n-1), (2, 1), \dots, (n-1, 1)$ ماتریس A متوالیا^۳ برابر صفر می‌گردند. برای صفر کردن هر عنصر سه دوران باید به صورت زیر متوالیا اعمال شوند. ابتدا یک دوران مناسب از طرف چپ بر A اعمال می‌شود تا عنصر را برابر صفر سازد. سپس همان دوران از طرف چپ بر B اعمال می‌شود. بالاخره دوران مناسب دیگری از طرف راست بر B اعمال می‌کنیم تا عنصر پر شده ناخواسته را در B مجدداً برابر صفر سازد. در انتهای اولین گام ماتریس A در ستون اولش هسنبرگی است، در حالی که B بالا مثلثی باقی می‌ماند:

$$A = \begin{bmatrix} * & * & \dots & * \\ * & * & \dots & * \\ \cdot & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & * & \dots & * \end{bmatrix}, \quad B = \begin{bmatrix} * & * & \dots & \dots & * \\ \cdot & * & \dots & \dots & * \\ \cdot & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & * \\ \cdot & \dots & \cdot & \cdot & * \end{bmatrix}$$

اکنون صفرها در ستون دوم A در جاهای مناسب تولید می‌شوند، در حالی که ساختار مثلثی B به طریق مشابه حفظ می‌شود.

فرآیند ادامه می‌یابد تا ماتریس A به یک ماتریس بالا هسنبرگی تبدیل شود و B بالا مثلثی

باشد.

مثال ۹-۳-۱

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 4 \\ 1 & 3 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$$

گام ۱: تشکیل Q_{23} برای صفر کردن a_{31} :

$$Q_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cdot/\sqrt{0.71} & \cdot/\sqrt{0.71} \\ 0 & -\cdot/\sqrt{0.71} & \cdot/\sqrt{0.71} \end{bmatrix}$$

$$A \equiv A^{(1)} = Q_{23}A = \begin{bmatrix} 1 & 2 & 3 \\ 1/\sqrt{4142} & 2/\sqrt{2426} & 2/\sqrt{9497} \\ 0 & 0 & -\cdot/\sqrt{0.71} \end{bmatrix}$$

گام ۲: به هنگام در آوردن B :

$$B \equiv B^{(1)} = Q_{23}B = \begin{bmatrix} 1 & 1 & 1 \\ 0 & \cdot/\sqrt{0.71} & 2/\sqrt{8284} \\ 0 & -\cdot/\sqrt{0.71} & 0 \end{bmatrix}$$

گام ۳: تشکیل Z_{23} برای صفر کردن b_{32} :

$$Z_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$B \equiv B^{(1)}Z_{23} = Q_{23}BZ_{23} = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 2/\sqrt{8284} & -\cdot/\sqrt{0.71} \\ 0 & 0 & \cdot/\sqrt{0.71} \end{bmatrix}$$

گام ۴: به هنگام در آوردن A :

$$A \equiv A^{(1)}Z_{23} = Q_{23}AZ_{23} = \begin{bmatrix} 1 & 3 & -2 \\ 1/4142 & 4/9497 & -4/2426 \\ \cdot & -\cdot/7071 & \cdot \end{bmatrix}$$

A بالا هسنبرگی است و B به شکل بالا مثلثی است.

۹-۳-۲ کاهش به شکل شور تعمیم یافته

در ابتدای این فرآیند، A و B را که از مرحله ۱ به دست آمده‌اند به ترتیب به شکل یک ماتریس بالا هسنبرگی و یک ماتریس بالا مثلثی در اختیار داریم. بدون از دست دادن کلیت مسأله می‌توانیم فرض کنیم که ماتریس A یک ماتریس بالا هسنبرگی کاهش‌ناپذیر باشد. اکنون ایده اساسی این است که یک گام QR ضمنی را بر AB^{-1} اعمال کنیم بدون آن که هرگز به طور صریح این ماتریس را به طور کامل تشکیل دهیم.

بنابراین یک گام QZ ، مشابه با یک گام QR ضمنی (الگوریتم ۸-۹-۱ را ملاحظه کنید) به صورت زیر خواهد بود:

- اولین ستون ماتریس $N = (C - \alpha_1 I)(C - \alpha_2 I)$ را تشکیل می‌دهیم، که در آن $C = AB^{-1}$ و α_1 و α_2 انتقالهای انتخاب شده مناسب هستند.
- یک ماتریس هاوس هولدر Q_1 به قسمی پیدا می‌کنیم که $Q_1 N e_1$ مضربی از e_1 باشد.
- $Q_1 A$ و $Q_1 B$ را تشکیل می‌دهیم.
- به طور همزمان $Q_1 A$ را به یک ماتریس بالا هسنبرگی A_1 و $Q_1 B$ را به یک ماتریس بالا مثلثی B_1 تبدیل می‌کنیم:

$$A_1 \equiv Q_1'(Q_1 A)Z: \quad \text{یک بالا هسنبرگی}$$

$$B_1 \equiv Q_1'(Q_1 B)Z: \quad \text{یک بالا مثلثی}$$

با استفاده از قضیه Q ضمنی (قضیه ۵-۵-۳) می‌توانیم نشان دهیم (تمرین ۵) که ماتریس $A_1 B_1^{-1}$ اساساً همان ماتریسی می‌باشد که ما با اعمال مستقیم یک گام QR ضمنی بر AB^{-1} به دست می‌آوریم.

سپس کاربرد پی در پی تعداد کمی از گامهای QZ یک ماتریس شبه مثلثی $R = Q^T A Z$ و یک

ماتریس بالا مثلثی $T = Q^T B Z$ نتیجه خواهد داد که با استفاده از آنها می توان به سادگی مقادیر ویژه را حساب کرد. چهار گام قبلی یک تکرار واحد از الگوریتم QZ را تشکیل می دهند. اکنون نشان می دهیم که چگونه این گامها را پیاده سازی کنیم.

محاسبه اولین ستون $(C - \alpha_1 I)(C - \alpha_2 I)$

تنگنای واقعی در پیاده سازی همه الگوریتم در محاسبه ستون اول $(C - \alpha_1 I)(C - \alpha_2 I)$ بدون تشکیل صریح $C = AB^{-1}$ می باشد. خوشبختانه این کار می تواند انجام شود. ابتدا توجه کنید که به واسطه بالا هسنبرگی بودن A و بالا مثلثی بودن B این ستون اول شامل حداکثر سه عنصر مخالف صفر در سه مکان اولش می باشد:

$$n_1 = Ne_1 = (C - \alpha_1 I)(C - \alpha_2 I)e_1 = \begin{bmatrix} x \\ y \\ z \\ \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

برای محاسبه x ، y ، z همه آنچه لازم است بشناسیم، دو ستون اول C می باشد که می توانند فقط با معکوس کردن زیر ماتریس اصلی پیشرو 2×2 ماتریس B^{-1} به دست آیند؛ همه B^{-1} لازم نیست محاسبه شود (تمرین ۳). بنابراین اگر c_1 و c_2 دو ستون اول $C = AB^{-1}$ باشند، آنگاه

$$(c_1, c_2) = (a_1, a_2) \begin{bmatrix} b_{11} & b_{12} \\ \cdot & b_{22} \end{bmatrix}^{-1}$$

که در آن a_1 و a_2 دو ستون اول A می باشند

محاسبه A_1 و B_1

چون n_1 ستون اول $N = (C - \alpha_1 I)(C - \alpha_2 I)$ حداکثر سه عنصر مخالف صفر دارد، ماتریس هاوس هولدر Q_1 که n_1 را به مضربی از e_1 تبدیل می کند دارای شکل زیر است

$$Q_1 = \begin{bmatrix} \hat{Q}_1 & \cdot \\ \cdot & I_{n-3} \end{bmatrix}$$

که در آن \hat{Q}_1 یک ماتریس هاوس هولدر 3×3 است. سپس ماتریسهای $Q_1 A$ و $Q_1 B$ دارای شکلهایی به صورت زیر، در اینجا نشان داده شده برای $n = 6$ ، خواهند بود، نماد \otimes یک پر شدن احتمالی را نشان می دهد.

$$A \equiv Q_1 A = \begin{pmatrix} * & * & * & * & * & * \\ * & * & * & * & * & * \\ \otimes & * & * & * & * & * \\ . & . & * & * & * & * \\ . & . & . & * & * & * \\ . & . & . & . & * & * \end{pmatrix}$$

$$B \equiv Q_1 B = \begin{pmatrix} * & * & * & * & * & * \\ \otimes & * & * & * & * & * \\ \otimes & \otimes & * & * & * & * \\ . & . & . & * & * & * \\ . & . & . & . & * & * \\ . & . & . & . & . & * \end{pmatrix}$$

همان طور که ملاحظه می‌کنید هم شکل هسنبرگی A و هم شکل مثلثی B از بین رفته‌اند، و اکنون در هر دوی آنها عناصر مخالف صفر ناخواسته وجود دارند، مکان $(۳, ۱)$ ماتریس $Q_1 A$ و سه عنصر مخالف صفر ناخواسته در مکانهای $(۲, ۱)$ ، $(۳, ۱)$ ، $(۳, ۲)$ ماتریس $Q_1 B$ یافت می‌شوند. کار ما اکنون شکار زیرکانه عناصر مخالف صفر ناخواسته و صفر کردن آنها با استفاده از تبدیلهای متعامد خواهد بود. برای انجام این عمل ابتدا یک ماتریس هوس هولدر Z_1 برای ایجاد صفر در مکانهای $(۳, ۱)$ و $(۳, ۲)$ ماتریس B می‌سازیم. سپس یک گیونز Z_2 برای ایجاد صفر در مکان $(۲, ۱)$ ماتریس B می‌سازیم. پس از آن A را به هنگام درمی‌آوریم

$$B \equiv Q_1 B Z_1 Z_2 = \begin{pmatrix} * & * & * & * & * & * \\ . & * & * & * & * & * \\ . & . & * & * & * & * \\ . & . & . & * & * & * \\ . & . & . & . & * & * \\ . & . & . & . & . & * \end{pmatrix}$$

$$A \equiv Q_1 A Z_1 Z_2 = \begin{pmatrix} * & * & * & * & * & * \\ * & * & * & * & * & * \\ \otimes & * & * & * & * & * \\ \otimes & \otimes & * & * & * & * \\ . & . & . & * & * & * \\ . & . & . & . & * & * \end{pmatrix}$$

توجه کنید که اکنون صف‌های ناخواسته در مکانهای (۱, ۳)، (۱, ۴)، و (۲, ۴) ماتریس A داریم. اکنون یک ماتریس هاوس هولدر Q_2 برای ایجاد صفر در مکانهای (۱, ۳) و (۱, ۴) ماتریس A می‌سازیم. سپس B را به هنگام درمی‌آوریم. اکنون داریم

$$A \equiv Q_2 Q_1 A Z_1 Z_2 = \begin{bmatrix} * & * & * & * & * & * \\ * & * & * & * & * & * \\ \cdot & * & * & * & * & * \\ \cdot & * & * & * & * & * \\ \cdot & \cdot & \cdot & * & * & * \\ \cdot & \cdot & \cdot & \cdot & * & * \end{bmatrix}$$

$$B \equiv Q_2 Q_1 B Z_1 Z_2 = \begin{bmatrix} * & * & * & * & * & * \\ \cdot & * & * & * & * & * \\ \cdot & * & * & * & * & * \\ \cdot & * & * & * & * & * \\ \cdot & \cdot & \cdot & \cdot & * & * \\ \cdot & \cdot & \cdot & \cdot & \cdot & * \end{bmatrix}$$

زیر ماتریسهای A و B به صورتی که هم اکنون نشان داده شده، دارای همان ساختار ماتریسهای اصلی، یعنی ماتریسهای $Q_1 A$ و $Q_1 B$ هستند. اکنون این فرایند کاهش دادن با این ماتریسها ادامه می‌یابد تا ساختار هسنبِرگی مثلثی (A, B) به صورت اول درآید. با در نظر داشتن این بحث اکنون یک تکرار از الگوریتم QZ را خلاصه می‌کنیم.

الگوریتم ۹-۳-۱ یک گام از الگوریتم QZ . با مفروض بودن A ، یک ماتریس بالا هسنبِرگی کاهش‌ناپذیر، و B یک ماتریس بالا مثلثی، گامهای زیر یک گام تکراری الگوریتم QZ را تشکیل می‌دهند؛ یعنی این گامها ماتریسهای متعامد Q و Z را به قسمی می‌سازند که $Q^T A Z$ یک ماتریس بالا هسنبِرگی و $Q^T B Z$ یک ماتریس بالا مثلثی باشند.

گام ۱: انتقالهای α_1 و α_2 را انتخاب کنید.

گام ۲: اولین ستون ماتریس $N = (C - \alpha_1 I)(C - \alpha_2 I)$ را که در آن $C = AB^{-1}$ ، بدون تشکیل

صریح B^{-1} حساب کنید. فرض کنید (c_1, c_2) دو ستون اول C باشند. آنگاه

$$(c_1, c_2) = (a_1, a_2) \begin{bmatrix} b_{11} & b_{12} \\ \cdot & b_{22} \end{bmatrix}^{-1}$$

سه عنصر مخالف صفر ستون اول N به صورت زیر ارائه می‌شوند

$$x = (c_{11} - \alpha_1)(c_{11} - \alpha_2) + c_{12}c_{21}$$

$$y = c_{21}(c_{11} - \alpha_2) + c_{21}(c_{22} - \alpha_1)$$

$$z = c_{21}c_{22}$$

$$\text{اولین ستون } N, \text{ بردار } n_1 = \begin{bmatrix} x \\ y \\ z \\ \vdots \\ \cdot \end{bmatrix} \text{ است.}$$

$$Q_1 n_1 = \begin{bmatrix} * \\ \cdot \\ \vdots \\ \cdot \end{bmatrix} \text{ یک ماتریس هاوس هولدر } Q_1 \text{ به قسمی پیدا کنید که}$$

گام ۴: $Q_1 A$ و $Q_1 B$ را تشکیل دهید.

گام ۵: $Q_1 B$ و $Q_1 A$ را به ترتیب به یک ماتریس بالا هسنبرگی A_1 و یک ماتریس بالا مثلثی B_1

تبدیل کنید. از هم ارزی متعامد به طریقی که قبلاً نشان داده شده استفاده کنید و از مزایای

ساختارهای خاص ماتریسها در ایجاد ماتریسهای متعامد Q_2 تا Q_{n-1} و Z_1 تا Z_{n-1}

بهره‌برداری کنید.

به دست آوردن ماتریسهای تبدیل

ماتریسهای تبدیل Q و Z به صورت زیر به دست می‌آیند. ماتریس Q :

$$Q = Q_1 Q_2 \dots Q_{n-1}$$

ماتریس Z :

$$Z = Z_1 Z_2 \dots Z_{n-1}$$

توجه کنید که Q همان سطر اول Q_1 را دارا می‌باشد.

شمارش اعمال ممیز شناور. یک گام QZ تقریباً به $13n^2$ عمل ممیز شناور نیاز دارد. اگر Q و

Z موردنیاز باشند آنگاه به ترتیب $5n^2$ و $8n^2$ عمل ممیز شناور (flops) دیگر لازم

می‌باشد.

انتخاب انتقالها

انتقالهای مضاعف α_1 و α_2 در یک گام QZ می توانند برابر مقادیر ویژه زیر ماتریس 2×2 پایینی ماتریس $C = AB^{-1}$ اختیار شوند. زیر ماتریس پایینی 2×2 ماتریس C می تواند بدون تشکیل صریح B^{-1} محاسبه شود (تمرین ۴).

الگوریتم ۹-۳-۲ الگوریتم QZ . ماتریسهای $n \times n$ حقیقی A و B مفروضند.

گام ۱: (A, B) را با یک هم ارزی متعامد به یک جفت هسنبرگی مثلثی تبدیل کنید:

$$A \equiv Q^T A Z: \quad \text{یک ماتریس بالا هسنبرگی}$$

$$B \equiv Q^T B Z: \quad \text{یک ماتریس بالا مثلثی}$$

گام ۲: الگوریتم ۹-۳-۱ را برای تولید $\{A_k\}$ و $\{B_k\}$ تکرار کنید، انتقالها را برای هر تکرار به صورت توصیف شده در فوق انتخاب کنید.

گام ۳: همگرایی دنباله های $\{A_k\}$ و $\{B_k\}$ را امتحان کنید.

$$\{A_k\} \rightarrow R \quad \text{شبه مثلثی}$$

$$\{B_k\} \rightarrow T \quad \text{مثلثی}$$

شمارش اعمال ممیز شناور. پیاده سازی گام ۱ تا ۳ تقریباً به $15n^3$ عمل ممیز شناور (flops) نیاز دارد. تشکیل Q و Z ، اگر مورد نیاز باشند، به ترتیب به $8n^3$ و $10n^3$ عمل ممیز شناور دیگر نیاز دارند. (تجربه نشان می دهد که با استراتژی انتقال فرانسیس تقریباً ۲ گام QZ برای هر مقدار ویژه لازم است).

خاصیتهای گرد کردن. الگوریتم تکرار QZ به پایداری الگوریتم تکرار QR است. می توان نشان داد که \hat{R} و \hat{S} محاسبه شده در روابط زیر صدق می کنند

$$Q_0^T (A + E) Z_0 = \hat{R}$$

$$Q_0^T (B + F) Z_0 = \hat{S}$$

در اینجا Q_0 و Z_0 متعامد هستند و

$$\|E\| \cong \mu \|A\|,$$

$$\|F\| \cong \mu \|B\|,$$

μ دقت ماشین می باشد.

۴-۹ تکرار معکوس برای بردارهای ویژه مداد (A, B)

به محض این که یک مقدار ویژه تقریبی λ محاسبه شود، بردار ویژه متناظر v می تواند مانند قبل با استفاده از تکرار معکوس محاسبه شود.

الگوریتم ۱-۴-۹

Step 1: Choose an initial eigenvector v_0 .

Step 2: For $k = 1, 2, \dots$ do until convergence

$$\text{Solve } (A - \lambda B)\hat{v}_k = Bv_{k-1}$$

$$v_k = \hat{v}_k / \|\hat{v}_k\|_2$$

تذکر MATCOM: الگوریتم ۱-۴-۹ در برنامه INVITRGN در MATCOM پیاده سازی شده است.

تذکری در مورد حل $(A - \lambda B)\hat{v}_k = Bv_{k-1}$

در حل $(A - \lambda B)\hat{v}_k = Bv_{k-1}$ صرفه جویی قابل ملاحظه ای می تواند با بهره گیری از ساختار هسنبرگی - مثلثی انجام پذیرد که جفت (A, B) در طی الگوریتم QZ به آن کاهش می یابد. توجه کنید که برای یک λ مفروض ماتریس $(A - \lambda B)$ یک ماتریس هسنبرگی است. بنابراین در هر تکرار فقط لازم است که یک دستگاه هسنبرگی حل شود.

مثال ۱-۴-۹

$$A = \begin{bmatrix} 3 & -1/5 & 0 \\ -1/5 & 3 & -1/5 \\ 0 & -1/5 & 1/5 \end{bmatrix}$$

$$B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

$$\begin{aligned} \lambda_1 &= (A - \lambda B) \text{ یافته ماتریس} \\ &= 1/9508 \end{aligned}$$

$$v_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$k = 1$: دستگاه زیر را نسبت به v_1 حل می‌کنیم:

$$(A - \lambda_1 B)\hat{v}_1 = Bv_1.$$

$$\hat{v}_1 = 10^4 \begin{bmatrix} 1/7.27 \\ -1/0.236 \\ 0/2435 \end{bmatrix}$$

$$v_1 = \hat{v}_1 / \|\hat{v}_1\| = \begin{bmatrix} 0/85.7 \\ -0/5114 \\ 0/1217 \end{bmatrix}$$

۹-۵ مسأله مقدار ویژه تعمیم یافته معین متقارن

بسیاری از مسائل کاربردی منجر به مسائل مقدار ویژه تعمیم یافته معین متقارن به شکل زیر می‌شوند

$$Ax = \lambda Bx$$

و

$$ABx = \lambda x$$

که در آنها A و B متقارن هستند، و حداقل یکی از آنها معین مثبت است. فرض خواهیم کرد که B معین مثبت باشد و فقط مسأله $Ax = \lambda Bx$ را ملاحظه می‌کنیم. هنگامی که B معین مثبت است، می‌توانیم

$$ABx = \lambda x \text{ را به شکل}$$

$$(A - \lambda B^{-1})(Bx) = 0.$$

بنویسیم. چون B^{-1} نیز معین مثبت است، مسأله $ABx = \lambda x$ به همان شکل مسأله $Ax = \lambda Bx$ می‌باشد.

مسأله معین متقارن $Ax = \lambda Bx$ به طور معمول در تحلیل ارتعاش سازه‌ها حل می‌شود.

فرکانسها، مدها، شکلهای مد

مقادیر ویژه به فرکانسهای طبیعی وابسته هستند، و اندازه و علامت هر عنصر یک بردار ویژه شکل ارتعاش را در هر لحظه از زمان تعیین می‌کند. بنابراین بردارهای ویژه شکلهای مد، یا به طور ساده مد نامیده می‌شوند.

«زبان مدها، شکلهای مد، فرکانسهای طبیعی اساس بحث پدیده ارتعاش دستگاههای پیچیده را تشکیل می‌دهند. یک صنعت کامل درباره مفهوم مدها شکل گرفته است» (Inman 1994, 196).

با یک خاصیت مهم (اما نه تعجب آور) مداد معین متقارن شروع می‌کنیم.

قضیه ۹-۵-۱ مداد معین متقارن $(A - \lambda B)$ دارای مقادیر ویژه حقیقی و بردارهای ویژه مستقل خطی است.

اثبات: چون B معین مثبت متقارن است، تجزیه چولسکی را می‌پذیرد.

$$B = LL^T$$

بنابراین از

$$Ax = \lambda Bx$$

داریم:

$$Ax = \lambda LL^T x$$

که نتیجه می‌دهد

$$L^{-1}A(L^T)^{-1}L^T x = \lambda L^T x$$

یا

$$Cy = \lambda y,$$

که در آن $y = L^T x$. ماتریس $C = L^{-1}A(L^T)^{-1}$ متقارن است؛ بنابراین λ حقیقی است. ادعای مربوط به بردارهای ویژه واضح است، زیرا یک ماتریس متقارن یک مجموعه از n بردار ویژه مستقل دارد (قضیه ۸-۲-۸)

یک بازه دربردارنده مقادیر ویژه یک مداد معین متقارن مقادیر ویژه یک مداد معین متقارن $A - \lambda B$ در بازه $[-\|B^{-1}A\|, \|B^{-1}A\|]$ قرار دارد (تمرین ۲۶ را ملاحظه کنید).

متعامد بودن بردارهای ویژه

اگر بردارهای ویژه y_1, y_2, \dots, y_n ماتریس C یکا متعامد انتخاب شوند، آنگاه به سادگی ملاحظه می‌شود که بردارهای ویژه $x_i, B -$ یکا متعامد خواهند بود؛ یعنی

$$x_i^T B x_j = \delta_{ij}, \quad i = 1, \dots, n$$

$$x_i^T B x_j = 0, \quad i \neq j$$

بعلاوه اگر x_i یک بردار ویژه نرمال شده باشد، آنگاه $x_i^T A x_i = c_{ii}$ ، $i = 1, \dots, n$ و

$$x_i^T A x_j = 0, \quad i \neq j$$

برای اثبات $x_i^T B x_j = 0$ توجه می‌کنیم که

$$x_i^T B x_j = x_i^T L L^T x_j = (L^T x_i)^T L^T x_j = y_i^T y_j = 0$$

(چون y_1, y_2, \dots, y_n متعامد هستند). روابط دیگر می‌توانند به سادگی ثابت شوند.

الگوریتمهای محاسباتی

۱-۵-۹ روش QZ برای مداد معین متقارن

الگوریتم QZ توصیف شده در بخش قبل برای مداد منظم $A - \lambda B$ می‌تواند البته برای یک مداد معین متقارن نیز به کار رود. ولیکن مشکل این است که هم متقارن و هم معین مثبت بودن مسأله به طور کلی از بین خواهند رفت. اکنون الگوریتم خاصی را برای یک مداد معین متقارن بیان می‌کنیم.

۲-۵-۹ الگوریتم QR چولسکی

اثبات قضیه ۱-۵-۹ فوراً یک روش محاسباتی برای پیدا کردن مقدارهای ویژه تعمیم یافته یک مداد معین متقارن ارائه می‌دهد.

الگوریتم ۱-۵-۹ الگوریتم QR چولسکی برای یک مداد معین متقارن. الگوریتم زیر مقادیر ویژه λ_i و بردارهای ویژه متناظر x_i ، $i = 1, \dots, n$ مداد معین متقارن (A, B) را حساب می‌کند.

گام ۱: تجزیه چولسکی B را با استفاده از الگوریتم ۴-۴-۶ پیدا کنید:

$$B = LL^T$$

گام ۲: $C = L^{-1}A(L^T)^{-1}$ را با بهره‌گیری از تقارن A تشکیل دهید.

گام ۳: مقادیر ویژه λ_i ، $i = 1, \dots, n$ بردارهای ویژه یکا متعامد y_i ، $i = 1, \dots, n$ ماتریس متقارن C را با استفاده از الگوریتم تکرار QR متقارن (الگوریتمهای ۸-۱۱-۲ و ۸-۱۰-۱) حساب کنید. $[\lambda_1, \lambda_2, \dots, \lambda_n]$ مقادیر ویژه جفت (A, B) هستند.

گام ۴: بردارهای ویژه x_i متناظر با مقادیر ویژه λ_i مداد (A, B) را با حل $L^T x_i = y_i$ ، $i = 1, \dots, n$ محاسبه کنید.

تذکر MATCOM: الگوریتم ۱-۵-۹ در برنامه CHOLQR در MATCOM پیاده‌سازی شده است.

پایداری الگوریتم

هنگامی که B خوش‌وضع است، الگوریتم پایدار است. ولیکن اگر B بدوضع و یا نزدیک به منفرد باشد، آنگاه L^{-1} نیز چنین است؛ و ماتریس C نمی‌تواند به طور دقیق محاسبه شود. بنابراین در این حالت مقادیر ویژه و بردارهای ویژه نادقیق خواهند بود.

بویژه می‌توان نشان داد (Golub and Van Loan MC (1989, 469)) که یک مقدار ویژه λ محاسبه شده و به دست آمده توسط الگوریتم مقدار ویژه واقعی ماتریس $(L^{-1}A(L^T)^{-1} + E)$ است که در آن $\|B^{-1}\|_2 \approx \mu \|A\|_2$. بنابراین بدوضع B به طور جدی بر مقادیر ویژه محاسبه شده اثر خواهد داشت، حتی اگر مقادیر ویژه خودشان خوش‌وضع باشند.

همان‌گونه که در بخش بعد ملاحظه خواهیم کرد، در بسیاری از کاربردهای عملی فقط یک تعداد کمی از مقادیر ویژه تعمیم یافته مورد نظر هستند. گاهی اوقات این مقادیر ویژه را می‌توان با استفاده از RSF مرتب شده ماتریس B به جای تجزیه چولسکی به صورت زیر با دقت معقول محاسبه کرد حتی اگر B بدوضع باشد.

الگوریتم ۹-۵-۲ الگوریتم QR شور برای یک مداد معین متقارن. الگوریتم زیر مقادیر ویژه و بردارهای ویژه یک مداد معین متقارن (A, B) را با استفاده از شکل شور حقیقی مرتب شده B محاسبه می‌کند. انتظار می‌رود کوچک‌ترین مقدار ویژه به طور دقیق تعیین شود.

گام ۱: شکل شور حقیقی B را محاسبه کنید و مقادیر ویژه B را از کوچک‌ترین به بزرگ‌ترین مرتب کنید.

$$U^T B U = D = \text{diag}(d_1, \dots, d_n)$$

گام ۲: ماتریس L را تشکیل دهید.

$$L = U D^{-\frac{1}{2}} = U \text{diag}(\sqrt{d_1}, \dots, \sqrt{d_n})$$

گام ۳: ماتریس C را تشکیل دهید.

$$C = L^{-1} A (L^T)^{-1}$$

گام ۴: مقادیر ویژه و بردارهای ویژه z_i ماتریس C را با استفاده از تکرار QR متقارن محاسبه کنید (مقادیر ویژه C مقادیر ویژه مداد (A, B) هستند).

از گام ۴ الگوریتم ۹-۵-۱ داریم

$$x_i = (L^T)^{-1}y_i, \quad i = 1, \dots, n$$

با قرار دادن $X = (x_1, x_2, \dots, x_n)$ و $Y = (y_1, y_2, \dots, y_n)$ آنگاه

$$X^TAX = Y^TL^{-1}A(L^{-1})^TY = Y^TCY = \text{diag}(c_1, c_2, \dots, c_n) = D$$

$$X^TBX = Y^TL^{-1}B(L^{-1})^TY = Y^TL^{-1}(LL^T)(L^{-1})^TY = Y^TY = I$$

بالاخره از دو معادله قبل داریم

$$AX = (X^T)^{-1}Y^TCY = BXD$$

اکنون می‌توانیم الگوریتم ۹-۵-۱ را از طریق قطری‌سازی همزمان A و B دوباره بیان کنیم.

الگوریتم ۹-۵-۳ قطری‌سازی همزمان یک مداد معین متقارن. با مفروض بودن یک ماتریس متقارن A و یک ماتریس معین مثبت متقارن B ، الگوریتم زیر یک ماتریس نامنفرد X به قسمی محاسبه می‌کند که X^TBX یک ماتریس واحد و X^TAX یک ماتریس قطری باشند.

گام ۱: تجزیه چولسکی B را محاسبه کنید: (الگوریتم ۶-۴-۴)

$$B = LL^T$$

گام ۲: ماتریس $C = L^{-1}A(L^T)^{-1}$ را با بهره‌گیری از تقارن A تشکیل دهید (C متقارن است).

گام ۳: الگوریتم تکرار QR متقارن (الگوریتم ۸-۱۱-۲) را در مورد C به کار ببرید، یک ماتریس متعامد Y به قسمی پیدا کنید که Y^TCY یک ماتریس قطری باشد.

گام ۴: ماتریس $X = (L^{-1T})Y$ را تشکیل دهید.

شمارش اعمال ممیز شناور. الگوریتم برای قطری‌سازی نیاز به $7n^3$ عمل ممیز شناور (flops) دارد.

تذکر **MATCOM**: الگوریتم ۹-۵-۳ در برنامه **SIMDIAG** در **MATCOM** پیاده‌سازی شده است.

مثال ۹-۵-۲ ماتریسهای زیر را در نظر بگیرید

$$B = \begin{bmatrix} 10 & 1 & 1 \\ 1 & 10 & 1 \\ 1 & 1 & 10 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

A متقارن و B معین مثبت متقارن است.

گام ۱: تجزیه چولسکی $B = LL^T$ را تشکیل می‌دهیم:

$$L = \begin{bmatrix} 3/1623 & \cdot & \cdot \\ \cdot/3162 & 3/1464 & \cdot \\ \cdot/3162 & \cdot/2860 & 3/1334 \end{bmatrix}$$

گام ۲: ماتریس $C = L^{-1}A(L^{-1})^T$ را تشکیل می‌دهیم:

$$C = \begin{bmatrix} \cdot/1000 & \cdot/1910 & \cdot/2752 \\ \cdot/1910 & \cdot/2636 & \cdot/3320 \\ \cdot/2752 & \cdot/3320 & \cdot/3864 \end{bmatrix}$$

گام ۳: یک ماتریس متعامد Y به قسمی به دست می‌آوریم که

$$Y^T C Y = \text{diag}(c_1, c_2, \dots, c_n)$$

$$Y = \begin{bmatrix} \cdot/4220 & -\cdot/8197 & -\cdot/3873 \\ \cdot/5684 & -\cdot/0936 & \cdot/8174 \\ \cdot/7063 & \cdot/5651 & -\cdot/4262 \end{bmatrix}$$

گام ۴: ماتریس X را تشکیل می‌دهیم.

$$X = (L^{-1})^T Y = \begin{bmatrix} \cdot/09409 & -\cdot/2726 & -\cdot/1361 \\ \cdot/1601 & -\cdot/0462 & \cdot/2722 \\ \cdot/2254 & \cdot/1803 & -\cdot/1361 \end{bmatrix}$$

گام ۵: تحقیق می‌کنیم که

$$X^T A X = \text{diag}(\cdot/8179, -\cdot/0679, 0)$$

$$X^T B X = \text{diag}(1, 1, 1) = \text{واحد}$$

ماتریس مدی و مدهای نرمال

در مسائل ارتعاش شامل ماتریسهای جرم و سختی، یعنی وقتی K ماتریس سختی و M ماتریس جرم در مسأله مقدار ویژه تعمیم یافته معین متقارن

$$Kx = \lambda Mx$$

می‌باشند، ماتریس P که به طور همزمان M و K را قطری می‌سازد و به ترتیب آنها را به یک ماتریس

واحد و یک ماتریس قطری تبدیل می‌کند، ماتریس مدی نامیده می‌شود، و ستونهای ماتریس P مدهای نرمال نامیده می‌شوند.

خارج قسمت ریلی تعمیم یافته

خارج قسمت ریلی تعریف شده برای یک ماتریس متقارن A در فصل ۸ (بخش ۸-۵-۳) می‌تواند به آسانی برای جفت معین متقارن (A, B) تعمیم داده شود.

تعریف ۹-۵-۱ عدد

$$\lambda = \frac{x^T A x}{x^T B x}, \quad (\|x\|_r = 1)$$

خارج قسمت ریلی تعمیم یافته نامیده می‌شود. می‌توان نشان داد که $\lambda_1 \leq \frac{x^T A x}{x^T B x} \leq \lambda_n$ ، که در آن λ_1 و λ_n به ترتیب کوچک‌ترین و بزرگ‌ترین مقادیر ویژه (A, B) هستند.

می‌توان از خارج قسمت ریلی برای محاسبه تقریبهایی برای مقادیر ویژه λ_k و بردارهای ویژه x_k برای مسأله تعمیم یافته معین مثبت، به صورت نشان داده شده در الگوریتم زیر استفاده کرد.

الگوریتم ۹-۵-۴ تکرار خارج قسمت ریلی تعمیم یافته

گام ۱: x_0 را به قسمی انتخاب کنید که $\|x_0\| = 1$.

گام ۲: برای $k = 0, 1, \dots$ تا رخ دادن همگرایی انجام دهید.

(الف): محاسبه کنید

$$\lambda_k = \frac{x_k^T A x_k}{x_k^T B x_k}$$

(ب): دستگاه زیر را نسبت به \hat{x}_{k+1} حل کنید.

$$(A - \lambda_k B) \hat{x}_{k+1} = B x_k$$

(ج): \hat{x}_{k+1} را نرمال کنید.

$$x_{k+1} = \frac{\hat{x}_{k+1}}{\|\hat{x}_{k+1}\|_2}$$

تذکر MATCOM: الگوریتم ۹-۵-۴ در برنامه GENRAYQT در MATCOM پیاده‌سازی شده است.

۹-۶ مسائل مقدار ویژه تعمیم یافته معین متقارن مطرح در ارتعاشهای سازه‌ها

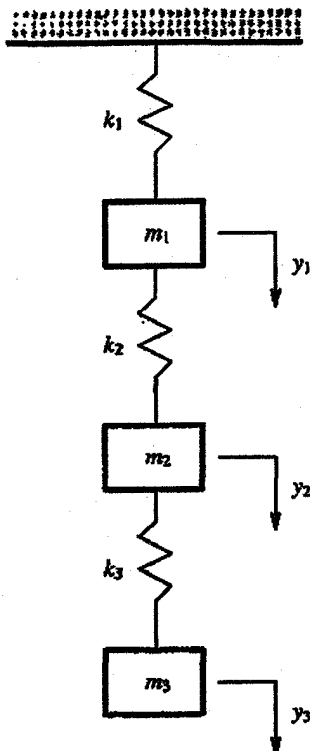
همان‌گونه که قبلاً بیان کردیم، تقریباً همه مسائل مقدار ویژه مطرح در ارتعاشهای سازه‌ها مسائل مقدار ویژه تعمیم یافته معین متقارن هستند. در فصل ۸ (بخش ۸-۳) چند مثال ارائه داده‌ایم. در این بخش مثالهای بیشتری ارائه خواهیم داد.

۹-۶-۱ یک مطالعه موردی: ارتعاش یک دستگاه جرم-فنر

دستگاه سه جرم - فنر نشان داده شده در شکل ۹-۱ را با ثابتهای فنر k_1 ، k_2 ، و k_3 در نظر بگیرید. (ماتریسهای سختی اغلب با استفاده از مشخصه‌های فیزیکی دستگاه تعیین می‌شوند. مقادیر ویژه و بردارهای ویژه سپس برای مطالعه رفتار دستگاه مورد استفاده قرار می‌گیرند.) ما مایلیم وضعیت دستگاه متناظر با مدهای مختلف ارتعاش را مطالعه کنیم.

۱: فرمول‌بندی مسأله به صورت یک مسأله مقدار ویژه تعمیم یافته معین متقارن

فرض کنید تغییر مکانهای جرمها از وضعیتهای تعادل ایستا به ترتیب توسط مختصات تعمیم یافته y_1 ، y_2 ، و y_3 تعریف شوند.



شکل ۹-۱ سه درجه آزادی

معادلات حرکت برای دستگاه می‌توانند به صورت

$$\begin{aligned} m_1 \ddot{y}_1 + (k_1 + k_2)y_1 - k_2 y_2 &= 0 \\ m_2 \ddot{y}_2 - k_2 y_1 + (k_2 + k_3)y_2 - k_3 y_3 &= 0 \\ m_3 \ddot{y}_3 - k_3 y_2 + k_3 y_3 &= 0 \end{aligned}$$

یا به شکل ماتریسی

$$\begin{bmatrix} m_1 & \cdot & \cdot \\ \cdot & m_2 & \cdot \\ \cdot & \cdot & m_3 \end{bmatrix} \begin{bmatrix} \ddot{y}_1 \\ \ddot{y}_2 \\ \ddot{y}_3 \end{bmatrix} + \begin{bmatrix} k_1 + k_2 & -k_2 & \cdot \\ -k_2 & k_2 + k_3 & -k_3 \\ \cdot & -k_3 & k_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

نوشته شود، یعنی

$$M\ddot{y} + Ky = 0$$

که در آن $M = \text{diag}(m_1, m_2, m_3)$ و

$$K = \begin{bmatrix} k_1 + k_2 & -k_2 & \cdot \\ -k_2 & k_2 + k_3 & -k_3 \\ \cdot & -k_3 & k_3 \end{bmatrix}$$

با فرض حرکت هماهنگ، می‌توانیم بنویسیم

$$y_1 = x_1 e^{i\omega t}$$

$$y_2 = x_2 e^{i\omega t}$$

$$y_3 = x_3 e^{i\omega t}$$

که در آن x_1, x_2, x_3 به ترتیب دامنه جرمهای m_1, m_2, m_3 و هستند و ω فرکانس طبیعی را نمایش می‌دهد. با جایگزین کردن این عبارتها به جای y_1, y_2, y_3 در معادلات حرکت و با توجه به این که

$$\ddot{y}_k = -\omega^2 x_k e^{i\omega t}, \quad k = 1, 2, 3$$

داریم:

$$\begin{aligned} -m_1 x_1 \omega^2 + (k_1 + k_2)x_1 - k_2 x_2 &= 0 \\ -m_2 x_2 \omega^2 - k_2 x_1 + (k_2 + k_3)x_2 + k_3 x_3 &= 0 \\ -m_3 x_3 \omega^2 - k_3 x_2 + k_3 x_3 &= 0 \end{aligned}$$

که می‌تواند به شکل ماتریسی زیر نوشته شود:

$$\begin{bmatrix} k_1 + k_2 & -k_2 & 0 \\ -k_2 & k_2 + k_3 & -k_3 \\ 0 & -k_3 & k_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \omega^2 \begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

یا

$$Kx = \lambda Mx$$

که در آن $\lambda = \omega^2$. مقادیر ویژه $\lambda_i = \omega_i^2$ ، $i = 1, 2, 3$ ، به ترتیب توان دوم فرکانسهای طبیعی مدهای اول، دوم، سوم ارتعاش هستند.

۲: حل مسأله مقدار ویژه تعمیم یافته با استفاده از الگوریتم QR چولسکی اجازه دهید اختیار کنیم $m_1 = 20000$ ، $m_2 = 30000$ و $m_3 = 40000$ و $k_1 = k_2 = k_3 = 10^9 \times 1/5$ آنگاه

$$M = \text{diag}(20000, 30000, 40000)$$

$$K = 10^9 \begin{bmatrix} 3 & -1/5 & 0 \\ -1/5 & 3 & -1/5 \\ 0 & -1/5 & 1/5 \end{bmatrix}$$

$$L = \begin{bmatrix} 141/4214 & \cdot & \cdot \\ \cdot & 173/2.51 & \cdot \\ \cdot & \cdot & 200/000 \end{bmatrix}$$

$$C = 10^5 \begin{bmatrix} 1/5 & -0.6124 & \cdot \\ -0.6124 & 1 & -0.4222 \\ \cdot & -0.4222 & 0.3750 \end{bmatrix}$$

مقادیر ویژه (K, M) برابر مقادیر ویژه C هستند که عبارتند از $(0.0860, 0.8382, 1.9508)$ و بردارهای ویژه فرکانسهای طبیعی عبارتند از $(0.9273, 2.8951, 4.4168)$. بردارهای ویژه عبارتند از

$$\begin{bmatrix} 0.0056 \\ -0.0034 \\ 0.0008 \end{bmatrix}, \begin{bmatrix} -0.0040 \\ -0.0035 \\ 0.0028 \end{bmatrix}, \begin{bmatrix} 0.0017 \\ 0.0031 \\ 0.0040 \end{bmatrix}$$

این بردارهای ویژه می‌توانند برای تعیین وضعیتهای مختلف دستگاه برای مدهای مختلف استفاده شوند؛ یعنی آنها می‌توانند برای دیدن این که چگونه سازه‌ها مرتعش می‌شوند هنگامی که در فرکانسهای طبیعی خاص تحریک می‌گردند، استفاده شوند.

تبصره: در محاسبات عملی بهتر است K را بدون تشکیل صریح آن به شکل حاصلضرب $K = BB^T$ حفظ کرد، که در آن

$$B = \begin{bmatrix} \sqrt{k_1} & -\sqrt{k_2} & 0 \\ 0 & \sqrt{k_2} & -\sqrt{k_3} \\ 0 & 0 & \sqrt{k_3} \end{bmatrix}$$

۹-۶-۲ یک مطالعه موردی: ارتعاش یک ساختمان

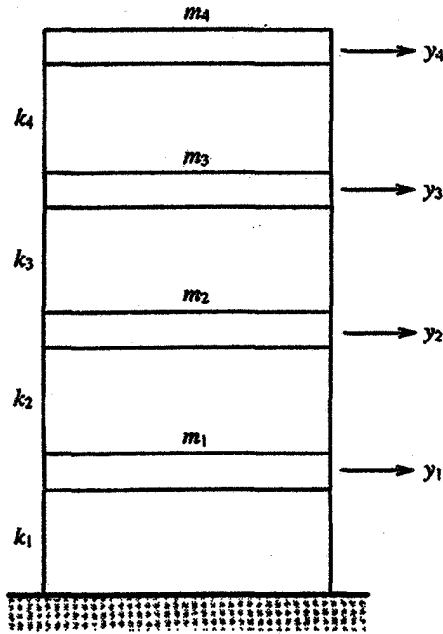
یک ساختمان بتونی محکم چهار طبقه، به صورت نشان داده شده در شکل ۹-۲ را در نظر بگیرید. کفها و سقف، که نسبتاً سفت و محکم هستند، توسط جرمهای توده شده m_1 تا m_4 نمایش داده می‌شوند که یک حرکت افقی بر اثر تغییر برشی ستونها دارند، و k_1 تا k_4 ثابتهای فنری معادل از ستونها هستند که همانند فنرها به طور موازی عمل می‌کنند.

ما مایلیم که وضعیت را به هنگامی که ساختمان در دو مد اول خود (که متناظر با دو کوچکترین مقدار ویژه می‌باشند) ارتعاش می‌کند، مورد مطالعه قرار دهیم.

۱. مسأله را به صورت یک مسأله مقدار ویژه تعمیم یافته معین متقارن بر حسب جرم و ماتریسهای

سختی فرمول‌بندی می‌کنیم:

$$Kx = \lambda Mx$$



شکل ۹-۲ شمای یک بنای چهار طبقه

همانند قبل، M قطری و K سه قطری است:

$$M = \text{diag}(m_1, m_2, m_3, m_4)$$

$$K = \begin{bmatrix} k_1 + k_2 & -k_2 & \cdot & \cdot \\ -k_2 & k_2 + k_3 & -k_3 & \cdot \\ \cdot & -k_3 & k_3 + k_4 & -k_4 \\ \cdot & \cdot & -k_4 & k_4 \end{bmatrix}$$

با اختیار کردن $m_1 = 5 \times 10^7$, $m_2 = 4 \times 10^7$, $m_3 = 3 \times 10^7$, و $m_4 = 2 \times 10^7$ و $k_1 = 10 \times 10^{14}$, $k_2 = 8 \times 10^{14}$, $k_3 = 6 \times 10^{14}$, $k_4 = 4 \times 10^{14}$ داریم:

$$K = 10^{14} \begin{bmatrix} 18 & -8 & \cdot & \cdot \\ -8 & 14 & -6 & \cdot \\ \cdot & -6 & 10 & -4 \\ \cdot & \cdot & -4 & 4 \end{bmatrix}$$

$$M = 10^7 \text{diag}(5, 4, 3, 2)$$

و

۲. مقادیر ویژه و بردارهای ویژه را با استفاده از الگوریتم چولسکی QR پیدا می‌کنیم

$$L = 10^3 \begin{bmatrix} 7/0.711 & \cdot & \cdot & \cdot \\ \cdot & 6/3246 & \cdot & \cdot \\ \cdot & \cdot & 5/4772 & \cdot \\ \cdot & \cdot & \cdot & 4/4721 \end{bmatrix}$$

$$C = 10^7 \begin{bmatrix} 2/6 & -1/7889 & \cdot & \cdot \\ -1/7889 & 3/5000 & -1/7321 & \cdot \\ \cdot & -1/7321 & 3/3333 & -1/6330 \\ \cdot & \cdot & -1/6330 & 2/000 \end{bmatrix}$$

مقادیر ویژه C که همان مقادیر ویژه مداد (K, M) هستند، عبارتند از:

$$10^7 \begin{bmatrix} 6/1432 \\ 1/8516 \\ 0/3435 \\ 4/0945 \end{bmatrix}$$

بردارهای ویژه (K, M) وابسته به این مقادیر ویژه عبارتند از:

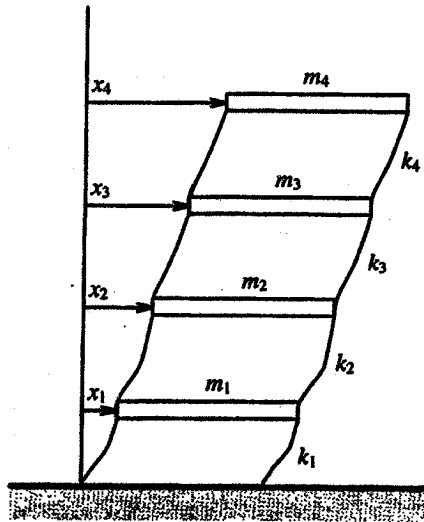
$$10^{-3} \begin{bmatrix} 0/0666 \\ -0/1058 \\ 0/0977 \\ -0/0472 \end{bmatrix}, \quad 10^{-3} \begin{bmatrix} -0/0785 \\ -0/0858 \\ 0/0104 \\ 0/1403 \end{bmatrix}$$

$$10^{-3} \begin{bmatrix} 0/0370 \\ 0/0753 \\ 0/1091 \\ 0/1318 \end{bmatrix}, \quad 10^{-3} \begin{bmatrix} 0/0896 \\ -0/0277 \\ -0/1085 \\ 0/1036 \end{bmatrix}$$

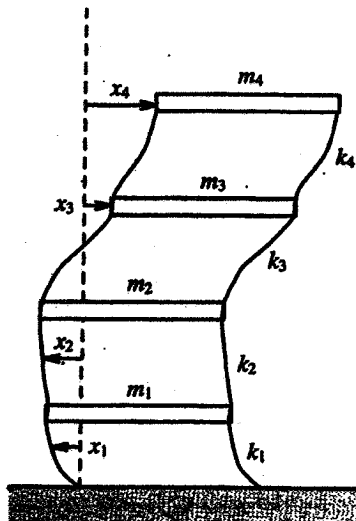
بردارهای ویژه متناظر با دو کوچک‌ترین مقدار ویژه $(0/3435)$ و $(1/8516)$ عبارتند از:

$$10^{-3} \begin{bmatrix} 0/0370 \\ 0/0753 \\ 0/1091 \\ 0/1318 \end{bmatrix}, \quad 10^{-3} \begin{bmatrix} -0/0785 \\ -0/0858 \\ 0/0104 \\ 0/1403 \end{bmatrix}$$

دو مد ارتعاش به ترتیب در شکل‌های ۳-۹ و ۴-۹ نشان داده شده‌اند.



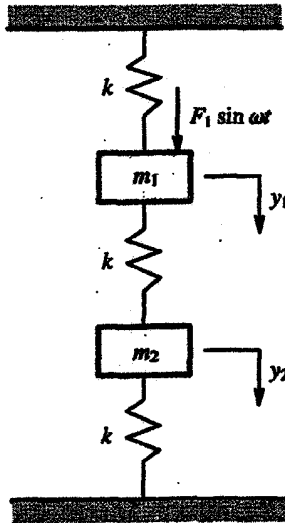
شکل ۳-۹ اولین مد ارتعاش ساختمان چهار طبقه



شکل ۴-۹ دومین مد ارتعاش ساختمان چهار طبقه

۹-۶-۳ ارتعاش هماهنگ اجباری

در دو مثال قبل ارتعاش دستگاههای بدون نیروی خارجی را ملاحظه کردیم. اکنون یک دستگاه با دو درجه آزادی تحریک شده توسط یک نیروی هماهنگ، به صورت نشان داده شده در شکل ۹-۵ را ملاحظه می‌کنیم.



شکل ۹-۵ ارتعاش اجباری یک دستگاه با دو درجه آزادی

پس معادلات دیفرانسیل حرکت برای این دستگاه به شکل زیر در می‌آیند

$$m_1 \ddot{y}_1 = -k(y_1 - y_2) - ky_1 + F_1 \sin \omega t$$

$$m_2 \ddot{y}_2 = -k(y_1 - y_2) - ky_2$$

یا

$$\begin{bmatrix} m_1 & \cdot \\ \cdot & m_2 \end{bmatrix} \begin{bmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{bmatrix} + \begin{bmatrix} 2k & -k \\ -k & 2k \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ \cdot \end{bmatrix} \sin \omega t \quad (9-6-1)$$

با فرض این که یک جواب به شکل

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sin \omega t$$

می‌باشد، با جایگذاری در معادلات به دست می‌آوریم:

$$(\gamma k - m_1 \omega^2)x_1 - kx_2 = F_1$$

$$-kx_1 + (\gamma k - m_2 \omega^2)x_2 = 0$$

به شکل ماتریسی داریم

$$\begin{bmatrix} (\gamma k - m_1 \omega^2) & -k \\ -k & (\gamma k - m_2 \omega^2) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ 0 \end{bmatrix}$$

این معادلات را می توانیم به سادگی حل کنیم و عبارتهای صریحی برای دامنه های x_1 و x_2 به دست آوریم:

$$x_1 = \frac{(\gamma k - m_2 \omega^2) F_1}{m_1 m_2 (\omega_1^2 - \omega^2)(\omega_2^2 - \omega^2)} \quad (2-6-9)$$

$$x_2 = \frac{k F_1}{m_1 m_2 (\omega_1^2 - \omega^2)(\omega_2^2 - \omega^2)} \quad (3-6-9)$$

که در آن ω_1 و ω_2 فرکانسهای مدی هستند. برای حالت خاص $m_1 = m_2 = m$ ، ω_1 و ω_2 به صورت زیر ارائه می شوند

$$\omega_1 = \sqrt{\frac{k}{m}} \quad \text{و} \quad \omega_2 = \sqrt{\frac{3k}{m}}$$

و x_1 و x_2 نیز به صورت زیر ارائه می شوند

$$x_1 = \frac{(\gamma k - m \omega^2) F_1}{m^2 (\omega_1^2 - \omega^2)(\omega_2^2 - \omega^2)} \quad (3-6-9)$$

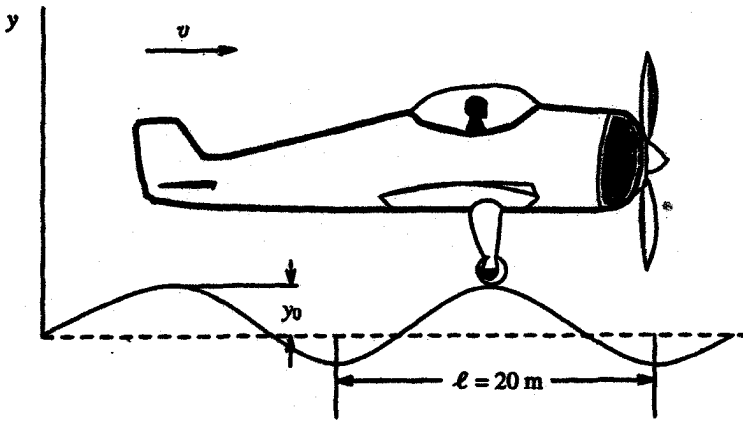
$$x_2 = \frac{k F_1}{m^2 (\omega_1^2 - \omega^2)(\omega_2^2 - \omega^2)} \quad (4-6-9)$$

از این فرمولها فوراً نتیجه گرفته می شود که هرگاه ω مساوی یا نزدیک به ω_1 یا ω_2 باشد، دامنه به طور دلخواه بزرگ می شود و وقوع تشدید را اعلام می کند. توجه کنید که در این حالت مخرج صفر است و یا نزدیک به صفر می باشد. این وضعیت برای یک مهندس بسیار هشدار دهنده است.

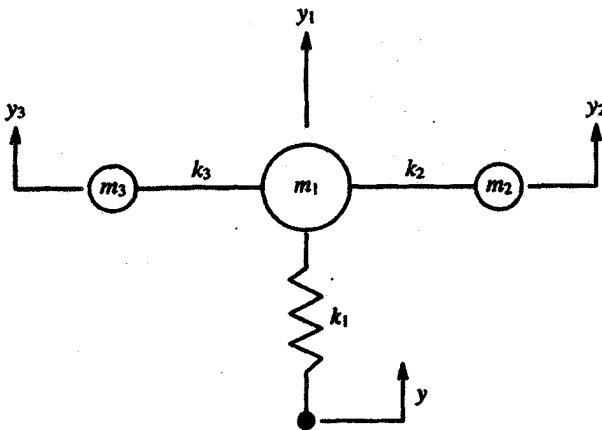
به بیان دیگر اگر فرکانس نیروی دوره ای تحمیل شده برابر یا نزدیک به یکی از فرکانسهای طبیعی یک دستگاه باشد، تشدید رخ می دهد، و وضعیتی است که خطرناک می باشد.

مثال ۹-۶-۱ برای نشان دادن اثرهای تحریک در تشدید یا نزدیکی تشدید، اجازه دهید فرود یک هواپیما را بر روی یک باند فرودگاه ناهموار در نظر بگیریم. هواپیما به طور شمایی در شکل ۹-۶ نشان داده شده است. فرض می کنیم که بدنه هواپیما و موتور با هم دارای یک جرم m_1 باشند. بالها توسط جرمهای توده شده m_2 و m_3 و سختیهای k_2 و k_3 مدل بندی شوند؛ k_1 سختی حرکت دنده های فرود و چرخها را نمایش می دهد. فرض می شود جرمهای چرخها و دنده های فرود در مقایسه با جرم

بدنه هواپیما و بالها قابل صرفنظر باشد. دستگاه در شکل ۷-۹ مدل‌بندی شده است.



شکل ۶-۹ فرود یک هواپیما بر روی یک باند فرودگاه



شکل ۷-۹ مدل هواپیما

باند فرودگاه توسط یک منحنی سینوسی، به صورت نشان داده شده در شکل ۶-۹، مدل‌بندی می‌شود. فرض کنید که خط تراز توسط $y = y_0 \sin \omega t$ توصیف شود، و فرض کنید که هواپیما تحت اثر یک نیروی ورودی $f_1 \sin \omega t$ قرار گیرد که در آن $f_1 = k_1 y_0$. معادلات حرکت برای این دستگاه با سه درجه آزادی و توصیف شده در فوق به صورت زیر می‌باشند.

$$\begin{bmatrix} m_1 & \cdot & \cdot \\ \cdot & m_2 & \cdot \\ \cdot & \cdot & m_3 \end{bmatrix} \begin{Bmatrix} \ddot{y}_1 \\ \ddot{y}_2 \\ \ddot{y}_3 \end{Bmatrix} + \begin{bmatrix} k_1 + k_2 + k_3 & -k_2 & -k_3 \\ -k_2 & k_2 & \cdot \\ -k_3 & \cdot & k_3 \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \\ y_3 \end{Bmatrix} = \begin{Bmatrix} f_1 \sin \omega t \\ \cdot \\ \cdot \end{Bmatrix}$$

که در آن ω_1 ، ω_2 ، و ω_3 حرکت را نسبت به زمین نشان می دهند.
فرض کنید

$$k_1 = 1/7 \times 10^5 \text{ N/m}$$

$$k_2 = k_3 = 6 \times 10^5 \text{ N/m}$$

$$m_1 = 1300 \text{ kg}$$

$$m_2 = m_3 = 300 \text{ kg}$$

فرکانسهای طبیعی به دست آمده از حل مسأله مقدار ویژه تعمیم یافته $Kx = \lambda Mx = \omega^2 Mx$ به صورت زیر می باشند

$$\omega_1 = 9/39 \text{ rad/s}$$

$$\omega_2 = 44/72 \text{ rad/s}$$

$$\omega_3 = 54/46 \text{ rad/s}$$

فرکانس تحمیلی ω به سرعت فرود v توسط $v = \omega l / 2\pi$ وابسته است. در نتیجه اگر داشته باشیم $\omega = \omega_1$ ، آنگاه

$$v = \frac{\omega_1 l}{2\pi} = \frac{9/39 \times 20 \text{ m/s}}{2\pi} = 29/8 \text{ m/s} = 107 \text{ km/h}$$

بنابراین اگر سرعت فرود برابر 107 km/h یا نزدیک به آن باشد، خطر تحریک در تشدید یا نزدیکی تشدید وجود دارد.

۹-۷ کاربردها برای جداسازی و کاهش مدل

اکنون ملاحظه نمودیم که چگونه مقدار ویژه تعمیم یافته در کاربردها مطرح می شوند. در این بخش چند کاربرد دیگر از مهندسی را ذکر خواهیم کرد. این شامل جدا سازی یک دستگاه مرتبه دو از معادلات دیفرانسیل و کاهش مدل می باشد.

۹-۷-۱ جداسازی یک دستگاه مرتبه دو

حالت ۱: دستگاه نامیرا

همان طور که قبلاً دیدیم مسائل ارتعاش در تحلیل سازه‌ای منجر به یک دستگاه همگن از معادلات دیفرانسیل مرتبه دو و به شکل زیر می گردد:

$$M\ddot{y} + Ky = 0 \quad (1-7-9)$$

که در آن

$$\ddot{y} = \frac{d^2 y}{dt^2}$$

و

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

یک بردار n بعدی است. ماتریسهای M و K ، همانند همیشه، ماتریسهای جرم و سختی هستند. با فرض این که این ماتریسها متقارن هستند و M معین مثبت می باشد، اکنون نشان می دهیم که چگونه روش متعامدسازی همزمان (الگوریتم ۹-۵-۳) می تواند به طور مفیدی برای حل دستگاه معادلات مرتبه دو مورد استفاده قرار گیرد.

ایده این است که دستگاه به n معادله جدا تبدیل شود تا هر کدام از این معادلات جدا بتوانند با استفاده از یک روش استاندارد حل شوند. فرض کنید P ماتریس مدی باشد به طوری که

$$P^T M P = I \quad (۲-۷-۹)$$

$$P^T K P = \Lambda = \text{diag}(\omega_1^2, \dots, \omega_n^2) \quad (۳-۷-۹)$$

فرض کنید $y = Pz$ ، در نتیجه دستگاه همگن

$$M\ddot{y} + Ky = 0$$

به صورت زیر در می آید

$$M P \ddot{z} + K P z = 0$$

سپس با پیش ضرب این معادله در P^T داریم

$$P^T M P \ddot{z} + P^T K P z = 0$$

یا

$$\ddot{z} + \Lambda z = 0$$

بنابراین دستگاه همگن به n معادله جدا تبدیل می شود:

$$\ddot{z}_i + \omega_i^2 z_i = 0, \quad i = 1, 2, \dots, n$$

یا

$$\ddot{z}_1 + \omega_1^2 z_1 = 0$$

$$\begin{aligned} \ddot{z}_1 + \omega_1^2 z_1 &= 0 \\ &\vdots \\ \ddot{z}_n + \omega_n^2 z_n &= 0 \end{aligned} \quad (4-7-9)$$

که در آن

$$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}, \quad \Lambda = \text{diag}(\omega_1^2, \omega_2^2, \dots, \omega_n^2)$$

اکنون جواب دستگاه اصلی می‌تواند از حل این n معادله جدا با استفاده از روشهای استاندارد و سپس بازیابی جواب اصلی y از رابطه زیر به دست آید

$$y = Pz$$

بنابراین اگر جوابهای دستگاه تبدیل یافته (۴-۷-۹) به صورت

$$z_i = A_i \cos \omega_i t + B_i \sin \omega_i t, \quad i = 1, 2, \dots, n$$

ارائه شوند، آنگاه جوابهای دستگاه اصلی (۱-۷-۹) به صورت زیر خواهند بود

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = P \begin{bmatrix} A_1 \cos \omega_1 t + B_1 \sin \omega_1 t \\ A_2 \cos \omega_2 t + B_2 \sin \omega_2 t \\ \vdots \\ A_n \cos \omega_n t + B_n \sin \omega_n t \end{bmatrix} \quad (5-7-9)$$

ثابتهای A_i و B_i از شرایط اولیه نظیر شرایط زیر تعیین خواهند شد

$$y_i|_{t=0} = \quad t = 0 \quad \text{تغییر مکان در زمان}$$

$$\dot{y}_i|_{t=0} = \quad \text{سرعت اولیه}$$

مثال ۱-۷-۹ دستگاه با سه جرم مثال ۱-۶-۹ را در نظر بگیرید. فرض کنید که دستگاه هنگامی که

$t = 0$ از حالت سکون رها شود و در معرض یک تغییر مکان قرار گیرد.

می‌خواهیم پاسخ زمانی نامیرای دستگاه را پیدا کنیم. شرایط اولیه عبارتند از:

$$y_0 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \dot{y}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

چون سرعتهای اولیه صفر هستند، به دست می‌آوریم

$$\dot{y} = P \begin{bmatrix} B_1 \omega_1 \\ B_2 \omega_2 \\ B_3 \omega_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

این رابطه نتیجه می‌دهد $B_1 = B_2 = B_3 = 0$.

مجدداً در $t = 0$ داریم:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = P \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

برای این مسأله ماتریس بردارهای ویژه را به خاطر آورید:

$$P = \begin{bmatrix} 0/0056 & -0/0040 & 0/0017 \\ -0/0034 & -0/0035 & 0/0031 \\ 0/0008 & 0/0028 & 0/0040 \end{bmatrix}$$

سپس جواب دستگاه خطی

$$P \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

عبارت است از:

$$A_1 = 6/1816$$

$$A_2 = 50/6264$$

$$A_3 = 705/2650$$

با جایگذاری این مقادیر A_1 ، A_2 ، و A_3 و مقادیر ω_1 ، ω_2 ، و ω_3 به دست آمده (در بخش

۹-۶-۱) در رابطه

$$y = Pz$$

یا

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = P \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = P \begin{bmatrix} A_1 \cos \omega_1 t \\ A_2 \cos \omega_2 t \\ A_3 \cos \omega_3 t \end{bmatrix}$$

مقادیر y_1 ، y_2 ، و y_3 را به دست می‌آوریم که پاسخ زمانی نامیرای دستگاههای قرار گرفته در معرض شرایط اولیه مفروض را ارائه می‌دهند.

حالت ۲: دستگاههای میرا

برخی میراییها، نظیر آنچه که حاصل مقاومت هوا یا مایع و اصطکاک جامد می باشد، در دستگاههای حقیقی وجود دارند. اجازه دهید اکنون دستگاههای میرای همگن را ملاحظه کنیم.

فرض کنید D ماتریس میرایی باشد. آنگاه معادلات حرکت دستگاه میرا به شکل زیر در می آید

$$M\ddot{y} + D\dot{y} + Ky = 0 \quad (6-7-9)$$

اگر D یک ترکیب خطی از M و K باشد؛ یعنی

$$D = \alpha M + \beta K \quad (7-7-9)$$

که در آن α و β ثابت هستند، آنگاه گفته می شود که دستگاه دارای میرایی تناسبی، یا ریلی می باشد. فرض کنید P ماتریس مدی باشد. پس داریم

$$\begin{aligned} P^T D P &= \alpha P^T M P + \beta P^T K P \\ &= \alpha I + \beta \Lambda \end{aligned}$$

فرض کنید که $z = P^T y$ ، بنابراین معادلات میرای همگن قبلی به n معادله جدا تبدیل می شوند:

$$\ddot{z}_i + (\alpha + \beta \omega_i^2) \dot{z}_i + \omega_i^2 z_i = 0, \quad i = 1, 2, \dots, n \quad (8-7-9)$$

در کاربردهای مهندسی مرسوم است که نسبت میرایی مدی i امین مد را توسط رابطه زیر تعریف کنند

$$\xi_i = \frac{\alpha + \beta \omega_i^2}{2\omega_i}$$

ξ_i معمولاً به صورت یک عدد کوچک بین صفر و یک اختیار می شود. متداول ترین مقادیر عبارتند از $0 \leq \xi \leq 0.05$ (Inman (1994, 1996)). ولیکن در برخی از کاربردها، مثلاً در طرح سازه های قابل انعطاف، $\{\xi_i\}$ به کوچکی 0.05 اختیار می شوند. از طرف دیگر برای جذب کننده شوک اتوماتیک، مقداری به بزرگی $\xi = 0.5$ نیز امکان پذیر است.

با تعریف بالا برای ξ_i ، معادلات جدا شده (8-7-9) به صورت زیر در می آیند

$$\ddot{z}_i + 2\xi_i \omega_i \dot{z}_i + \omega_i^2 z_i = 0, \quad i = 1, 2, \dots, n$$

جوابهای این معادلات سپس به صورت زیر داده می شوند

$$z_i = e^{-\xi_i \omega_i t} (A_i \cos \omega_i \sqrt{1 - \xi_i^2} t + B_i \sin \omega_i \sqrt{1 - \xi_i^2} t), \quad i = 1, 2, \dots, n$$

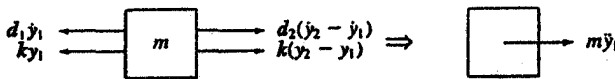
در اینجا ثابتهای A_i و B_i باید از شرایط اولیه داده شده تعیین شوند.

اکنون دستگاه اصلی می‌تواند با حل جداگانه این n معادله جدا شده حل شود و سپس جواب اصلی از $y = Pz$ به دست آید.

تبصره: توجه کنید که برای قطری سازی همزمان M ، K ، و D ، میرایی تناسبی دقیقاً یک شرط خاص است. یک شرط کلی تر $DM^{-1}K = KM^{-1}D$ می‌باشد (مرجع (Inman (1989) ملاحظه شود).

مثال ۹-۷-۲ مفید بودن میرایی تناسبی. دستگاه با دو درجه آزادی (DOF) نشان داده شده در شکل ۹-۸ را ملاحظه کنید. در اینجا نشان خواهیم داد که چرا میرایی تناسبی مفید است.

با در نظر گرفتن یک دیاگرام جسم آزاد برای هر جرم، معادلات حرکت دستگاه تدوین می‌یابند.
برای جرم m :



معادله حرکت برای این جرم به صورت زیر است

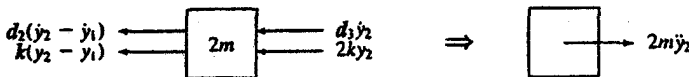
$$-d_1 \dot{y}_1 + d_r (\dot{y}_r - \dot{y}_1) - ky_1 + k(y_r - y_1) = m \ddot{y}_1$$

یا

$$m \ddot{y}_1 + (d_1 + d_r) \dot{y}_1 - d_r \dot{y}_r + \gamma ky_1 - ky_r = 0$$

یا

برای جرم $2m$:



معادله حرکت برای این جرم به صورت زیر است

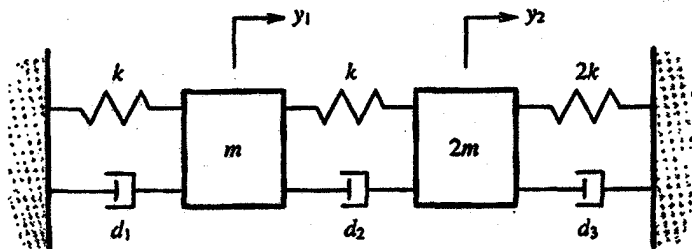
$$-d_r (\dot{y}_r - \dot{y}_1) - d_r \dot{y}_r - k(y_r - y_1) - \gamma ky_r = 2m \ddot{y}_r$$

یا

$$2m \ddot{y}_r - d_r \dot{y}_1 + (d_r + d_r) \dot{y}_r - ky_1 + \gamma ky_r = 0$$

بنابراین برای تمام دستگاه خواهیم داشت

$$\begin{bmatrix} m & 0 \\ 0 & 2m \end{bmatrix} \begin{bmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{bmatrix} + \begin{bmatrix} d_1 + d_2 & -d_2 \\ -d_2 & d_2 + d_3 \end{bmatrix} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \end{bmatrix} + \begin{bmatrix} 2k & -k \\ -k & 2k \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$



شکل ۹-۸

قرار دهید $k = 2$ و $m = 5$ ، و فرض کنید که $D = \alpha M + \beta K$. برای این دستگاه با دو درجه آزادی، ماتریس جرم به صورت زیر است

$$M = \begin{bmatrix} m & 0 \\ 0 & 2m \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}$$

و ماتریس سختی به شکل زیر می باشد

$$K = \begin{bmatrix} 2k & -k \\ -k & 2k \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}$$

و ماتریس میرایی نیز عبارت است از

$$D = \begin{bmatrix} d_1 + d_2 & -d_2 \\ -d_2 & d_2 + d_3 \end{bmatrix}$$

دو حالت زیر را در نظر بگیرید

حالت ۱: $d_3 = 2, d_2 = 1, d_1 = 1$

$$\begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} = \alpha \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix} + \beta \begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}$$

$$2 = 5\alpha + 4\beta \quad (9-7-9)$$

$$-1 = -2\beta \quad (10-7-9)$$

$$3 = 10\alpha + 6\beta \quad (11-7-9)$$

معادلات (۹-۷-۹)، (۹-۷-۱۰)، و (۹-۷-۱۱) فقط با $\beta = \frac{1}{4}$ و $\alpha = 0$ برقرار می‌باشند. بنابراین این یک حالت میرایی تناسبی است.

حالت ۲: $d_3 = 1, d_2 = 4, d_1 = 2$

$$\begin{bmatrix} 6 & -4 \\ -4 & 5 \end{bmatrix} = \alpha \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix} + \beta \begin{bmatrix} 4 & -2 \\ -2 & 6 \end{bmatrix}$$

$$6 = 5\alpha + 4\beta \quad (12-7-9)$$

$$-4 = -2\beta \quad (13-7-9)$$

$$5 = 10\alpha + 6\beta \quad (14-7-9)$$

معادلات (۹-۷-۱۲)، (۹-۷-۱۳)، و (۹-۷-۱۴) نمی‌توانند با هیچ مقداری از α و β برقرار شوند. بنابراین میرایی تناسبی ممکن نیست. بعلاوه در این حالت به سادگی می‌توان امتحان کرد که شرط کلی قطری‌سازی همزمان، یعنی $DM^{-1}K = KM^{-1}D$ نقض می‌گردد.

بنابراین در حالت اول، معادلات حرکت می‌توانند با استفاده از $y = Pz$ جدا شوند، و مدهای حقیقی و فرکانسهای طبیعی حقیقی به دست آیند. ولیکن در حالت دوم، چنین جدا سازی امکان پذیر نیست.

دستگاههای میرای تحت تحریک نیرو

هنگامی که یک دستگاه میرا تحت یک نیروی خارجی F قرار می‌گیرد، معادلات حرکت به صورت زیر ارائه می‌شوند.

$$M\ddot{y} + D\dot{y} + Ky = F(t) = \begin{Bmatrix} F_1(t) \\ F_2(t) \\ \vdots \\ F_n(t) \end{Bmatrix} \quad (15-7-9)$$

با فرض این که M معین مثبت متقارن بوده و K متقارن است، و میرایی تناسبی می‌باشد، به راحتی از بحثهای قبل ملاحظه می‌کنیم که معادلات قبلی می‌توانند با استفاده از قطری سازی همزمان جدا شوند.

فرض کنید $P = (p_{ij})$ ماتریس مدی باشد. آنگاه

$$P^T F = \begin{pmatrix} p_{11} & p_{21} & \dots & p_{n1} \\ p_{12} & p_{22} & \dots & p_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1i} & p_{2i} & \dots & p_{ni} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1n} & p_{2n} & \dots & p_{nn} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_i \\ \vdots \\ F_n \end{pmatrix}$$

سپس معادلات جدا شده به صورت زیر ارائه خواهند شد.

$$\ddot{z}_i + \gamma \zeta_i \omega_i \dot{z}_i + \omega_i^2 z_i = p_{1i} F_1 + p_{2i} F_2 + \dots + p_{ni} F_n, \quad i = 1, 2, \dots, n$$

یا

$$\ddot{z}_i + \gamma \zeta_i \omega_i \dot{z}_i + \omega_i^2 z_i = E_i(t) \quad (۱۶-۷-۹)$$

$$.i = 1, 2, \dots, n, E_i(t) = \sum_{j=1}^n p_{ji} F_j$$

تابع $E_i(t)$ تابع تحریک، یا تابع نیروی i امین مد نامیده می شود. اگر هر نیروی F_i به صورت

$$F_i = f_i s(t)$$

نوشته شود، آنگاه

$$E_i(t) = s(t) \sum_{j=1}^n p_{ji} f_j$$

تعریف ۱-۷-۹ عبارت

$$\sum_{j=1}^n p_{ji} f_j$$

عامل شرکت مد برای i امین مد نامیده می شود.

به محض آن که معادلات جدا شده نسبت به z_i ها حل شوند، جوابهای معادلات اصلی را می توانیم

از رابطه

$$y = Pz$$

یا

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = P \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$$

به دست آوریم. جوابهای معادلات جدا شده

$$\ddot{z}_i + 2\zeta_i \omega_i \dot{z}_i + \omega_i^2 z_i = E_i(t)$$

به طبیعت نیروی $F(t)$ بستگی دارند. برای مثال هنگامی که نیرو از نوع نیروی شوک، نظیر یک زمین لرزه می باشد، ما به طور نرمال علاقه مند به پاسخهای ماکزیمم هستیم، و مقادیر ماکزیمم z_1, z_2, \dots, z_n می توانند از پاسخهای یک معادله تک با یک درجه آزادی به دست آیند (مثال بخش ۹-۷-۳ را ملاحظه کنید).

۹-۷-۲ کاهش یک مدل بزرگ

بسیاری از کاربردها منجر به یک دستگاه خیلی بزرگ از معادلات دیفرانسیل مرتبه دو می شوند:

$$M\ddot{y} + D\dot{y} + Ky = F(y, t)$$

بنابراین این معادله از لحاظ نظری از بعد نامتناهی و در عمل از بعد خیلی بزرگ است. یک مدل با مرتبه کاهش یافته و تولید شده توسط عنصر متناهی می تواند یک دستگاه از مرتبه دو بزرگ باشد. بعد مسأله در یک چنین کاربردی می تواند از مرتبه چندین هزار باشد. طبیعتاً جواب یک دستگاه بزرگ به یک جواب مسأله مقدار ویژه تعمیم یافته خیلی بزرگ منجر خواهد شد. متأسفانه روشهای عددی مؤثر برای محاسبه یک تعداد زیادی از مقادیر ویژه و بردارهای ویژه یک مسأله مقدار ویژه تعمیم یافته واقعا وجود ندارند و خیلی خوب توسعه نیافته اند (بخش ۹-۹ را ملاحظه کنید). بنابراین طبیعی است که به حل مسأله توسط ساختن مدلی با مرتبه کاهش یافته به کمک یک تعداد کمی از مقادیر ویژه و بردارهای ویژه فکر کنیم که امکان محاسبه آنها وجود داشته باشد. یک چنین فکری مبتنی بر این فرض است که در مثالهای زیادی پاسخ سازه اساساً بستگی به چند مقدار ویژه اول (فرکانسهای پایین) دارد، معمولاً مدهای بالاتر تحت تحریک قرار نمی گیرند.

اکنون نشان می دهیم که چگونه محاسبات می توانند فقط با استفاده از شناخت تعداد کمی از مقادیر ویژه و بردارهای ویژه ساده شوند.

فرض کنید که تحت فرض معمول که M و K متقارن و از مرتبه n هستند و M معین مثبت می باشد، ما توانسته ایم فقط چند مد نرمال اول را محاسبه کنیم، شاید m تا از آنها، که $m \ll n$. فرض کنید ماتریس این مدهای نرمال $P_{n \times m}$ باشد. آنگاه

$$P^T M P = I_{m \times m}$$

$$P^T K P = \Lambda_{m \times m} = \text{diag}(\omega_1^2, \dots, \omega_m^2)$$

با قرار دادن $y = Pz$ و با فرض میرایی تناسبی، دستگاه n معادله دیفرانسیل (۹-۷-۱۷) سپس به m معادله زیر کاهش می‌یابد

$$\ddot{z}_i + 2\zeta_i \omega_i \dot{z}_i + \omega_i^2 z_i = E_i(t), \quad i = 1, 2, \dots, m$$

که در آن E_i مؤلفه i ام بردار $P^T F$ می‌باشد. به محض این که این تعداد کم از معادلات حل شوند، تغییر مکان هر جرم تحت نیروی خارجی می‌تواند از فرمول زیر محاسبه شود.

$$y_i = Pz_i, \quad i = 1, 2, \dots, n$$

گاهی اوقات فقط ماکزیمم مقدار تغییر مکان مورد نظر می‌باشد.

تعداد زیادی از گروه‌های ارتعاش در صنعت و ارتش از تقریب زیر برای به دست آوردن ماکزیمم مقدار y_i استفاده می‌کنند (Thomson (1992):

$$|y_i|_{\max} = |p_i z_{1(\max)}| + \sqrt{\sum_{j=2}^m |p_i z_{j(\max)}|^2}$$

که در آن p_i ، i امین ستون P محاسبه شده در z_i می‌باشد.

۹-۷-۳ یک مطالعه موردی: خسارت بالقوه بر یک ساختمان ناشی از یک زمین لرزه فرض کنید مایل باشیم که ماکزیمم مطلق پاسخهای ساختمان چهار طبقه مثال ملاحظه شده در بخش ۹-۲-۶ را به هنگامی که تحت یک زمین لرزه قوی قرار می‌گیرد، با استفاده از پاسخ معلوم ساختمان که ناشی از زمین لرزه قبلی باشد، پیدا کنیم. ما فقط دو مد اول (مدهای متناظر با دو کوچک‌ترین فرکانس) را در محاسبات استفاده خواهیم کرد.

در شکل ۹-۹، y_i ها، $i = 1, \dots, 4$ ، تغییر مکان هر طبقه را نسبت به محمل حرکت، که حرکتش توسط y_0 نمایش داده می‌شود، نشان می‌دهند.

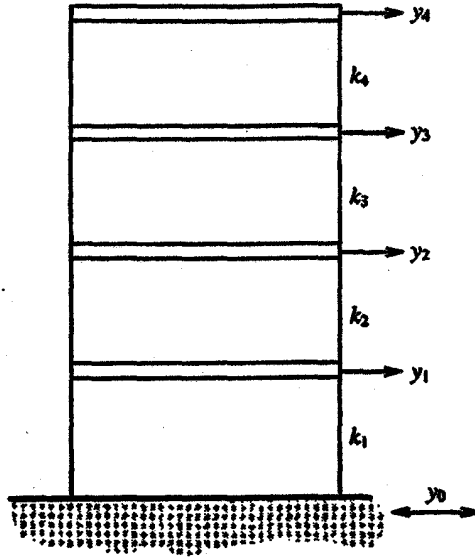
در این حالت معادلات مد نرمال جدا شده به شکل مدی می‌توانند به صورت زیر نوشته شوند

$$\ddot{z}_i + 2\zeta_i \omega_i \dot{z}_i + \omega_i^2 z_i = -E_i \ddot{y}_0, \quad i = 1, 2$$

که در آن

\ddot{y}_0 = شتاب مطلق محمل حرکت

عامل شرکت مد از مد انتخاب شده p_i حاصل از تحریک محمل $E_i = \sum_{j=1}^4 p_{ji} m_j$



شکل ۹-۹ ساختمان تحت تأثیر یک زمین لرزه

p_{ji} ها مختصات مد شرکت کننده p_i هستند؛ یعنی

$$P = (p_1, p_2) = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \\ p_{31} & p_{32} \\ p_{41} & p_{42} \end{bmatrix}$$

که در آن p_1 و p_2 دو مد شرکت کننده انتخاب شده هستند.

فرض کنید R_1 و R_2 ماکزیمم پاسخهای نسبی $z_1(\max)$ و $z_2(\max)$ به دست آمده از آزمایش قبلی باشند. آنگاه می توانیم اختیار کنیم

$$z_1(\max) = E_1 R_1$$

$$z_2(\max) = E_2 R_2$$

فورا از این مشاهدات نتیجه می گیریم

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} (\max) = E_1 R_1 \begin{bmatrix} p_{11} \\ p_{21} \\ p_{31} \\ p_{41} \end{bmatrix} + E_2 R_2 \begin{bmatrix} p_{12} \\ p_{22} \\ p_{32} \\ p_{42} \end{bmatrix}$$

با استفاده از داده های مثال بخش ۹-۶-۲ داریم

بردار ویژه متناظر با کوچک ترین مقدار ویژه $p_1 =$

$$= 10^{-3} \begin{bmatrix} 0/0.370 \\ 0/0.753 \\ 0/1.091 \\ 0/1318 \end{bmatrix}$$

بردار ویژه متناظر با دومین کوچک ترین مقدار ویژه $p_2 =$

$$= 10^{-3} \begin{bmatrix} -0/0.785 \\ -0/0.858 \\ 0/0.104 \\ 0/14.3 \end{bmatrix}$$

$$E_1 = m_1 p_{11} + m_2 p_{21} + m_3 p_{31} + m_4 p_{41} = 10^4 \times (1/0.772)$$

$$E_2 = m_1 p_{12} + m_2 p_{22} + m_3 p_{32} + m_4 p_{42} = 10^3 \times (-4/2417)$$

فرض کنید که $R_1 = 1/5 \text{ in}$ و $R_2 = 0/25 \text{ in}$. به دست می آوریم

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} (\max) = 10^{-3} \times 10^4 (1/0.772)(1/5) \begin{bmatrix} 0/0.370 \\ 0/0.753 \\ 0/1.091 \\ 0/1318 \end{bmatrix}$$

$$+ 10^{-3} \times 10^3 (-4/2417)(0/25) \begin{bmatrix} -0/0.785 \\ -0/0.858 \\ 0/0.104 \\ 0/14.3 \end{bmatrix}$$

$$= \begin{bmatrix} 0/5979 \\ 1/2169 \\ 1/7636 \\ 2/1292 \end{bmatrix} + \begin{bmatrix} 0/0.833 \\ 0/0.910 \\ -0/0.110 \\ -0/1488 \end{bmatrix}$$

$$= \begin{bmatrix} 0/6812 \text{ in} \\ 1/3.79 \text{ in} \\ 1/7526 \text{ in} \\ 1/98.05 \text{ in} \end{bmatrix}$$

بنابراین ماکزیمم مطلق تغییر مکان (نسبت به محمل حرکت) طبقه اول برابر $0/6812 \text{ in}$ ، طبقه دوم برابر $1/3079 \text{ in}$ می باشند، و غیره. شکل ۹-۱۰ را ملاحظه کنید.

ماکزیمم مطلق تغییر مکانهای نسبی از اضافه کردن جملات با استفاده از قدر مطلق مقادیرشان به دست می آیند.

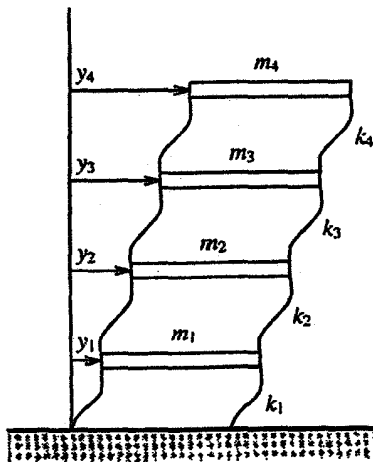
$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} (\text{abs, max}) = \begin{pmatrix} 0/6812 \text{ in} \\ 1/3079 \text{ in} \\ 1/7746 \text{ in} \\ 2/2780 \text{ in} \end{pmatrix}$$

تذکر: سهم مد شرکت کننده دوم در پاسخها در مقایسه با سهم مد اول کوچک است.

میانگین ماکزیمم تغییر مکان نسبی جرما

ماکزیمم مطلق تغییر مکان نسبی جرما برای ما یک کران بالا برای بزرگترین تغییر مکانهای نسبی که جرما می توانند داشته باشند، فراهم می سازد و در نتیجه به ما کمک می کند تا پارامترهای طرح را انتخاب کنیم. راه حل عملی دیگر برای چنین اندازه یی در الفبای مهندسی استفاده از جذر مجموع توانهای دوم همان جملات بوده است که متوسط ماکزیمم مقادیر تغییر مکان نسبی را ارائه می دهد.

$$(V_i)_{\text{average max}} = \sqrt{(E_1 R_1 P_{i1})^2 + (E_2 R_2 P_{i2})^2 + \dots + (E_k R_k P_{ik})^2}$$



شکل ۹-۱۰ ماکزیمم مطلق تغییر مکان

برای مثال قبل، $k = 2$ ، و متوسط ماکزیمم تغییر مکانهای نسبی به صورت زیر ارائه می شود

$$(y_1)_{\text{average max}} = \sqrt{(E_1 R_1 P_{11})^2 + (E_2 R_2 P_{12})^2} = 0.9975 \text{ in}$$

$$(y_2)_{\text{average max}} = \sqrt{(E_1 R_1 P_{21})^2 + (E_2 R_2 P_{22})^2} = 1.5610 \text{ in}$$

و غیره.

۸-۹ مسأله مقدار ویژه درجه دو

تاکنون ما دستگاه مرتبه دو

$$M\ddot{y} + D\dot{y} + Ky = 0 \quad (1-8-9)$$

را تحت فرض میرایی تناسبی یا مدی ملاحظه نمودیم. این فرض به ما اجازه استفاده از قطری سازی همزمان را بر روی M ، K ، و D می دهد (بخش ۹-۷-۱ حالت ۲ را ملاحظه کنید)، که البته همان طور

که در بخش ۹-۵ ملاحظه نمودیم به قیمت حل یک مسأله مقدار ویژه تعمیم یافته تمام می شود.

در حالت میرایی کلی، با فرض این که جوابها به شکل $y = x e^{\lambda t}$ می باشند، که در آن x یک بردار

ثابت است ما مسأله درجه دوم زیر را خواهیم داشت

$$(\lambda^2 M + \lambda D + K)x = 0 \quad (2-8-9)$$

این مسأله دارای $2n$ مقدار ویژه است و $2n$ بردار ویژه متناظر با آنها وجود دارد. دو طریق موجود برای حل این مسأله مقدار ویژه درجه دو را توصیف می کنیم و نشان می دهیم که چگونه فرکانسها و نرخهای میرایی مدی را از مقادیر ویژه استخراج نماییم. همانند قبل فرض می کنیم که M معین مثبت متقارن باشد.

طریق ۱: کاهش به یک مسأله مقدار ویژه استاندارد

می توانیم مسأله (۲-۸-۹) را دوباره به صورت

$$\begin{bmatrix} \cdot & I \\ -M^{-1}K & -M^{-1}D \end{bmatrix} \begin{bmatrix} x \\ \lambda x \end{bmatrix} = \lambda \begin{bmatrix} x \\ \lambda x \end{bmatrix}$$

یا

$$Au = \lambda u \quad (3-8-9)$$

بنویسیم که در آن

$$A = \begin{bmatrix} \cdot & I \\ -M^{-1}K & -M^{-1}D \end{bmatrix}, \quad u = \begin{bmatrix} x \\ \lambda x \end{bmatrix} \quad (4-8-9)$$

بنابراین ما مسأله مقدار ویژه درجه دو (۹-۸-۲) را به مسأله مقدار ویژه استاندارد با مرتبه $2n \times 2n$ (۹-۸-۳) کاهش داده‌ایم. که در آن A و u توسط (۹-۸-۴) ارائه می‌شوند.
 اگر دستگاه مرتبه دو مدل یک سازه مرتعش باشد، آنگاه فرکانسهای طبیعی و نرخهای میرایی مدی می‌توانند از مقادیر ویژه به صورت زیر محاسبه شوند.

محاسبه فرکانسها و نرخهای میرایی

فرض کنید $\lambda_k = \alpha_k + i\beta_k$ ؛ یعنی α_k و β_k به ترتیب قسمت‌های حقیقی و موهومی مقدار ویژه مختلط λ_k باشند. آنگاه فرکانسهای طبیعی ω_k به صورت زیر ارائه می‌شوند

$$\omega_k = \sqrt{\alpha_k^2 + \beta_k^2}, \quad k = 1, 2, \dots, n \quad (5-8-9)$$

(Inman (1994, 462)). توجه کنید که مقادیر ویژه λ_k به صورت جفت‌های مزدوج مختلط ظاهر می‌شوند. نسبت‌های میرایی مدی به صورت زیر ارائه می‌شوند

$$\xi_k = \frac{-\alpha_k}{\sqrt{\alpha_k^2 + \beta_k^2}} \quad (6-8-9)$$

طریق ۲: کاهش به یک مسأله مقدار ویژه تعمیم یافته متقارن
 چندین راه دیگر برای بازنویسی (۹-۸-۲) وجود دارد. یکی از آنها عبارت است از

$$\begin{bmatrix} D & K \\ K & \cdot \end{bmatrix} \begin{bmatrix} \lambda x \\ x \end{bmatrix} = \lambda \begin{bmatrix} -M & \cdot \\ \cdot & K \end{bmatrix} \begin{bmatrix} \lambda x \\ x \end{bmatrix}$$

یا

$$Bz = \lambda Cz \quad (7-8-9)$$

که در آن

$$B = \begin{bmatrix} D & K \\ K & \cdot \end{bmatrix}, \quad C = \begin{bmatrix} -M & \cdot \\ \cdot & K \end{bmatrix}, \quad z = \begin{bmatrix} \lambda x \\ x \end{bmatrix} \quad (8-8-9)$$

رابطه (۹-۸-۷) یک مسأله مقدار ویژه تعمیم یافته متقارن است.
 در اینجا B و C ، در حالت کلی، نامعین هستند، بنابراین بردارهای ویژه می‌توانند مختلط باشند.

فرکانسهای طبیعی و نرخهای میرایی مدی می توانند از مقادیر ویژه همانند طریق ۱ استخراج شوند. خلاصه. مسأله درجه دو

$$(\lambda^2 M + \lambda D + K)x = 0$$

می تواند یا به مسأله مقدار ویژه نامتقارن استاندارد

$$Au = \lambda u$$

تبدیل شود که در آن

$$A = \begin{bmatrix} \cdot & I \\ -M^{-1}K & -M^{-1}D \end{bmatrix}, \quad u = \begin{bmatrix} x \\ \lambda x \end{bmatrix}$$

یا به مسأله مقدار ویژه تعمیم یافته متقارن

$$Bz = \lambda Cz$$

تبدیل شود که در آن

$$B = \begin{bmatrix} D & K \\ K & \cdot \end{bmatrix}, \quad C = \begin{bmatrix} -M & \cdot \\ \cdot & K \end{bmatrix}, \quad z = \begin{bmatrix} \lambda x \\ x \end{bmatrix}$$

به محض این که مقادیر ویژه و بردارهای ویژه محاسبه شوند، فرکانسهای طبیعی و نرخهای میرایی مدی یک سازه مرتعش که مدل ریاضی آن دستگاه مرتبه دو (۹-۸-۱) می باشد، می توانند از مقادیر ویژه (۹-۸-۳) یا (۹-۸-۷) محاسبه شوند.

یک پیغام احتیاط در مورد استفاده از طریق ۱: اگر M بدوضع باشد، تشکیل صریح ماتریس A ارائه شده توسط (۹-۸-۴) به وسیله محاسبه M^{-1} از لحاظ محاسباتی فجیع خواهد بود. A و همچنین مقادیر ویژه و بردارهای ویژه به طور نادقیق محاسبه خواهند شد. همچنین توجه کنید که A نامتقارن است.

تذکر MATCOM: طریق ۱ و ۲ به ترتیب در برنامه های QUADEIG1 و QUADEIG2 در MATCOM پیاده سازی شده اند.

مثال ۹-۸-۱ مقادیر ویژه A و بردارهای ویژه مسأله مقدار ویژه درجه دو (۹-۸-۲) را با ماتریسهای زیر پیدا کنید.

$$M = \text{diag}(1, 2), \quad D = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 2 \\ 2 & \cdot \end{bmatrix}$$

طریق ۱: ماتریس A را تشکیل می‌دهیم

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -2 & -2 & -2 \\ -1 & 0 & -1 & -1 \end{bmatrix}$$

مقادیر ویژه A و بردارهای ویژه آن عبارتند از: -2 ، 1 ، -1 ، و -1 .

$$\begin{bmatrix} 0/4000 \\ 0/2000 \\ -0/8000 \\ -0/4000 \end{bmatrix}, \begin{bmatrix} 0/5000 \\ -0/5000 \\ 0/5000 \\ -0/5000 \end{bmatrix}, \begin{bmatrix} 0/6696 \\ 0/2273 \\ -0/6696 \\ -0/2273 \end{bmatrix}, \begin{bmatrix} 0/3486 \\ -0/6152 \\ -0/3486 \\ 0/6152 \end{bmatrix}$$

بنابراین مقادیر ویژه و بردارهای ویژه مسأله مقدار ویژه درجه دو (۹-۸-۲) عبارتند از: -2 ، 1 ، -1 ، و -1

$$\begin{bmatrix} 0/4000 \\ 0/2000 \end{bmatrix}, \begin{bmatrix} 0/5000 \\ -0/5000 \end{bmatrix}, \begin{bmatrix} 0/6696 \\ 0/2273 \end{bmatrix}, \begin{bmatrix} 0/3486 \\ -0/6122 \end{bmatrix}$$

طریق ۲: ماتریس B و C را تشکیل می‌دهیم

$$B = \begin{bmatrix} 2 & 2 & 1 & 2 \\ 2 & 2 & 2 & 0 \\ 1 & 2 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix}$$

مقادیر ویژه و بردارهای ویژه مسأله مقدار ویژه تعمیم یافته $Bz = \lambda Cz$ عبارتند از:

$$1, -2, -1, -1 \text{ و } \begin{bmatrix} -0/5000 \\ 0/5000 \\ -0/5000 \\ 0/5000 \end{bmatrix}, \begin{bmatrix} -0/8000 \\ -0/4000 \\ 0/4000 \\ 0/2000 \end{bmatrix}, \begin{bmatrix} -0/5000 \\ -0/5000 \\ 0/5000 \\ 0/5000 \end{bmatrix}, \begin{bmatrix} 0/6747 \\ 0/2117 \\ -0/6747 \\ -0/2117 \end{bmatrix}$$

بنابراین مقدارهای ویژه مسأله مقدار ویژه درجه دو (۹-۸-۲) عبارتند از: 1 ، -2 ، -1 ، و -1 ؛ و

بردارهای ویژه به صورت زیر می‌باشند

$$\begin{bmatrix} -0/5000 \\ 0/5000 \end{bmatrix}, \begin{bmatrix} 0/4000 \\ 0/2000 \end{bmatrix}, \begin{bmatrix} 0/5000 \\ 0/5000 \end{bmatrix}, \begin{bmatrix} -0/6747 \\ -0/2117 \end{bmatrix}$$

۹-۹ مسائل مقدار ویژه تعمیم یافته متقارن برای ماتریسهای ساخت یافته و بزرگ

همان گونه که در چندین مطالعه موردی قبلی ملاحظه نمودیم، در وضعیتهای عملی زیادی ماتریسهای A و B ساخت یافته هستند: حالت‌های سه قطری و نواری کاملاً متداول می‌باشند. متأسفانه الگوریتم چولسکی، هنگامی که بر چنین مسائلی ساخت یافته‌ای اعمال می‌شود، اغلب تنگ بودن را از بین می‌برد. حتی اگر A و B نواری باشند ماتریس $C = L^{-1}A(L^T)^{-1}$ در حالت کلی، پر خواهد بود. بنابراین روشهای قطری‌سازی همزمان برای ماتریسهای بزرگ و تنگ عملی نیستند.

۹-۹-۱ روش دنباله استورم برای A و B سه قطری

اکنون روشی را بیان می‌کنیم که می‌تواند مورد استفاده قرار گیرد هنگامی که A و B هر دو سه قطری متقارن و B معین مثبت می‌باشد. روش تعمیمی از الگوریتم دنباله استورم برای مسأله مقدار ویژه متقارن توصیف شده در فصل ۸ می‌باشد و از شکلهای سه قطری A و B بهره می‌گیرد. فرض کنید

$$A = \begin{bmatrix} \alpha_1 & \beta_1 & & & \cdot \\ \beta_1 & \ddots & & & \\ & \ddots & \ddots & & \\ & & \ddots & \beta_{n-1} & \\ \cdot & & & \beta_{n-1} & \alpha_n \end{bmatrix}, \quad B = \begin{bmatrix} \alpha'_1 & \beta'_1 & & & \cdot \\ \beta'_1 & \ddots & & & \\ & \ddots & \ddots & & \\ & & \ddots & \beta'_{n-1} & \\ \cdot & & & \beta'_{n-1} & \alpha'_n \end{bmatrix}$$

دنباله چندجمله‌ایهای $\{p_r(\lambda)\}$ را به صورت زیر تعریف می‌کنیم

$$p_0(\lambda) = 1 \tag{۱-۹-۹}$$

$$p_1(\lambda) = \alpha_1 - \lambda\alpha'_1 \tag{۲-۹-۹}$$

$$p_r(\lambda) = (\alpha_r - \lambda\alpha'_r)p_{r-1}(\lambda) - (\beta_{r-1} - \lambda\beta'_{r-1})^2 p_{r-2}(\lambda) \tag{۳-۹-۹}$$

$$r = 2, 3, \dots, n$$

سپس می‌توان نشان داد (تمرین ۲۲) که این چندجمله‌ایها یک دنباله استورم تشکیل می‌دهند. سپس مقادیر ویژه و مداد $(A - \lambda B)$ توسط صفرهای $p_n(\lambda)$ ارائه می‌شوند. صفرها را می‌توان توسط روش تنصیف یا هر روش ریشه‌یابی مناسب دیگر برای چندجمله‌ایها پیدا کرد. برای اثبات الگوریتم کتاب Wilkinson (1965, 340-341) را ملاحظه کنید.

تذکر MATCOM: روش قبل در برنامه GENSTURM در MATCOM پیاده‌سازی شده است.

مثال ۹-۹-۱

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 5 \end{bmatrix}$$

$$p_0(\lambda) = 1, \quad p_1(\lambda) = 1 - \lambda$$

$$p_2(\lambda) = (2 - 4\lambda)(1 - \lambda) - (0 - \lambda)^2 \times 1 \\ = 2\lambda^2 - 6\lambda + 2$$

$$p_3(\lambda) = (3 - 5\lambda)(2\lambda^2 - 6\lambda + 2) - (-\lambda)^2(1 - \lambda) \\ = -14\lambda^3 + 38\lambda^2 - 28\lambda + 6$$

ریشه‌های $p_3(\lambda) = 0$ عبارتند از $0.3964, 0.6472, 1.6708$. می‌توان به سادگی امتحان کرد که اینها مقادیر ویژه تعمیم یافته مداد $(A - \lambda B)$ هستند.

محاسبه یک بردار ویژه

به محض این که یک مقدار ویژه محاسبه شود، بردار ویژه متناظر می‌تواند با استفاده از تکرار معکوس با بهره‌گیری کامل از شکلهای سه‌قطری A و B به صورت زیر محاسبه شود.

فرض کنید λ تقریب اولیه یک بردار ویژه متناظر با یک مقدار ویژه محاسبه شده λ باشد. آنگاه

For $i = 0, 1, 2, \dots$ do until convergence occurs.

Step 1: Solve for x_{i+1} : $(A - \lambda B)x_{i+1} = y_i$

Step 2: Form $y_{i+1} = Bx_{i+1}$

تبصره‌ها:

۱. در حدود دو تکرار برای هر بردار ویژه کافی می‌باشد.

۲. توجه کنید که دستگاه $(A - \lambda B)x_{i+1} = y_i$ را می‌توان با بهره‌گیری از شکلهای سه‌قطری A و B حل کرد.

۹-۹-۲ الگوریتم لانزوس برای مداد $A - \lambda B$

در این بخش نشان می‌دهیم که چگونه روش لانزوس متقارن توصیف شده در فصل ۸ برای یک ماتریس A می‌تواند به مسأله مقدار ویژه تعمیم یافته $Ax = \lambda Bx$ توسعه یابد. روش لانزوس که در اینجا توصیف می‌شود برای تخمین تعداد کمی از مقادیر ویژه اکسترمال یک مسأله مقدار ویژه تعمیم یافته معین متقارن تنک مفید است. به خاطر داشته باشید که مسائل مقدار ویژه تعمیم یافته در عمل متداول

هستند و مانند بیشتر مسائل عملی آنها تنک می‌باشند. روش لانزوس که فقط به ضربهای ماتریس- بردار وابسته است برای حفظ تنک بودن مناسب می‌باشد.
فرض کنید $B = LL^T$ تجزیه چولسکی B باشد.

الگوریتم ۹-۹-۱ الگوریتم لانزوس برای مداد معین متقارن. با مفروض بودن یک ماتریس متقارن A و یک ماتریس معین مثبت متقارن B ، الگوریتم لانزوس زیر یک ماتریس سه قطری متقارن

$$T_j = \begin{bmatrix} \alpha_1 & \beta_1 & & & \cdot \\ \beta_1 & \ddots & & & \\ & \ddots & \ddots & & \\ & & & \beta_{j-1} & \\ \cdot & & & \beta_{j-1} & \alpha_j \end{bmatrix}$$

و یک ماتریس متعامد $V_j = (v_1, v_2, \dots, v_j)$ به قسمی می‌سازد که

$$V_j^T A V_j = T_j$$

$$V_j^T B V_j = I_{j \times j}$$

Step 1: Choose v_1 such that $v_1^T B v_1 = 1$

Step 2: Set $\beta_0 = 1, r_0 = v_1, v_0 = 0$

Step 3: For $i = 1, 2, \dots, j$ do

If $i > 1$, solve: $Bv_i = r_{i-1}/\beta_{i-1}$

$$\alpha_i = v_i^T (A v_i - B v_{i-1})$$

$$r_i = A v_i - \beta_{i-1} B v_{i-1} - \alpha_i B v_i$$

$$\beta_i = \|L^{-1} r_i\|_2$$

تذکرها:

۱. بردارهای لانزوس در این حالت B - یکا متعامد هستند:

$$v_i^T B v_i = 1$$

$$v_i^T B v_j = 0, \quad i \neq j$$

۲. برای $j = n$:

$$V^T B V = I_{n \times n}$$

$$V^T A V = T_{n \times n}$$

یک ماتریس سه قطری متقارن

۳. اگر (θ, v) یک جفت مقدار ویژه T_j باشد، آنگاه (θ, V_j) جفت ریتز متناظر برای مداد (A, B) می باشد.

تبصره های محاسباتی

۱. توجه کنید که $L^{-1}r_i$ می تواند با حل دستگاه پایین مثلثی زیر محاسبه شود

$$Ly_1 = r_1$$

همچنین رابطه

$$Bv_1 = \frac{r_{i-1}}{\beta_{i-1}}$$

با رابطه زیر معادل است

$$L^T v_i = L^{-1} \left(\frac{r_{i-1}}{\beta_{i-1}} \right)$$

زیرا $B = LL^T$.

بنابراین برای پیاده سازی طرح به زیر برنامه های زیر نیاز داریم: (۱) یک زیر برنامه تجزیه چولسکی برای یک ماتریس تنک، (۲) زیر برنامه هایی برای دستگاه های مثلثی تنک، و (۳) یک زیر برنامه برای ضرب ماتریس - بردار.

۲. در هر تکرار، بردار βv_i برای استفاده در تکرار بعدی باید حفظ شود.

تذکر MATCOM: الگوریتم ۹-۹-۱ در برنامه LANSYMGN در MATCOM پیاده سازی شده است.

۹-۹-۳ برآورد مقادیر ویژه

حقیقت امر: برای z به اندازه کافی بزرگ مقادیر ویژه اکسترم ماتریسهای سه قطری

$$T_j = \begin{bmatrix} \alpha_1 & \beta_1 & & \cdot \\ \beta_1 & \ddots & \ddots & \\ & \ddots & \ddots & \beta_j \\ \cdot & & \beta_j & \alpha_j \end{bmatrix}$$

برخی مقادیر ویژه مداد $(A - \lambda B)$ را تقریب خواهند زد.

مثال ۹-۹-۲

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

$$B = \text{diag}(1, 2, 3)$$

$$L = \text{diag}(1, 1/4142, 1/7321)$$

$$\beta_0 = 1, \quad v_1 = (1, 0, 0)^T \quad r_0 = v_1$$

(توجه کنید که $v_1^T B v_1 = 1$.)

$i = 1$

$$\alpha_1 = v_1^T (A v_1 - \beta v_0) = 1$$

$$r_1 = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$$

$$\beta_1 = 2/2361$$

$i = 2$

$$v_2 = \begin{bmatrix} 0 \\ \cdot \\ \cdot/4472 \\ \cdot/4472 \end{bmatrix}$$

$$\alpha_2 = v_2^T (A v_2 - \beta v_1) = 3/2000$$

$$T_2 = \begin{bmatrix} \alpha_1 & \beta_1 \\ \beta_1 & \alpha_2 \end{bmatrix} = \begin{bmatrix} 1 & 2/2361 \\ 2/2361 & 3/2000 \end{bmatrix}$$

مقادیر ویژه T_2 عبارتند از $-0/3920$ ، $4/5919$. مقادیر ویژه (A, B) عبارتند از $4/6013$ ، $4/4347$ و 0 . (توجه کنید که بزرگترین مقدار ویژه T_2 ، $4/5919$ یک تقریب معقول از بزرگترین مقدار ویژه $4/6013$ مداد $(A - \lambda B)$ می باشد)

$i = 3$

$$v_3 = \begin{bmatrix} \cdot \\ -\cdot/5477 \\ -\cdot/3651 \end{bmatrix}$$

$$a_3 = v_3^T (Av_3 - \beta v_3) = \cdot/0.2323$$

بنابراین

$$T_3 = T = \begin{bmatrix} \alpha_1 & \beta_1 & \cdot \\ \beta_1 & \alpha_2 & \beta_2 \\ \cdot & \beta_2 & \alpha_3 \end{bmatrix} = \begin{bmatrix} 1 & 2/2361 & \cdot \\ 2/2361 & 3/2000 & \cdot/2449 \\ \cdot & \cdot/2449 & -\cdot/0.2323 \end{bmatrix}$$

مقادیر ویژه T عبارتند از -0.4347 ، 0.0000 ، و 0.4613 . و جفت‌های ویژه T_3 عبارتند از:

$$\left[-\cdot/3920, \begin{bmatrix} \cdot/8489 \\ -\cdot/5285 \end{bmatrix} \right] \text{ و } \left[4/5919, \begin{bmatrix} \cdot/5285 \\ \cdot/8489 \end{bmatrix} \right]$$

جفت‌های ریتز عبارتند از

$$\left[-\cdot/3920, \begin{bmatrix} \cdot/8489 \\ -\cdot/2363 \\ -\cdot/2363 \end{bmatrix} \right] \text{ و } \left[4/5919, \begin{bmatrix} \cdot/5285 \\ \cdot/3796 \\ \cdot/3796 \end{bmatrix} \right]$$

۹-۱۰ روش لانزوس برای مقادیر ویژه مداد $A - \lambda B$ در یک بازه

در کاربردهای زیادی نظیر ارتعاش و تحلیل سازه‌ای، ما اغلب علاقمند به محاسبه یک تعداد خاص از مقادیر ویژه در قسمت‌های خاصی از طیف هستیم. برای انجام این امر ما الگوریتم لانزوس را برای مسأله مقدار ویژه متقارن تعمیم یافته به طور اجمالی بیان می‌کنیم. الگوریتم توسط اریکسون و روه (Ericsson and Ruhe (1980)) ارائه شده است.

الگوریتم ۹-۱۰-۱ روش لانزوس برای مقادیر ویژه در یک بازه (الگوریتم اریکسون و روه). با مفروض بودن (A, B) متقارن و B معین مثبت متقارن، الگوریتم زیر مقادیر ویژه تعمیم یافته λ مداد $(A - \lambda B)$ را در یک بازه مشخص $[\alpha, \beta]$ محاسبه می‌کند. در الگوریتم از نماد گذارهای زیر استفاده می‌کنیم:

$$S(v, \delta) = (v, \delta) \text{ مجموعه مقادیر ویژه در } (v, \delta)$$

$$C(v, \delta) = (v, \delta) \text{ مجموعه مقادیر ویژه همگرا شده در } (v, \delta)$$

قرار دهید

$$\mu_1 = \alpha, \quad C = \phi \text{ (مجموعه تهی)}$$

برای $i = 1, 2, \dots$ تا برقراری رابطه $|C(\alpha, \beta)| = |S(\alpha, \beta)|$ انجام دهید.
گام ۱: $(A - \mu_i B)$ را تجزیه کنید:

$$(A - \mu_i B) = LDL^T$$

تعداد عناصر قطری منفی D را بشمارید. $|S(\alpha, \mu_i)|$ را برابر این تعداد قرار دهید.

گام ۲: یک بردار تصادفی v_1 به قسمی انتخاب کنید که $v_1^T B v_1 = 1$

گام ۳: الگوریتم لانزوس متقارن را در مورد ماتریس

$$W^T(A - \mu_i B)^{-1}W$$

اعمال کنید که در آن

$$B = WW^T$$

بردار آغازین لانزوس را برابر v_1 اختیار کنید.

گام ۴: فرض کنید v_1, v_2, \dots, v_r مقادیر ویژه همگرا شده از گام ۳ باشند. آنگاه $\frac{1}{v_s} + \mu_i$ را به ازای $s = 1, \dots, r$ به C اضافه کنید.

گام ۵: اگر $|C(\alpha, \mu_i)| < |S(\alpha, \mu_i)|$ ، آنگاه یک بردار جدید v_1 متعامد بر بردارهای ویژه همگرا شده اختیار کنید و گامهای ۳ و ۴ را تکرار کنید. در غیر این صورت یک انتقال جدید μ_{i+1} تعیین کنید و ادامه دهید.

تبصره‌ها: توجه کنید که در اعمال لانزوس متقارن بر ماتریس

$$W^T(A - \mu_i B)^{-1}W$$

ما به یک ضرب ماتریس - بردار به شکل

$$y = W^T(A - \mu_i B)^{-1}Wx$$

نیاز داریم که می‌تواند به صورت زیر به دست آید.

گام ۱: $z = Wx$ را محاسبه کنید.

گام ۲: دستگاه زیر را نسبت به p حل کنید:

$$(A - \mu_i B)p = z$$

این کار می‌تواند با استفاده از تجزیه محاسبه شده در گام ۱ الگوریتم انجام شود.

گام ۳: $y = W^T p$ را محاسبه کنید.

برای جزئیات بیشتر در مورد پیاده‌سازی و اثبات این الگوریتم مقاله Ericsson and Ruhe (1980)

را ملاحظه کنید.

۱۱-۹ مرور و خلاصه

این فصل به متداول‌ترین مسأله مقدار ویژه مطرح در مهندسی، یعنی مسأله مقدار ویژه تعمیم یافته حاوی دو ماتریس: $Ax = \lambda Bx$ ، اختصاص داده شده است.

۱. نتایج وجود: مشابه‌های شور و شور حقیقی مسأله مقدار ویژه معمولی برای مسأله مقدار ویژه تعمیم یافته نیز وجود دارند (قضیه‌های ۹-۲-۱ و ۹-۲-۲).

۲. الگوریتم QZ : مورد استفاده‌ترین الگوریتم برای مسأله مقدار ویژه تعمیم یافته الگوریتم QZ است که شکل شور حقیقی مسأله (A, B) را می‌سازد. الگوریتم در دو مرحله انجام می‌شود. در مرحله ۱ جفت (A, B) به یک جفت هسنبیگی مثلثی کاهش می‌یابد. در مرحله ۲ جفت هسنبیگی مثلثی با استفاده از تکرار ضمنی QR در مورد AB^{-1} ، باز هم به شکل شور حقیقی تعمیم یافته کاهش می‌یابد. B^{-1} هرگز به طور صریح تشکیل نمی‌شود. الگوریتم تقریباً به $15n^3$ عمل ممیز شناور (flops) نیاز دارد و پایدار پسر و است.

۳. مسأله مقدار ویژه تعمیم یافته متقارن: تقریباً همه مقادیر ویژه مطرح در مهندسی سازه و ارتعاش به شکل

$$Kx = \lambda Mx$$

می‌باشند که در آن M معین مثبت متقارن بوده و K متقارن می‌باشد. این مسأله مقدار ویژه تعمیم یافته معین مثبت متقارن نامیده می‌شود. به واسطه اهمیت این مسأله، در اینجا آن را با جزئیات بیشتر مورد مطالعه قرار دادیم.

چندین مطالعه موردی از مهندسی ارتعاش در بخش ۹-۶ ارائه شدند تا نشان دهند که چگونه این مسأله در کاربردهای عملی مهم مطرح می‌شود. این شامل موارد زیر است.

(الف) ارتعاش یک دستگاه جرم - فنر آزاد

(ب) ارتعاش یک ساختمان

(ج) ارتعاش هماهنگ اجباری یک دستگاه جرم - فنر

فرکانسهای طبیعی و دامنه‌های یک دستگاه مرتعش به ترتیب به مقادیر ویژه و بردارهای ویژه وابسته هستند. اگر فرکانس نیروی دوره‌ای تحمیل شده برابر یا تقریباً برابر با یکی از فرکانسهای طبیعی دستگاه شود، آنگاه تشدید رخ می‌دهد، و وضعیت کاملاً نگران‌کننده است.

فرو ریختن پل دره تا کوما درایالت واشنگتن و پل برایتن در انگلستان موجب می‌شوند که وابسته بودن به چنین پدیده‌ای را کاملاً باور داشته باشیم (فصل ۸ را ملاحظه کنید).

(الف) البته روش QZ می‌تواند برای حل یک مسأله مقدار ویژه معین مثبت متقارن استفاده شود. ولیکن به طور کلی هم تقارن و هم معین مثبت بودن از بین خواهند رفت.

(ب) یک روش حافظ تقارن الگوریتم QR چولسکی است. این دربخش ۹-۵-۱ توصیف شده است. دقت حاصل از این الگوریتم می‌تواند به طوری آسب ببیند اگر ماتریس B در

$$Ax = \lambda Bx$$

بد وضع باشد.

در این بخش یک تغییر از این روش را که اثر ویلکینسون بوده و کوچک‌ترین مقادیر ویژه را با دقت معقول محاسبه می‌کند، توضیح دادیم. روش یک شکل شور حقیقی مرتب از B به جای تجزیه چولسکی می‌سازد.

۴. قطری‌سازی همزمان و کاربردها: اساساً الگوریتم QR چولسکی برای مسأله معین متقارن

$$Ax = \lambda Bx$$

یک ماتریس نامفرد P می‌سازد که A و B را همزمان به شکلهای قطری با هم‌نهشتی تبدیل می‌کند.

$$P^TAP = \text{قطری}$$

$$P^TBP = I$$

این قطری‌سازی همزمان A و B نامیده می‌شود. در ارتعاش و سایر کاربردهای مهندسی این تجزیه تجزیه مدی نامیده می‌شود و ماتریس P یک ماتریس مدی نامیده می‌شود. روش قطری‌سازی همزمان روش خیلی مفیدی در مهندسی عملی است. کاربردهای آن شامل موارد زیر است.

(الف) جداسازی یک دستگاه معادلات دیفرانسیل مرتبه دو

$$M\ddot{y} + D\dot{y} + Ky = 0$$

به n معادله مستقل

$$\ddot{z}_i + (\alpha + \beta\omega_i^2)\dot{z}_i + \omega_i^2 z_i = 0, \quad i = 1, 2, \dots, n$$

اگر $DM^{-1}K = KM^{-1}D$ باشد.

(ب) ساختن یک مدل با مرتبه کاهش یافته از یک دستگاه خیلی بزرگ از دستگاههای مرتبه دو. تفکیک و کاهش مدل مسلماً راههای خیلی مفیدی برای راهاندازی یک دستگاه مرتبه دو بزرگ از معادلات دیفرانسیل می باشند. ولیکن متأسفانه روش قطری سازی همزمان برای مسائل بزرگ و تنگ عملی نیست. از طرف دیگر مسائل کاربردی زیادی نظیر طراحی ساختارهای فضایی بزرگ منجر به مسائل مقدار ویژه معین متقارن خیلی بزرگ و تنگ می شوند.

روش قطری سازی همزمان تقارن را حفظ می کند ولیکن سایر خاصیت‌های دیگر قابل بهره برداری موجود در داده‌های مسأله را از بین می برد. (توجه کنید که اغلب مسائل بزرگ عملی تنگ هستند و حفظ تنگ بودن یک امر مهم برای تدوین کننده‌های الگوریتمها برای صرفه جویی در ذخیره سازی و نیازهای محاسباتی است).

۵. مسأله مقدار ویژه درجه دو: مسأله مقدار ویژه $(\lambda^2 M + \lambda D + K)x = 0$ دربخش ۹-۸ بحث شد. مانشان دادیم که مسأله را می توان یا به یک مسأله نامتقارن $2n \times 2n$ استاندارد (۹-۸-۳) یا به یک مسأله مقدار ویژه تعمیم یافته متقارن $2n \times 2n$ (۹-۸-۷) کاهش داد. همچنین نشان دادیم که چگونه فرکانسها و نرخهای میرایی مدی یک دستگاه مرتعش را که توسط دستگاه مرتبه دو ارائه می شوند، استخراج کنیم به محض این که مقادیر ویژه به دست آمدند (معادلات ۹-۸-۵ و ۹-۸-۶).

۶. روشهای لانزوس و دنباله استورم: شرح مختصری از روش دنباله استورم برای مسأله مقدار ویژه تعمیم یافته با ماتریسهای سه قطری A و B ارائه دادیم. مسائل سه قطری معین متقارن در کاربردهای زیادی مطرح می شوند. همچنین یک توصیف خیلی خلاصه از روش لانزوس متقارن را برای مسأله مقدار ویژه تعمیم یافته ارائه دادیم.

این فصل با یک الگوریتم لانزوس که اثر اریکسون و روه بوده به پایان می رسد. این الگوریتم یک تعداد مشخص از مقادیر ویژه واقع در یک بازه را برای یک مسأله مقدار ویژه تعمیم یافته معین متقارن پیدا می کند (بخش ۹-۱۰).

۷. تحلیل حساسیت مسأله مقدار ویژه تعمیم یافته: در مورد این مطلب هیچ بحثی نکردیم. قطع نظر از جزئیات، اگر x و y بردارهای ویژه با طول واحد باشند به طوری که $Ax = \lambda Bx$ و $y^* = \lambda^* B^*$ ، آنگاه λ بدو وضع در نظر گرفته می شود اگر

$$(y^* Ax)^2 + (y^* Bx)^2$$

کوچک باشد. برای جزئیات مرجع (1978) Stewart را ملاحظه کنید.

تمرینها:

بخش ۹-۲

۱. نشان دهید که اگر A و B دارای یک بردار پوچ مشترک باشند، آنگاه معادله مشخصه تعمیم یافته متحد با صفر است:

$$\det(A - \lambda B) = 0.$$

۲. (الف) ثابت کنید که اگر A و B ماتریسهای $n \times n$ باشند، آنگاه $\det(A - \lambda B)$ یک چندجمله‌ای از درجه حداکثر n می‌باشد.

(ب) ثابت کنید که درجه $\det(A - \lambda B)$ برابر n است اگر و فقط اگر B نامنفرد باشد.

بخشهای ۹-۳ و ۹-۴

۳. نشان دهید که محاسبه ماتریس Q_1 در گام آغازین QZ فقط بستگی به معکوس کردن زیر ماتریس اصلی پیشرو 2×2 ماتریس مثلثی B دارد.

۴. نشان دهید که انتقالهای α_1 و α_2 در یک گام QZ ، که مقادیر ویژه زیر ماتریس اصلی 2×2 پایینی ماتریس $C = AB^{-1}$ می‌باشند، می‌توانند بدون تشکیل کامل B^{-1} حساب شوند. راهنمایی: محاسبه فقط به زیر ماتریس 3×3 پایینی راست B^{-1} بستگی دارد.

۵. با استفاده از قضیه Q ضمنی نشان دهید که ماتریس Q در گام QZ دارای همان سطر اول ماتریس Q_1 می‌باشد.

۶. تعداد اعمال ممیز شناور را برای فرایندهای زیر پیدا کنید.

(الف) کاهش هسبرگی - مثلثی در بخش ۹-۳-۱ با و بدون انباشتن ماتریسهای تبدیل.

(ب) یک تکرار از الگوریتم QZ (الگوریتم ۹-۳-۱).

(ج) کاهش (A, B) به شکل شور تعمیم یافته (الگوریتم ۹-۳-۲).

۷. یک الگوریتم مبتنی بر حذفی گاوس با محورگیری جزئی برای کاهش هسبرگی - مثلثی به دست آورید. تعداد اعمال ممیز شناور آن چقدر است؟

۸. بردارهای ویژه جفت (A, B) چگونه به بردارهای ویژه جفت به طور متعامد هم ارز (\tilde{A}, \tilde{B}) وابسته هستند، که در آن

$$\tilde{A} = UAV$$

$$\tilde{B} = UB^V$$

بخش ۹-۵

۹. الگوریتم QR چولسکی (الگوریتم ۹-۵-۱) مداد معین متقارن را با استفاده از تجزیه شور B دوباره انجام دهید.

$$V^T B V = D$$

۱۰. کاهش هسنبِریگی - مثلثی (A, B) را با B منفرد ملاحظه کنید، نشان دهید که در این حالت اگر بعد فضای پوچ AB برابر k باشد، آنگاه ساختار هسنبِریگی - مثلثی شکل زیر را اختیار می‌کند.

$$A = \begin{bmatrix} A_{11} & A_{12} \\ \cdot & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} \cdot & B_{12} \\ \cdot & B_{22} \end{bmatrix}$$

که در آن A_{11} یک ماتریس بالا مثلثی $k \times k$ و A_{22} یک ماتریس بالا هسنبِریگی، و B_{22} یک ماتریس $(n-k) \times (n-k)$ بالا مثلثی و نامنفرد هستند. چگونه این مطلب در فرایند کاهش به شکل شور تعمیم یافته می‌تواند کمک کند.

۱۱. تعداد اعمال ممیز شناور (flops) را برای الگوریتم QR چولسکی مداد معین متقارن پیدا کنید (الگوریتم ۹-۵-۱).

۱۲. فرض کنید A و B معین مثبت باشند. نشان دهید که الگوریتم برای قطری‌سازی همزمان (الگوریتم ۹-۵-۳) تقریباً به $7n^3$ عمل ممیز شناور (flops) نیاز دارد.

۱۳. (تکرار متعامد تعمیم یافته): الگوریتم تکراری زیر را که معروف به تکرار متعامد تعمیم یافته است، در نظر بگیرید.

(a) Choose an $n \times m$ orthonormal matrix Q_0 such that $Q_0^T Q_0 = I_{m \times m}$

(b) For $k = 0, 1, 2, \dots$ do

(i) Solve for $Z_k: BZ_k = AQ_{k-1}$

(ii) Find QR factorization of $Z_k: Z_k = Q_k R_k$

این الگوریتم را در مورد ماتریسهای زیر اعمال کنید.

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 5 \\ 1 & 4 & 5 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 1.0 & 1 & 1 & 1 \\ 1 & 1.0 & 1 & 1 \\ 1 & 1 & 1.0 & 1 \\ 1 & 1 & 1 & 1.0 \end{bmatrix}$$

۱۴. با مفروض بودن

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1.0 & 1 & 0 \\ 1 & 1.0 & 1 \\ 0 & 1 & 1.0 \end{bmatrix} \quad (\text{الف})$$

$$A = \begin{bmatrix} 1.0 & 1 & 1 \\ 1 & 1.0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix} \quad (\text{ب})$$

مقادیر ویژه و بردارهای ویژه را برای هر جفت بالا با استفاده از روشهای زیر پیدا کنید.

(i) تکرار QZ و در پی آن تکرار معکوس (الگوریتم ۹-۳-۲ و الگوریتم ۹-۴-۱).

(ii) تکرار خارج قسمت ریلی تعمیم یافته (الگوریتم ۹-۵-۴).

(iii) روشهای قطری سازی همزمان (الگوریتم ۹-۵-۳).

بخشهای ۹-۶ تا ۹-۹

۱۵. فرض کنید که یک پل با خرپا دارای یک فرکانس طبیعی 5 Hz باشد (از آزمایش قبلی معلوم می باشد). همچنین فرض کنید که بر اثر یک وسیله نقلیه 900000 کیلوگرمی در حدود 2 mm در وسط خم شود. فرکانسهای طبیعی پل و وسیله نقلیه را حساب کنید.

۱۶. برای معادله حرکت تحت نیروی $F_1 e^{i\omega t}$

$$m\ddot{y} + ky = F_1 e^{i\omega t}$$

دامنه حالت پایای ارتعاش را پیدا کنید و وضعیتی را که می تواند موجب تشدید شود، تعیین کنید.

۱۷. مسأله جرم - فنر شکل ۹-۱۱ را در نظر بگیرید

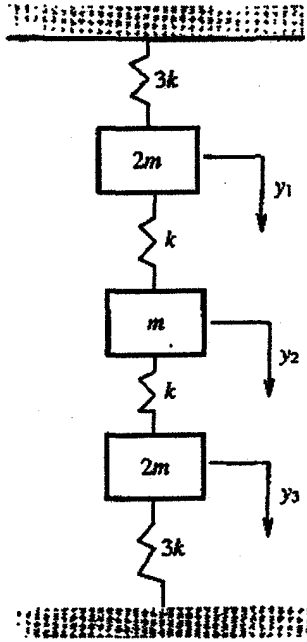
(الف) معادلات حرکت را تعیین کنید

(ب) مسأله مقدار ویژه تعمیم یافته را به شکل

$$Kx = \lambda Mx$$

بنویسید و سپس فرکانسهای طبیعی و مدهای ارتعاش را تعیین کنید.

(ج) شکل دستگاه را در هر مد پیدا کنید.



شکل ۹-۱۱

۱۸. ساختمان چهار طبقه رسم شده در شکل ۹-۱۲ را در نظر بگیرید: با مفروض بودن

$$m_1 = 1/0 \times 10^5 \text{ kg}$$

$$m_2 = 0/8 \times 10^5 \text{ kg}$$

$$m_3 = 0/5 \times 10^5 \text{ kg}$$

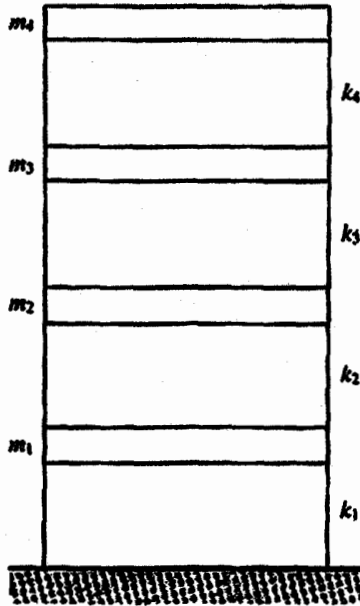
$$m_4 = 0/6 \times 10^5 \text{ kg}$$

$$k_1 = 15 \times 10^8 \text{ N/m}, k_2 = 12 \times 10^8 \text{ N/m}$$

$$k_3 = 15 \times 10^8 \text{ N/m}, k_4 = 10 \times 10^8 \text{ N/m}$$

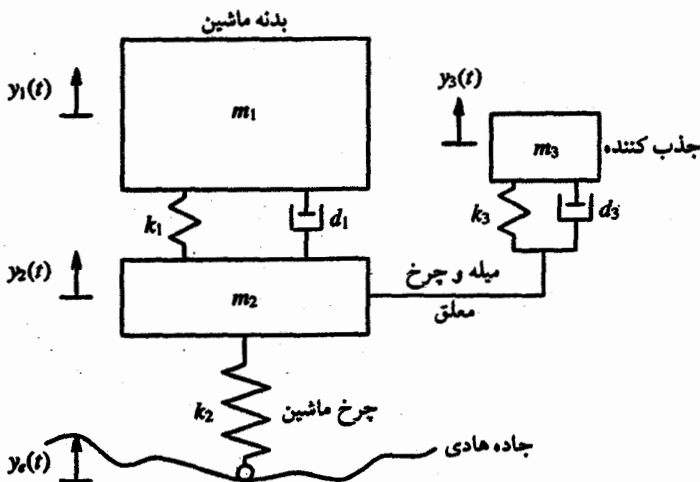
ماکزیمم دامنه هر طبقه را برای یک تغییر مکان افقی پایه ۳ میلی متری با یک دوره زمانی ۰/۲۵ ثانیه، با فرض شرایط اولیه صفر، پیدا کنید.

$$\ddot{y}_i = \frac{-4\pi^2 y_i}{T^2} \quad \text{راهنمایی:}$$



شکل ۹-۱۲

۱۹. دیاگرام زیر را در نظر بگیرید که یک اتومبیل معلق با جذب کننده ارتعاش می باشد (این مسأله از کتاب *Linear Vibrations* by P. C. Müller and W. O. Schiehlen, (1985, 226) گرفته شده است).



شکل ۹-۱۳ اتومبیل معلق با جذب کننده ارتعاش

با مفروض بودن

$$m_1 = 1200 \text{ kg}$$

$$m_2 = 80 \text{ kg}$$

$$m_3 = 20 \text{ kg}$$

$$k_1 = 300 \text{ N/cm}$$

$$k_2 = 3200 \text{ N/cm}$$

$$k_3 = 600 \text{ N/cm}$$

پاسخهای مختلف فرکانس دامنه را با مقادیر مختلف میرایی جذبکننده

$$d_3 = 0, 300, 600, 1000 \text{ Ns/m} \text{ (فرض کنید که } d_1 = 0 \text{)}$$

بخشهای ۸-۹ و ۹-۹

۲۰. مقادیر ویژه و بردار مداد درجه دو $(MA^2 + KA + D)x = 0$ را با استفاده از هر دو طریق ۱ و ۲ بخش ۸-۹ پیدا کنید و نتیجه‌ها را مقایسه کنید.

$$M = \text{diag}(1, 2, 3), \quad K = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

۲۱. معادلات (۵-۸-۹) و (۶-۸-۹) را برای فرکانسها و نرخ میرایی مدی به دست آورید.

۲۲. نشان دهید که چند جمله‌ایهای تعریف شده توسط (۱-۹-۹) تا (۳-۹-۹) تشکیل یک دنباله استورم می‌دهند.

۲۳. مقادیر ویژه را برای جفت (A, B) پیدا کنید که در آن

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 10 & 1 & 0 \\ 1 & 10 & 1 \\ 0 & 1 & 10 \end{bmatrix}$$

برای این منظور از روش دنباله استورم توصیف شده در بخش ۱-۹-۹ استفاده کنید.

۲۴. مقادیر ویژه و بردار جفت (A, B) تمرین ۱۳ را با استفاده از الگوریتم لانزوس متقارن (الگوریتم ۱-۹-۹) برآورد کنید.

۲۵. مقادیر ویژه و بردار جفت (A, B) تمرین ۲۳ را در بازه $(0, 0/3)$ پیدا کنید.

۲۶. ثابت کنید که مقادیر ویژه ومداد معین متقارن $A - \lambda B$ در بازه $[-\|B^{-1}A\|, \|B^{-1}A\|]$ قرار دارند. این نتیجه را در مورد جفت (A, B) تمرین ۲۳ به کار ببرید.

۲۷. مسأله پیدا کردن λ ها را به قسمی در نظر بگیرید که $\det(\lambda^k I + \lambda^{k-1} A_{k-1} + \dots + A_0)$ منفرد باشد، در اینجا A_0, A_1, \dots, A_{p-1} ماتریس هستند. نشان دهید که این مسأله معادل با پیدا کردن مقادیر ویژه ماتریس همراه بلوکی زیر است.

$$C = \begin{bmatrix} \cdot & I & \cdot & \dots & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \vdots & & \vdots & \vdots & \vdots & \vdots \\ \vdots & & \vdots & \vdots & \vdots & \vdots \\ \vdots & & \vdots & \vdots & \cdot & I \\ -A_0 & -A_1 & \dots & \dots & \dots & -A_{k-1} \end{bmatrix}$$

این ادعا را با ساختن یک مثال از خودتان با سه ماتریس A_0, A_1, A_2 و هر سه از مرتبه ۳ تحقیق کنید. (از MATLAB استفاده کنید).

برنامه‌های MATLAB و مسائل

تذکر: شما به **backsub** از MATCOM نیاز دارید.

۱. یک برنامه MATLAB به نام **hesstri** برای کاهش یک جفت ماتریس (A, B) به یک جفت هسنبرگی - مثلثی (H, T) بنویسید:

$$[H, T] = \text{hesstri}(A, B)$$

برنامه خود را با تولید تصادفی A و B هر کدام از مرتبه ۱۵، آزمایش کنید.

۲. (الف) یک برنامه MATLAB به نام **qztri** برای پیاده سازی یک تکرار الگوریتم **QZ** بنویسید:

$$[A1, B1] = \text{qztri}(A, B)$$

(ب) اکنون یک تکرار از **qritrdsi** (تکرار **QR** ضمنی با انتقال مضاعف از فصل ۸) را در مورد $C = AB^{-1}$ به کار ببرید:

$$[C] = \text{qritrdsi}(C)$$

(ج) ماتریس $D = A1 * B1^{-1}$ را محاسبه کنید که در آن $(A1, B1)$ جفت هسنبرگی - مثلثی

به دست آمده از گام (الف) می باشد.
 (د) C و D را برای تحقیق این که آنها یکی هستند، مقایسه کنید.

داده های آزمون

$A =$ یک ماتریس 15×15 بالا هسنبرگی کاهش ناپذیر و به طور تصادفی تولید شده
 $B =$ یک ماتریس 15×15 بالا مثلثی با همه عناصر برابر ۱، بجز پنج عنصر قطری آن و هر
 یک برابر 10^{-5}

۳. با استفاده از `lynsyspp` از فصل ۶ یا دستور `MATLAB` \backslash و `hesstri`، یک برنامه `MATLAB` به نام `invitrng` بنویسید تا یک بردار ویژه u متناظر با یک تقریب مفروض از یک مقدار ویژه λ از مداد (A, B) را حساب کند:

$$[u] = \text{invitrng}(A, B, \lambda, V_0)$$

برنامه خود را با استفاده از ماتریسهای A و B به طور تصادفی تولید شده، هر کدام از مرتبه ۱۵، آزمایش کنید، و سپس نتیجه را با آنچه که از اجرای دستور `MATLAB` زیر به دست می آورید، مقایسه کنید:

$$[U, D] = \text{eig}(A, B)$$

۴. (منظور از این تمرین مقایسه دقت راههای مختلف پیدا کردن مقادیر ویژه و بردارهای ویژه یک مداد معین متقارن $(K - \lambda M)$ است)

(الف) با استفاده از دستورهای `MATLAB` نظیر `chol`، `inv`، `eig` و \backslash (یا `backsub` از فصل ۳) یک برنامه `MATLAB` به نام `cholqr` بنویسید تا الگوریتم `QR` چولسکی را برای مداد معین متقارن $(K - \lambda M)$ (الگوریتم ۹-۵-۱) پیاده سازی کند.

$$[V_1, D_1] = \text{cholqr}(K, M)$$

در اینجا V_1 ماتریسی است که بردارهای ویژه را به صورت ستونهایش در بردارد و D_1 یک ماتریس قطری حاوی مقادیر ویژه مداد معین متقارن $(K - \lambda M)$ می باشد.

(ب) برنامه `eig(K, M)` از `MATLAB` را برای محاسبه $[V_2, D_2]$ اجرا کنید

$$[V2, D2] = \text{eig}(K, M)$$

(ج) برنامه $\text{eig}(\text{inv}(M) * K)$ از MATLAB را برای محاسبه $[V3, D3]$ اجرا کنید

$$[V3, D3] = \text{eig}(\text{inv}(M) * K)$$

(د) نتایج (الف)، (ب) و (ج) را مقایسه کنید.

داده‌های آزمون

$$M = \text{diag}(m, m, \dots, m)_{50 \times 50}$$

$$K = \begin{bmatrix} k & -k & \cdot & \cdot & \dots & \cdot \\ -k & 2k & -k & \cdot & \dots & \vdots \\ \cdot & \ddots & \ddots & \ddots & \ddots & \cdot \\ \vdots & \ddots & \ddots & \ddots & \ddots & \cdot \\ \vdots & \ddots & \ddots & \ddots & \ddots & -k \\ \cdot & \dots & \dots & \cdot & -k & 2k \end{bmatrix}_{50 \times 50}$$

$$m = 100$$

$$k = 10^8$$

۵. با استفاده از برنامه‌های chol ، inv ، eig از MATLAB یک برنامه MATLAB به نام simdiag بنویسید تا به طور همزمان یک جفت ماتریس (M, K) را قطری سازد، که در آن M معین مثبت متقارن و K متقارن هستند. (الگوریتم ۹-۵-۳):

$$(I, D) = \text{simdiag}(M, K)$$

داده‌های آزمون

از U و M مربوط به مطالعه موردی بر روی ارتعاش یک ساختمان در بخش ۹-۶-۲ استفاده کنید.

۶. با استفاده از simdiag ، مثال ۹-۷-۱ از کتاب را به طور کامل حل کنید.

۷. [منظور از این تمرین مقایسه دقت طریقه‌های مختلف حل مسأله مقدار ویژه درجه دو

$$(\lambda^2 M + \lambda D + K)x = 0$$

(الف) (طریق ۱ بخش ۸-۹) ماتریس

$$A = \begin{bmatrix} \cdot & I \\ -M^{-1}K & -M^{-1}D \end{bmatrix}$$

را با استفاده از دستورهای `eig`، `inv`، و `zeros` تشکیل دهید. سپس مقادیر ویژه و بردارهای

ویژه A را با استفاده از دستور `MATLAB`، `[V1, D1] = eig(A)`، محاسبه کنید.

(ب) (طریق ۲ از بخش ۸-۹) از برنامه `cholqr` برای محاسبه مقادیر ویژه استفاده کنید.

$$[V2, D2] = \text{cholqr}(A, B)$$

که در آن

$$A = - \begin{bmatrix} D & K \\ K & \cdot \end{bmatrix}, \quad B = \begin{bmatrix} M & \cdot \\ \cdot & K \end{bmatrix}$$

(ج) از دستور `MATLAB`، `[V3, D3] = eig(A, B)`، استفاده کنید که در آن ماتریسهای A و B به

صورت ارائه شده در قسمت (ب) می‌باشند.

نتیجه‌های این سه طریق را از نظر دقت، تعداد اعمال ممیز شناور، زمان سپری شده مقایسه

کنید. (مقادیر ویژه و بردارهای ویژه به دست آمده در (ج) را می‌توان به عنوان مقادیر ویژه و

بردارهای ویژه دقیق اختیار کرد.

داده‌های آزمون

از همان M و K ارائه شده در مسأله ۴ و $D = 10^{-4} \times M$ استفاده کنید.

۸. فرکانسهای طبیعی و نرخهای میرایی مدی را با استفاده از نتایج (۵-۸-۹) و (۶-۸-۹) کتاب و با M ، K

و D به همان صورت مسأله ۷ پیدا کنید.

۹. یک برنامه `MATLAB`، مبتنی بر الگوریتم ۹-۹-۲، به نام `lansymgn` بنویسید تا جفت (A, B)

معین متقارن را به (TJ, II) تبدیل کند، که در آن TJ یک ماتریس سه قطری متقارن $J \times J$ زبده و II

ماتریس واحد $J \times J$ زمی باشد، $n \leq J$.

$$[\alpha, \beta] = \text{lansymgn}(A, B)$$

در اینجا α یک بردار حاوی عناصر قطری TJ و β یک بردار حاوی عناصر خارج قطری TJ است.

داده‌های آزمون و آزمایش

اختیار کنید $M = K$ ، $B = M$ ، که در آن M و K همان ماتریسهای ارائه شده در تمرین ۷ هستند. برنامه **lansymgn** را با استفاده از $z = 10, 20, 30, 40, 50$ اجرا کنید. بگویید چگونه مقادیر ویژه اکسترم TJ برای هر مقدار z مقادیر ویژه مداد $(A - \lambda B)$ را تقریب می‌زنند. مقادیر ویژه را با استفاده از دستور MATLAB زیر پیدا کنید

$$[U, D] = \text{eig}(A, B)$$

تذکر: برخی از برنامه‌ها که از شما خواسته شده است بنویسید در MATCOM وجود دارند.

پیشنهادها برای مطالعه بیشتر

تقریباً همه کتابهای مربوط به ارتعاش و مهندسی سازه‌ای به طور صریح یا ضمنی بحث می‌کنند که چگونه مسائل مقدار ویژه تعمیم یافته در این کاربردها مطرح می‌شوند. برخی از این کتابهای معروف در نوشته‌های ارتعاش عبارتند از

Inman, Daniel J. *Vibration with Control, Measurement, and Stability*. Englewood Cliffs, N. J.: Prentice Hall, 1989.

Inman, Daniel J. *Engineering Vibration*. Englewood Cliffs, N. J.: Prentice Hall, 1994.

James, M. L., G. M. Smith, J. C. Wolford, and P. W. Whaley. *Vibration of Mechanical and Structural Systems*, 2d ed. New York: Harper & Row, 1989.

Müller, P. C., and W. O. Schiehlen. *Linear Vibrations*, Dordrecht, Netherlands: Martinus Nijhoff Publications, 1985.

Steidel, Robert F., Jr. *An Introduction to Mechanical Vibrations*. 2d ed. New York: Wiley, 1979.

Thomson, W. T. *Theory of Vibrations with Applications*, 4th ed. Englewood Cliffs, N. J.: Prentice Hall, 1992.

الگوریتم تکرار QZ با جزئیات بیشتر در کتابهای زیر بحث شده‌اند.

Golub, and Van Loan (1989, Chapter 7),

Stewart (1973, Chapter 7),

Watkins (1991, Chapter 5).

(مقاله اصلی (Moler and Stewart (1973) در این زمینه مقاله ارزشمندی است). طرز ارائه ما از این الگوریتم خیلی نزدیک به طرز ارائه آن در (Stewart (1973) است.

برای نتایجی در مورد تحلیل اختلال مسأله مقدار ویژه مراجع زیر را ملاحظه کنید.

Stewart (1978, 1979b), Boley (1990), Stewart and Sun (1990).

برای نظریه همگرایی تکرار خارج قسمت ریلی با یک بردار آغازین دلخواه x ، مرجع Parlett (1980, 319-320) و مقاله جدید (Pantazis and Szyld (1994) را ملاحظه کنید. برای کاربردهای مسأله مقدار ویژه تعمیم‌یافته مقارن در مورد مهندسی زمین لرزه، کتاب زیر را ملاحظه کنید. *Introduction to Earthquake Engineering* (Okamoto, 1984)

برای کاربردهای مسأله مقدار ویژه تعمیم‌یافته برای جواب معادلات ریگاتی جبری مراجع زیر را ملاحظه کنید.

Arnold and Laub (1984), Laub and Sandell (1980), Van Dooren (1981a, 1981b), Mehrmann (1988)

برای کاربردها در مورد دیگر مسائل کنترل مرجع (Inman (1989) را ملاحظه کنید.

روشی کاراتر از روش تکرار QR چولسکی برای محاسبه مقادیر ویژه تعمیم‌یافته یک مداد معین مقارن برای ماتریسهای نواری توسط کرافورد ((Crawford (1973) ارائه شده است. روش کرافورد در آخرین نسخه LAPACK پیاده‌سازی شده است. کارهای جدید نظیر (Wang and Zhao (1991) و (Kaufman (1993) را نیز ملاحظه کنید. برای نظریه گرشگورین مسأله مقدار ویژه تعمیم‌یافته مرجع (Stewart (1975) را ملاحظه کنید.

ویلیامز و لوب ((Williams and Laub (1992) مثلثی‌سازی همزمان ماتریسهای M ، D ، و K دستگاه مرتبه دو $M\ddot{y} + D\dot{y} + Ky = 0$ را ملاحظه نمودند.

برای جنبه‌های متعادل‌سازی مسأله مقدار ویژه تعمیم‌یافته مرجع (Ward (1981) را ملاحظه کنید. فصل ۱۵ کتاب پارلت *The Symmetric Eigenvalue Problem* (1980b)، یک منبع غنی از دانستنیها در زمینه مسائل مقدار ویژه تعمیم‌یافته است. همچنین مراجع (Saad (1992) و

Stewart (1976a) را ملاحظه کنید.

الگوریتم لانزوس (الگوریتم ۹-۹-۱) در حالتی که B متفرد است نیاز به فرمول‌بندی دوباره دارد. مرجع (Saad (1992, 298-299) را برای یک چنین فرمول‌بندی دوباره از الگوریتم لانزوس ملاحظه کنید (الگوریتم ۹-۳). برای الگوریتم‌های لانزوس بلوکی برای مسائل مقدار ویژه تعمیم یافته مرجع Grimes, Lewis, and Simon (1994) را ملاحظه کنید. کارهای بیشتری بر روی الگوریتم ۹-۱۰-۱ از زمانی که این الگوریتم ارائه گردیده، انجام شده است: به طور برجسته در آنها ماتریس B لازم نیست تجزیه پذیر باشد. مرجع (Grimes, Lewis, and Simon (1994) را برای جزئیات بیشتر ملاحظه کنید. همچنین مرجع (Ruhe (1994) را ملاحظه کنید.

برای اطلاعات بیشتر در مورد تجزیه شور تعمیم یافته یک مداد دلخواه مرجع (Demmel and Kagström (1993) را ملاحظه کنید.

مسأله مقدار ویژه درجه دو (QEP) در کتابها کمتر بحث شده است. ولیکن مسائل سازه‌ای بسیاری منجر به QEP می‌شوند. شیوه نرمال این است که ماتریس میرای D نادیده گرفته می‌شود یا فرض می‌شود میرایی تناسبی است، و سپس مسائل مقدار ویژه تعمیم یافته حاصل حل می‌شوند. برای یک روش لانزوس برای پیدا کردن یک تعداد کمی مقادیر ویژه اکستریمال یک QEP مرجع (Parlett and Chen (1990) را ملاحظه کنید. در مرجع (Datta and Rincon (1993) یک الگوریتم برای پایدارسازی یک دستگاه کنترل مدل شده توسط یک معادله دیفرانسیل ارائه شده است و این الگوریتم نیاز به شناخت صریح مقادیر ویژه و بردارهای ویژه مداد درجه دو وابسته ندارد. مرجع (Datta (1994) را در این زمینه ملاحظه کنید.

تجزیه مقدار تکین (SVD)

۱-۱۰ مقدمه

قبلاً دیده‌ایم که هر ماتریس A را می‌توان به شکل

$$A = U\Sigma V^*$$

تجزیه کرد که در آن U و V یکانی هستند و Σ یک ماتریس شبه قطری است. اگر A حقیقی باشد، U و V متعامد حقیقی هستند. ما فرض خواهیم کرد که A در طول این فصل حقیقی باشد. این تجزیه مقدار تکین یا SVD ماتریس A نامیده می‌شود. SVD به طور نزدیکی به تجزیه ویژه AA^T وابسته می‌باشد. SVD دارای یک تاریخ طولانی و جذاب است. حداقل پنج ریاضیدان مشهور و کلاسیک به نامهای ای. بلترامی^۲ (۱۸۳۵-۱۸۹۹)، سی. جردن^۳ (۱۸۲۱-۱۸۳۸)، جی. سیلوستر^۴ (۱۸۹۷-۱۸۱۴)، ای. اشمیت^۵ (۱۸۷۶-۱۹۵۹)، اچ. ویل^۶ (۱۸۸۵-۱۹۵۵) در توسعه نظریه SVD نقش داشته‌اند. برخی جزئیات سهم این ریاضیدانها را در SVD می‌توان در مقاله جالب و جدید (1993b) G. W. Stewart پیدا کرد. همچنین کتاب (1991) Horn and Johnson حاوی یک تاریخ زیبا از SVD است.

1. The Singular Value Decomposition

3. C. Jordan

5. E. Schmidt

2. E. Beltrami

4. J. Sylvester

6. H. Weyl

در سالهای اخیر SVD یک وسیله محاسباتی کارا برای حل طیف وسیعی از مسائل مطرح در بسیاری از کاربردهای عملی شده است. راز استفاده از SVD در این کاربردها در این حقیقت است که این کاربردها نیاز به شناخت رتبه یک ماتریس، تقریبهای یک ماتریس با استفاده از ماتریسهای با رتبههای کمتر، پایههای یکا متعامد برای فضاها سطری و ستونی یک ماتریس همچنین برای متممهای متعامد آنها، و تصویرهای متعامد بر روی زیر فضاها دارند. در بیشتر حالتها این محاسبات باید تحت وجود ناخالصیها در دادهها (که نوفه نامیده می شوند) انجام شوند، همان گونه که در بخشهای بعد در این فصل خواهیم دید SVD برای این محاسبات خیلی مؤثر است. ولیکن توانایی به طور مؤثر راه اندازی این مسائل با استفاده از SVD مسلماً بستگی به این دارد که چگونه SVD به طور مؤثر بتواند محاسبه شود. از این رو جوامع مهندسی و علوم کاربردی به مقدار زیادی مرهون آنالیزدانان معاصر نظیر ژن اچ. گولوب، ویلیام کاهان، و سی. رینسک هستند که به ما نشان داده اند که چگونه SVD را به یک طریق کارا و پایدار عددی محاسبه کنیم.

این فصل به مطالعه SVD اختصاص داده شده است. خاصیتهای مهم، برخی از کاربردها، روشهای محاسباتی موجود به صورت زیر مورد بحث قرار خواهند گرفت. قضیه وجود SVD (بخش ۱۰-۲)، یک رابطه بین مقادیر تکین و مقادیر ویژه بخش (۱۰-۳)، یک کاربرد از SVD در نوار قلبی جنین (ECG) (بخش ۱۰-۴)، حساسیت مقادیر تکین (بخش ۱۰-۵)، کاربردهای SVD در مسائل وابسته به رتبه، محاسبه پایههای یکا متعامد، تقریبهای ماتریسی، و غیره (بخش ۱۰-۶)، محاسبه ماتریس وارپانس - کوواریانس (بخش ۱۰-۷)، استفاده از SVD در مسائل کمترین توانهای دوم و شبه معکوس (بخش ۱۰-۸)، الگوریتم گولوب - کاهان - رینسک، و الگوریتم دمل - کاهان (بخش ۱۰-۹)، و SVD تعمیم یافته بخش (۱۰-۱۰).

بازار مهم زیر و مفاهیم قبلاً توسعه یافته برای یادگیری و فهم روان مطالب این فصل مورد نیاز خواهند بود: مفاهیم نرم و نتایج وابسته (بخش ۱-۷)، مفهوم رتبه (بخش ۱-۳-۳)، پایههای یکا متعامد (بخشهای ۱-۳-۶ و ۱-۳-۶) و تصویرها (بخشهای ۱-۳-۶ و ۱-۳-۷)، ایجاد صفر با ماتریسهای هاوس هولدر و گینوز (بخشهای ۵-۴ و ۵-۵)، تجزیه QR یک ماتریس نامنفرد (بخش ۵-۴-۲)، گزارهها و روشها برای مسائل کمترین توانهای دوم و شبه معکوس (فصل ۷)، نتایج اساسی در مورد مقادیر ویژه و بردارهای ویژه (بخش ۸-۲)، و قضیه بوئر - فایک (قضیه ۸-۷-۱) و خاصیت اختلال کلی مسأله مقدار ویژه متقارن (قضیه ۸-۱۱-۲)، و تکرار QR ضمنی (بخش ۸-۹-۶).

۲-۱۰ قضیه تجزیه مقدار تکین

قضیه ۱۰-۲-۱ فرض کنید A یک ماتریس حقیقی $m \times n$ باشد، آنگاه ماتریسهای متعامد V و U وجود دارند به قسمی که

$$U^T A V = \begin{bmatrix} \Sigma_1 & \cdot \\ \cdot & \cdot \end{bmatrix} = \Sigma \quad (1-2-10)$$

که در آن Σ_1 یک ماتریس قطری نامنفرد است. عناصر قطری Σ همگی نامنفی هستند و می‌توانند به ترتیب ناصعودی مرتب شوند. تعداد عناصر قطری مخالف صفر Σ برابر رتبه ماتریس A است.

اثبات: ماتریس $A^T A$ را در نظر بگیرید. این ماتریس یک ماتریس $n \times n$ نیمه معین مثبت متقارن است؛ بنابراین مقادیر ویژه آن نامنفی هستند. مقادیر ویژه $A^T A$ را به $\lambda_1 = \sigma_1^2, \lambda_2 = \sigma_2^2, \dots, \lambda_n = \sigma_n^2$ نمایش می‌دهیم. فرض کنید که $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ و

$$\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_n = 0$$

قبلاً دیده‌ایم که یک ماتریس متقارن دارای یک مجموعه بردارهای ویژه یکا متعامد است (قضیه ۸-۲-۸). مجموعه بردارها ویژه یکا متعامد $A^T A$ متناظر با λ_1 تا λ_n را به v_1, \dots, v_n نمایش می‌دهیم؛ یعنی v_1 تا v_n یکا متعامد هستند و در رابطه زیر صدق می‌کنند

$$A^T A v_i = \sigma_i^2 v_i, \quad i = 1, \dots, n$$

پس

$$v_i^T A^T A v_i = \sigma_i^2 \neq 0, \quad i = 1, \dots, r \quad (2-2-10)$$

و

$$v_i^T A^T A v_j = 0, \quad i = 1, \dots, r; j \neq i \quad (3-2-10)$$

می‌نویسیم

$$V_1 = (v_1, \dots, v_r)$$

$$V_2 = (v_{r+1}, \dots, v_n)$$

که در آن v_1 تا v_r بردارهای ویژه متناظر با مقادیر ویژه مخالف صفر λ_1 تا λ_r و v_{r+1}, \dots, v_n بردارهای ویژه متناظر با مقادیر ویژه صفر هستند. پس

$$V_2^T A^T A V_2 = V_2^T A^T A (v_{r+1}, v_{r+2}, \dots, v_n) = V_2^T (0, 0, \dots, 0) = 0$$

این ایجاب می‌کند که $AV_r = 0$ ، یا

$$Av_k = 0, \quad k = r + 1, r + 2, \dots, n \quad (۴-۲-۱۰)$$

اکنون یک مجموعه بردارهای مخالف صفر $\{u_i\}$ را به صورت زیر تعریف می‌کنیم

$$u_i = \frac{1}{\sigma_i} Av_i, \quad i = 1, \dots, r \quad (۵-۲-۱۰)$$

بردارهای u_i ، $i = 1, \dots, r$ ، یک مجموعه متعامد تشکیل می‌دهند، زیرا

$$\begin{aligned} u_i^T u_j &= \frac{1}{\sigma_i} (Av_i)^T \frac{1}{\sigma_j} (Av_j) \\ &= \frac{1}{\sigma_i \sigma_j} (v_i^T A^T Av_j) \\ &= \begin{cases} 0 & \text{اگر } i \neq j \\ 1 & \text{اگر } i = j \end{cases} \end{aligned} \quad (۶-۲-۱۰)$$

تعریف می‌کنیم $U_1 = (u_1, \dots, u_r)$ و $U_2 = (u_{r+1}, \dots, u_m)$ را به قسمی انتخاب می‌کنیم که $U = (U_1, U_2)$ متعامد باشد. سپس برای هر $k > r$ ، داریم:

$$u_k^T Av_i = \sigma_i u_k^T u_i = 0, \quad i = 1, \dots, r \quad (Av_i = \sigma_i u_i \text{ داریم } (۵-۲-۱۰))$$

(بر طبق متعامد بودن بردارهای U)، و بر طبق (۴-۲-۱۰) داریم:

$$u_k^T Av_i = 0, \quad i = r + 1, \dots, n$$

قرار می‌دهیم $V = (V_1, V_2)$ آنگاه

$$U^T AV = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_m^T \end{bmatrix} A(v_1, \dots, v_n) = \begin{bmatrix} \frac{1}{\sigma_1} v_1^T A^T \\ \frac{1}{\sigma_2} v_2^T A^T \\ \vdots \\ \frac{1}{\sigma_r} v_r^T A^T \\ u_{r+1}^T \\ \vdots \\ u_m^T \end{bmatrix} A(v_1, \dots, v_n) \quad (۷-۲-۱۰)$$

$$= \begin{bmatrix} \frac{1}{\sigma_1} \cdot \sigma_1^2 & & & & \\ & \frac{1}{\sigma_2} \cdot \sigma_2^2 & & & \\ & & \ddots & & \\ & & & \frac{1}{\sigma_r} \cdot \sigma_r^2 & \\ & & & & \cdot \\ & & & & & \cdot \end{bmatrix} = \Sigma = \begin{bmatrix} \Sigma_1 & \cdot \\ \cdot & \cdot \end{bmatrix} \quad (۱۰-۲-۱۰)$$

که در آن $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_r)$.

□ حکم مربوط به رتبه واضح است، زیرا $\text{rank}(A) = \text{rank}(U\Sigma V^T) = \text{rank}(\Sigma) = r$.

تجزیه $A = U\Sigma V^T$ به تجزیه مقدار تکین (SVD) ماتریس A معروف است.

تعریف ۱۰-۲-۱۰ عناصر قطری ماتریس Σ مقادیر تکین ماتریس A نامیده می‌شوند. اعداد $\sigma_1, \sigma_2, \dots, \sigma_r$ مقادیر تکین مثبت A هستند.

تعریف ۱۰-۲-۲ ستونهای U بردارهای تکین چپ و ستونهای V بردارهای تکین راست نامیده می‌شوند.

منحصر به فرد بودن تجزیه مقدار تکین

به تعداد $k = \min(m, n)$ مقدار تکین برای A وجود دارد. فرض کنید رتبه A برابر r باشد. پس r مقدار تکین مثبت وجود دارد. اینها ریشه‌های دوم مثبت مقادیر ویژه مخالف صفر $A^T A$ (یا AA^T) هستند. بقیه $(k - r)$ مقدار تکین، اگر $k < r$ ، صفر هستند. بنابراین مقادیر تکین منحصر به فرد هستند ولیکن بردارهای تکین منحصر به فرد نیستند. برای مثال اگر A دارای یک مقدار تکین چندگانه $\sigma > 0$ باشد، آنگاه ستونهای ماتریس V می‌تواند به صورت هر پایهٔ یکا متعامد از فضای تولید شده توسط بردارهای ویژه وابسته به مقدار ویژهٔ چندگانه $\sigma^2 = \lambda$ از $A^T A$ انتخاب شود.

مثال ۱۰-۲-۱۰ مقادیر تکین و بردارهای تکین ماتریس زیر را پیدا کنید.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}$$

۱. محاسبه مقادیر تکین. مقادیر ویژه $A^T A$ عبارتند از $۴۲/۸۶۰۰$ و $۰/۱۴۰۰$. بنابراین مقادیر

تکین عبارتند از:

$$\sigma_1 = \sqrt{۴۲/۸۶۰۰} = ۶/۵۴۶۸, \quad \sigma_2 = \sqrt{۰/۱۴۰۰} = ۰/۳۷۴۲$$

۲. محاسبه V . ماتریس V_1 بردارهای ویژه وابسته به مقادیر ویژه $A^T A$ برابر است با

$$V_1 = \begin{bmatrix} 0/5696 & -0/8219 \\ 0/8219 & 0/5696 \end{bmatrix}$$

چون $r = 2$ ، داریم $V = V_1$.

۳. محاسبه U .

$$u_1 = \frac{1}{\sigma_1} A v_1 = \begin{bmatrix} 0/3381 \\ 0/5506 \\ 0/7632 \end{bmatrix} \quad u_2 = \frac{1}{\sigma_2} A v_2 = \begin{bmatrix} 0/8480 \\ 0/1735 \\ -0/5009 \end{bmatrix}$$

u_3 را به قسمی اختیار می‌کنیم که $U = (u_1, u_2, u_3)$ یکانی باشد.

$$u_3 = \begin{bmatrix} 0/4082 \\ -0/8165 \\ 0/4082 \end{bmatrix}$$

۴. ماتریسهای U ، Σ ، و V تعریف کننده SVD ماتریس A به صورت زیر ارائه می‌شوند

$$U = \begin{bmatrix} 0/3381 & 0/8480 & 0/4082 \\ 0/5506 & 0/1735 & -0/8165 \\ 0/7632 & -0/5009 & 0/4082 \end{bmatrix}_{3 \times 3}$$

$$V = \begin{bmatrix} 0/5696 & -0/8219 \\ 0/8219 & 0/5696 \end{bmatrix}_{2 \times 2} \quad \Sigma = \begin{bmatrix} 6/5468 & \cdot \\ \cdot & 0/3742 \\ \cdot & \cdot \end{bmatrix}$$

یک قرارداد

برای بقیه این فصل فرض خواهیم کرد، بدون از دست دادن کلیت مسأله، که $m \geq n$ ، زیرا اگر $m < n$ باشد، SVD ماتریس A^T را ملاحظه خواهیم کرد، و اگر SVD ماتریس A^T برابر $U \Sigma V^T$ باشد، آنگاه SVD ماتریس A برابر $V \Sigma^T U^T$ خواهد بود.

ما همچنین این قرارداد را دنبال خواهیم کرد که مقادیر تکین به ترتیب ناصعودی باشند. بنابراین $\sigma_1 = \sigma_{\max}$ بزرگ‌ترین مقدار تکین و $\sigma_n = \sigma_{\min}$ کوچک‌ترین مقدار تکین هستند. بعلاوه ما مجموعه مقادیر تکین A را توسط $\sigma(A)$ نمایش خواهیم داد.

۱۰-۳ یک رابطه بین تجزیه مقدار تکین و تجزیه‌های ویژه

در اثبات قضیه ۱۰-۲-۱ از این حقیقت استفاده کردیم که مقادیر تکین A ریشه‌های دوم نامنفی مقادیر ویژه $A^T A$ هستند (در حقیقت همان گونه که آنها تعریف شده‌اند). قضیه زیر نشان می‌دهد که چگونه SVD ماتریس A به تجزیه‌های ویژه AA^T و $A^T A$ وابسته است.

قضیه ۱۰-۳-۱ فرض کنید $A = U\Sigma V^T$ تجزیه مقدار تکین یک ماتریس A از مرتبه $m \times n$ باشد ($m \leq n$) فرض کنید r رتبه ماتریس A باشد. آنگاه

$$V^T(A^T A)V = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2, 0, \dots, 0)_{(n \times n)} \quad ۱.$$

$$U^T(AA^T)U = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2, 0, \dots, 0)_{(m \times m)} \quad ۲.$$

اثبات:

$$\begin{aligned} A^T A &= (V\Sigma^T U^T)(U\Sigma V^T) \\ &= V\Sigma^T \Sigma V^T \quad (۱-۳-۱۰) \\ &= V\Sigma' V^T \end{aligned}$$

که در آن Σ' یک ماتریس قطری $n \times n$ با $\sigma_1^2, \dots, \sigma_r^2, 0, \dots, 0$ به عنوان عناصر قطری می‌باشد.

بنابراین

$$\begin{aligned} V^T A^T A V &= \Sigma' \quad (۲-۳-۱۰) \\ &= \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2, 0, \dots, 0)_{(n \times n)} \end{aligned}$$

به طور مشابه

$$\begin{aligned} AA^T &= U\Sigma\Sigma^T U^T \quad (۳-۳-۱۰) \\ &= \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2, 0, \dots, 0)_{(m \times m)} \end{aligned}$$

□

تذکر: از قضیه ۱۰-۳-۱ فوراً نتیجه می‌شود که

۱. بردارهای تکین راست v_1, v_2, \dots, v_n بردارهای ویژه ماتریس $A^T A$ هستند.
۲. بردارهای تکین چپ u_1, u_2, \dots, u_m بردارهای ویژه ماتریس AA^T هستند.
۳. $\sigma_1^2, \dots, \sigma_r^2$ مقادیر ویژه مخالف صفر $A^T A$ و AA^T هستند.

مثال ۱۰-۳-۱ فرض کنید

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

SVD ماتریس A عبارت است از

$$A = U\Sigma V^T = \begin{bmatrix} 0/5774 & -0/8165 & 0/0000 \\ 0/5774 & 0/4082 & -0/7071 \\ 0/5774 & 0/4082 & 0/7071 \end{bmatrix} \begin{bmatrix} 2/4495 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0/7071 & -0/7071 \\ 0/7071 & 0/7071 \end{bmatrix}$$

به سادگی می توان تحقیق کرد که $V^T A^T A V = \text{diag}(6, 0)$ و $U^T A A^T U = \text{diag}(6, 0, 0)$

نتیجه ۱۰-۳-۱ فرض کنید A یک ماتریس متقارن با مقادیر ویژه $\lambda_1, \lambda_2, \dots, \lambda_n$ باشد. آنگاه مقادیر تکین A عبارتند از $|\lambda_i|, i = 1, \dots, n$.

اثبات: چون

$$A = A^T \\ A^T A = A^2$$

بنابراین، از قضیه ۱۰-۳-۱ ملاحظه می کنیم که n مقدار تکین A ریشه های دوم نامنفی n مقدار ویژه A^2 هستند. چون مقادیر ویژه A^2 مقادیر $\lambda_1^2, \dots, \lambda_n^2$ هستند، پس نتیجه مورد نظر را خواهیم داشت. □
نتیجه ۱۰-۳-۲ یک ماتریس A از مرتبه $n \times n$ نامنفرد است اگر و فقط اگر همه مقادیر تکین آن مخالف صفر باشند.

اثبات: می دانیم که $\det(A^T A) = (\det(A))^2$. بنابراین A نامنفرد است اگر و فقط اگر $A^T A$ نامنفرد باشد. چون یک ماتریس نامنفرد است اگر و فقط اگر مقادیر ویژه اش همگی مخالف صفر باشند، اکنون اثبات بلافاصله از قضیه ۱۰-۳-۱ نتیجه می شود. □

قضیه ۱۰-۳-۲ فرض کنید A یک ماتریس $m \times n$ ($m \geq n$) باشد و فرض کنید C یک ماتریس $(m+n) \times (m+n)$ تعریف شده به صورت زیر باشد

$$C = \begin{bmatrix} \cdot_{m \times m} & * A \\ A^T & \cdot_{n \times n} \end{bmatrix}$$

فرض کنید $\sigma_1, \sigma_2, \dots, \sigma_n$ مقادیر تکین A باشند. آنگاه مقادیر ویژه C مقادیر $\sigma_1, \sigma_2, \dots, \sigma_n, -\sigma_1, -\sigma_2, \dots, -\sigma_n$ و صفر با مرتبه تکرار $m - n$ خواهند بود.

اثبات: فرض کنید

$$A = U\Sigma V^T$$

با نوشتن

$$U = (U_1(m \times m) \quad U_2(m \times (m-n)))$$

$$\Sigma = \begin{bmatrix} \Sigma & & \\ & 1(n \times n) & \\ & & \cdot \\ & & & (m-n) \times n \end{bmatrix}$$

و تعریف ماتریس

$$P = \begin{bmatrix} \tilde{U}_1 & -\tilde{U}_1 & U_2 \\ \tilde{V} & \tilde{V} & \cdot \\ & & (m-n) \times n \end{bmatrix}$$

که در آن

$$\tilde{U}_1 = \frac{1}{\sqrt{2}} U_1$$

و

$$\tilde{V} = \frac{1}{\sqrt{2}} V$$

آنگاه به سادگی ملاحظه می‌کنیم

$$P^T C P = \begin{bmatrix} \Sigma_1 & \cdot & \cdot \\ \cdot & -\Sigma_1 & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$$

که نشان می‌دهد که مقادیر ویژه C عبارتند از مقادیر $\sigma_1, \sigma_2, \dots, \sigma_n, -\sigma_1, -\sigma_2, \dots, -\sigma_n$ و صفر با مرتبه تکرار $m - n$ خواهند بود.

□

مثال ۱۰-۳-۲ فرض کنید

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}$$

آنگاه با استفاده از مثال ۱۰-۲-۱ داریم:

$$P = \begin{bmatrix} 0/2391 & 0/5996 & -0/2391 & -0/5996 & 0/4082 \\ 0/3894 & 0/1227 & -0/3894 & -0/1227 & -0/8165 \\ 0/5397 & -0/3542 & -0/5397 & 0/3542 & 0/4082 \\ 0/4082 & -0/5812 & 0/4082 & -0/5812 & 0 \\ 0/5812 & 0/4028 & 0/5812 & 0/4082 & 0 \end{bmatrix}$$

$$P^T C P = \text{diag}(6/5468, 0/3742, -6/5468, -0/3742, 0)$$

مقادیر ویژه مثبت C عبارتند از $6/5468$ ، $0/3742$ ، اینها مقادیر تکین مخالف صفر A هستند. (مثال ۱۰-۲-۱ را ملاحظه کنید).

۱۰-۴ کاربردهای SVD

SVD یک وسیله مؤثر در راه‌اندازی مسائل مهم گوناگون مطرح در زمینه‌های کاربردی با تنوع وسیع، نظیر نظریه کنترل، پردازش سیگنال و تصویر، هویت‌شناسی و برآورد، ساخت گفتار، الگوشناسی، تحلیل سری زمانی، نظریه شبکه الکتریکی و مهندسی زیست پزشکی می‌باشد.

جنبه‌های نظریه کنترل و مسائل هویت‌شناسی که نیاز به استفاده از SVD دارند شامل مسائل مربوط به کنترل‌پذیری و مشاهده‌پذیری، تحقق مدل‌های فضای حالت، کنترل H - نامتناهی، متعادل‌سازی، پایدارسازی بازخورد قوی، کاهش مدل، و غیره می‌باشند.

در پردازش سیگنال و گفتار SVD می‌تواند به عنوان یک فیلتر در نظر گرفته شود که برآوردی از یک سیگنال را از داده‌های نوفه تولید کند. برای مثال، هنگامی که یک شخص صحبت می‌کند، در حدود ۵۰٪ از زمان گفتاری موجود نیست. بنابراین اگر نوفه‌های زمینه وجود داشته باشند که از یک پنکه، یک ماشین مرتعش، و غیره بیایند، آنگاه این اختلالها در سیگنال میکروفن به هنگام وجود نداشتن گفتار، غالب خواهند بود. در چنین وضعیتی نسبت سیگنال گفتار بر نوفه‌های زمینه باید زیادت‌ر باشد، و SVD می‌تواند استفاده شود و به طور مؤثری این کار را انجام دهد.

در کاربردهای پردازش تصویر، SVD به طور معمول برای فشرده‌سازی تصویر، برای برداشتن خشها در یک عکس، و غیره استفاده می‌شود. برای جزئیات بیشتر مرجع Jain (1989) را ملاحظه کنید.

در مهندسی زیست پزشکی SVD یک نقش مهم در به دست آوردن یک نوار معنی‌دار از قلب جنین (FECG) از روی نوار قلب مادر (MECG) بازی می‌کند. مرجع Vandewalle and De Moor (1988) را ملاحظه کنید.

بعلاوه SVD مؤثرترین وسیله در حل مسائل کمترین توانهای دوم و مسائل کمترین توانهای دوم

تعمیم یافته است. در این بخش یک کاربرد زندگی واقعی از SVD را در مهندسی زیست پزشکی ارائه می دهیم.

نوار قلب جنین (FECG) (از مرجع Callaerts, De Moor, Vandewalle and Sansen(1990) گرفته شده است). مسأله گرفتن نوار قلب یک جنین را با جایگذاری الکترودهای پوستی در قلب و شکم مادر حامله در نظر بگیرید. نوار قلب مادر (MECG) به طور واضح مشاهده نوار قلب جنین را از بین می برد، زیرا نقش سیگنالهای قلب مادر به مراتب قویتر از نقش سیگنالهای قلب جنین است. بنابراین هدف آشکار نمودن نوار قلب جنین (FECG) خواهد بود در حالی که به طور همزمان از نوار قلب مادر در خصوص نوفه ها جلوگیری به عمل آید.

فرض کنید p سیگنال سنجش $m_1(t), \dots, m_p(t)$ وجود داشته باشند. فرض کنید این سنجشها در یک بردار به نام بردار سنجش $m(t)$ به طور مرتب قرار گیرند.

$$m(t) = (m_1(t), m_2(t), \dots, m_p(t))^T \quad (1-4-10)$$

فرض کنید r سیگنال منبع $s_1(t), \dots, s_r(t)$ وجود داشته باشند که به طور مرتب در بردار سیگنال منبع $s(t)$ قرار گیرند.

$$s(t) = (s_1(t), s_2(t), \dots, s_r(t))^T \quad (2-4-10)$$

به طور واضح سیگنالهای سنجش توسط یک سیگنال نوفه جمعی تباہ می شوند، و یک رابطه بین سیگنالهای سنجش و سیگنالهای منبع وجود دارد. می توان فرض کرد که این رابطه خطی است (مرجع (Vandewalle and De Moor(1988))، و در حقیقت هر سیگنال سنجش $m_i(t)$ را می توان به صورت ترکیب خطی از r سیگنال منبع $s_j(t)$ و سیگنال نوفه جمعی $n_i(t)$ نوشت. این منجر به معادلات زیر می شود

$$\begin{aligned} m_1(t) &= t_{11}s_1(t) + t_{12}s_2(t) + \dots + t_{1r}s_r(t) + n_1(t) \\ m_2(t) &= t_{21}s_1(t) + t_{22}s_2(t) + \dots + t_{2r}s_r(t) + n_2(t) \\ &\vdots \\ m_p(t) &= t_{p1}s_1(t) + t_{p2}s_2(t) + \dots + t_{pr}s_r(t) + n_p(t) \end{aligned} \quad (3-4-10)$$

یا

$$m(t) = Ts(t) + n(t) \quad (4-4-10)$$

که در آن

$$T = (t_{ij}), \quad \text{و} \quad n(t) = (n_1(t), n_2(t), \dots, n_r(t))^T \quad (5-4-10)$$

ماتریس T ماتریس انتقال نامیده می‌شود و بستگی به هندسه بدن، وضعیت‌های الکترودها و منبعها و میزان رساناییهای بافتهای بدن دارد.

اکنون مسأله به دست آوردن یک برآورد از سیگنالهای منبع $s(t)$ با شناخت فقط $m(t)$ است، و از آن برآورد، جداسازی برآورد سیگنالهای منبع چنین خواهد بود.

فرض کنید هر سنجش از q نمونه تشکیل شود. پس سنجشها می‌توانند در یک ماتریس از مرتبه $p \times q$ ذخیره شوند. شکل معادله ماتریسی (۴-۴-۱۰) به صورت $M = TS + N$ است.

اکنون نشان می‌دهیم که SVD ماتریس M می‌تواند برای به دست آوردن برآوردهای سیگنالهای منبع مورد استفاده قرار گیرد. فرض کنید SVD ماتریس M به شکل

$$M = U\Sigma V^T \quad (6-4-10)$$

باشد. آنگاه ماتریس \hat{S} از مرتبه $p \times q$ و تعریف شده به صورت

$$\hat{S} = U^T M \quad (7-4-10)$$

شامل p برآورد از سیگنالهای منبع خواهد بود. سپس نیاز خواهیم داشت که برآوردهای سیگنالهای منبع چنین را از \hat{S} استخراج کنیم؛ اجازه دهید این را \hat{S}_F بنامیم.

ماتریس مقدار تکین Σ ماتریس M را به صورت زیر افزایش می‌کنیم.

$$\Sigma = \begin{bmatrix} \Sigma_M & \cdot & \cdot \\ \cdot & \Sigma_F & \cdot \\ \cdot & \cdot & \Sigma_\cdot \end{bmatrix} \quad (8-4-10)$$

در اینجا Σ_M حاوی r_m بزرگترین مقادیر تکین وابسته به قلب مادر است، Σ_F حاوی r_f مقدار تکین است، که آنها کوچکترین مقادیر تکین بوده و وابسته به قلب چنین می‌باشند، Σ_\cdot حاوی بقیه مقادیر تکین و وابسته به سروصداها و غیره می‌باشند.

فرض کنید $U = (U_M, U_F, U_\cdot)$ یک افزایش متوافق از U باشد. آنگاه به طور واضح داریم:

$$\hat{S} = U^T M = \begin{bmatrix} U_M^T \\ U_F^T \\ U_\cdot^T \end{bmatrix} M \quad (9-4-10)$$

$$= \begin{bmatrix} U_M^T M \\ U_F^T M \\ U_{\cdot}^T M \end{bmatrix} = \begin{bmatrix} \hat{S}_M \\ \hat{S}_F \\ \hat{S}_{\cdot} \end{bmatrix}$$

بنابراین $\hat{S}_F = U_F^T M$.

به محض این که \hat{S}_F تعیین شود، می‌توانیم همچنین یک ماتریس F حاوی فقط سهمهای جنین در هر سیگنال اندازه‌گیری شده را به شکل زیر بسازیم.

$$F = U_F \hat{S}_F = \sum_{i=r_m+1}^{r_m+r_j} \sigma_i u_i v_i^T$$

در اینجا u_i و v_i ستونهای U و V هستند و σ_i ، i امین مقدار تکین M است. سیگنالها در \hat{S}_F سیگنالهای اصلی جنین نامیده می‌شوند.

روش قبلی خودکار شده است، و یک الگوریتم تطبیقی و متصل به شبکه برای محاسبه ماتریس U طراحی شده است. برای جزئیات بیشتر روش و نتایج آزمون، مقاله Callaerts, De Moor, Vandewalle and Sansen(1990) را ملاحظه کنید. توجه کنید که اگر SVD ماتریس M به صورت

$$M = U \Sigma V^T \\ = (U_1, U_2) \begin{bmatrix} \Sigma_1 & \cdot \\ \cdot & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

ارائه شود، آنگاه S می‌تواند توسط $\hat{S} = U_1^T M$ برآورد شود.

۵-۱۰ حساسیت مقادیر تکین

چون توانهای دوم مقادیر تکین A دقیقاً مقادیر ویژه ماتریس متقارن حقیقی $A^T A$ هستند و قبلاً ملاحظه کرده‌ایم که مقادیر ویژه یک ماتریس متقارن نسبت به اختلالهای کوچک حساس نیستند، همان وضع را باید همچنین برای مقادیر تکین انتظار داشت.

در حقیقت نتیجه مهم زیر برقرار است. این نتیجه مشابه نتیجه بیان شده در مورد حساسیت مقادیر ویژه یک ماتریس متقارن در فصل ۸ (قضیه ۸-۱۱-۲) می‌باشد. اساساً نتیجه بیان می‌کند که، نظیر مقادیر ویژه یک ماتریس متقارن، مقادیر تکین یک ماتریس خوش - وضع هستند.

قضیه ۱۰-۵-۱ قضیه اختلال برای مقادیر تکین. فرض کنید A و $B = A + E$ دو ماتریس

$m \times n$ ($m \geq n$) باشند. فرض کنید σ_i ، $i = 1, \dots, n$ و $\tilde{\sigma}_i$ ، $i = 1, \dots, n$ به ترتیب مقادیر تکین ماتریسهای A و B و به ترتیب ناصعودی باشند. آنگاه به ازای هر i داریم $\|\tilde{E}\|_p$:
 اثبات: اثبات بر رابطه جالب بین مقادیر تکین ماتریس A و مقادیر ویژه ماتریس متقارن زیر (نمایش داده شده در قضیه ۱۰-۳-۲) مبتنی است.

$$\tilde{A} = \begin{bmatrix} \cdot & A \\ A^T & \cdot \end{bmatrix}$$

در آنجا نشان داده شد که مقادیر ویژه ماتریس \tilde{A} عبارتند از $\sigma_1, \dots, \sigma_k, -\sigma_1, \dots, -\sigma_k$ ، که در آن σ_1 تا σ_k مقادیر تکین مخالف صفر A هستند. البته بقیه مقادیر ویژه \tilde{A} صفر هستند. با تعریف

$$\tilde{B} = \begin{bmatrix} \cdot & B \\ B^T & \cdot \end{bmatrix}, \quad \tilde{E} = \begin{bmatrix} \cdot & E \\ E^T & \cdot \end{bmatrix}$$

آنگاه داریم $\tilde{B} - \tilde{A} = \tilde{E}$.

مقادیر ویژه \tilde{B} و \tilde{E} به ترتیب به مقادیر تکین B و E به همان طریقی وابسته هستند که مقادیر ویژه \tilde{A} به مقادیر تکین A وابسته هستند. اکنون نتیجه قضیه ۱۰-۵-۱ از به کاربردن قضیه ۱۱-۱-۲ در مورد \tilde{A}, \tilde{B} و \tilde{E} بلافاصله حاصل می‌گردد. □

مثال ۱۰-۵-۱

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}, \quad E = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot / \dots 2 \end{bmatrix} \quad (10-5-1)$$

مقادیر تکین A عبارتند از

$$\sigma_1 = 14/5576$$

$$\sigma_2 = 1/0372$$

$$\sigma_3 = 0/0000$$

(۱۰-۵-۲)

مقادیر تکین $A + E$ عبارتند از

$$\tilde{\sigma}_1 = 14/5576$$

$$\tilde{\sigma}_2 = 1/0372$$

$$\tilde{\sigma}_3 = 0/0000$$

$$\|E\|_F = 0.0002 \quad (3-5-10)$$

$$|\tilde{\sigma}_1 - \sigma_1| \cong 10^{-4} < \|E\|_F$$

$$|\tilde{\sigma}_2 - \sigma_2| \cong < \|E\|_F$$

$$|\tilde{\sigma}_3 - \sigma_3| \cong < \|E\|_F$$

اکنون یک نتیجه (بدون اثبات) در مورد اختلال مقادیر تکین ارائه می‌دهیم که از نرم فروبینوس به جای نرم دو استفاده می‌کند. نتیجه قضیه ۱۰-۵-۲ بعداً در به دست آوردن یک نتیجه در مورد نزدیک بودن A به یک ماتریس با رتبه پایین‌تر استفاده می‌شود. اثبات شبیه به اثبات قضیه ۱۰-۵-۱ است. مرجع (Horn and Johnson (1985, 419) را ملاحظه کنید.

قضیه ۱۰-۵-۲ فرض کنید $A, E, \sigma_i, \tilde{\sigma}_i$ و $i = 1, \dots, n$ همان مقادیر ارائه شده در قضیه ۱۰-۵-۱ باشند، آنگاه

$$\sqrt{\sum_{i=1}^n (\tilde{\sigma}_i - \sigma_i)^2} \leq \|E\|_F$$

۱۰-۶ تجزیه مقدار تکین و ساختار یک ماتریس

SVD می‌تواند به طور خیلی مؤثری برای محاسبه برخی خواص مهم وابسته به ساختار یک ماتریس، نظیر رتبه آن (در حقیقت، نزدیک بودن به رتبه - ناقص، که یک موضوع عملی مهم است)، نرم اقلیدسی، عدد وضعیت، پایه یکا متعامد برای فضای پوچ و برد آن استفاده شود. این کاربردها را در این بخش بحث خواهیم کرد.

۱۰-۶-۱ نرمها و عدد وضعیت

قضیه ۱۰-۶-۱ فرض کنید $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ مقدار تکین یک ماتریس A از مرتبه $m \times n$ باشند. آنگاه

$$\|A\|_F = \sigma_1 = \sigma_{\max} \quad 1.$$

$$\|A\|_F = (\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2)^{\frac{1}{2}} \quad 2.$$

$$\|A^{-1}\|_F = \frac{1}{\sigma_n} = \frac{1}{\sigma_{\min}} \quad 3.$$

۴. اگر A یک ماتریس $n \times n$ نامنفرد باشد، آنگاه $\text{cond}_F(A) = \|A\|_F \|A^{-1}\|_F = \frac{\sigma_{\max}}{\sigma_{\min}}$

اثبات:

$$\|A\|_r = \|U\Sigma V^T\|_r = \|\Sigma\|_r = \max_i(\sigma_i) = \sigma_1 \quad ۱.$$

(توجه کنید که نرم ۲ و نرم فروبینوس تحت ضربهای ماتریسی متعامد پایا هستند).

$$\|A\|_F = \|U\Sigma V^T\|_F = \|\Sigma\|_F = (\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2)^{\frac{1}{2}} \quad ۲.$$

۳. با اختیار کردن SVD ماتریس A^{-1} ، و با توجه به قضیه ۱۰-۲-۱ که بزرگترین مقدار تکین

A^{-1} برابر $\frac{1}{\sigma_n}$ می باشد، نتیجه از ۱ حاصل می شود. (توجه کنید که اگر A معکوس پذیر باشد، آنگاه $\sigma_n \neq 0$).

۴. از تعریف $\text{cond}_r(A)$ و استفاده از ۱ و ۳ نتیجه مورد نظر حاصل می شود.

تذکر: برای هر ماتریس A از مرتبه $m \times n$ ، تعریف می کنیم $\text{cond}_r(A) = \frac{\sigma_{\max}}{\sigma_{\min}}$. هنگامی که A از رتبه ناقص است، $\sigma_{\min} = 0$ ، و می گوییم که $\text{cond}(A)$ بی نهایت است.

مثال ۱۰-۶-۱

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 6 & 7 & \sqrt{999} \end{bmatrix}$$

$$\sigma_1 = 14/5570,$$

$$\sigma_2 = 1/0.375$$

$$\sigma_3 = 0/0001$$

$$\|A\|_r = \sigma_1 = 14/5570 \quad ۱.$$

$$\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2} = 14/5940 \quad ۲.$$

$$\text{cond}(A) = \sigma_1/\sigma_3 = 1/0.993 \times 10^5 \quad ۳.$$

۱۰-۶-۲ پایه های یکا متعامد و تصویرها

در فصل ۵ ما نشان دادیم که چگونه پایه های یکا متعامد را برای برد و فضای پوچ یک ماتریس A با استفاده از تجزیه QR بسازیم. در این بخش ما نشان می دهیم که چگونه بردارهای تکین می توانند برای ساختن این پایه ها استفاده شوند.

فرض کنید $A = U\Sigma V^T$ تجزیه مقدار تکین یک ماتریس $m \times n$ باشد. فرض کنید $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ مقادیر تکین مثبت A باشند، آنگاه از قضیه ۱۰-۲-۱ (قضیه SVD) داریم:

$$Av_i = \sigma_i \mu_i, \quad i = 1, \dots, r$$

(۱۰-۶-۱۰)

$$Av_i = 0, \quad i = r + 1, \dots, n$$

به طور مشابه با اختیار کردن SVD ماتریس $A^T = V\Sigma^T U^T$ داریم:

$$A^T u_i = \sigma_i v_i, \quad i = 1, \dots, r$$

$$A^T u_i = 0, \quad i = r + 1, \dots, m$$

بنابراین

$$R(A) = \text{span}\{u_1, \dots, u_r\}$$

$$N(A) = \text{span}\{v_{r+1}, \dots, v_n\}$$

$$R(A^T) = \text{span}\{v_1, \dots, v_r\}$$

$$N(A^T) = \text{span}\{u_{r+1}, \dots, u_m\}$$

ستونهای U متناظر با مقادیر تکین مخالف صفر A یک پایه یکا متعامد برای $R(A)$ تشکیل می‌دهند؛ ستونهای U متناظر با مقادیر تکین صفر یک پایه یکا متعامد برای $N(A^T)$ تشکیل می‌دهند، و غیره.

تصویرهای متعامد

وقتی که پایه‌های یکا متعامد برد $R(A)$ و فضای پوچ $N(A)$ ماتریس A به دست آیند. تصویرهای متعامد می‌توانند به سادگی محاسبه شوند. بنابراین اگر U و V را به صورت

$$U = (U_1, U_2), \quad V = (V_1, V_2)$$

افراز کنیم، که در آنها U_1 و V_1 از r ستون اول U و V تشکیل می‌شوند، آنگاه می‌توان تصویرها را به صورت زیر محاسبه کرد.

محاسبه تصویرها با استفاده از SVD

$$R(A) \text{ تصویر بر روی } = U_1 U_1^T \quad ۱.$$

$$N(A) \text{ تصویر بر روی } = V_2 V_2^T \quad ۲.$$

$$N(A^T) \text{ تصویر بر روی } = U_2 U_2^T \quad ۳.$$

$$R(A^T) \text{ تصویر بر روی } = V_1 V_1^T \quad ۴.$$

مثال ۱۰-۶-۲

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{pmatrix}$$

$$\sigma_1 = 14/5576,$$

$$\sigma_2 = 1/0.372,$$

$$\sigma_3 = 0.$$

$$U = \begin{pmatrix} 0.2500 & 0.8371 & 0.4867 \\ 0.4852 & 0.3267 & -0.8111 \\ 0.8378 & -0.4379 & 0.3244 \end{pmatrix}$$

$$V = \begin{pmatrix} 0.4625 & -0.7870 & 0.4082 \\ 0.5706 & -0.0882 & -0.8165 \\ 0.6786 & 0.6106 & 0.4082 \end{pmatrix}$$

یک پایه برای فضای پوچ A عبارت از

$$V_2 = \begin{bmatrix} 0.4082 \\ -0.8165 \\ 0.4082 \end{bmatrix}$$

یک پایه یکا متعامد برای برد A عبارت از

$$U_1 = \begin{bmatrix} 0.2500 & 0.8371 \\ 0.4852 & 0.3267 \\ 0.8370 & -0.4379 \end{bmatrix}$$

(اکنون خودتان چهار تصویر را حساب کنید.)

تذکر MATCOM: برنامه ORTHPROJ در MATCOM تصویرهای قبل را حساب می‌کند.

۱۰-۶-۳ رتبه و رتبه - ناقص یک ماتریس

پیدا کردن رتبه یک ماتریس $m \times n$ ، بویژه تعیین نامنفرد بودن یک ماتریس مربعی کارهای خیلی مهمی هستند که اغلب در جبر خطی و کاربردهای مهم بسیاری مطرح می‌شوند. البته واضح‌ترین و کم هزینه‌ترین راه برای تعیین رتبه یک ماتریس مثلثی سازی ماتریس با استفاده از اعمال حذفی گاوس و سپس پیدا کردن رتبه ماتریس بالا مثلثی کاهش یافته است. ولیکن متأسفانه این یک راه خیلی قابل اطمینان در حساب ممیز شناور نیست. در عمل مهم‌تر است، همان‌گونه که خواهیم دید، که تعیین کنیم

که آیا ماتریس مفروض نزدیک به یک ماتریس با یک رتبه معین است، و بویژه بدانیم که آیا یک ماتریس از رتبه کامل نزدیک به یک ماتریس رتبه - ناقص است. روش حذفی گاوس که از تبدیلات مقدماتی استفاده می‌کند، ممکن است بر اثر خطای گردکردن عددی یک ماتریس با رتبه - ناقص را به یک ماتریس با رتبه کامل تبدیل کند. برای مثال یک ماتریس $n \times n$ بالا مثلثی با ۱ در امتداد قطر اصلی و ۱- در سایر جاها در نظر بگیرید. در فصل ۶ دیدیم که $\text{cond}_{\infty}(A) = 2^{n-1}$. بنابراین A خیلی نزدیک به یک ماتریس منفرد است هنگامی که n بزرگ است. ولیکن با نگاه کردن به این ماتریس هیچ اطلاعی در مورد آن نمی‌توان به دست آورد.

راه دیگر برای تعیین رتبه و رتبه ناقص که قابل اطمینان‌تر از روش حذفی گاوس است، تجزیه QR با محورگیری ستونی می‌باشد. همان‌گونه که در فصل ۵ دیدیم، این روش یک ماتریس جایگشت P و یک ماتریس متعامد Q به قسمی می‌سازد که

$$Q^TAP = \begin{bmatrix} R_1 & R_2 \\ \cdot & \cdot \end{bmatrix} \quad (10-6-2)$$

و R_1 بالا مثلثی و نامنفرد و بعد r ماتریس R_1 برابر رتبه ماتریس A است. ولیکن نتیجه فوق به طور قابل اطمینانی نزدیکی A را به یک ماتریس رتبه ناقص تعیین نمی‌کند. برای جزئیات بیشتر، بحثهای مربوطه به آشکارسازی رتبه توسط تجزیه QR را در فصل ۵ ملاحظه کنید.

چون تعداد مقادیر تکین مخالف صفر رتبه یک ماتریس را تعیین می‌کند، می‌توانیم بگوئیم که یک ماتریس A به طور دلخواه نزدیک به یک ماتریس رتبه کامل است: فقط هر مقدار تکین صفر را با یک عدد کوچک ϵ عوض کنید. بنابراین معنای بیشتری خواهد داشت اگر بدانیم که یک ماتریس نزدیک به یک ماتریس از یک رتبه معین است تا این که بدانیم رتبه آن چند است. تجزیه مقدار تکین به طور دقیق جواب این سؤال را می‌دهد.

قابل اطمینان‌ترین راه برای تعیین رتبه و نزدیکی به یک ماتریس رتبه ناقص استفاده از SVD می‌باشد. فرض کنید A دارای رتبه r باشد، یعنی $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ و $\sigma_{r+1} = \dots = \sigma_n = 0$. آنگاه سؤال این است: چقدر A از یک ماتریس با رتبه k ، $k < r$ ، فاصله دارد. قضیه‌های زیر می‌توانند برای پاسخ دادن به این سؤال مورد استفاده قرار گیرند. قضیه ۱۰-۶-۲ معمولاً معروف به قضیه اکارت - یانگ ((Eckart-Young, 1939)) است.

قضیه ۱۰-۶-۲ فرض کنید $A = U\Sigma V^T$ SVD ماتریس A باشد، و فرض کنید $\text{rank}(A)$ برابر

• $r > 2$ باشد. فرض کنید $k \leq r$. ماتریس

$$A_k = U \Sigma_k V^T$$

را تعریف می‌کنیم که در آن

$$\Sigma_k = \left[\begin{array}{ccc|ccc} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & & & & \\ & & & \sigma_k & & \\ \hline & & & & & \end{array} \right], \quad (\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > 0)$$

آنگاه (۱) $\text{rank}(A_k) = k$. (۲) فرض کنید B یک ماتریس دلخواه از رتبه k باشد، آنگاه $\|B - A\|_F \geq \|A_k - A\|_F$.

اثبات: حکم مربوط به رتبه واضح است، زیرا

$$\text{rank}(A_k) = \text{rank}(U \Sigma_k V^T) = \text{rank}(\Sigma_k) = k$$

برای اثبات قسمت دوم، نشان می‌دهیم که اگر B هر ماتریس دلخواه دیگر از رتبه k باشد، آنگاه

$$\|B - A\|_F^2 \geq \|A_k - A\|_F^2$$

برای دیدن این مطلب، بیاید مقادیر تکین B را به $\phi_1 \geq \phi_2 \geq \dots \geq \phi_n$ نمایش دهیم. آنگاه چون B دارای رتبه k است داریم $\phi_{k+1} = \phi_{k+2} = \dots = \phi_n = 0$. سپس بر طبق قضیه ۱۰-۵-۲ داریم:

$$\square \quad \|B - A\|_F^2 \geq \sum_{i=1}^n |\phi_i - \sigma_i|^2 \geq \sigma_{k+1}^2 + \dots + \sigma_n^2 = \|A_k - A\|_F^2$$

اکنون با استفاده از نرم ۲ نتیجه‌ای مشابه را ارائه می‌دهیم. روند اثبات با جزئی اختلاف نظیر اثبات ۱۰-۶-۲ می‌باشد.

قضیه ۱۰-۶-۳ فرض کنید A و A_k همان ماتریسهای قضیه ۱۰-۶-۲ باشند. A_k دارای رتبه k است.

۲. فاصله A_k از A (بر حسب نرم ۲): $\|A - A_k\|_2 = \sigma_{k+1}$.

۳. میان همه ماتریسهای رتبه k ، A_k نزدیک‌ترین به A است، یعنی

$$\min_{\text{rank}(B)=k} \|A - B\|_2 = \|A - A_k\|_2$$

اثبات ۱: واضح است که $\text{rank}(A_k) = k$ ؛ زیرا $\text{rank}(A_k) = \text{rank}(U \Sigma_k V^T) = \text{rank}(\Sigma_k) = k$

(توجه کنید که $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > 0$.)

اثبات ۲: چون $A - A_k = U(\Sigma - \Sigma_k)V^T$ داریم

$$\|A - A_k\|_F = \|U(\Sigma - \Sigma_k)V^T\|_F = \|(\Sigma - \Sigma_k)\|_F = \sigma_{k+1}$$

بنابراین فاصله بین A و A_{k+1} برابر σ_{k+1} می‌باشد.

اثبات ۳: برای اثبات (۳)، نشان می‌دهیم که اگر $B \in \mathbb{R}^{m \times n}$ هر ماتریس دیگری از رتبه k باشد، آنگاه

$\|A - B\|_F \geq \sigma_{k+1}$ ؛ یعنی A_k نزدیک‌ترین به A در میان همه ماتریسهای دیگر از رتبه k است.

چون B دارای رتبه k است، فضای پوچ B ، $N(B)$ دارای بعد $n - k$ است. اکنون فضای

$S = \text{span}\{v_1, \dots, v_{k+1}\}$ را در نظر بگیرید که در آن v_1 تا v_{k+1} بردارهای تکین راست A هستند.

چون $N(B)$ و S هر دو زیر فضاهای \mathbb{R}^n هستند و چون مجموع ابعاد آنها بزرگتر از n است، اشتراک آنها

باید تهی نباشد. فرض کنید z یک بردار واحد واقع در این اشتراک باشد. آنگاه چون

$z \in \text{span}\{v_1, \dots, v_{k+1}\}$ ، اسکالرهایی (نه همگی برابر صفر) وجود دارند به طوری که

$$z = c_1 v_1 + c_2 v_2 + \dots + c_{k+1} v_{k+1}$$

بعلاوه چون v_1, v_2, \dots, v_{k+1} یکا متعامد هستند، ما باید داشته باشیم

$$|c_1|^2 + |c_2|^2 + \dots + |c_{k+1}|^2 = 1. \text{ همچنین چون } z \text{ متعلق به } N(B) \text{ است، داریم } Bz = 0.$$

نتیجه

$$(A - B)z = Az = \sum_{i=1}^{k+1} c_i A v_i = \sum_{i=1}^{k+1} \sigma_i c_i u_i$$

چون u_1, \dots, u_{k+1} یکا متعامد هستند، داریم

$$\|(A - B)z\|_F^2 = \sum_{i=1}^{k+1} \|\sigma_i c_i\|_F^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} \|c_i\|_F^2 = \sigma_{k+1}^2$$

□ بنابراین $\|(A - B)z\|_F \geq \frac{\|(A - B)z\|_F}{\|z\|_F} \geq \sigma_{k+1}$ (چون $\|z\|_F = 1$)

نتیجه ۱۰-۶-۱ فرض کنید A یک ماتریس $m \times n$ با رتبه کامل باشد و فرض کنید $r = \min(m, n)$

فرض کنید $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ مقادیر تکین A باشند. فرض کنید C ماتریس

$m \times n$ دیگری باشد به قسمی که $\|C - A\|_F < \sigma_r$. آنگاه C همچنین دارای رتبه کامل

است.

نتیجه ۱۰-۶-۲ فاصله نسبی یک ماتریس نامنفرد A به نزدیک‌ترین ماتریس منفرد C برابر $\frac{1}{\text{cond}_r(A)}$ می‌باشد، یعنی

$$\frac{\|C - A\|_r}{\|A\|_r} = \frac{1}{\text{cond}_r(A)}$$

مثال ۱۰-۶-۳ فرض کنید

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{pmatrix}, \quad \sigma_1 = 6/5468, \quad \sigma_2 = 0/3742$$

فرض کنید k برابر ۱ باشد. آنگاه

$$A_1 = U \Sigma_1 V^T = \begin{pmatrix} 1/2608 & 1/8193 \\ 2/534 & 2/9630 \\ 2/8460 & 4/1067 \end{pmatrix}$$

میان همه ماتریسهای رتبه ۱، A_1 نزدیک‌ترین به A است.

فاصله یک ماتریس از نزدیک‌ترین ماتریس با رتبه پایین‌تر

نتیجه قبلی بیان می‌کند که کوچک‌ترین مقدار تکین مخالف صفر فاصله A را از نزدیک‌ترین ماتریس با رتبه پایین‌تر ارائه می‌دهد. بویژه برای یک ماتریس $n \times n$ نامنفرد A ، σ_n اندازه فاصله A را تا نزدیک‌ترین ماتریس منفرد ارائه می‌دهد.

بنابراین برای این که بدانیم که آیا یک ماتریس A از رتبه r به اندازه کافی به یک ماتریس با رتبه پایین‌تر نزدیک است به کوچک‌ترین مقدار تکین مخالف صفر σ_r نگاه می‌کنیم. اگر این مقدار خیلی کوچک باشد، آنگاه ماتریس به یک ماتریس از رتبه $r - 1$ خیلی نزدیک است، زیرا یک اختلال با اندازه‌ای به کوچکی $|\sigma_r|$ یک ماتریس از رتبه $r - 1$ تولید خواهد کرد. در حقیقت یک چنین اختلالی برابر $u_r \sigma_r v_r^T$ می‌باشد.

$$A = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & 2 & \cdot \\ \cdot & \cdot & \cdot / \dots \dots \dots 4 \end{bmatrix}$$

$$\text{rank}(A) = 3, \quad \sigma_3 = \cdot / \dots \dots \dots 4$$

$$U = \begin{bmatrix} \cdot & 1 & \cdot \\ 1 & \cdot & \cdot \\ \cdot & \cdot & 1 \end{bmatrix}$$

$$V = \begin{bmatrix} \cdot & 1 & \cdot \\ 1 & \cdot & \cdot \\ \cdot & \cdot & 1 \end{bmatrix}$$

$$A' = A - u_3 \sigma_3 v_3^T = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & 2 & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$$

$$\text{rank}(A') = 2$$

اختلال لازم $u_3 \sigma_3 v_3^T$ برای منفرد ساختن A خیلی کوچک است:

$$10^{-6} \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot / 4 \end{bmatrix}$$

تقریبهایی توسط ماتریسهای متعامد

فرض کنید A یک ماتریس $n \times n$ باشد. در این بخش ما مسأله پیدا کردن یک ماتریس متعامد Q را ملاحظه می‌کنیم به قسمی که

$$\|A - Q\|_F$$

می‌نیمم شود.

می‌توان نشان داد که اگر

$$A = U \Sigma V^T$$

تجزیه مقدار تکین A باشد، آنگاه

$$Q = UV^T$$

تعمیمی از این مسأله در آنالیز عاملها در آمار مطرح می شود. با مفروض بودن ماتریسهای حقیقی و $n \times n$ بعدی A و B یک ماتریس متعامد Q پیدا کنید به قسمی که

$$\|A - BQ\|_F$$

می نیمم شود.

دوباره SVD می تواند برای حل این مسأله استفاده شود. می توان نشان داد (مرجع Green 1952) که اگر

$$B^T A = U \Sigma V^T$$

SVD ماتریس $B^T A$ باشد، آنگاه

$$Q = UV^T$$

برای تعمیم بیشتری از مسأله مرجع Rao(1980) را ملاحظه کنید.

۱۰-۶-۴ یک بسط حاصلضرب خارجی A و پیامدهای آن

در کاربردهای بسیاری اندازه ماتریس A آن چنان بزرگ است که ذخیره کردن A ممکن است به قسمت اعظمی از کل فضای حافظه موجود نیاز داشته باشد.

شناخت مقادیر تکین مخالف صفر و بردارهای تکین راست و چپ وابسته می تواند به ذخیره کردن ماتریس، همراه با صرفه جویی قابل توجهی در فضای حافظه، کمک کند.

فرض کنید A یک ماتریس $m \times n$ ($m \geq n$) باشد و $\sigma_1, \dots, \sigma_r$ مقادیر تکین مخالف صفر A باشند. فرض کنید بردارهای تکین راست و چپ وابسته به ترتیب v_1, \dots, v_r و u_1, \dots, u_r باشند. آنگاه قضیه زیر نشان می دهد که A می تواند به صورت مجموع r ماتریس، هر کدام از رتبه ۱، نوشته شود. چون یک ماتریس رتبه ۱ از مرتبه $m \times n$ می تواند فقط با استفاده از $(m+n)$ مکان حافظه به جای mn مکان مورد نیاز برای ذخیره سازی یک ماتریس دلخواه ذخیره شود، این امر صرفه جویی قابل توجهی را نتیجه خواهد داد اگر r به طور قابل توجهی کمتر از n باشد.

قضیه ۱۰-۶-۴ فرض کنید A یک ماتریس $m \times n$ باشد و فرض کنید $\sigma_1, \sigma_2, \dots, \sigma_r > 0$ مقدار تکین مخالف صفر A باشند. آنگاه

$$A = \sum_{j=1}^r \sigma_j u_j v_j^T$$

اثبات: اثبات فوراً از رابطه

$$A = U\Sigma V^T$$

□ و با توجه به این که σ_{r+1} تا σ_n برابر صفر هستند، نتیجه می شود.

تعریف ۱۰-۶-۱ نمایش $A = \sum_{j=1}^r \sigma_j u_j v_j^T$ بسط حاصلضرب خارجی A نامیده می شود.

مثال ۱۰-۶-۵

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}$$

$$\sigma_1 = 14/5576$$

$$\sigma_2 = 1/0.372$$

$$\sigma_3 = 0, \quad r = 2$$

$$u_1 = \begin{bmatrix} 0/2500 \\ 0/4852 \\ 0/8379 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0/8371 \\ 0/3267 \\ -0/4389 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0/4625 \\ 0/5706 \\ 0/6786 \end{bmatrix}, \quad v_2 = \begin{bmatrix} -0/7870 \\ -0/0882 \\ 0/6106 \end{bmatrix}$$

$$\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}$$

یک پیامد مهم

قضیه ۱۰-۶-۴ دارای یک پیامد عملی مهم است. در کاربردهای عملی معمولاً ماتریس A دارای یک تعداد زیادی مقادیر تکین کوچک است. فرض کنید $(n - k)$ مقدار تکین کوچک از ماتریس A وجود داشته که قابل صرف نظر باشند. آنگاه ماتریس $\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_k u_k v_k^T$ یک تقریب خیلی خوب از A است، و یک چنین تقریبی می تواند در کاربردها کافی باشد. برای مثال، در پردازش تصویر رقمی، حتی هنگامی که k خیلی کوچک تر از n انتخاب می شود تصویر رقمی متناظر با A_k می تواند خیلی به تصویر اصلی نزدیک باشد. برای جزئیات بیشتر مرجع Andrews and Hunt (1977) را ملاحظه کنید. اگر A یک ماتریس $n \times n$ باشد، ذخیره A_k نیاز به

$(2nk + k)$ مکان خواهد داشت، در مقایسه با n^2 مکان مورد نیاز برای A ، صرفه جویی زیادی را نتیجه خواهد داد.

۱۰-۶-۵ رتبه عددی

همان گونه که اکنون ملاحظه کردیم، در کاربردهای عملی که نیاز به مقادیر تکین دارند، ما باید بدانیم چه موقع یک مقدار تکین محاسبه شده را باید برابر صفر بپذیریم. البته اگر از مرتبه "صفرهای گرد کردن" (ε ماشین) باشد، می توانیم آن را برابر صفر اعلام کنیم.

ولیکن، اگر داده‌ها خطای نسبی بزرگی وارد کنند، باید آن را نیز در نظر گرفت. یک معیار عملی به صورت زیر می باشد.

یک مقدار تکین محاسبه شده برابر صفر پذیرفته می شود اگر کوچک تر یا مساوی $\|A\|_{\infty} 10^{-t}$ باشد. در اینجا عناصر A تا t رقم صحیح هستند.

با تعریف تولرانس $\delta = 10^{-t} \|A\|_{\infty}$ برای یک مقدار تکین صفر، می توانیم قرارداد زیر را در نظر بگیریم (Golub and Van Loan (1989, 247)):

A دارای "رتبه عددی" r است اگر مقادیر تکین محاسبه شده $\bar{\sigma}_1, \bar{\sigma}_2, \dots, \bar{\sigma}_r$ در رابطه زیر صدق کنند:

$$\bar{\sigma}_1 \geq \bar{\sigma}_2 \geq \dots \geq \bar{\sigma}_r > \delta \geq \bar{\sigma}_{r+1} \geq \dots \geq \bar{\sigma}_n \quad (10-6-4)$$

بنابراین برای تعیین رتبه عددی یک ماتریس A ، فقط مقادیر تکین "بزرگ" A را می شماریم. اگر این تعداد برابر r باشد، آنگاه A دارای رتبه عددی r خواهد بود.

تبصره: توجه کنید که پیدا کردن رتبه عددی یک ماتریس خدعه آمیز خواهد بود اگر اختلاف مناسبی بین یک مجموعه از مقادیر تکین وجود نداشته باشد.

۱۰-۷ محاسبه ماتریس وارینانس - کوواریانس

در فصل ۷ یک روش برای محاسبه ماتریس وارینانس - کوواریانس $(A^T A)^{-1}$ با استفاده از تجزیه QR ماتریس A بیان کردیم. توجه می کنیم که در اینجا این ماتریس همچنین می تواند فوراً محاسبه شود به محض آن که مقادیر تکین و بردارهای تکین راست ماتریس A محاسبه شده باشند.

محاسبه $(A^T A)^{-1}$ با استفاده از SVD

فرض کنید

$$A = U \Sigma V^T$$

SVD ماتریس A باشد: آنگاه عناصر ماتریس $C = (A^T A)^{-1}$ توسط

$$c_{ij} = \sum_{k=1}^n \frac{v_{ik} v_{jk}}{\sigma_k^2}$$

ارائه می‌شوند، که در آن $n = \text{rank}(A)$. توجه کنید که $(C = (A^T A)^{-1} = V(\Sigma^T \Sigma)^{-1} V^T$

مثال ۱۰-۷-۱

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}$$

$$c_{11} = \frac{v_{11}^2}{\sigma_1^2} + \frac{v_{12}^2}{\sigma_2^2} = 4/833$$

$$c_{12} = \frac{v_{11} v_{21}}{\sigma_1^2} + \frac{v_{12} v_{22}}{\sigma_2^2} = -3/3333$$

$$c_{22} = \frac{v_{21}^2}{\sigma_1^2} + \frac{v_{22}^2}{\sigma_2^2} = 2/3333$$

تذکر MATCOM: برنامه COVSVD در MATCOM روش فوق را برای محاسبه ماتریس وارینانس-کووارینانس پیاده‌سازی می‌کند.

۱۰-۸ تجزیه مقدار تکین، مسأله کمترین توانهای دوم، و شبه معکوس

در فصل ۷ ملاحظه نموده‌ایم که مسأله کمترین توانهای دوم با رتبه کامل می‌تواند به طور مؤثری با استفاده از تجزیه QR حل شود، اما محورگیری برای انجام حالت رتبه ناقص مورد نیاز خواهد بود. اندکی پیش ملاحظه نموده‌ایم که مقادیر تکین A مورد نیاز هستند تا به طور قابل اطمینانی نزدیکی به ناقص بودن رتبه تعیین گردد. اکنون ملاحظه خواهیم کرد که SVD همچنین یک وسیله مؤثر برای حل مسأله کمترین توانهای دوم در هر دو حالت رتبه کامل و رتبه ناقص است.

۱-۸-۱۰ SVD و مسأله کمترین توانهای دوم

به طور خلاصه در فصل ۷ بحث کردیم که چگونه یک مسأله کمترین توانهای دوم را با استفاده از SVD حل می‌کنیم. در این جا طرز عمل را به طور کامل توضیح می‌دهیم.

یک بار دیگر مسأله کمترین توانهای دوم را ملاحظه کنید: x یی را پیدا کنید که $\|Ax - b\|_2 = \|r\|_2$ می‌نیمم باشد.

فرض کنید $A = U\Sigma V^T$ ، SVD ماتریس A باشد، آنگاه داریم:

$$\begin{aligned} \|r\|_2 &= \|(U\Sigma V^T x - b)\|_2 \\ &= \|U(\Sigma V^T x - U^T b)\|_2 \\ &= \|\Sigma y - b'\|_2 \end{aligned}$$

که در آن $V^T x = y$ و $U^T b = b'$. بنابراین استفاده از SVD ماتریس A مسأله کمترین توانهای دوم مربوط به یک ماتریس کامل A را به یک مسأله کمترین توانهای دوم با یک ماتریس قطری Σ کاهش می‌دهد:

لا را به قسمی پیدا کنید که

$$\|\Sigma y - b'\|_2$$

می‌نیمم باشد.

حل مسأله کاهش یافته ساده است. داریم

$$\|\Sigma y - b'\|_2 = \sum_{i=1}^k \|\sigma y_i - b'_i\|_2 + \sum_{i=k+1}^m |b'_i|^2$$

که در آن k تعداد مقادیر تکین مخالف صفر A است. بنابراین بردار

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

که $\|\Sigma y - b'\|_2$ را می‌نیمم می‌سازد به صورت زیر ارائه می‌شود:

$$y_i = \begin{cases} \frac{b'_i}{\sigma_i}, & \text{اگر } \sigma_i \neq 0 \\ \text{دلخواه}, & \text{اگر } \sigma_i = 0 \end{cases}$$

(توجه کنید که به هنگام $k < n$ ، تا y_{k+1} تا y_n در عبارت قبل ظاهر نمی‌شوند، بنابراین هیچ اثری بر مانده ندارند). البته وقتی که γ محاسبه شد، جواب را می‌توان از رابطه $x = Vy$ به دست آورد.

چون متناظر با هر مقدار تکین σ_i "صفر"، γ_i می‌تواند به طور دلخواه انتخاب شود، در حالت رتبه-ناقص ما جوابهای بیشماری را برای مسأله کمترین توانهای دوم خواهیم داشت. مثالهایی وجود دارند که این ناقص بودن رتبه واقعا" مطلوب است زیرا یک دسته غنی از جوابها را فراهم می‌سازد که ممکن است برای بهینه سازی برخی جنبه‌های دیگر مسأله اصلی مورد استفاده قرار گیرند. البته در حالت رتبه کامل جواب کمترین توانهای دوم منحصر به فرد است.

بحث قبل را می‌توان در الگوریتم زیر خلاصه کرد.

الگوریتم ۱۰-۸-۱ جوابهای کمترین توانهای دوم با استفاده از SVD

گام ۱: SVD ماتریس A را پیدا کنید

$$A = U\Sigma V^T$$

$$\text{گام ۲: بردار } b' = U^T b = \begin{bmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_n \end{bmatrix} \text{ را تشکیل دهید.}$$

$$\text{گام ۳: بردار } y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \text{ را به صورت زیر حساب کنید.}$$

$$y_i = \begin{cases} \frac{b'_i}{\sigma_i}, & \text{اگر } \sigma_i \neq 0 \\ \text{دلخواه}, & \text{اگر } \sigma_i = 0 \end{cases}$$

گام ۴: دسته جوابهای کمترین توانهای دوم را به صورت زیر محاسبه نمایید

$$x = Vy$$

(توجه کنید که در حالت رتبه کامل، این دسته جواب فقط یک عضو دارد.)

شمارش اعمال ممیز شناور. با استفاده از روش گولوب - کاهان - رینسک که بعداً بیان خواهد شد، تقریباً $2mn^2 + 2n^3$ عمل ممیز شناور (flops) برای حل مسأله کمترین توانهای دوم مورد نیاز خواهد بود اگر A یک ماتریس $m \times n$ و $m \geq n$ باشد. (در به دست آوردن این تعداد اعمال ممیز شناور توجه کنید که بردار کامل b' لازم نیست محاسبه شود، فقط ستونهای U که متناظر با مقادیر تکین مخالف صفر هستند در محاسبات لازم می‌باشند).

یک عبارت برای جواب کمترین توانهای دوم با می نیم نرم

از گام ۳ واضح است که در حالت رتبه ناقص، جواب کمترین توانهای دوم با می نیم نرم ۲ آن جوابی است که با قرار دادن $v_i = 0$ به دست می‌آید هرگاه $\sigma_i = 0$ باشد. بنابراین از قبل عبارت زیر را برای جواب با می نیم نرم ۲ داریم.

جواب کمترین توانهای دوم با می نیم نرم ۲ با استفاده از SVD

$$x = \sum_{i=1}^k \frac{u_i^T b}{\sigma_i} v_i \quad (1-8-10)$$

که در آن k برابر رتبه عددی A بوده و u_i و v_i به ترتیب i امین ستون U و V می‌باشند.

مثال ۱-۸-۱۰

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 1 & 2 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} 6 \\ 9 \\ 6 \end{bmatrix}$$

$$\sigma_1 = 7.5358 \quad .1$$

$$\sigma_2 = 0.4597$$

$$\sigma_3 = 0$$

A از رتبه ناقص است

$$U = \begin{bmatrix} 0.4956 & 0.5044 & 0.7071 \\ 0.7133 & -0.7008 & -0.0000 \\ 0.4956 & 0.5044 & -0.7071 \end{bmatrix}$$

$$V = \begin{bmatrix} 0.3208 & -0.8546 & 0.4082 \\ 0.5470 & -0.1847 & -0.8165 \\ 0.7732 & 0.4853 & 0.4082 \end{bmatrix}$$

$$b' = U^T b = (12/3667, -0/2547, 0)^T \quad 2.$$

$$y = (1/6411, -0/5541, 0) \quad 3.$$

۴. جواب کمترین توانهای دوم با می نیمم نرم ۲ عبارت است از $Vy = (1, 1, 1)^T$

تبصره‌های محاسباتی: فرض کنید A دارای رتبه $r \leq \min(m, n)$ باشد. آنگاه در محاسبه جوابهای کمترین توانهای دوم با استفاده از SVD ماتریس A از طریق الگوریتم ۱۰-۸-۱، ما فقط نیاز به محاسبه y_1 تا y_r داریم که اینها می‌توانند فقط با استفاده از r ستون اول ماتریس U محاسبه شوند. بنابراین در این حالت لازم نیست که تمام U ها محاسبه شوند. به طور مشابه در محاسبه جواب با می نیمم نرم فقط r ستون اول V مورد نیاز می‌باشند.

بعلاوه همان گونه که در بخش ۱۰-۹-۱ خواهیم دید، ماتریسهای U و V به صورت حاصلضرب ماتریسهای هاوس هولدر و گینوز محاسبه می‌شوند، بنابراین بردار b' می‌تواند با به هنگام درآوری حاصلضرب b در یک ماتریس هاوس هولدر و یا یک ماتریس گینوز محاسبه گردد، وقتی که این ماتریس اخیر محاسبه شود، در نتیجه به طور کامل از تشکیل صریح U پرهیز می‌شود.

این تبصره‌ها باید در محاسبات عملی در نظر گرفته شوند تا صرفه جویی مورد انتظار حاصل گردد. در محاسبه U و V هر اندازه صرفه جویی شود باز هم حل یک مسأله کمترین توانهای دوم با استفاده از SVD به مراتب گران‌تر از استفاده از تجزیه QR یا هر روش دیگر بحث شده در فصل ۷ است. (بخش ۱۰-۹-۳ را ملاحظه کنید).

همچنین توجه کنید که در محاسبات عملی r را باید برابر رتبه عددی A که توسط معیار ارائه شده در بخش ۱۰-۶-۵ تعیین می‌شود، اختیار کرد.

تذکر **MATCOM**: الگوریتم ۱۰-۸-۱ در برنامه LSQRSVD در **MATCOM** پیاده سازی شده است. همچنین **MATCOM** دارای یک برنامه به نام **MINNMSVD** برای محاسبه جواب با می نیمم نرم دو با استفاده از SVD است.

۱۰-۸-۲ حل یک دستگاه خطی با استفاده از تجزیه مقدار تکین

ایده استفاده از SVD در حل مسأله کمترین توانهای دوم می‌تواند به آسانی برای تعیین این که آیا یک دستگاه خطی $Ax = b$ دارای یک جواب است، به کار برده شود، اگر دارای جواب است چگونه باید تعیین گردد.

بنابراین اگر

$$A = U\Sigma V^T$$

آنگاه

$$Ax = b$$

معادل است با

$$\Sigma y = b'$$

که در آن $y = V^T x$ و $b' = U^T b$.

بنابراین دستگاه $Ax = b$ سازگار است، اگر و فقط اگر دستگاه قطری $\Sigma y = b'$ سازگار باشد (که امتحان کردن آن ساده است)، و یک جواب $Ax = b$ می‌تواند ابتدا با حل دستگاه قطری $\Sigma y = b'$ و سپس بازیابی x از $x = V^T y$ محاسبه شود. ولیکن این راه خیلی گران‌تر از روشهای حذفی گاوس و تجزیه QR خواهد بود. به این دلیل به طور کلی SVD در عمل برای حل یک دستگاه خطی استفاده نمی‌شود.

۱۰-۸-۳ SVD و شبه معکوس

در فصل ۷ ملاحظه کردیم که اگر A یک ماتریس $m \times n$ ($m \geq n$) و بارتبه کامل باشد، آنگاه شبه معکوس A توسط رابطه زیر ارائه می‌شود

$$A^+ = (A^T A)^{-1} A^T$$

یک تعریف رسمی از شبه معکوس هر ماتریس A (چه دارای رتبه کامل باشد چه نباشد) می‌تواند به صورت زیر ارائه شود.

چهار خاصیت شبه معکوس

شبه معکوس یک ماتریس A از مرتبه $m \times n$ یک ماتریس X از مرتبه $n \times m$ بوده و در شرایط زیر صدق می‌کند

$$AXA = A \quad ۱.$$

$$XAX = X \quad ۲.$$

$$(AX)^T = AX \quad ۳.$$

$$(XA)^T = XA \quad ۴.$$

شبه معکوس یک ماتریس همیشه وجود دارد و منحصر به فرد است. اکنون نشان می‌دهیم که SVD یک عبارت زیبا برای شبه معکوس فراهم می‌سازد.

فرض کنید $A = U\Sigma V^T$ ، SVD ماتریس A باشد؛ آنگاه به آسانی می‌توان تحقیق کرد که ماتریس

$$V\Sigma^+U^T$$

که در آن

$$\Sigma^+ = \text{diag}\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \dots, \frac{1}{\sigma_r}, 0, \dots, 0\right) \in \mathbb{R}^{m \times n}$$

در همه چهار شرط صدق می‌کند، و بنابراین شبه معکوس A است. توجه کنید که این عبارت برای شبه معکوس بر A^{-1} انطباق می‌یابد اگر A نامنفرد باشد، زیرا

$$\begin{aligned} A^{-1} &= (A^T A)^{-1} A^T && (۳-۸-۱۰) \\ &= (V\Sigma^T U^T U \Sigma V^T)^{-1} V \Sigma^T U^T \\ &= V \Sigma^{-1} (\Sigma^T)^{-1} V^T V \Sigma^T U^T \\ &= V \Sigma^{-1} U^T \end{aligned}$$

(توجه کنید که در این حالت $\Sigma^+ = \Sigma^{-1}$).

فرایند محاسبه شبه معکوس A^+ ماتریس A با استفاده از SVD ماتریس A به صورت زیر خلاصه می‌شود.

الگوریتم ۱۰-۸-۲ محاسبه شبه معکوس با استفاده از SVD

گام ۱: SVD ماتریس A را پیدا کنید:

$$A = U\Sigma V^T$$

گام ۲: ماتریس

$$\Sigma^+ = \left[\begin{array}{ccc|c} \frac{1}{\sigma_1} & & & \cdot \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_r} \\ \hline & & & \cdot \end{array} \right]$$

را محاسبه کنید که در آن $\sigma_1, \dots, \sigma_r$ مقادیر تکین مخالف صفر A هستند.

گام ۳: $A^+ = V\Sigma^+U^T$ را محاسبه کنید.

مثال ۱۰-۸-۲ SVD یک ماتریس A به صورت

$$A = \begin{bmatrix} \cdot & \cdot & -1 \\ -1 & \cdot & \cdot \\ \cdot & -1 & \cdot \end{bmatrix} \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & 2 & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{6}{9} & -\frac{6}{9} \\ -\frac{6}{9} & \frac{1}{3} & -\frac{6}{9} \\ -\frac{6}{9} & -\frac{6}{9} & \frac{1}{3} \end{bmatrix}$$

می باشد. پس

$$\Sigma^+ = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & \frac{1}{2} & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$$

$$U^T = \begin{bmatrix} \cdot & -1 & \cdot \\ \cdot & \cdot & -1 \\ -1 & \cdot & \cdot \end{bmatrix}$$

$$V^T = V$$

بنابراین شبه معکوس $A^+ = V\Sigma^+U^T$ به صورت زیر خواهد بود.

$$\begin{aligned}
 A^+ &= \begin{bmatrix} \frac{1}{3} & -\frac{6}{9} & -\frac{6}{9} \\ -\frac{6}{9} & \frac{1}{3} & -\frac{6}{9} \\ -\frac{6}{9} & -\frac{6}{9} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{bmatrix} \\
 &= \begin{bmatrix} 0 & -\frac{1}{3} & \frac{1}{3} \\ 0 & \frac{2}{3} & -\frac{1}{6} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}
 \end{aligned}$$

شبه معکوس و مسأله کمترین توانهای دوم از $(1-8-10)$ و $(2-8-10)$ نتیجه می شود که جواب با می نیمم نرم ۲ مسأله کمترین توانهای دوم به دست آمده از طریق SVD در بخش قبل برابر A^+b می باشد که در آن A^+ شبه معکوس A است.

مثال ۱۰-۸-۳

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 1 & 2 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} 6 \\ 9 \\ 6 \end{bmatrix}$$

$$A^+ = \begin{bmatrix} -.9167 & 1.3333 & -.9167 \\ -.1667 & .3333 & -.1667 \\ .5833 & -.6667 & .5833 \end{bmatrix}$$

جواب کمترین توانهای دوم با می نیمم نرم ۲ برابر است با

$$A^+b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

۹-۱۰ محاسبه تجزیه مقدار تکین

همان‌گونه که قبلاً ذکر کردیم، کارایی مقادیر تکین و بردارهای تکین برای محاسبه رتبه عددی، عدد وضعیت، پایه‌های یکا متعامد برای فضاهاى برد و پوچ، و فاصله یک ماتریس از یک ماتریس با رتبه پایین‌تر چیزی است که تجزیه مقادیر تکین را در کاربردهای عملی زیادی بسیار جذاب می‌سازد. بنابراین مهم است که بدانیم چگونه SVD می‌تواند به یک طریق مؤثر عددی محاسبه شود. در این بخش این مهم‌ترین جنبه SVD را مورد بحث قرار می‌دهیم.

۱-۹-۱۰ محاسبه SVD از تجزیه مقدار ویژه $A^T A$

ملاحظه کرده‌ایم (قضیه ۱۰-۳-۱) که مقادیر تکین یک ماتریس A دقیقاً ریشه‌های دوم نامنفی مقادیر ویژه ماتریس متقارن $A^T A$ هستند. بنابراین طبیعی است که به محاسبه مقادیر تکین توسط پیدا کردن مقادیر ویژه ماتریس متقارن $A^T A$ فکر کنیم. ولیکن این یک راه مؤثر عددی نیست زیرا همان‌گونه که در فصل ۷ دیدیم، برخی اطلاعات حیاتی ممکن است بر اثر خطای گرد کردن در طی محاسبه $A^T A$ از بین برود.

مثال ساده زیر این پدیده را نشان می‌دهد

مثال ۱۰-۹-۱

$$A = \begin{bmatrix} 1/0001 & 1/0000 \\ 1/0000 & 1/0001 \end{bmatrix}$$

مقادیر تکین A عبارتند از $2/00010$ و $0/00001$.

اکنون داریم

$$A^T A = \begin{bmatrix} 2/0002 & 2/0002 \\ 2/0002 & 2/0002 \end{bmatrix}$$

مقادیر ویژه $A^T A$ (با ۴ رقم حسابی) عبارتند از صفر و $4/0004$.

۲-۹-۱۰ الگوریتم گولوب-کاهان-رینسک

امروزه الگوریتم زیر یک الگوریتم استاندارد برای محاسبه مقادیر تکین و بردارهای تکین است. الگوریتم در دو فاز انجام می‌شود: فاز I یک رویه متناهی است و فاز II یک رویه تکراری است. فاز I. ماتریس A از مرتبه $m \times n$ ($m \geq n$) به یک ماتریس بالا دوقطری توسط هم‌ارزی متعامد زیر تبدیل می‌شود.

$$U^T AV = \begin{bmatrix} B \\ \cdot \end{bmatrix} = \tilde{B} \quad (2-9-10)$$

در اینجا B یک ماتریس $n \times n$ دوقطری ارائه شده به صورت زیر می‌باشد

$$\begin{bmatrix} b_{11} & b_{12} & \cdot & \dots & \cdot \\ \cdot & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \cdot \\ \vdots & & & & b_{n-1,n} \\ \cdot & \dots & \dots & \cdot & b_{nn} \end{bmatrix}$$

فاز II. سپس با استفاده از الگوریتم تکرار QR ضمنی ماتریس دوقطری B باز هم توسط هم‌ارزی متعامد به یک ماتریس قطری Σ کاهش می‌یابد

$$U_1^T B V_1 = \Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$$

سپس تجزیه مقدار تکین A به صورت

$$U^T AV = \begin{bmatrix} \Sigma \\ \cdot \end{bmatrix}$$

ارائه می‌شود که در آن $U = U_0 \text{diag}(U_1, I_{(m-n)})$ و $V = V_0 V_1$

تبصره‌ها: ایده دوقطری سازی A اولین بار در مقاله مشهور (Golub and Kahan (1965) چاپ شد. ایده استفاده از QR ضمنی بر روی ماتریس دوقطری از گولوب ((Golub (1968) می‌باشد: توصیف قطعی همراه با برنامه‌هایی به زبان Algol در مقاله (Golub and Reinsch (1970) ارائه گردید. یک برنامه فترن که الگوریتم را پیاده‌سازی می‌کند در مقاله (Businger and Golub (1969) ارائه شد. تغییری از فاز I، که در حالت‌های خاصی کارا تر است، در مراجع (Chan (1982a, 1982b) ارائه شد. اخیراً تغییری از فاز II توسط دمل و کاهان ((Demmel and Kahan (1990) پیشنهاد شده است. همچنین الگوریتم جدید و سریعتر دیگری توسط فرناندو و پارلت ((Fernando and Parlett (1994) اخیراً پیشنهاد شده است.

اثر اختلال در مقادیر تکین یک ماتریس دوقطری

نتایج زیر نشان می‌دهند که مقادیر تکین یک ماتریس دوقطری در برابر اختلال‌های کوچک غیر حساس

می‌باشند. برای اثبات قضیه ۱۰-۹-۱ مرجع Demmel and Kahan (1990) را ملاحظه کنید.

قضیه ۱۰-۹-۱ فرض کنید $B = (b_{ij})$ یک ماتریس دوقطری $n \times n$ باشد. همچنین فرض کنید $\Delta B = (\delta b_{ij})$ دوقطری باشد. بعلاوه فرض کنید که $b_{ii} + \delta b_{ii} = \alpha_{2i-1} b_{ii}$ و $\alpha_j \neq 0, b_{ii+1} + \delta b_{ii+1} = \alpha_{2i} b_{ii+1}$.

فرض کنید $\bar{\alpha} = \prod_{i=1}^{2n-1} \max(|\alpha_i|, |\alpha_i^{-1}|)$ بعلاوه فرض کنید $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ مقادیر تکین $B + \Delta B$ باشند. آنگاه

$$\frac{\sigma_i}{\bar{\alpha}} \leq \sigma'_i \leq \bar{\alpha} \sigma_i, \quad i = 1, \dots, n \quad (4-9-10)$$

کاهش به شکل دوقطری

ماتریسهای U_0 و V_0 در فاز I به صورت حاصلضرب حداکثر $2n - 2$ ماتریس هاوس هولدر به صورت زیر ساخته می‌شوند

$$U_0 = U_{0,1} U_{0,2} \dots U_{0,n} \quad (5-9-10)$$

و

$$V_0 = V_{0,1} V_{0,2} \dots V_{0,n-2} \quad (6-9-10)$$

اجازه دهید ساختن $U_{0,1}, V_{0,1}, U_{0,2}, V_{0,2}$ و نقش آنها را در فرایند دوقطری سازی با $m = 5$ و $n = 4$ نشان دهیم.

ابتدا یک ماتریس هاوس هولدر $U_{0,1}$ به قسمی ساخته می‌شود که

$$A^{(1)} = U_{0,1} A = \begin{pmatrix} * & * & * & * \\ \cdot & * & * & * \\ \cdot & * & * & * \\ \cdot & * & * & * \\ \cdot & * & * & * \end{pmatrix}$$

سپس یک ماتریس هاوس هولدر $V_{0,1}$ به قسمی ساخته می‌شود که

$$A^{(2)} = A^{(1)} V_{.1} = \begin{bmatrix} * & * & \cdot & \cdot \\ \cdot & * & * & * \\ \cdot & * & * & * \\ \cdot & * & * & * \\ \cdot & * & * & * \end{bmatrix} = \left[\begin{array}{cc|cc} * & * & \cdot & \cdot \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \end{array} \right] A'$$

اکنون فرایند با $A^{(2)}$ تکرار می‌شود؛ یعنی ماتریسهای $U_{.2}$ و $V_{.2}$ طوری ساخته می‌شوند که

$$U_{.2} A^{(2)} V_{.2} = \begin{bmatrix} * & * & \cdot & \cdot \\ \cdot & * & * & \cdot \\ \cdot & \cdot & * & * \\ \cdot & \cdot & * & * \\ \cdot & \cdot & * & * \end{bmatrix}$$

البته، در این گام ما با ماتریس A' از مرتبه 4×3 به جای ماتریس $A^{(2)}$ کار خواهیم کرد. بنابراین ماتریسهای متعامد $U'_{.2}$ و $V'_{.2}$ ابتدا طوری ساخته می‌شوند که

$$U'_{.2} A' V'_{.2} = \begin{bmatrix} * & * & \cdot \\ \cdot & * & * \\ \cdot & * & * \\ \cdot & * & * \end{bmatrix}$$

سپس $U_{.2}$ و $V_{.2}$ از روی $U'_{.2}$ و $V'_{.2}$ به طریق معمول، یعنی با جای دادن آنها در ماتریسهای واحد با مرتبه‌های مناسب، ساخته می‌شوند. فرایند تا به دست آمدن ماتریس دو قطری ادامه می‌یابد.

تذکر **MATCOM**: رویه دو قطری سازی قبلی گولوب و کاهان (Golub and Kahan 1965) در برنامه **BIDIAG** از **MATCOM** پیاده سازی شده است.

مثال ۱۰-۹-۲

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}$$

گام ۱:

$$U_{.1} = \begin{bmatrix} -./1474 & -./4423 & -./8847 \\ -./4423 & ./8295 & -./3410 \\ -./8847 & -./3410 & ./3180 \end{bmatrix}$$

$$A^{(1)} = U_{.1}A = \begin{bmatrix} -6/7823 & -1/2567 & -9/7312 \\ . & ./461 & ./923 \\ . & -./9077 & -1/1154 \end{bmatrix}$$

گام ۲:

$$V_{.1} = \begin{bmatrix} 1 & . & . \\ . & -./6470 & -./7625 \\ . & -./7625 & ./6470 \end{bmatrix}$$

$$A^{(2)} = A^{(1)}V_{.1} = \begin{bmatrix} -6/7823 & 12/7620 & . \\ . & -./1002 & ./245 \\ . & 1/9716 & -./4824 \end{bmatrix}$$

گام ۳:

$$U_{.2} = \begin{bmatrix} 1 & . & . \\ . & -./508 & ./9987 \\ . & ./9987 & ./508 \end{bmatrix}$$

$$B = U_{.2}A^{(2)} = U_{.2}A^{(1)}V_{.1} = U_{.2}U_{.1}AV_{.1}$$

$$= \begin{bmatrix} -6/7823 & 12/7620 & . \\ . & 1/9741 & -./4830 \\ . & . & . \end{bmatrix}$$

توجه کنید که از عبارت قبلی B فوراً نتیجه می شود که صفر یک مقدار تکین A است. همچنین فقط سه تبدیل هاوس هولدر مورد نیاز بودند.

پیدا کردن SVD یک ماتریس دو قطری

فرايند گونه‌ای از تکرار QR است. با شروع با ماتریس دو قطری B از مرتبه $n \times n$ و به دست آمده در فاز I ، کافی است یک دنباله از ماتریسهای دو قطری $\{B_i\}$ به طریقی بسازیم که هر B_i با احتمال زیاد دارای عناصر خارج قطری کوچک تری از ماتریس قبلی باشد. تکرار i ام معادل با به کار بردن QR متقارن

ضمنی، توصیف شده در فصل ۸، با یک انتقال و یلکینسون در مورد ماتریس سه قطری متقارن $B_i^T B_i$ البته بدون تشکیل صریح حاصلضرب $B_i^T B_i$ می باشد. فرض می شود که ماتریسهای سه قطری مورد عمل کاهش ناپذیر باشند (توجه کنید الگوریتم QR متقارن ضمنی با ماتریسهای کاهش ناپذیر کار می کند)؛ در غیر این صورت ما باید با مسائل SVD تفکیک شده کار کنیم. برای مثال اگر $b_{k, k+1} = 0$ ، آنگاه B می تواند به صورت مجموع مستقیم دو ماتریس دو قطری B_1 و B_2 نوشته شود، و $\sigma(B) = \sigma(B_1) \cup \sigma(B_2)$.

بعلاوه فرض می کنیم که هیچ کدام از عناصر قطری B صفر نباشند. اگر $b_{kk} = 0$ به ازای k یی، آنگاه می توان عنصر بالا قطری متناظر $b_{k, k+1}$ را با پیش ضرب در یک دنباله از دورانهای گیونز مناسب صفر کرد؛ بنابراین مسأله می تواند دوباره به دو مسأله تفکیک شده تبدیل شود (تمرین ۲۵). بنابراین یک عنصر صفر بر روی قطر یا بالای قطر یک موهبت است؛ این منجر به تفکیک سازی مسأله می شود.

در اینجا فقط یک گام تکرار از روش توصیف می شود. برای توصیف مفصل الگوریتم مرجع (Golub and Reinsch (1970) را ملاحظه کنید. برای ساده نمودن نمادگذاری، فرض کنید

$$B = \begin{bmatrix} \alpha_1 & \beta_2 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & \beta_n \\ & & 0 & \alpha_n \end{bmatrix} \quad (7-9-10)$$

یک ماتریس دو قطری $n \times n$ باشد. پس انتقال و یلکینسون σ برای ماتریس متقارن $B^T B$ برابر آن مقدار ویژه λ از زیر ماتریس گوشه راست 2×2 ماتریس $B^T B$:

$$\begin{bmatrix} \alpha_{n-1}^2 + \beta_n^2 & \alpha_{n-1} \beta_n \\ \alpha_{n-1} \beta_n & \alpha_n^2 + \beta_n^2 \end{bmatrix} \quad (8-9-10)$$

می باشد که نزدیک به $\alpha_n^2 + \beta_n^2$ است. (بخش ۸-۱۱-۲ را ملاحظه کنید).

۱. σ را به صورت نشان داده شده در بخش ۸-۱۱-۲ محاسبه می کنیم، و سپس یک دوران گیونز J_1'

به قسمی محاسبه می کنیم که

$$J_1' \begin{bmatrix} \alpha_1^2 - \sigma \\ \alpha_1 \beta_2 \end{bmatrix} = \begin{bmatrix} * \\ 0 \end{bmatrix}$$

ماتریس J_1 را به صورت زیر تشکیل می‌دهیم

$$J_1 = \begin{bmatrix} J'_1 & \cdot \\ \cdot & I_{n-2} \end{bmatrix}$$

۲. J_1 را از طرف راست بر B اعمال می‌کنیم؛ یعنی به صورت زیر عمل می‌نماییم

$$B \equiv BJ_1$$

این عمل موجب پر شدن مکان (۲, ۱) ماتریس B می‌شود. یعنی خواهیم داشت

$$B \equiv BJ_1 = \begin{bmatrix} * & * & & & & \\ + & * & * & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & & \\ & & & & * & \\ & & & & & * \end{bmatrix} \quad (10-9-9)$$

که در آن $+$ یک مقدار پر شده و نشان می‌دهد. (مقدار پر شده در مکان (۲, ۱) قرار دارد). اکنون ایده این است که عنصر مخالف صفر $+$ را که در زیر قطر قرار دارد با به کار بردن دورانهای گینوز با یک ترتیب مناسب، به صورت نشان داده در زیر، به انتهای ماتریس ببریم.

۳.

$$B \equiv J_2 B = \begin{bmatrix} * & * & + & & & \\ & * & \ddots & \ddots & & \\ & & \ddots & \ddots & * & \\ & & & \ddots & * & \\ & & & & & * \end{bmatrix}, \quad (\text{پرشدهگی در مکان } (1, 3))$$

$$B \equiv BJ_3 = \begin{bmatrix} * & * & & & & \\ & * & & & & \\ & * & \ddots & \ddots & & \\ & + & \ddots & \ddots & & \\ & & & \ddots & * & \\ & & & & * & \end{bmatrix}, \quad (\text{پرشدهگی در مکان } (3, 2))$$

$$B \equiv J_p B \begin{bmatrix} * & * & & & \\ & * & * & + & \\ & & \ddots & \ddots & \\ & & & \ddots & * \\ & & & & * \end{bmatrix}, \quad (\text{پرشدگی در مکان } (2, 4))$$

$$B \equiv B J_\delta \begin{bmatrix} * & * & & & \\ & * & * & & \\ & & * & \ddots & \\ & & & + & \ddots & * \\ & & & & \ddots & * \end{bmatrix}, \quad (\text{پرشدگی در مکان } (4, 3) \text{ و غیره})$$

توجه کنید که عناصر (۲, ۱) و (۳, ۲) و غیره با پیش ضرب صفر می‌شوند در صورتی که عناصر (۱, ۳)، (۲, ۴) و غیره با یک پس ضرب صفر می‌شوند.

در انتهای یک تکرار یک ماتریس دوقطری جدید \bar{B} به طور متعامد هم‌ارز با ماتریس دوقطری اصلی خواهیم داشت:

$$\bar{B} = (J_{2n-2} \dots J_p J_r) B (J_1 J_3 \dots J_{2n-3})$$

مثال ۱۰-۹-۳

$$B = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

۱. انتقال ویلکینسون: $\sigma = 15/0.828$

$$J_1 = \begin{bmatrix} -0.9901 & 0.1406 & 0 \\ -0.1406 & -0.9901 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$B \equiv B J_1 = \begin{bmatrix} -1/2713 & -1/8395 & 0 \\ -0.2812 & -1/98.1 & 3 \\ 0 & 0 & 1 \end{bmatrix}, \quad (\text{پرشدگی در مکان } (2, 1))$$

۳.

$$J_2 = \begin{bmatrix} -. / 9764 & -. / 216. & . \\ . / 216. & -. / 9764 & . \\ . & . & 1 \end{bmatrix}$$

$$B \equiv J_2 B = \begin{bmatrix} 1/3.2. & 2/2238 & -. / 648. \\ . & 1/5361 & -2/9292 \\ . & . & 1 \end{bmatrix}, \quad (\text{پرشدگی در مکان } ((1, 3))$$

$$J_3 = \begin{bmatrix} 1 & . & . \\ . & . / 96.1 & . / 2797 \\ . & -. / 2797 & . / 96.1 \end{bmatrix}$$

$$B \equiv B J_3 = \begin{bmatrix} 1/3.2. & 2/3163 & . \\ . & 2/2942 & -2/3825 \\ . & -. / 2797 & . / 96.1 \end{bmatrix}, \quad (\text{پرشدگی در مکان } ((3, 2))$$

۴.

$$J_4 = \begin{bmatrix} 1 & . & . \\ . & . / 9926 & -. / 121. \\ . & . / 121. & . / 9926 \end{bmatrix}$$

$$B \equiv J_4 B = \begin{bmatrix} 1/3.2. & 2/3163 & . \\ . & 2/3112 & -2/4812 \\ . & . & . / 6646 \end{bmatrix}$$

تقلیل و همگرایی

هم اکنون یک گام تکراری از الگوریتم SVD دو قطری را توضیح دادیم. همان گونه که قبلاً متذکر شدیم، با به کار بردن الگوریتم به طور تکراری، یک دنباله از ماتریسهای با عناصر قطری $\alpha_1^{(k)}, \dots, \alpha_n^{(k)}$ و عناصر بالای قطری $\beta_1^{(k)}, \dots, \beta_n^{(k)}$ می‌سازیم.

از همگرایی مرتبه دو الگوریتم تکرار ضمنی با انتقال ویلکینسون توصیف شده در الگوریتم ۸-۱۱-۲ نتیجه می‌شود که $\beta_n^{(k)}$ به طور مرتبه دو به صفر همگرا می‌گردد (همگرایی در عمل بیشتر شبیه مرتبه سه است).

با قرار دادن $\beta_i^{(k)}$ برابر صفر (که بر طبق معیار نتیجه می شود) ماتریس تقلیل می یابد و الگوریتم بر روی یک یا چند ماتریس از مرتبه $n - 1$ یا کمتر تکرار می شود.

گولوب و رینسک (Golub and Reinsch (1970)) اظهار داشته اند که «آزمایشهای عددی نشان می دهند که متوسط تعداد کامل تکرارهای QR بر روی ماتریس دو قطری معمولاً کمتر از دو برای هر مقدار تکین است».

معیارها برای نادیده گرفتن عناصر قطری و غیر قطری

۱. یک عنصر غیر قطری β_i را می پذیریم که برابر صفر باشد اگر

$$|\beta_i| \leq \varepsilon(|\alpha_i| + |\alpha_{i-1}|)$$

که در آن ε مضرب کوچکی از μ دقت ماشین می باشد.

۲. یک عنصر قطری α_i را می پذیریم که برابر صفر باشد اگر

$$|\alpha_i| \leq \varepsilon(|\beta_i| + |\beta_{i+1}|)$$

که در آن ε مضرب کوچکی از μ دقت ماشین می باشد.

شمارش اعمال ممیز شناور. اهم هزینه روش از هزینه فاز I تشکیل می گردد. تعداد اعمال ممیز شناور برآورد شده برابر $\frac{9}{4}n^3 + 2mn^2 + 2m^2n$ است. این تعداد شامل هزینه U ، Σ ، و V می باشد. کاربردهایی وجود دارند (برای مثال، کمترین توانهای دوم) که در آنها همه سه ماتریس به طور صریح مورد نیاز نمی باشند. یک جدول زیبا از تعداد اعمال ممیز شناور مختلف الگوریتم SVD گولوب - کاهان - رینسک و الگوریتم SVD چان^۱ (توصیف شده در بخش بعد) برای نیازهای مختلف U ، Σ ، و V در مرجع (Golub and Van Loan (1989, P. 239)) ارائه شده است. اگر فقط مقادیر تکین مورد نیاز باشند، آنگاه تعداد اعمال ممیز شناور برآورده شده برابر $\frac{2}{3}n^3 - 2mn^2$ می باشد.

خاصیت گرد کردن. می توان نشان داد (Golub and Van Loan (1989), (Bjorck, (1994b)) که مقادیر تکین محاسبه شده، تولید شده توسط الگوریتم گولوب - کاهان - رینسک، مقادیر تکین واقعی $A + E$ هستند، که در آن

$$\|E\|_F \leq c(m, n)\mu \|A\|_F$$

در اینجا $c(m, n)$ یک ثابت وابسته به m و n است. بنابراین الگوریتم پایدار پسرو است. دقت مقادیر تکین محاسبه شده

بعلاوه فرض کنید $\bar{\sigma}_i$ یک مقدار تکین محاسبه شده باشد؛ آنگاه

$$|\sigma_i - \bar{\sigma}_i| \leq c(m, n) \|A\|_F = c(m, n) \mu \sigma_{\max}$$

این نتیجه بیان می‌کند که مقادیر تکین محاسبه شده نمی‌توانند از مقادیر تکین واقعی بیشتر از مقدار $\delta = c(m, n) \mu \sigma_{\max}$ تفاوت داشته باشند.

۱۰-۹-۳ SVD چان

تی - چان (T. Chan (1982a)) مشاهده کرد که الگوریتم گولوب - کاهان - رینسک برای محاسبه SVD، توصیف شده در بخش آخر، می‌تواند در حالت $m \gg n$ بهبود یابد اگر ماتریس A ابتدا به QR تجزیه شود و سپس دوقطری سازی بر روی ماتریس بالا مثلثی کوچک‌تر حاصل از تجزیه انجام پذیرد. طبیعتاً بهبود از این حقیقت ناشی می‌شود که کار لازم برای دوقطری سازی ماتریس بالا مثلثی R به مراتب کمتر از کار لازم برای دوقطری سازی ماتریس A می‌باشد. البته به محض آن که SVD ماتریس R به دست آید، می‌توان به سادگی SVD ماتریس A را به دست آورد. بنابراین SVD چان را می‌توان به صورت زیر توصیف کرد.

۱. تجزیه QR ماتریس A را پیدا کنید:

$$Q^T A = \begin{bmatrix} R_1 \\ \cdot \\ \cdot \end{bmatrix} \quad (10-9-10)$$

۲. SVD ماتریس R_1 را با استفاده از الگوریتم گولوب - کاهان - رینسک پیدا کنید:

$$R_1 = X \Sigma Y^T \quad (11-9-10)$$

سپس مقادیر تکین A دقیقاً مقادیر تکین R_1 هستند. ماتریسهای بردارهای تکین U و V توسط رابطه زیر ارائه می‌شوند

$$U = Q \text{diag}(X, I_{m-n}), \quad V = Y \quad (12-9-10)$$

شمارش اعمال ممیز شناور. در مقایسه با $\frac{9}{4}n^3 + 2mn^2 + 2m^2n$ عمل ممیز شناور (flops) مورد نیاز توسط الگوریتم SVD گولوب - کاهان - رینسک، SVD چان تقریباً به $11n^3 + 2m^2n$ عمل ممیز شناور (flops) برای محاسبه Σ ، U ، و V نیاز دارد. واضح است که هنگامی که $m \gg n$

صرفه جویی با SVD چنان وجود خواهد داشت.

تبصره‌ها: SVD چنان نقش مهمی در بسیاری از کاربردهای نظیر نظریه کنترل، که در آن نوعاً $m \gg n$ بازی می‌کند. ایده مثلثی‌سازی A قبل از انجام دوقطری سازی برای SVD ماتریس A در مقاله Lawson and Hanson (1974, 119) در زمینه مسائل کمترین توانهای دوم ذکر شده است.

مثال ۱۰-۹-۴

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 4 & 5 \end{bmatrix}$$

۱. تجزیه QR ماتریس A عبارت است از

$$Q = \begin{bmatrix} -0.2182 & -0.8165 & -0.5345 \\ -0.4364 & -0.4082 & 0.8018 \\ -0.8729 & 0.4082 & -0.2673 \end{bmatrix}$$

$$R_1 = \begin{bmatrix} -4/5826 & -6/1101 \\ . & -0.8165 \end{bmatrix}$$

۲. SVD ماتریس R_1 عبارت است از

$$R_1 = XSY^T$$

$$X = \begin{bmatrix} -0.9963 & 0.0856 \\ -0.0856 & -0.9963 \end{bmatrix}$$

$$Y = \begin{bmatrix} 0.5956 & -0.8033 \\ 0.8033 & 0.5956 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 7/6656 & . \\ . & 0.4881 \end{bmatrix}$$

۳. تجزیه مقدار تکین $A = USV^T$ به صورت زیر است. مقادیر تکین A عبارتند از $7/6656$ و

0.4881

$$U = \begin{bmatrix} 0.2873 & 0.7948 & -0.5345 \\ 0.4698 & 0.3694 & 0.8018 \\ 0.8347 & -0.4814 & -0.2673 \end{bmatrix}$$

$$V = Y$$

شمارش اعمال ممیز شناور برای مسأله کمترین توانهای دوم با استفاده از SVD تعداد اعمال ممیز شناور برای حل مسأله کمترین توانهای دوم خطی ($m \geq n$) با استفاده از الگوریتم SVD محاسبه شده توسط الگوریتم گولوب - کاهان - رینسک تقریباً برابر $2mn^2 + 2n^3$ است. اگر SVD چنان استفاده شود تعداد اعمال ممیز شناور تقریباً برابر $\frac{11}{4}n^3 + mn^2$ می باشد. یادآور می شویم که تعداد اعمال ممیز شناور از طریق معادلات نرمال برای مسأله کمترین توانهای دوم برابر $mn^2/2 + n^3/6$ و با استفاده از تجزیه QR (با ماتریسهای هاوس هولدن) برابر $mn^2 - n^3/3$ می باشد. مجدداً یک جدول زیبا از روشهای کمترین توانهای دوم در خصوص کارایی آنها در مرجع Golub and Van Loan (1989, P. 248) ارائه شده است. مرجع Chan (1982a) را نیز ملاحظه کنید. در شمارش فوق فرض شده است که به طور متوسط ۲ تکرار QR برای هر مقدار تکین مورد نیاز می باشد.

۱۰-۹-۴ الگوریتم QR دمل - کاهان با انتقال صفر^۱

اخیراً در یک مقاله، دمل و کاهان (Demmel and Kahan (1990)) یک الگوریتم جدید ارائه داده اند که همه مقادیر تکین یک ماتریس دو قطری را با دقتی که داده ها اجازه می دهند، حساب می کند. آنها این الگوریتم را تکرار QR با انتقال صفر نامیدند. دلیل این نام این است که الگوریتم جدید متناظر با الگوریتم گولوب - کاهان - رینسک می باشد هنگامی که انتقال صفر است. الگوریتم جدید مبتنی بر رعایت این اصل است که به هنگام انتقال صفر حذفی بر اثر تفریق رخ ندهد؛ بنابراین مقادیر تکین خیلی کوچک را تقریباً می توان با همان دقتی که داده ها اجازه می دهند، پیدا کرد.

در حقیقت اثر انتقال صفر قابل توجه است. این به صورت زیر ثابت می شود. فرض کنید B به همان شکل (۱۰-۹-۷) باشد. فرض کنید J'_1 دوران گیونز باشد به طوری که

$$(J'_1) \begin{bmatrix} \alpha_1 \\ \alpha_1 \beta_2 \end{bmatrix} = \begin{bmatrix} * \\ 0 \end{bmatrix}$$

پس

$$J_1 = \begin{bmatrix} J'_1 & \cdot \\ \cdot & I_{n-2} \end{bmatrix}$$

و

$$BJ_1 = \begin{bmatrix} * & \cdot & & & & \\ + & * & * & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & & \\ & & & & \ddots & * \\ & & & & & * \end{bmatrix}$$

که همانند قبل یک پرشدگی در مکان (۲, ۱) رخ می‌دهد، ولیکن عنصر (۱, ۲) به جای این که مخالف صفر باشد، اکنون برابر صفر می‌باشد. این صفر در طول بقیه الگوریتم تکثیر می‌یابد و این کلید مؤثر بودن این الگوریتم است.

اکنون اجازه دهید J_2 را از طرف چپ بر BJ_1 اعمال کنیم تا عنصر مخالف صفر در مکان (۲, ۱) را صفر کند. پس

$$B \equiv J_2 BJ_1 = \begin{bmatrix} * & * & + & & & \\ & * & \ddots & & & \\ \cdot & & \ddots & & * & \\ & & & \ddots & & * \end{bmatrix}$$

اکنون یک پرشدگی در مکان (۱, ۳)، همانند الگوریتم گولوب - کاهان - رینسک، وجود دارد. اکنون به سادگی می‌توان تحقیق کرد که رتبه زیر ماتریس 2×2

$$\begin{bmatrix} b_{12} & b_{13} \\ b_{22} & b_{23} \end{bmatrix}$$

از ماتریس $J_2 BJ_1$ برابر ۱ است، سپس نتیجه می‌شود که هنگامی که J_3 بر $J_2 BJ_1$ از سمت راست اعمال می‌شود تا عنصر (۱, ۳) را صفر کند، عنصر (۲, ۳) را نیز صفر خواهد کرد، یعنی خواهیم داشت

$$B \equiv J_3 BJ_2 J_1 = \begin{bmatrix} * & * & \cdot & & & \\ & * & \cdot & & & \\ & + & \ddots & * & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & * \\ \cdot & & & & & * \end{bmatrix}$$

بنابراین در مقایسه با همان مرحله الگوریتم گولوب - کاهان - رینسک اکنون یک صفر اضافی بر روی قطر فوقانی داریم. این پدیده ادامه می‌یابد. در حقیقت دوران J_4 عنصر (۳, ۴) و همچنین عنصر (۲, ۴) را صفر می‌کند و غیره.

تیسره: در اینجا فقط اثر یک انتقال صفر را بر روی الگوریتم گلوب - کاهان - رینسک ثابت کرده ایم. برای جزئیات بیشتر در مورد الگوریتم ما خواننده را به مقاله Demmel and Kahan (1990) ارجاع می دهیم.

مثال ۱۰-۹-۵

۱.

$$B = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

$$J_1 = \begin{bmatrix} 0/4472 & -0/8944 & 0 \\ 0/8944 & 0/4472 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

۲.

$$B \equiv BJ_1 = \begin{bmatrix} 2/2361 & 0 & 0 \\ 1/7889 & -0/8944 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

(توجه کنید که یک پرشدگی در مکان (۲, ۱) وجود دارد، اما عنصر (۱, ۲) صفر است).

۳.

$$J_2 = \begin{bmatrix} 0/7809 & 0/6244 & 0 \\ -0/6247 & 0/7809 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$B \equiv J_2 BJ_1 = \begin{bmatrix} 2/8636 & 0/5588 & 1/8741 \\ 0 & 0/6984 & 2/3426 \\ 0 & 0 & 1 \end{bmatrix}$$

توجه کنید که رتبه ماتریس 2×2 زیر برابر ۱ است.

$$\begin{bmatrix} 0/5588 & 1/8741 \\ 0/6984 & 2/3426 \end{bmatrix}$$

$$J_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0/2857 & -0/9583 \\ 0 & 0/9583 & 0/2857 \end{bmatrix}$$

$$B \equiv J_2 B J_1 J_3 = \begin{bmatrix} 2/8636 & 1/9557 & \cdot \\ \cdot & 2/4445 & \cdot \\ \cdot & \cdot/9583 & \cdot/2857 \end{bmatrix}$$

(توجه کنید که هر دو عنصر (۲، ۳) و (۱، ۳) صفر هستند).

.۴

$$J_4 = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & \cdot/9310 & \cdot/3650 \\ \cdot & \cdot\cdot/3650 & \cdot/9310 \end{bmatrix}$$

$$B \equiv J_4 J_2 B J_1 J_3 = \begin{bmatrix} 2/8636 & 1/9557 & \cdot \\ \cdot & 2/6256 & \cdot/1043 \\ \cdot & \cdot & \cdot/2660 \end{bmatrix}$$

توجه کنید که مقادیر تکین آخرین ماتریس B عبارتند از $3/8990$ ، $1/9306$ ، و $0/2657$ ، که همان مقادیر تکین ماتریس اصلی B هستند.

نرخ همگرایی. همگرایی آخرین عنصر خارج قطر به صفر خطی و با ضریب ثابت $\sigma_{n-1}^2/\sigma_n^2$ می باشد. اگر یک دسته از m مقدار تکین کوچک جدا از بقیه مقادیر تکین یافت شود، آنگاه همگرایی خطی و با ضریب ثابت $\sigma_{n-m+1}^2/\sigma_{n-m}^2$ خواهد بود.

دقت مقادیر ویژه محاسبه شده. فرض کنید $\bar{\sigma}_i$ ، $i = 1, \dots, n$ مقادیر تکین ماتریس \bar{B} محاسبه شده توسط به کار بردن یک تکرار QR ضمنی با انتقال صفر در مورد B باشد. فرض کنید $\bar{\sigma}_1 \geq \bar{\sigma}_2 \geq \dots \geq \bar{\sigma}_n$ توسط دمل و کاهان (Demmel and Kahan (1990)) نشان داده شده است که اگر

$$w \approx \epsilon n^2 \mu < 1$$

آنگاه

$$|\sigma_i - \bar{\sigma}_i| \leq \frac{w}{1-w} \sigma_i, \quad i = 1, \dots, n$$

بعلاوه اگر بعد از k تکرار مقادیر تکین ماتریس دو قطری B_k مقادیر $\sigma_{k1} \geq \dots \geq \sigma_{kn}$ باشند، آنگاه هنگامی که $w \approx \epsilon n^2 \mu < 1$ ، داریم:

$$|\sigma_i - \sigma_{ki}| \leq \left[\frac{1}{(1-w)^k} - 1 \right] \cdot \sigma_i \approx \epsilon k n^2 \mu \cdot \sigma_i$$

نتیجه قبل اساساً بیان می‌کند که «خطای نسبی در مقادیر تکین محاسبه شده می‌توانند فقط با توان دوم بعد ماتریس رشد کنند». این یک نتیجه نسبتاً بدبینانه است. نویسندگان مقاله نتیجه‌گرد کردن دیگری ارائه داده‌اند که بیان می‌کند که در نزدیکی همگرایی‌های الگوریتم "خطاها به هیچ وجه انباشته نمی‌شوند و خطا در مقادیر محاسبه شده α_i و β_i توسط $c'\mu$ کراندار می‌شود، در اینجا c' ثابت نسبتاً کوچک دیگری است.

یک کران خطا بر روی مقادیر تکین محاسبه شده به شکل زیر است.

$$\frac{c'\eta\mu}{1 - c'\eta\mu}$$

در اینجا c' ثابت نسبتاً کوچک دیگری است. برای جزئیات بیشتر و اثباتهای این نتایج، خواننده را به مقاله Demmel and Kahan (1990) ارجاع می‌دهیم.

۵-۹-۱۰ الگوریتم dqds فرناندو - پارلت

اخیراً فرناندو و پارلت (Fernando and Parlett (1994)) الگوریتمی به نام dqds (تفاضل خارج قسمت دیفرانسیل با انتقال^۱) ارائه داده‌اند که همه مقادیر تکین را با حداکثر دقت نسبی محاسبه می‌کند. متأسفانه در اینجا این الگوریتم مهم را نمی‌توانیم توصیف کنیم زیرا مطالب لازم برای توصیف این الگوریتم در این کتاب ارائه نشده‌اند. برای توصیف الگوریتم، مقایسه‌ای با سایر الگوریتمهای SVD، و مراجع الگوریتم اصلی تفاضل خارج قسمت اچ - روتیس هاوزر خواننده را به مقاله Fernando and Parlett (1994) ارجاع می‌دهیم.

۶-۹-۱۰ الگوریتم تنصیف

برای خاتمه دادن به این بخش، متذکر می‌شویم که روش تنصیف برای مسأله مقدار ویژه متقارن بحث شده در فصل ۸ را می‌توان برای محاسبه مقادیر تکین یک ماتریس تطبیق داد. این ایده برای اولین بار در مرجع Golub and Kahan (1965) بیان شد.

توجه کنید که مقادیر تکین ماتریس

$$B = \begin{bmatrix} b_{11} & b_{12} & & & \\ & \ddots & \ddots & & \\ & & \ddots & & \\ & & & b_{n-1,n} & \\ \cdot & & & & b_{nn} \end{bmatrix}$$

$m \times n$ و $p \times n$ ($m \geq n$) باشند. آنگاه ماتریسهای متعامد U و V ، و یک ماتریس نامنفرد W و ماتریسهای قطری C و D به ترتیب از مرتبه‌های $m \times n$ و $p \times n$ وجود دارند به قسمی که

$$U^T A W = C = \text{diag}(c_1, \dots, c_n), \quad c_i \geq 0$$

$$V^T B W = D = \text{diag}(d_1, \dots, d_q), \quad d_i \geq 0$$

در اینجا $q = \min(p, n)$ ؛ $d_1 \geq \dots > d_r > d_{r+1} = \dots = d_q = 0$ ؛ $r = \text{rank}(B)$. عناصر

$$\left\{ \frac{c_1}{d_1}, \frac{c_2}{d_2}, \dots, \frac{c_r}{d_r} \right\}$$

مقادیر تکین تعمیم یافته A و B نامیده می‌شوند. مقادیر تکین تعمیم یافته متناظر با d های صفر نامتناهی هستند.

۱۱-۱۰ مرور و خلاصه

مطالب زیر مهمترین نتایج بحث شده در این فصل هستند.

۱. وجود و یکتایی SVD: تجزیه مقدار تکین (SVD) یک ماتریس A همیشه وجود دارد (قضیه

۱-۲-۱۰)

$$A = U \Sigma V^T$$

مقادیر تکین (عناصر قطری Σ) منحصر به فرد هستند، اما U و V منحصر به فرد نیستند.

۲. روابط بین مقادیر تکین و بردارهای تکین و مقادیر ویژه: مقادیر تکین A ریشه‌های دوم نامنفی

مقادیر ویژه ATA (یا AA^T) هستند. (قضیه ۱-۳-۱۰؛ همچنین قضیه ۱-۳-۱۰ را ملاحظه کنید).

۳. حساسیت مقادیر تکین: مقادیر تکین در برابر اختلاهای کوچک حساس نیستند

(قضیه ۱-۵-۱۰).

۴. کاربردهای SVD: مقادیر تکین و بردارهای تکین ماتریس A مفید هستند و قابل اعتمادترین

ابزار برای تعیین رتبه (عددی) و رتبه ناقص A ؛ پیدا کردن پایه‌های یکا متعامد برای برد و فضای

پوچ A ، پیدا کردن فاصله A از ماتریس دیگری با رتبه پایین تر (بویژه نزدیکی به منفرد بودن

یک ماتریس نامنفرد)؛ حل مسائل کمترین توانهای دوم با رتبه کامل و رتبه ناقص؛ و پیدا کردن

شبه معکوس A (بخشهای ۱۰-۶ الی ۱۰-۸) می‌باشند.

این تواناییهای قابل توجه و این حقیقت که مقادیر تکین در برابر اختلالهای کوچک حساس

نیستند، SVD را در یک تنوع وسیعی از عرصه‌های کاربردی، نظیر نظریه کنترل دستگاهها، پردازش سیگنال، کاربردهای آماری، مهندسی زیست پزشکی، پردازش تصویر، یک وسیله ضروری ساخته‌اند.

۵. محاسبه SVD: مورد استفاده‌ترین راه برای محاسبه SVD ماتریس A الگوریتم گولوب - کاهان - رینسک است. این الگوریتم در دو فاز انجام می‌شود. در فاز I ماتریس A با هم‌ارزی متعامد به یک ماتریس دوقطری کاهش می‌یابد، و در فاز II باز هم ماتریس دوقطری توسط تشابه متعامد با استفاده از تکرار QR ضمنی با انتقال و یلکینسون به یک ماتریس قطری کاهش می‌یابد. تغییری از این روش موسوم به تکرار QR - صفر - انتقال، یا تکرار QR با یک انتقال صفر توسط دمل و کاهان در ۱۹۹۰ پیشنهاد شده است. روش دمل و کاهان همه مقادیر تکین یک ماتریس دوقطری را با دقت نسبتاً بالا محاسبه می‌کند. باز هم روش دیگری اخیراً توسط فرناندو و پارلت پیشنهاد شده است (بخش ۱۰-۹-۵).

تمرینها:

بخشهای ۱۰-۲ و ۱۰-۳

۱. (الف) فرض کنید A یک ماتریس $n \times n$ باشد و فرض کنید ماتریسهای U و V متعامد باشند. آنگاه از تعریف مقادیر تکین استفاده کنید ثابت کنید که مقادیر تکین A و U^TAV یکی هستند.

(ب) چگونه بردارهای تکین A با بردارهای تکین U^TAV مرتبط هستند؟

(ج) اگر A معین مثبت متقارن باشد، ثابت کنید که مقادیر ویژه A همان مقادیر تکین ماتریس A هستند.

۲. (نسخه اقتصادی از SVD) فرض کنید A از مرتبه $m \times n$ ($m \geq n$) باشد. فرض کنید $\text{rank}(A) = r \leq n$. آنگاه ثابت کنید که ماتریسهای $\hat{U}_{m \times r}$ و $\hat{V}_{n \times r}$ با ستونهای یکا متعامد و یک ماتریس قطری نامنفرد $S_{r \times r}$ وجود دارند به قسمی که

$$A = \hat{U}S(\hat{V})^T$$

چرا این نسخه، نسخه اقتصادی نامیده می‌شود؟

۳. فرض کنید σ یک مقدار تکین A با مرتبه تکرار l باشد؛ یعنی $\sigma_i = \sigma_{i+1} = \dots = \sigma_{i+l-1}$. فرض کنید $U\Sigma V^T$ یک تجزیه مقدار تکین A باشد. آنگاه نشان دهید که چگونه می‌توان \hat{U} و \hat{V} را از روی

U و V به قسمی ساخت که $\tilde{U}\Sigma(\tilde{V})^T$ نیز یک SVD از ماتریس A باشد.

۴. (الف) قضیه ۱۰-۳-۱ را از قضیه ۱۰-۳-۲ نتیجه بگیرید.

(ب) با مفروض بودن

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

مقادیر تکین σ_1 و σ_2 ماتریس A را با محاسبه مقادیر ویژه $A^T A$ پیدا کنید. سپس ماتریس متعامد P را به قسمی پیدا کنید که

$$P^T S P = \text{diag}(\sigma_1, \sigma_2, -\sigma_1, -\sigma_2, 0)$$

که در آن

$$S = \begin{bmatrix} \circ_{3 \times 3} & A \\ A^T & \circ_{2 \times 2} \end{bmatrix}$$

۵. با استفاده از اثبات ساختنی قضیه ۱۰-۲-۱، SVD ماتریسهای زیر را پیدا کنید:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad A = (1 \ 2 \ 3)$$

$$A = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad A = \text{diag}(1, 0, 2, 0, -5)$$

$$A = \begin{bmatrix} 1 & 1 \\ \varepsilon & 0 \\ 0 & \varepsilon \end{bmatrix}, \quad \text{که در آن } \varepsilon \text{ کوچک است.}$$

بخشهای ۱۰-۵ تا ۱۰-۸

۶. (الف) رتبه، $\| \cdot \|_2$ ، $\| \cdot \|_F$ ، $\text{cond}(A)$ ، و پایه‌های یکا متعامد را برای فضای پوچ و برد هر یک

از ماتریسهای تمرین ۵ پیدا کنید. تصویرهای متعامد را بر روی برد و فضای پوچ A و

متمم‌های متعامد آنها پیدا کنید.

(ب) با استفاده از SVD ماتریس A ، ثابت کنید که برای هر ماتریس A از مرتبه $m \times n$ داریم:

$$\text{rank}(A^T A) = \text{rank}(A A^T) = \text{rank}(A) = \text{rank}(A^T) \quad (i)$$

(ii) $A A^T$ و $A^T A$ دارای مقادیر ویژه مخالف صفر یکسان هستند.

(iii) اگر بردارهای ویژه u_1 و u_2 ماتریس $A^T A$ متعامد باشند، آنگاه $A u_1$ و $A u_2$ نیز متعامد هستند.

۷. (الف) فرض کنید A یک ماتریس $m \times n$ با مقادیر تکین $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$ باشد. فرض

کنید v_1, \dots, v_n بردارهای تکین راست A باشند. آنگاه ثابت کنید که $\|Ax\|_2 / \|x\|_2$

ماکزیمم می‌باشد اگر $x = v_1$ و می‌نیمم می‌باشد اگر $x = v_n$.

(ب) فرض کنید U یک ماتریس متعامد باشد. آنگاه ثابت کنید که

$$\|AU\|_2 = \|A\|_2$$

و

$$\|AU\|_F = \|A\|_F$$

(ج) فرض کنید A یک ماتریس $m \times n$ باشد. با استفاده از SVD ماتریس A ثابت کنید که

$$\text{cond}_r(A^T A) = (\text{cond}_r(A))^2 \quad (i)$$

$$\|A^T A\|_r = \|A\|_r^2 \quad (ii)$$

$$\text{cond}_r(A) = \text{cond}_r(U^T A V) \quad (iii)$$

در اینجا U و V متعامد هستند.

(د) فرض کنید $\text{rank}(A_{m \times n}) = r$ و فرض کنید $B_{m \times r}$ ماتریس به دست آمده از حذف $n - r$

ستون A باشد. آنگاه ثابت کنید که $\text{cond}_r(B) \leq \text{cond}_r(A)$.

۸. ثابت کنید که اگر A یک ماتریس $m \times n$ با رتبه r باشد و اگر B یک ماتریس $m \times n$

دیگری باشد که در رابطه $\|A - B\|_2 \leq \sigma_r$ صدق کند، آنگاه B حداقل دارای رتبه r می‌باشد.

۹. (الف) اثبات دیگری از قضیه ۱۰-۶-۳ ارائه دهید.

(ب) با مفروض بودن

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

یک ماتریس B با رتبه یک و نزدیکترین به A در نرم دو پیدا کنید. $\|A - B\|_2$ چند است؟

بردار x و $\|x\|_2$ را هم با استفاده از SVD ماتریس A و هم شبه معکوس A به دست آورید.

۱۹. با مفروض بودن

$$u = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad v = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

A^+ را پیدا کنید که در آن $A = uv^T$.

۲۰. فرض کنید

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 3 \\ 14 \end{bmatrix}$$

(الف) یک جواب کمترین توانهای دوم x پیدا کنید: $\min_x \|Ax - b\|_2$ و $\|x\|_2$ چقدر است؟

(ب) جواب با می نیمم نرم دو x و $\|x\|_2$ را پیدا کنید.

بخش ۱۰-۹

۲۱. فرض کنید B یک ماتریس بالا مثلثی دو قطری بوده و دارای یک مقدار تکین چندگانه باشد. آنگاه

ثابت کنید که B باید یا دارای یک صفر بر روی قطر و یا بر روی قطر فوقانی باشد.

۲۲. خانواده‌ای از ماتریسهای دو قطری به شکل زیر را در نظر بگیرید

$$B(\eta) = \begin{bmatrix} 1 - \eta & \beta(1 + \eta) & & \\ & \ddots & \ddots & \\ & & \ddots & \beta(1 + \eta) \\ & & & 1 - \eta \end{bmatrix}$$

$\beta \gg 1$ می توان نشان داد (Demmel and Kahan (1990)) که کوچک ترین مقدار تکین $B(\eta)$

تقریباً برابر $(1 - (2n-1)\eta)\beta^{1-n}$ می باشد. با اختیار کردن $\beta = 10^6$ و استفاده از $n = 20$ و

$\eta = 0$ ، این نتیجه را تحقیق کنید.

۲۳. فرض کنید

$$A = \begin{bmatrix} 1 & 0.9999 \\ 2 & 1/9999 \\ 3 & 2/9999 \end{bmatrix}$$

(الف) SVD ماتریس A را با استفاده از الگوریتم گولوب - کاهان - رینسک پیدا کنید.

(ب) SVD ماتریس A را با استفاده از الگوریتم دمل - کاهان پیدا کنید.

(ج) نتایج (الف) و (ب) را با هم مقایسه کنید.

۲۴. نشان دهید که تعداد اعمال ممیز شناور برای الگوریتم گولوب - کاهان - رینسک و SVD چنان به

ترتیب تقریباً برابرند با $\frac{9}{4}n^3 + 4mn^2 + 2m^2n + 11n^3$ و $2m^2n + 11n^3$. همچنین تعداد اعمال ممیز شناور را برای الگوریتم دمل - کاهان محاسبه کنید.

۲۵. فرض کنید B یک ماتریس بالا دو قطری با مقادیر $\alpha_1, \dots, \alpha_n$ بر روی قطر و β_1, \dots, β_n بر روی

قطر فوقانی باشد. ثابت کنید که اگر $\alpha_k = 0$ یا $\alpha_j \neq 0$ و $\beta_j \neq 0$ و $j = k + 1, \dots, n$ آنگاه

پیش ضرب B در $(n - k)$ دوران گیسونز، ماتریس B را به دو ماتریس تفکیک خواهد کرد.

برنامه‌های MATLAB و مسائل

تذکر: شما به برنامه‌های `varcovar`، `parpiv`، `housqr` از `MATCOM` نیاز دارید.

۱. (هدف از این تمرین مطالعه حساس بودن (حساس نبودن) مقادیر تکین یک ماتریس است.)

از دستورهای `MATCOM` نظیر `svd` و `norm` استفاده کنید، و نامساویهای قضیه‌های ۱۰-۵-۱ و

۱۰-۵-۲ را به ترتیب تحقیق کنید.

داده‌های آزمون

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & .99 & 1 & 1 \\ 0 & 0 & .99 & 1 \\ 0 & 0 & 0 & .99 \end{bmatrix} \quad ۱.$$

۲. $A =$ ماتریس دو قطری ویلکینسون از مرتبه ۲۰.

در هر حالت، یک E مناسب بسازید به قسمی که $(A + E)$ با A در عنصر $(n, 1)$ فقط توسط $\varepsilon = 10^{-5}$ تفاوت داشته باشد. (توجه کنید که مقادیر ویژه هر دو ماتریس بد وضع هستند).

۲. یک ماتریس 5×15 با استفاده از دستور MATLAB، $A = \text{rand}(15, 5)$ به طور تصادفی تولید کنید. $S = \text{svd}(A)$ را پیدا کنید. مقادیر $\|A\|_2$ و $\|A\|_F$ ، و عدد وضعیت A را نسبت به نرم ۲ با استفاده از عناصر S محاسبه کنید، و سپس نتایج خود را به ترتیب با نتایج حاصل از دستورهایی MATLAB، $\text{norm}(a)$ ، $\text{norm}(A, 'fro')$ و $\text{cond}(A)$ مقایسه کنید.

۳. الف) پایه‌های یکا متعامد را برای برد و فضای پوچ یک ماتریس A به صورتهای زیر حساب کنید.

(i) با استفاده از $[U, S, V] = \text{svd}(A)$ از MATLAB (از طریق SVD:

بخش ۱۰-۶-۲)

(ii) با استفاده از housqr یا housqrp از فصل ۵، به صورت مناسب (از طریق QR:

بخش ۵-۶)

(iii) با استفاده از دستورهایی $\text{orth}(A)$ و $\text{null}(A)$ از MATLAB

(ب) نتایج (i)، (ii) و (iii) و تعداد اعمال متمیز شناور را برای هر الگوریتم مقایسه کنید.

داده‌های آزمون

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} \quad .1$$

۲. $A =$ یک ماتریس به طور تصادفی تولید شده از مرتبه ۱۰.

۴. رتبه هر یک از ماتریسهای زیر را با استفاده از (i) دستور rank از MATLAB (که از مقادیر تکین A استفاده می‌کند) و (ii) برنامه‌های housqr (تجزیه QR هاوس هولدر با محورگیری) و parpiv از فصل ۵ محاسبه کنید؛ نتایج را با هم مقایسه نمایید.

داده‌های آزمون

۱. (ماتریس کاهان)

$$A = \text{diag}(1, s, \dots, s^{n-1}) \begin{bmatrix} 1 & -c & -c & \dots & -c \\ 0 & 1 & -c & \dots & -c \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & -c \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

که در آن $n = 100$ و $s = 0/6$ ، $c = 0/2$ ، $c, s > 0$ ، $c^2 + s^2 = 1$

۲. A یک ماتریس 10×15 و تولید شده به صورت $A = xy^T$ می‌باشد، که در آن

$$x = \text{round}(10 * \text{rand}(15, 1))$$

$$y = \text{round}(10 * \text{rand}(10, 1))$$

۵. (الف) با استفاده از دستور $\text{rand}(6)$ از MATLAB یک ماتریس A از مرتبه ۶ به طور تصادفی

تولید کنید. اکنون دستور MATLAB زیر را اجرا کنید.

$$[U, S, V] = \text{svd}(A)$$

قرار دهید $s(4, 4) = 0$ ؛ ماتریس $B = U * S * V'$ را محاسبه کنید. $\text{rank}(B)$ چند

است؟

(ب) یک ماتریس C از رتبه ۶ به صورت زیر بسازید:

$$C = \text{qr}(\text{rand}(6))$$

از دستور $\text{norm}(A, 'fro')$ از MATLAB برای محاسبه نرم فروبنیوس استفاده کنید و تحقیق

$$\|C - A\|_F^2 \geq \|B - A\|_F^2$$

(ج) فاصله B از A چقدر است؟

(د) یک ماتریس D از مرتبه ۲ پیدا کنید که به A نزدیک‌ترین باشد.

(این تمرین بر قضیه‌های ۱۰-۶-۲ و ۱۰-۶-۳ مبتنی است)

۶. فرض کنید

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0/0100 & 1 & 1 \\ 0 & 0 & 0/0100 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

فاصله A را از نزدیک‌ترین ماتریس منفرد پیدا کنید. اختلالی بیابید که A را منفرد سازد. اندازه این اختلال را با $|\sigma_r|$ مقایسه کنید.

۷. فرض کنید $A = U\Sigma V^T$ تجزیه مقدار تکین (SVD) یک ماتریس 10×15 به طور تصادفی تولید شده، $A = \text{rand}(15, 10)$ ، و به دست آمده با استفاده از دستور $[U, S, V] = \text{svd}(A)$ از MATLAB باشد. همه عناصر زیر را برابر صفر قرار دهید:

$$s(8, 8) = s(9, 9) = s(10, 10) = 0$$

ماتریس $B = U * S * V'$ را محاسبه کنید.

بهترین تقریب ماتریس B را به شکل

$$B \approx \sum_{i=1}^r x_i y_i^T$$

به قسمی پیدا کنید که داشته باشیم: می‌نیم $\|B - \sum_{i=1}^r x_i y_i^T\|$ ، که در آن x_i و y_i بردار هستند و r رتبه ماتریس B می‌باشد.

۸. برای ماتریسهای A و B تمرین ۷، یک ماتریس یکانی Q به قسمی پیدا کنید که $\|A - BQ\|_F$ را می‌نیم سازد. (راهنمایی: $Q = UV^T$ ، که در آن $A^T B = U\Sigma V^T$). (از دستور MATLAB برای حل این مسأله استفاده کنید).

۹. (الف) یک برنامه MATLAB به نام `covsvd` برای محاسبه $(A^T A)^{-1}$ با استفاده از تجزیه مقدار تکین بنویسید. از آن برای پیدا کردن $(A^T A)^{-1}$ برای ماتریس هیلبرت 8×8 استفاده کنید و نتایج حاصل و تعداد اعمال ممیز شناور را با نتایج به دست آمده از اجرای دستور `varcovar` از MATCOM مقایسه کنید.

(ب) از `covsvd` استفاده کنید و شبه معکوس A را پیدا کنید و نتیجه حاصل را با نتیجه به دست آمده از اجرای دستور `pinv` از MATLAB مقایسه کنید. (A همان ماتریس قسمت (الف) می‌باشد).

۱۰. فرض کنید A یک ماتریس هیلبرت 10×10 باشد و بردار b به قسمی تولید شود که همه عناصر بردار x جواب دستگاه $Ax = b$ برابر ۱ باشد. $Ax = b$ را با استفاده از SVD ماتریس A حل کنید. و دقت، تعداد اعمال ممیز شناور، و زمان سپری شده را با آنچه از دستور `linsyspp` از فصل ۶ به دست می‌آید، مقایسه کنید.

۱۱. فرض کنید $A = \text{rand}(10, 3)$ ، و

$$X = \text{pinv}(A)$$

با استفاده از MATLAB تحقیق کنید که A در همه چهار شرط شبه معکوس: $AXA = A$ ،

$$(XA)^T = XA, (AX)^T = AX, XAX = X$$

۱۲. یک برنامه MATLAB به نام **chansvd**، با استفاده از دستورهایی **qr** و **svd** از MATLAB،

برای پیاده‌سازی الگوریتم SVD بهبود یافته چان و ارائه شده در بخش ۱۰-۹-۳ بنویسید:

$$[U, S, V] = \text{chansvd}(A)$$

برنامه خودتان را با یک ماتریس 5×30 به طور تصادفی تولید شده $A = \text{rand}(30, 5)$ اجرا

کنید و تعداد اعمال متمیز شناور و زمان سپری شده را با آنچه از دستور $\text{svd}(A)$ به دست می‌آید، مقایسه کنید.

۱۳. یک برنامه MATLAB به نام **bidiag** برای دوقطری سازی یک ماتریس A بنویسید

(بخش ۱۰-۹-۲):

$$[B] = \text{bidiagd}(A)$$

در اینجا B یک ماتریس دو قطری است. برنامه خودتان را با استفاده از $A = \text{rand}(15, 10)$

امتحان کنید.

۱۴. (هدف از این تمرین مقایسه سه طریق مرتبط برای پیدا کردن مقادیر تکین یک ماتریس دوقطری

می‌باشد.) یک برنامه MATLAB برای پیاده‌سازی الگوریتم دمل - کاهان برای محاسبه مقادیر

تکین یک ماتریس دوقطری بنویسید.

$$[s] = \text{dksvd}(A)$$

در اینجا s بردار حاوی مقادیر تکین A است.

آزمایش: قرار دهید $A = \text{rand}(50, 20)$. $[U, S, V] = \text{svd}(A)$ را محاسبه کنید. قرار دهید

$$B = U * S * V', S(10, 10) = S(9, 9) = S(8, 8) = 10^{-5}$$

bidiag را بر روی B اجرا کنید.

$$C = \text{bidiag}(B)$$

مقادیر تکین C را با استفاده از **svd** (i)، **dksvd** (ii) و **chansvd** (iii) محاسبه کنید و نتایج را

از نظر دقت بویژه (برای مقادیر تکین کوچک) تعداد اعمال ممیز شناور، و زمان سپری شده، مقایسه کنید.

تذکر: برخی از برنامه‌های MATLAB را که از شما خواسته شده بنویسید، هم اکنون در MATCOM موجود هستند، اما نوشتن برنامه‌ها توسط خودتان ایده خوبی است.

پیشنهادها برای مطالعه بیشتر

اثبات ما برای قضیه SVD (قضیه ۱۰-۲-۱) متداول تر است. برای اثباتی متفاوت و زیبا از این قضیه مهم و کلاسیک مرجع (1994) Pan and Sigmon را ملاحظه کنید. همچنین مرجع Horn and Johnson (1991) را در این زمینه ملاحظه کنید.

همان گونه که قبلاً ذکر شد، کاربردهای SVD متنوع هستند. کتابهایی در هر زمینه از کاربردها که در آنها SVD یک نقش مهم بازی می کند، وجود دارند. برای کاربردهای SVD در مورد مسأله کنترل کلاسیک مقاله قدیمی تر (1980) Klema and Laub را ملاحظه کنید.

SVD همچنین نقش مهمی در نظریه کنترل مدرن، بویژه در نظریه کنترل قوی و H -نامتناهی ایفا می کند. خواننده‌های کنجکاو را به تعداد روزافزون مقاله‌های جالب در این زمینه‌ها ارجاع می دهیم و برای کاربردهای SVD در مورد مسأله تخصیص قطب قوی مرجع Kautsky, Nichols, and Van Dooren (1985) را ملاحظه کنید.

برای کاربردهای SVD در مورد دستگاه هویت شناسی و مسائل پردازش سیگنال مقاله مروری جالب زیر را ملاحظه کنید.

"Variety of Applications of Singular Value Decomposition", by Joos Vandewalle and Bart De Moor (1988).

دو کتاب مهم در زمینه SVD و پردازش سیگنال عبارتند از:

(1) *SVD and Signal Processing, Algorithms, Applications, and Architecture* edited by Ed. F. Deprettere (1988),

(2) *SVD and Signal Processing II, Algorithms, Analysis and Applications*, edited by R. Vaccaro (1991).

برای کاربردهای SVD در مورد پردازش تصویر کتاب

Fundamentals of Digital Image Processing by A. K. Jain (1989).

و مرجع زیر را ملاحظه کنید.

Digital Image Processing by H. C. Andrews and B. R. Hunt (1977).

ماتریسهای زیادی که در پردازش سیگنال و کنترل، و کاربردهای نظریه دستگاهها مطرح می‌شوند، نظیر ماتریسهای هانکل یا توپلیتز ساخت یافته هستند. پیدا کردن الگوریتمهای قابل اطمینان و کارا برای انواع مختلف محاسبات که بتوانند از ساختارهای این ماتریسها بهره‌برداری کنند، یک مسأله مبارزه‌طلبانه برای محققین است. کارهای زیادی انجام شده است، و اکنون یک زمینه فعال تحقیق می‌باشد. برای برخی مقاله‌های جالب در این زمینه کتاب زیر را ملاحظه کنید.

Linear Algebra in Signal, Systems, and Control, edited by B. N. Datta et al. (1988).

همچنین مقاله جدید (1991) Van Dooren را ملاحظه کنید.

برای برخی کاربردهای آماری، مقاله‌های مروری و عالی (1969) Golub و Hammarling (1985) را ملاحظه کنید.

بعضی در مورد کاربردهای ریاضی SVD، نظیر پیدا کردن رتبه عددی، نزدیکی به منفرد بودن، و پایه‌های یکا متعامد برای برد و فضای پوچ در همه کتابهای جبر خطی عددی موجود هستند. بویژه کتابهای (1989) Van Loan و (1991) Watkins این مفاهیم را خیلی خوب مورد بحث قرار داده‌اند. برای بعضی در مورد حساسیت مقادیر تکین مراجع (1979a) Stewart و (1991) Stewart را ملاحظه کنید.

برای جنبه‌های محاسباتی مقادیر تکین و بردارهای تکین مقاله اصلی (1965) Golub and Kahan، و مقاله‌های جدید (1990) Demmel and Kahan و (1994) Fernando and Parlett ارزشمند هستند. روال SGEVD از LAPACK همه یا بخشی از مقادیر تکین و بردارهای تکین یک ماتریس نامتقارن کلی را محاسبه می‌کند. کاهش به شکل دوقطری توسط زیر برنامه SGBRD انجام می‌شود، در حالی که SVD یک ماتریس دوقطری توسط زیر برنامه SBDSQR محاسبه می‌شود.

اخیراً الگوریتم QR ضمنی بخش ۱۰-۹-۲ مربوط به ماتریسهای دوقطری در مقاله (1994b) Chandrasekaran and Ipson برای پیدا کردن SVD ماتریسهای مثلثی توسعه یافته است.

یک تعمیم طبیعی از روش ژاکوبی برای محاسبه مقادیر ویژه یک ماتریس متقارن الگوریتم کوگبت لیانتز^۱ کلاسیک برای محاسبه تجزیه مقدار تکین است. برای مقاله‌های جدید بیشتر مقاله‌های (1987) Charlier and Van Dooren، (1988) Charlier, Vanbegin, and Van Dooren،

- Bai (1988) ، Paige (1986b) را ملاحظه کنید. برای یک فهرست جامع مرجع
Golub and Van Loan (1989, 485) را ملاحظه کنید.
- SVD تعمیم یافته یک تعمیم مستقیم از SVD است. برای مفاهیم گونه های دیگر SVD و تعمیم های
مختلف دیگر مراجع (De Moor and Zha(1991) و De Moor and Van Dooren (1992) را
ملاحظه کنید.
- برای SVD حاصل ضرب دو ماتریس مراجع Heath, Laub, Paige, and Ward (1986) و
De Moor (1992) را ملاحظه کنید.
- برای روابط بین SVD تعمیم یافته و مسأله مقدار ویژه تعمیم یافته مرجع Kagstrom (1985) را
ملاحظه کنید.

نمونه کوچکی از تحلیل خطای گرد کردن

در اینجا نمونه‌های کوچکی از تحلیل خطای گرد کردن در محاسبات ماتریسی را برای خواننده ارائه می‌دهیم. برای این منظور تحلیلهای پسرو برخی محاسبات اساسی نظیر جوابهای دستگاههای مثلثی، تجزیه LU با استفاده از فرآیند حذفی گاوس، و جواب یک دستگاه خطی را ارائه خواهیم داد. اجازه دهید یادآور شویم که منظور ما از تحلیل خطای پسرو تحلیلی است که نشان دهد که جواب محاسبه شده توسط الگوریتم یک جواب دقیق از یک مسأله اختلال یافته است. هنگامی که مسأله اختلال یافته به مسأله اصلی نزدیک می‌باشد، می‌گوییم که الگوریتم پایدار پسرو است.

۱-۱۱ قوانین اساسی حساب ممیز شناور

ابتدا قوانین اساسی حساب ممیز شناور را که در مطالب بعدی مورد نیاز خواهند بود، برای خواننده یادآور می‌شویم. این قوانین در فصل ۲ به دست آمده‌اند. فرض کنید $|\delta| \leq \mu$ ، که در آن μ دقت ماشین است. آنگاه

$$\text{fl}(x \pm y) = (x \pm y)(1 + \delta) \quad (1-1-11)$$

$$\text{fl}(xy) = xy(1 + \delta) \quad (2-1-11)$$

$$\text{fl}\left(\frac{x}{y}\right) = \left(\frac{x}{y}\right)(1 + \delta) \quad y \neq 0 \text{ اگر} \quad (3-1-11)$$

گاه‌گاهی از رابطه

$$\text{fl}(x * y) = \left(\frac{x * y}{1 + \delta} \right)$$

استفاده خواهیم کرد که * هر یک از اعمال حسابی +، -، ×، یا ÷ را نشان خواهد داد.

۱۱-۲ تحلیل خطای پسر و برای طرح حذفی پسر و و جایگزینی پسر و

حالت ۱. دستگاه پایین مثلثی

حل دستگاه پایین مثلثی

$$Ly = b \quad (11-2-1)$$

را با استفاده از طرح حذفی پسر و در نظر بگیرید که در آن

$$L = (l_{ij}), \quad b = (b_1, \dots, b_n)^T \quad (11-2-2)$$

و

$$y = (y_1, \dots, y_n)^T$$

ما از δ برای نشان دادن مقدار محاسبه شده δ استفاده خواهیم کرد.

گام ۱:

$$\hat{y}_1 = \text{fl}\left(\frac{b_1}{l_{11}}\right) = \frac{b_1}{l_{11}(1+\delta)} \quad (\text{با استفاده از (11-1-4)})$$

که در آن

$$|\delta_1| \leq \mu$$

این نتیجه می‌دهد

$$l_{11}(1 + \delta_1) \hat{y}_1 = b_1$$

یا

$$\hat{l}_{11} \hat{y}_1 = b_1$$

که در آن

$$\hat{l}_{11} = l_{11}(1 + \delta_1)$$

این نشان می‌دهد که \hat{y}_1 جواب دقیق معادله‌ای است که ضریب آن یک عدد نزدیک به l_{11}

می‌باشد.

گام ۲: به طور مشابه

$$\hat{y}_2 = fl \left[\frac{b_2 - l_{21} \hat{y}_1}{l_{22}} \right] = fl \left[\frac{b_2 - fl(l_{21} \hat{y}_1)}{l_{22}} \right] \quad (3-2-11)$$

$$= \frac{(b_2 - l_{21} \hat{y}_1 (1 + \delta_{11})) (1 + \delta_{22})}{l_{22} (1 + \delta_2)} \quad (4-2-11)$$

(با استفاده از (۲-۱-۱۱) و (۴-۱-۱۱))، در اینجا $|\delta_{11}|$ ، $|\delta_{21}|$ ، و $|\delta_2|$ همگی کمتر یا مساوی μ هستند. رابطه (۴-۲-۱۱) را می‌توان به صورت زیر نوشت

$$l_{21}(1 + \delta_{11})(1 + \delta_{22})\hat{y}_1 + l_{22}(1 + \delta_2)\hat{y}_2 = b_2(1 + \delta_{22})$$

یا

$$l_{21}(1 + \varepsilon_{21})\hat{y}_1 + l_{22}(1 + \varepsilon_{22})\hat{y}_2 = b_2$$

که در آن

$$\varepsilon_{21} = \delta_{11}, \quad \varepsilon_{22} = \frac{\delta_2 - \delta_{22}}{1 + \delta_{22}}$$

(با حذف δ_{11} ، δ_{22} ، که کوچک است). بنابراین می‌توان گفت که \hat{y}_1 و \hat{y}_2 در رابطه

$$\hat{l}_{21}\hat{y}_1 + \hat{l}_{22}\hat{y}_2 = b_2 \quad (5-2-11)$$

صدق می‌کند که در آن

$$\hat{l}_{21} = l_{21}(1 + \varepsilon_{21}) \quad \text{و} \quad \hat{l}_{22} = l_{22}(1 + \varepsilon_{22})$$

گام k : مراحل قبلی را می‌توان به سادگی تعمیم داد و می‌توان گفت که در گام k ام، مجهولهای y_k تا y_k در رابطه

$$\hat{l}_{k1}\hat{y}_1 + \hat{l}_{k2}\hat{y}_2 + \dots + \hat{l}_{kk}\hat{y}_k = b_k \quad (6-2-11)$$

صدق می‌کنند که در آن $\hat{l}_{kj} = l_{kj}(1 + \varepsilon_{kj})$ ، $j = 1, \dots, k$.

فرایند می‌تواند تا $k = n$ ادامه یابد.

بنابراین ملاحظه می‌کنیم که \hat{y}_1 تا \hat{y}_n محاسبه شده در دستگاه مثلثی اختلال یافته زیر

صدق می‌کنند

$$\hat{l}_{11}\hat{y}_1 = b_1$$

$$\hat{l}_{21}\hat{y}_1 + \hat{l}_{22}\hat{y}_2 = b_2$$

⋮

$$\hat{l}_{n1}\hat{y}_1 + \hat{l}_{n2}\hat{y}_2 + \dots + \hat{l}_{nn}\hat{y}_n = b_n$$

که در آن $\hat{l}_{kj} = l_{kj}(1 + \varepsilon_{kj})$ ، $k = 1, \dots, n$ ، $j = 1, \dots, k$. توجه کنید که $\varepsilon_{11} = \delta_1$.

این معادلات را می‌توان به شکل ماتریسی زیر نوشت

$$\hat{L}\hat{y} = (L + \Delta L)\hat{y} = b \quad (7-2-11)$$

در اینجا ΔL یک ماتریس پایین مثلثی است که عنصر (j, i) آن برابر $(\Delta L)_{ij} = l_{ij}\varepsilon_{ij}$ می‌باشد.

با شناخت کرانها برای ε_{ij} ها، کرانها برای $(\Delta L)_{ij}$ ها می‌توانند به سادگی محاسبه شوند. برای

مثال، اگر n به اندازه کافی کوچک باشد به طوری که $\frac{1}{10} < n\mu$ ، آنگاه

$$|\varepsilon_{kj}| \leq 1/06(k - j + 2)\mu \quad (\text{فصل ۲ بخش ۲-۵ را ملاحظه کنید}).$$

$$|(\Delta L)_{ij}| \leq 1/06(i - j + 2)\mu |l_{ij}| \quad (8-2-11)$$

مطالب فوق را می‌توان در قضیه زیر خلاصه کرد.

قضیه ۱-۲-۱۱ جواب محاسبه شده \hat{y} برای دستگاه پایین مثلثی $n \times n$ ، $Ly = b$ ، به دست آمده با

روش حذفی پیشرو، در دستگاه مثلثی اختلال یافته

$$(L + \Delta L)\hat{y} = b$$

صدق می‌کند که در آن عناصر ΔL در رابطه (۸-۲-۱۱) صدق می‌کنند؛ با فرض این که $\frac{1}{10} < n\mu$.

حالت ۲. دستگاه بالا مثلثی

تحلیل خطای گرد کردن حل دستگاه بالا مثلثی

$$Ux = c$$

با استفاده از جایگزینی پسرو مشابه حالت ۱ است. در این مورد قضیه زیر را بیان

می‌کنیم.

قضیه ۲-۲-۱۱ فرض کنید U یک ماتریس بالا مثلثی $n \times n$ و c یک بردار باشد. آنگاه جواب

محاسبه شده \hat{x} برای دستگاه

$$Ux = c$$

با استفاده از جایگزینی پسرو در دستگاه

$$(U + \Delta U) \bar{x} = c \quad (9-2-11)$$

صدق می‌کند که در آن

$$|(\Delta U)_{ij}| \leq 1/06(j-i+2)\mu |u_{ij}| \quad (10-2-11)$$

با فرض این که $n\mu < \frac{1}{10}$.

۳-۱۱ آنالیز خطای پسرو برای مثلثی سازی با استفاده از روش حذفی گاوس

طرز عمل در اینجا خیلی نزدیک به طرز عمل ارائه شده در مراجع (Ortega (1990) و Forsythe and Moler (1967) است.

یادآور می‌شویم که فرایند مثلثی سازی با استفاده از روش حذفی گاوس از $n-1$ گام تشکیل می‌شود. در گام k ، ماتریس $A^{(k)}$ ، که در k ستون اول مثلثی است، ساخته می‌شود؛ یعنی

$$A^{(k)} = \begin{bmatrix} a_{11}^{(k)} & \dots & \dots & a_{1n}^{(k)} \\ & \ddots & & \vdots \\ & & a_{kk}^{(k)} & \dots & a_{kn}^{(k)} \\ & & \vdots & \ddots & \vdots \\ & & a_{nk}^{(k)} & \dots & a_{nn}^{(k)} \end{bmatrix}$$

ماتریس نهایی، $A^{(n-1)}$ ، مثلثی است. فرض خواهیم کرد که مقادیر $a_{ij}^{(k)}$ اعداد محاسبه شده باشند. ابتدا اجازه دهید محاسبه عناصر $A^{(1)}$ از A را در اولین گام تحلیل نماییم. فرض کنید ضربگرهای محاسبه شده مقادیر

$$-\bar{m}_{i1}, \quad i = 2, 3, \dots, n$$

باشند. آنگاه

$$\bar{m}_{i1} = \text{fl}\left(\frac{a_{i1}}{a_{11}}\right) = \frac{a_{i1}}{a_{11}}(1 + \delta_{i1})$$

که در آن

$$|\delta_{i1}| \leq \mu$$

بنابراین خطای $e_{i_1}^{(0)}$ در قرار دادن $a_{i_1}^{(1)}$ برابر صفر به صورت زیر می‌باشد.

$$\begin{aligned} e_{i_1}^{(0)} &= a_{i_1}^{(1)} - a_{i_1} + \widehat{m}_{i_1} a_{11} \\ &= 0 - a_{i_1} + \left[\frac{a_{i_1}}{a_{11}} \right] (1 + \delta) a_{11} \\ &= \delta_{i_1} a_{11} \end{aligned} \quad (2-3-11)$$

اکنون اجازه دهید خطاها را در محاسبه عناصر $a_{ij}^{(1)}$ ماتریس $A^{(1)}$ پیدا کنیم. عناصر محاسبه شده $a_{ij}^{(1)}$ ، $j = 2, \dots, n$ به صورت زیر ارائه می‌شوند.

$$\begin{aligned} a_{ij}^{(1)} &= \text{fl}(a_{ij} - \text{fl}(\widehat{m}_{i_1} a_{1j})) \\ &= (a_{ij} - \text{fl}(\widehat{m}_{i_1} a_{1j}))(1 + \alpha_{ij}^{(1)}) \\ &= [a_{ij} - \widehat{m}_{i_1} a_{1j}(1 + \beta_{ij}^{(1)})](1 + \alpha_{ij}^{(1)}), \quad i, j = 2, \dots, n \end{aligned} \quad (3-3-11)$$

که در آن

$$\left| \alpha_{ij}^{(1)} \right| \leq \mu, \quad \left| \beta_{ij}^{(1)} \right| \leq \mu$$

رابطه (۳-۳-۱۱) را می‌توان به صورت زیر بازنویسی کرد.

$$a_{ij}^{(1)} = (a_{ij} - \widehat{m}_{i_1} a_{1j}) + e_{ij}^{(0)}, \quad i, j = 2, \dots, n \quad (4-3-11)$$

که در آن

$$e_{ij}^{(0)} = \frac{\alpha_{ij}^{(1)} a_{ij}^{(1)}}{1 + \alpha_{ij}^{(1)}} - \widehat{m}_{i_1} a_{1j} \beta_{ij}^{(1)}, \quad i, j = 2, \dots, n \quad (5-3-11)$$

با توجه به این که سطر اول $A^{(1)}$ همان سطر اول A می‌باشد، از (۴-۳-۱۱) و (۵-۳-۱۱) داریم:

$$A^{(1)} = A - L^{(0)}A + E^{(0)} \quad (6-3-11)$$

که در آن

$$L^{(0)} = \begin{bmatrix} \cdot & \cdot & \dots & \cdot \\ \widehat{m}_{21} & \cdot & \dots & \cdot \\ \vdots & & \ddots & \\ \widehat{m}_{n1} & \cdot & \dots & \cdot \end{bmatrix}, \quad E^{(0)} = \begin{bmatrix} \cdot & \cdot & \dots & \cdot \\ e_{21}^{(0)} & \dots & \dots & e_{2n}^{(0)} \\ \vdots & & & \\ e_{n1}^{(0)} & \dots & \dots & e_{nn}^{(0)} \end{bmatrix}$$

به طور مشابه در انتهای گام ۲ خواهیم داشت.

$$A^{(2)} = A^{(1)} - L^{(1)}A^{(1)} + E^{(1)} \quad (7-3-11)$$

که در آن $L^{(1)}$ و $E^{(1)}$ به طور مشابه تعریف می‌شوند.

با جایگذاری (۶-۳-۱۱) در (۷-۳-۱۱)، داریم:

$$\begin{aligned} A^{(2)} &= A^{(1)} - L^{(1)}A^{(1)} + E^{(1)} \\ &= A - L^{(0)}A + E^{(0)} - L^{(1)}A^{(1)} + E^{(1)} \end{aligned}$$

با ادامه این راه، می‌توانیم بنویسیم

$$\begin{aligned} A^{(n-1)} + L^{(0)}A^{(1)} + \dots + L^{(n-2)}A^{(n-2)} \\ = A + E^{(0)} + E^{(1)} + \dots + E^{(n-2)} \end{aligned} \quad (8-3-11)$$

چون

$$L^{(k-1)} = \begin{bmatrix} \cdot & \cdot & \dots & \cdot \\ & \cdot & & \\ \vdots & \widehat{m}_{k+1,k} & \ddots & \\ \vdots & \vdots & & \vdots \\ \cdot & \widehat{m}_{n,k} & \dots & \cdot \end{bmatrix}$$

داریم

$$L^{(k)}A^{(k)} = L^{(k)}A^{(n-1)}, \quad k = 0, 1, 2, \dots, n-2 \quad (9-3-11)$$

بنابراین از (۸-۳-۱۱) و (۹-۳-۱۱) به دست می‌آوریم

$$\begin{aligned} A^{(n-1)} + L^{(0)}A^{(n-1)} + L^{(1)}A^{(n-1)} + \dots + L^{(n-2)}A^{(n-1)} \\ = A + E^{(0)} + E^{(1)} + \dots + E^{(n-2)} \end{aligned} \quad (10-3-11)$$

$$A + E^{(0)} + E^{(1)} + \dots + E^{(n-2)} \quad (11-3-11)$$

$$= (I + L^{(0)} + L^{(1)} + \dots + L^{(n-2)})A^{(n-1)}$$

اکنون با توجه به این که

$$I + L^{(0)} + L^{(1)} + \dots + L^{(n-2)} = \begin{bmatrix} 1 & \cdot & \cdot & \dots & \cdot \\ \hat{m}_{21} & 1 & \cdot & \dots & \cdot \\ \hat{m}_{31} & \hat{m}_{32} & 1 & \dots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{m}_{n1} & \hat{m}_{n2} & \dots & \hat{m}_{n,n-1} & 1 \end{bmatrix}$$

$$= \hat{L} \text{ (محاسبه شده } L) \quad (12-3-11)$$

و U (محاسبه شده) $\hat{U} = A^{(n-1)}$ و با نمایش $E^{(0)} + E^{(1)} + \dots + E^{(n-2)}$ توسط E ، از (11-3-11) و (12-3-11) داریم

$$A + E = \hat{L}\hat{U} \quad (13-3-11)$$

که در آن ماتریسهای $E^{(0)}, \dots, E^{(n-2)}$ به صورت زیر ارائه می‌شوند.

$$E^{(k-1)} = \begin{bmatrix} \cdot & \cdot & \cdot & \dots & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \dots & \cdot \\ \vdots & \vdots & \vdots & e_{k+1,k}^{(k-1)} & \dots & e_{k+1,n}^{(k-1)} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \cdot & \cdot & \dots & e_{n,k}^{(k-1)} & \dots & e_{n,n}^{(k-1)} \end{bmatrix}, \quad k = 1, 2, \dots, n-1 \quad (11-3-14)$$

$$e_{ik}^{(k-1)} = a_{ik}^{(k-1)} \delta_{ik}, \quad i = k+1, \dots, n \quad (15-3-11)$$

$$e_{ij}^{(k-1)} = \frac{\alpha_{ij}^{(k)}}{1 + \alpha_{ij}^{(k)}} a_{ij}^{(k)} - \hat{m}_{ik} a_{kj}^{(k-1)} \beta_{ij}^{(k)}, \quad i, j = k + 1, \dots, n \quad (16-3-11)$$

و

$$|\delta_{ik}| \leq \mu, \quad \left| \alpha_{ij}^{(k)} \right| \leq \mu \quad (17-3-11)$$

و

$$\left| \beta_{ij}^{(k)} \right| \leq \mu \quad (18-3-11)$$

ما این بحث را در قضیه زیر بیان می‌کنیم:

قضیه ۱-۳-۱۱ ماتریسهای بالا و پایین مثلثی محاسبه شده \hat{L} و \hat{U} تولید شده توسط روش حذفی گاوس در رابطه

$$A + E = \hat{L} \hat{U}$$

صدق می‌کنند، که در آن $\hat{U} = A^{(n-1)}$ و \hat{L} ماتریس پایین مثلثی ضریب‌های محاسبه شده است:

$$\hat{L} = \begin{bmatrix} 1 & \cdot & \dots & \dots & \cdot \\ \hat{m}_{21} & 1 & \cdot & \dots & \cdot \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \hat{m}_{n1} & \hat{m}_{n2} & \dots & \hat{m}_{n,n-1} & 1 \end{bmatrix}$$

مثال ۱-۳-۱۱ با استفاده از حساب دو رقمی در محاسبه L و U ماتریس خطای E را به قسمی پیدا

کنید که $A + E = \hat{L} \hat{U}$

$$A = \begin{bmatrix} 0/21 & 0/35 & 0/11 \\ 0/11 & 0/81 & 0/22 \\ 0/33 & 0/22 & 0/39 \end{bmatrix}$$

گام ۱:

$$a_{22}^{(1)} = 0/81 - 0/52 \times 0/35 = 0/63$$

$$\hat{m}_{21} = -\frac{0/11}{0/21} = -0/52$$

$$a_{23}^{(1)} = 0/22 - 0/52 \times 0/11 = 0/16$$

$$\hat{m}_{31} = \frac{0/33}{0/21} = 1/57$$

$$a_{32}^{(1)} = 0/22 - 1/57 \times 0/35 = -0/33$$

$$a_{33}^{(1)} = 0/39 - 1/57 \times 0/11 = 0/22$$

$$A^{(1)} = \begin{bmatrix} 0/21 & 0/35 & 0/11 \\ \cdot & 0/63 & 0/16 \\ \cdot & -0/33 & 0/22 \end{bmatrix}$$

$$e_{21}^{(\cdot)} = 0 - [0/11 - 0/52 \times 0/21] = -0/0008$$

$$e_{22}^{(\cdot)} = 0/63 - [0/81 - 0/52 \times 0/35] = 0/0020$$

$$e_{23}^{(\cdot)} = 0/10 - [0/22 - 0/52 \times 0/11] = -0/0020$$

$$e_{31}^{(\cdot)} = 0 - [0/33 - 1/57 \times 0/21] = -0/0003$$

$$e_{32}^{(\cdot)} = -0/33 - [0/22 - 1/57 \times 0/35] = -0/0005$$

$$e_{33}^{(\cdot)} = 0/22 - [0/39 - 1/57 \times 0/11] = 0/0027$$

$$E^{(\cdot)} = \begin{bmatrix} \cdot & \cdot & \cdot \\ -0/0008 & 0/0020 & -0/0028 \\ -0/0003 & -0/0005 & 0/0027 \end{bmatrix}$$

گام ۲:

$$\hat{m}_{22} = -\frac{0/33}{0/63} = -0/52 \quad a_{22}^{(2)} = 0/22 + 0/52 \times 0/16 = 0/30$$

$$A^{(2)} = \begin{bmatrix} 0/21 & 0/35 & 0/11 \\ \cdot & 0/63 & 0/16 \\ \cdot & \cdot & 0/30 \end{bmatrix} = \hat{U}$$

$$e_{22}^{(1)} = 0 - [-0/33 + 0/52 \times 0/63] = 0/0024$$

$$e_{23}^{(1)} = -0/30 - [0/22 + 0/52 \times 0/16] = -0/0032$$

$$E^{(1)} = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & 0/0024 & -0/0032 \end{bmatrix}$$

بنابراین

$$E = E^{(\cdot)} + E^{(1)} = \begin{bmatrix} \cdot & \cdot & \cdot \\ -0/0008 & 0/0020 & -0/0028 \\ -0/0003 & 0/0019 & -0/0005 \end{bmatrix}$$

چون

$$\hat{L} = \begin{bmatrix} 1 & \cdot & \cdot \\ 0/52 & 1 & \cdot \\ 1/57 & -0/52 & 1 \end{bmatrix}$$

به راحتی می توانیم تحقیق کنیم که $\hat{L}\hat{U} = A + E$.

کرائنها برای عناصر E

اکنون ارزیابی می کنیم که چقدر عناصر ماتریس خطای E می توانند بزرگ باشند. برای این منظور فرض می کنیم که از محورگیری در روش حذفی گاوس استفاده شده باشد، در نتیجه $|\hat{m}_{ik}| \leq 1$. از فصل ۳ و ۵ یادآور می شویم که عامل رشد ρ به صورت زیر تعریف می شود

$$\rho = \frac{\max_{i, j, k} |a_{ij}^{(k)}|}{\max_{i, j} |a_{ij}|}$$

فرض کنید $a = \max_{i, j} |a_{ij}|$. پس از (۱۱-۳-۱۱) الی (۱۱-۳-۱۸) داریم:

$$|e_{ik}^{(k-1)}| \leq \rho \mu, \quad k = 1, 2, \dots, n-1, i = k+1, \dots, n$$

و به ازای $(k = 1, 2, \dots, n-1)$ ، $i, j = k+1, \dots, n$ داریم:

$$|e_{ij}^{(k-1)}| \leq \frac{\mu}{1-\mu} |a_{ij}^{(k)}| + \mu |a_{ij}^{(k-1)}| \leq \frac{2}{1-\mu} \rho \mu \quad (|\hat{m}_{ik}| \leq 1 \text{ چون})$$

با نمایش $\mu/(1-\mu)$ توسط η خواهیم داشت

$$E = |E^{(0)} + \dots + E^{(n-2)}|$$

$$\leq |E^{(0)}| + \dots + |E^{(n-2)}|$$

$$\leq \rho \eta \left[\begin{array}{cccc} \cdot & \cdot & \dots & \cdot \\ 1 & 2 & \dots & 2 \\ \vdots & \vdots & & \vdots \\ 1 & 2 & \dots & 2 \end{array} \right] + \left[\begin{array}{cccc} \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & 1 & 2 & \dots & 2 \\ \vdots & \vdots & \vdots & & \vdots \\ \cdot & 1 & 2 & \dots & 2 \end{array} \right]$$

$$+ \dots + \left[\begin{array}{cccc} \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \dots & \cdot & \cdot \\ \vdots & \vdots & & \vdots & \vdots \\ \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \dots & 1 & 2 \end{array} \right]$$

$$\leq \rho \eta \left[\begin{array}{cccc} \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & 2 & 2 & \dots & 2 \\ 1 & 3 & 4 & \dots & 4 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 3 & 5 & \dots & 2n-2 \end{array} \right]$$

تبصره: نامساویهای (۱۱-۳-۱۹) عنصر به عنصر برقرارند. فوراً می‌توانیم یک کران برحسب نرم‌ها به دست آوریم. بنابراین

$$\|E\|_{\infty} \leq apn(1 + 3 + \dots + (2n - 2))$$

$$\leq apn^2 \eta \quad (20-3-11)$$

۴-۱۱ تحلیل خطای پسرو برای حل $Ax = b$

اکنون برای ارائه یک تحلیل خطای گرد کردن پسرو برای حل $Ax = b$ ، با استفاده از مثلثی‌سازی روش حذفی گاوس و ادامه آن یا حذفی پیشرو و جایگزینی پسرو، آماده هستیم.

ابتدا از قضیه ۱۱-۳-۱، می‌دانیم که مثلثی‌سازی A با استفاده از روش حذفی گاوس ماتریسهای \hat{L} و \hat{U} را به قسمی نتیجه می‌دهد که $A + E = \hat{L}\hat{U}$.

سپس این ماتریسهای \hat{L} و \hat{U} باید برای حل دستگاههای زیر استفاده شوند

$$\hat{L}y = b$$

$$\hat{U}x = y$$

از قضیه ۱۱-۲-۱ و قضیه ۱۱-۲-۲ می‌دانیم که جوابهای محاسبه شده \hat{y} و \hat{x} این دو دستگاه مثلثی در معادلات زیر صدق می‌کنند

$$(\hat{L} + \Delta L)\hat{y} = b \quad \text{و} \quad (\hat{U} + \Delta U)\hat{x} = \hat{y}$$

از این معادلات داریم

$$(\hat{U} + \Delta U)\hat{x} = (\hat{L} + \Delta L)^{-1}b$$

یا

$$(\hat{L} + \Delta L)(\hat{U} + \Delta U)\hat{x} = b$$

یا

$$(A + F)\hat{x} = b \quad (11-4-4)$$

که در آن

$$F = E + (\Delta L)\hat{U} + \hat{L}(\Delta U) + (\Delta L)(\Delta U) \quad (2-4-11)$$

(توجه کنید که $\hat{L}\hat{U} = A + E$.)

کرانه‌های برای F

با گرفتن $\|\cdot\|_\infty$ از (۱۱-۴-۲) خواهیم داشت

$$\|F\|_\infty \leq \|E\|_\infty + \|\Delta L\|_\infty \|\hat{U}\|_\infty + \|\hat{L}\|_\infty \|\Delta U\|_\infty + \|\Delta L\|_\infty \|\Delta U\|_\infty$$

اکنون عبارتهایی را برای $\|\Delta L\|_\infty$ ، $\|\Delta U\|_\infty$ ، $\|\hat{L}\|_\infty$ ، و $\|\hat{U}\|_\infty$ به دست می‌آوریم. چون

$$\hat{L} = \begin{bmatrix} 1 & & & \\ \hat{m}_{21} & \ddots & & \\ \vdots & & \ddots & \\ \hat{m}_{n1} & \dots & \hat{m}_{n,n-1} & 1 \end{bmatrix}$$

از (۱۱-۲-۸) به دست می‌آوریم

$$|\Delta L| \leq 1/0.6\mu \begin{bmatrix} 2 & & & \\ 3|\hat{m}_{21}| & 2 & & \\ \vdots & & \ddots & \\ (n+1)|\hat{m}_{n1}| & \dots & 3|\hat{m}_{n,n-1}| & 2 \end{bmatrix} \quad (۱۱-۴-۳)$$

با فرض این که محورگیری جزئی انجام شده باشد، یعنی $|\hat{m}_{ik}| \leq 1$ ، $k = 1, 2, \dots, n-1$ ، $i = k+1, \dots, n$ داریم:

$$\|\hat{L}\|_\infty \leq n \quad (۱۱-۴-۴)$$

$$\|\Delta L\|_\infty \leq \frac{n(n+2)}{2} (1/0.6)\mu \quad (۱۱-۴-۵)$$

به طور مشابه

$$\|\hat{U}\|_\infty \leq nap \quad (۱۱-۴-۶)$$

(توجه کنید که $U = A^{(n-1)}$) و از (۱۱-۲-۱۰) داریم

$$\|\Delta U\|_\infty \leq \frac{n(n+2)}{2} 1/0.6ap\mu \quad (۱۱-۴-۷)$$

(توجه کنید که $\max \|u_{ij}\|_\infty \leq ap$). همچنین یادآور می‌شویم که

$$\|E\|_\infty \leq n^2 ap \frac{\mu}{1-\mu} \quad (۱۱-۴-۸)$$

فرض کنید که $1 < \mu(1/0.6)^2 \frac{(n+3)^2}{4}$ ، آنگاه

$$\|\Delta L\|_{\infty} \cdot \|\Delta U\|_{\infty} \leq n^2 a \rho \mu \quad (9-4-11)$$

سپس با استفاده از (۴-۴-۱۱) الی (۸-۴-۱۱) در (۲-۴-۱۱) خواهیم داشت

$$\begin{aligned} \|F\|_{\infty} &\leq \|E\|_{\infty} + \|\Delta L\|_{\infty} \|\hat{U}\|_{\infty} + \|\hat{L}\|_{\infty} \|\Delta U\|_{\infty} \\ &+ \|\Delta L\|_{\infty} \|\Delta U\|_{\infty} \\ &\leq n^2 a \rho \frac{\mu}{1-\mu} + (1/0.6) n^2 (n+3) a \rho \mu + n^2 \rho a \mu \quad (10-4-11) \end{aligned}$$

چون $1/0.6 < \frac{1}{1-\mu}$ ، و $\|A\|_{\infty} \leq a$ از (۱۰-۴-۱۱) می توان نوشت

$$\|F\|_{\infty} \leq (1/0.6) (n^3 + 5n^2) \rho \|A\|_{\infty} \mu \quad (11-4-11)$$

با نادیده گرفتن جملات حاوی $n^2 \mu$ ، این بحث منجر به قضیه زیر می شود.

قضیه ۱-۴-۱۱ جواب محاسبه شده دستگاه

$$Ax = b$$

با استفاده از روش حذفی گاوس با محورگیری جزئی در یک دستگاه اختلال یافته

$$(A + F)\hat{x} = b$$

صدق می کند، که در آن F توسط رابطه (۲-۴-۱۱) تعریف می شود و $\|F\|_{\infty} \leq cn^3 \rho \|A\|_{\infty} \mu$ ، که در آن c یک ثابت از مرتبه واحد است.

تبصره ها:

۱. کران قبلی برای F به طور فاحشی زیاد برآورده شده است. در عمل این کران خیلی بندرت برای F حاصل می گردد. ویلکینسون ((Wilkinson (1965)) بیان نموده است که در عمل $\|F\|_{\infty}$ همیشه کوچکتر یا مساوی $\|A\|_{\infty} n \mu$ است.
۲. با استفاده از (۸-۲-۱۱)، (۱۰-۲-۱۱) و (۱۴-۳-۱۱) الی (۱۹-۳-۱۱)، همچنین می توانیم یک کران عنصر به عنصر برای F به دست آوریم (تمرین ۴).

۴-۱۱ مرور و خلاصه

در این فصل ما تحلیلهای خطای پسرو را برای موارد زیر ارائه داده ایم.

۱. دستگاههای پایین و بالا مثلثی با استفاده از روش حذفی پیشرو و جایگزینی پسرو (قضیه‌های ۱۱-۲-۱ و ۱۱-۲-۲).

۲. تجزیه LU با استفاده از روش حذفی گاوس با محورگیری جزئی (قضیه ۱۱-۳-۱).

۳. مسأله دستگاه خطی $Ax = b$ با استفاده از روش حذفی گاوس با محورگیری جزئی و ادامه آن با حذفی پیشرو و جایگزینی پسرو (قضیه ۱۱-۴-۱).

کرانه‌ها را برای ماتریس خطای E در هر حالت به دست آورده‌ایم، $(۱۱-۳-۱۱)$ ، $(۱۱-۴-۱۱)$ ، $(۱۰-۲-۱۱)$ ، $(۸-۲-۱۱)$.

در اینجا ما فقط سعی کرده‌ایم نمونه‌های کوچکی از تحلیل گرد کردن را، همان گونه که عنوان فصل اظهار می‌دارد، برای خواننده ارائه دهیم. نتایج این فصل هم اکنون برای خواننده معلوم هستند. این نتایج قبلاً در کتاب بدون اثبات بیان شده‌اند. ما سعی کرده‌ایم اثباتهای رسمی را در اینجا ارائه دهیم.

تکرار می‌کنیم که این نتایج بیان می‌کنند که روشهای حذفی پیشرو و جایگزینی پسرو برای دستگاههای مثلثی پایدار پسرو هستند، در حالی که پایداری فرایند حذفی گاوس برای تجزیه LU ، و بنابراین برای مسأله دستگاه خطی $Ax = b$ با استفاده از این فرایند، بستگی به اندازه عامل رشد دارد.

تمرینها:

۱. با استفاده از $\beta = ۱۰$ و $t = ۲$ تجزیه LU را با استفاده از روش حذفی گاوس (بدون محورگیری)

برای ماتریسهای زیر حساب کنید، و ماتریس خطای E را در هر حالت به قسمی پیدا کنید که

$$A + E = LU$$

$$A = \begin{bmatrix} ۳ & ۴ \\ ۵ & ۶ \end{bmatrix} \quad (\text{الف})$$

$$\begin{bmatrix} ۰/۲۵ & ۰/۷۹ \\ ۰/۰۱ & ۰/۱۲ \end{bmatrix} \quad (\text{ب})$$

$$\begin{bmatrix} ۱۰ & ۹ \\ ۸ & ۵ \end{bmatrix} \quad (\text{ج})$$

$$\begin{bmatrix} ۴ & ۱ \\ ۱ & ۲ \end{bmatrix} \quad (\text{د})$$

$$\begin{bmatrix} ۰/۰۱ & ۰/۰۵ \\ ۰/۰۳ & ۰/۰۱ \end{bmatrix} \quad (\text{ه})$$

۲. اکنون فرض کنید که محورگیری جزئی در محاسبه تجزیه LU هر یک از ماتریسهای تمرین ۱ استفاده شده باشد. ماتریس خطای E را در هر حالت پیدا کنید و کرانه‌های عناصر E پیش‌بینی شده توسط (۱۱-۳-۱۹) را با خطاهای واقعی مقایسه کنید.

۳. مسائل حل دستگاهای خطی

$$Ax = b$$

را با استفاده از روش حذفی گاوس با محورگیری جزئی و با هر یک از ماتریسهای تمرین ۱ اختیار کردن $b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ در هر حالت در نظر بگیرید. F را در هر حالت به قسمی پیدا کنید که جواب محاسبه شده x در معادله

$$(A + F)x = b$$

صدق نماید. کرانه‌های پیش‌بینی شده توسط (۱۱-۴-۱۱) را با خطاهای واقعی مقایسه کنید.

۴. از (۱۱-۲-۸)، (۱۱-۲-۱۰) و (۱۱-۳-۱۳) الی (۱۱-۳-۱۹) استفاده کنید و یک کران جمله به جمله برای F در قضیه ۱۱-۴-۱۱ پیدا کنید:

$$(A + F)\tilde{x} = b$$

همچنین یک کران برای مانده پیدا کنید.

۵. از قضیه‌های ۱۱-۲-۱ و ۱۱-۲-۲ شما چه می‌توانید در مورد پایداری پسر و فرآیندهای حذفی پسر و جایگزینی پسر برای دستگاهای پایین و بالا مثلثی بگویید.

۶. از (۱۱-۳-۲۰) نتیجه بگیرید که پایداری پسر و روش حذفی گاوس اساساً توسط اندازه عامل رشد ρ تعیین می‌شود.

۷. مسأله محاسبه چند جمله‌ای

$$p(\alpha) = a_n \alpha^n + a_{n-1} \alpha^{n-1} + \dots + a_0$$

را با تقسیم ترکیبی در نظر بگیرید:

$$p_n = a_n$$

$$p_{i-1} = \text{fl}(\alpha p_i + a_{i-1}), \quad i = n, n-1, \dots, 1$$

پس $p_0 = p(\alpha)$. نشان دهید که

$$p_0 = a_n(1 + \delta_n)\alpha^n + a_{n-1}(1 + \delta_{n-1})\alpha^{n-1} + \dots + a_0(1 + \delta_0)$$

یک کران برای هر δ_i ، $i = 0, 1, \dots, n$ پیدا کنید. از روی کرانه‌ها حاصل درباره پایداری پسر و

الگوریتم چه می‌توانید بگویید؟

۸. نشان دهید که الگوریتم چولسکی (الگوریتم ۶-۴-۴) یک ماتریس پایین مثلثی \bar{L} به قسمی محاسبه می‌کند که $A + E = \bar{L}(\bar{L})^T$. یک کران برای عناصر ماتریس E بیابید. از اینرو ثابت کنید که الگوریتم چولسکی پایدار است.

پیشنهادها برای مطالعه بیشتر

معتبرترین کتاب برای یادگیری جزئیات تحلیلهای خطای پسرو الگوریتمها برای محاسبات ماتریسی کتاب (1965) *Wilkinson's Algebraic Eigenvalue Problem* است.

کتاب دیگر ویلکینسون در این زمینه (1963) *Rounding Errors in Algebraic Process* نیز یک منبع خوب از دانستنیها است. کتابهای آنالیز عددی،

A Second Course, by James Ortega (1990),

Computer Solution of Linear Algebraic Systems, by Forsythe and Moler (1967)

همچنین دارای بحثهای عمیقی در مورد مطالب این فصل هستند. هم اکنون یک کتاب کامل مختص تحلیل خطا توسط ان. جی. هایام (N. J. Higham) در دست تهیه است (در آینده نزدیک چاپ خواهد شد).

مقدمه مختصری برای MATLAB

A-1 برخی اطلاعات اساسی در مورد MATLAB

A-1-1 MATLAB چیست؟

MATLAB نشان دهنده کلمات MATrix LABoratory است. یک بسته نرم‌افزاری محاوره‌ای برای حل مسائل مطرح در محاسبات علمی و مهندسی است. MATLAB حاوی برنامه‌هایی برای همه محاسبات ماتریسی، نظیر جواب دستگاه‌های خطی، انواع تجزیه‌های ماتریسی، جوابهای مسائل کمترین توانهای دوم، محاسبات مقادیر ویژه و بردارهای ویژه، محاسبه مقدار تکین و بردار تکین است. رایج‌ترین بسته MATLAB همچنین حاوی برنامه‌هایی برای انواع زیاد دیگری از محاسبات، شامل قابلیت‌های ترسیم دوبعدی است.

A-1-2 ورود و خروج از MATLAB

در بیشتر دستگاهها دستور `matlab` به شما اجازه ورود به MATLAB را می‌دهد. دستور `exit` برای خروج از MATLAB به کار می‌رود.

A-1-3 دو دستور مهم: `demo` و `help`

دو دستور مهم `demo` و `help` می‌باشند. با تایپ دستور `help` شما فهرست همه توابع MATLAB و تواناییهای دیگر MATLAB را به دست خواهید آورد. تایپ `help` و به دنبال آن یک نام تابع

MATLAB از فهرست به شما اطلاعات خاص بیشتری را در مورد آن تابع خاص ارائه می‌دهد.

مثال A-1-1

```
>> help norm
```

NORM Matrix or Vector norm.

For matrices ..

NORM(X) is the Largest singular value of X, max (svd(X)).

NORM(X, 2) is the same as NORM(X).

NORM(X, 1) is the 1-norm of X, the largest column sum.

$$= \max (\text{sum}(\text{abs}(\{X\}))).$$

NORM(X, inf) is the infinity norm of X, the largest row sum,

$$= \max (\text{sum}(\text{abs}(\{X'\}))).$$

NORM(X, 'inf') is same as NORM(X,inf).

NORM(X, fro) is the F-norm, sqrt(sum(diag(X'*X))).

NORM(X, P) is available for matrix X only if P is 1, 2, inf

or 'fro'.

For vectors

$$\text{NORM}(V, P) = \text{sum}(\text{abs}(V)^P)^{(1/P)}.$$

$$\text{NORM}(V) = \text{norm}(V, 2).$$

$$\text{NORM}(V, \text{inf}) = \max(\text{abs}(V)).$$

$$\text{NORM}(V, -\text{inf}) = \min(\text{abs}(V)).$$

در MATLAB، اگر X دارای مؤلفه‌های مختلط، z ، باشد آنگاه

$$\text{abs}(z) = \sqrt{z \cdot \text{conj}(z)} \text{ و برابر } \text{abs}(\text{real}(z)) + \text{abs}(\text{imag}(z)) \text{ نیست}$$

که در برخی نسخه‌های قدیمی‌تر MATLAB استفاده می‌شد.

همچنین CONDEST، RCOND، COND و NORMEST را ملاحظه کنید.

DEMO به شما یاد می‌دهد که چگونه از تابع‌های MATLAB استفاده کنید، مثلاً چگونه مقادیر

ماتریس را در یک ماتریس وارد کنید، چگونه ترانهاده آن را پیدا کنید، چگونه رتبه یک ماتریس را پیدا

کنید و غیره.

A-1-۲ مورد استفاده‌ترین اعمال و توابع MATLAB

برخی اعمال ماتریسی اساسی

+	جمع
-	تفریق
*	حاصلضرب ماتریسی
.*	حاصلضرب آرایه‌ای
^	توان
\	(ممیز برعکس) یا تقسیم چپ
/	(ممیز) یا تقسیم راست
&	AND منطقی
	OR منطقی
~	NOT منطقی

توابع ماتریسی: جبر خطی عددی

تحلیل ماتریسی

cond	عدد وضعیت ماتریس
norm	نرم برداری یا نرم ماتریسی
rcond	برآوردگر معکوس وضعیت LINPACK
rank	تعداد سطرها یا ستونهای مستقل خطی
det	دترمینان
trace	مجموع عناصر قطری
null	فضای پوچ
orth	متعامد سازی
rref	شکل سطری پلکانی کاهش یافته

معادلات خطی

\ /	حل دستگاه خطی؛ از دستور "help slash" استفاده کنید.
chol	تجزیه چولسکی
lu	عاملها از روش حذفی گاوس

معکوس ماتریس	inv
تجزیه متعامد - مثلثی	qr
حذف یک ستون از تجزیه QR	qrdelete
وارد کردن یک ستون در تجزیه QR	qrinsert
شبه معکوس	pinv
کمترین توانهای دوم به هنگام معلوم بودن کوواریانس	lscov

مقادیر ویژه و مقادیر تکین

مقادیر ویژه و بردارهای ویژه	eig
چند جمله‌ای مشخصه	poly
شکل هسبرگی	hess
تجزیه مقدار تکین	svd
مقادیر ویژه تعمیم یافته	qz
شکل قطری بلوکی حقیقی به شکل قطری مختلط	rsf2csf
شکل قطری مختلط به شکل قطری بلوکی حقیقی	cdf2rdf
تجزیه شور	schur
مقیاس کردن قطری برای بهبود دقت مقدار ویژه	balance

با استفاده از **help** شما می‌توانید اطلاعاتی در مورد هر یک از روالهای فوق به دست آورید. یک مثال در زیر ارائه شده است.

مثال A-۱-۲

```
>> help lu
LU Factors from Gaussian elimination.
[L, U] = LU(X) stores an upper triangular matrix in U a
"psychologically lower triangular matrix," i.e., a product of
lower triangular and permutation matrices, in L, so that
X = L*U.
[L, U, P] = LU(X) returns lower triangular matrix L, upper
triangular matrix U, and permutation matrix P so that P*X
= L*U.
By itself, LU(X) returns the output from LINPACK'S ZGEFA
routine.
```

A-1-5 مثالهای عددی

وارد کردن یک ماتریس و یک بردار در MATLAB:

```
>> A = [1 3 5 ; 2 4 6; 1 3 9]
```

```
A =
```

```
1 3 5
2 4 6
1 3 9
```

```
>> b = [1 1 1]'
```

```
b =
```

```
1
1
1
```

حل دستگاه $Ax = b$

```
>> x = A\b;
```

```
x =
```

```
-0.5000
0.5000
0
```

محاسبه معکوس A :

```
>> inv(A)
```

```
ans =
```

```
-2.2500    1.5000    0.2500
 1.5000   -0.5000   -0.5000
-0.2500         0    0.2500
```

محاسبه مقادیر ویژه A :

```
>> eig(A)
```

```
ans =
```

```
-0.3246
12.3246
2.0000
```


کاهش A به ماتریس بالا هسنبرگی:

```
>> hess(A)
ans =
    1.0000   -4.9193    3.1305
   -2.2361    8.6000   -6.2000
         0    -3.2000    4.4000
```

پیدا کردن رتبه A :

```
>> rank(A)
ans =
    3
```

محاسبه نرم دو A :

```
>> norm(A)
ans =
   13.3538
```

محاسبه عدد وضعیت A (نسبت به نرم دو):

```
>> cond(A)
ans =
   42.1539
```

محاسبه تجزیه QR ماتریس A با استفاده از ماتریسهای هاوس هولدر:

>> $[q,r] = qr(A)$ تجزیه QR ماتریس A را ارائه خواهد داد

```
q =
   -0.4082    0.5774   -0.7071
   -0.8165   -0.5774   -0.0000
   -0.4082    0.5774    0.7071

r =
   -2.4495   -5.7155  -10.6145
         0    1.1547    4.6188
         0         0    2.8284
```

محاسبه تجزیه LU ماتریس A با استفاده از محورگیری جزئی:

```
>> [l,u] = lu(A)
```

```
l =
```

```
    0.5000    1.0000    0
    1.0000    0    0
    0.5000    1.0000    1.0000
```

```
u =
```

```
    2    4    6
    0    1    2
    0    0    4
```

A - ۲ نوشتن برنامه‌های خودتان با استفاده از دستوره‌های MATLAB

نوشتن برنامه‌های خودتان با استفاده از دستوره‌های MATLAB ساده است. ابتدا برخی از متداول‌ترین استفاده‌ها را فهرست نموده و سپس مثالهایی ارائه خواهیم داد.

ما شما را ترغیب می‌کنیم که *MATLAB Primer* را که توسط پرفسور کرمیت سیگمون^۱ از دانشگاه فلوریدا تهیه شده به دست آورید. این نسخه یک مقدمه کامل‌تر برای استفاده کارا از MATLAB ارائه می‌دهد. این نسخه را می‌توانید از طریق آدرس پستی زیر تهیه نمایید

CRC Press, Inc.

2000 Corporate Blvd., N.W.

Boca Raton, FL 33431

USA

یا از طریق تلفن

1-800-272-7737

قاره آمریکا، بجز فلوریدا

1-407-994-0555

فلوریدا و خارج قاره آمریکا

A-۲-۱ برخی اعمال ارتباطی در MATLAB

< کوچکتر از

> بزرگتر از

<= کوچکتر از یا مساوی با

>= بزرگتر از یا مساوی با

== مساوی

۲-۲-A برخی توابع مربوط به ساختمان ماتریسی

ماتریس واحد	eye
ماتریس صفر	zeros
ماتریس تولید شده تصادفی با عناصر بین صفر و ۱	rand
عنصر ماکزیمم	max
ایجاد یا استخراج یک قطر	diag
قسمت بالا مثلثی یک ماتریس	triu
قسمت پایین مثلثی یک ماتریس	tril
ماتریس هیلبرت	hilb
ماتریس توپلیتز	toeplitz
بعد یک ماتریس را ارائه می دهد	size
قدر مطلق عناصر یک ماتریس یا یک بردار	abs

مثال ۱-۲-A

۱. $\text{rand}(5, 3)$ یک ماتریس تصادفی 5×3 تولید خواهد کرد.
۲. اگر x یک بردار باشد، $\text{diag}(x)$ ماتریس قطری با عناصر x بر روی قطر می باشد؛ $\text{diag}(A)$ بردار متشکل از قطر A است.
۳. $\text{hilb}(5)$ ماتریس هیلبرت 5×5 تولید خواهد کرد.
۴. $\text{max}(A(:, 2))$ ماکزیمم مقدار ستون دوم A را ارائه خواهد داد؛ $\text{max}(\text{max}(A))$ بزرگترین عنصر همه ماتریس A را ارائه خواهد داد.

۳-۲-A نماد دو نقطه (:)

نماد (:): برای نشان دادن قراردادی یک سطر یا ستون کامل است.

مثال ۲-۲-A

$A(:, 2)$ تمام ستون دوم A را نمایش می دهد.

$A(1 : 3, :)$ سه سطر اول A را نمایش می دهد.

$A(:, [2, 5, 7])$ ستونهای دوم، پنجم، و هفتم A را نمایش می دهد.

عبارت $A(:, [2, 5, 7]) = B(:, 1 : 3)$ سه ستون اول B را در ستونهای ۲، ۵، ۷ ماتریس A

قرار می دهد.

A-۲-۴ دستوره‌های `for`، `while`، و `if`

دستوره‌های `for`، `while`، و `if` سودمندترین دستورها در نوشتن برنامه‌های MATLAB برای الگوریتمهای ماتریسی هستند. این دستورها را در مثالهای زیر توضیح می‌دهیم.

A-۲-۵ محاسبه تعداد اعمال ممیز شناور و زمان سپری شده یک الگوریتم

برای محاسبه تعداد اعمال ممیز شناور یک الگوریتم `flops(0)` را بلافاصله قبل از اجرای الگوریتم قرار دهید و `flops` را بلافاصله بعد از کامل شدن الگوریتم تایپ کنید.

مثال A-۲-۳

```
flops(0)
```

```
x = a\b
```

```
flops
```

کل اعمال ممیز شناور را برای حل دستگاه خطی با یک ماتریس مفروض a و یک بردار مفروض b ارائه می‌دهد.

تابع `cputime` زمان CPU بی‌راکه توسط فرایند MATLAB از زمان شروع فرایند MATLAB استفاده شده، بر حسب ثانیه برمی‌گرداند. برای مثال دستور

```
t = cputime; your-operation; cputime - t
```

`cputime` را برای اجرای اعمال شما برمی‌گرداند. چون نسخه PC نرم‌افزار MATLAB یک تابع `cputime` ندارد، MATCOM حاوی یک برنامه `cputime.m` می‌باشد. اگر نرم‌افزار MATLAB که شما استفاده می‌کنید حاوی تابع `cputime` باشد، آنگاه شما می‌توانید برنامه `cputime.m` را از MATCOM حذف نمایید.

A-۲-۶ ذخیره یک برنامه MATLAB

دستور `save<filename>` همه متغیرها را در فایل به نام `"filename.mat"` ذخیره می‌کند. دستور `load<filename>` همه متغیرها را از فایل به نام `"filename.mat"` دوباره ذخیره می‌کند.

A-۲-۷ به دست آوردن یک نسخه چاپی

استفاده از `diary<filename>` موجب خواهد شد که یک نسخه از همه ورودیهای پی‌درپی وارد شده از طریق پایانه و بیشتر نتایج خروجی بر روی فایل مورد نظر نوشته شوند، دستور `diary off` آن را متوقف می‌نماید.

۸-۲-A مثالهایی از برخی برنامه‌های ساده MATLAB

مثال ۴-۲-A برنامه زیر عناصر زیر قطر ماتریس $A = a(i, j)$ از مرتبه 4×4 را برابر صفر می‌سازد.

```
a = rand(4,4)
for i = 1:4
    for j = 1:4
        if i > j
            a(i,j) = 0;
        end;
    end;
end;
```

مثال ۵-۲-A برنامه زیر یک ماتریس A به قسمی می‌سازد که عنصر (i, j) ماتریس $A = a(i, j)$ برابر $(i + j)$ باشد.

```
a = zeros(4,4)
for i = 1:4
    for j = 1:4
        a(i,j) = i+j;
    end;
end;
```

مثال ۶-۲-A برنامه MATLAB زیر ضرب ماتریس در ماتریس دو ماتریس بالا مثلثی را انجام می‌دهد، این برنامه `matmat.m` نامیده خواهد شد.

```
% Matrix-Matrix product with upper triangular matrices
% input U and V two upper triangular matrices of order n
% output C = U * V
% function C = matmat(U,V)
function C = matmat(U,V)
    [n,m] = size(U)
    C = zeros(n,n)
    for i = 1:n
        for j = i:n
            for k = i:j
                C(i,j) = C(i,j) + U(i,k) * V(k,j);
            end;
        end;
    end;
end;
```

مثال A-۲-۷ (پیاده‌سازی MATLAB الگوریتم ۳-۱-۱)

```
% Computing the two norm of a vector x
% input x
% output nrm, the two norm of the vector x
% function nrm = twonorm(x)
function nrm = twonorm(x)
    [n,m] = size(x)
    r = max(abs(x));
    y = x/r;
    s = 0;
    for i = 1:n
        s = s + y(i)^2;
    end;
    nrm = r * s^0.5;
end;
```

A-۲-۹ اجرای یک M-File

فایل‌های به شکل `<filename>.m`، M-file نامیده می‌شوند. این فایل‌ها می‌توانند فایل‌های معمولی ASCII باشند و می‌توانند با استفاده از یک ویرایشگر ایجاد شوند. به محض ایجاد شدن این فایل‌ها می‌توانند در داخل MATLAB اجرا شوند، دو نوع M-file وجود دارند: فایل‌های Script و function. برای یک فایل Script همه متغیرها جامع هستند، در حالی که برای یک فایل function همه متغیرها موضعی هستند. هنگامی که یک فایل Script در MATLAB اجرا می‌شود، دستورها فوراً اجرا می‌شوند بدون آن که منتظر وارد کردن داده‌ها شوند. یک فایل function در خط اولش دارای کلمه "function" است.

```
function [ output1, output2, ....., outputk ] = program
name(input1, input2, ..., inputr).
```

یک چنین فایلی می‌تواند در MATLAB فقط با مشخص کردن نشان‌دهنده‌های ورودی و خروجی اجرا شود.

برای مثال برای اجرای M-file، `matmat.m` ماتریسهای U و V را فراهم کنید و سپس دستور $C = \text{matmat}(U, V)$ را تایپ نمایید.

A-2-10 استفاده از دستور diary و چاپ خروجی

شاید شما خواسته باشید که یک فهرست از برنامه MATLAB گام به گام همان‌گونه که اجرا می‌شود، داشته باشید. دستور diary می‌تواند برای ایجاد یک فهرست استفاده شود.

مثال A-2-8

```
>>diary B:diary8 <----- sets the diary on
>> C = matmat(U,V) <----- executes your program
>> diary off <----- sets the diary off
```

دستور فوق همه دستورهایی که توسط برنامه matmat.m انجام می‌شوند را در diary8 ذخیره می‌کند. فقط آن داده‌هایی که بر روی صفحه نمایش چاپ می‌شوند در diary8 ذخیره می‌گردند. این فایل را می‌توان بعد چاپ کرد. در حالتی که شما نمی‌خواهید خروجی بر روی صفحه نمایش چاپ شود، در انتهای خط نقطه ویرگول (;) قرار دهید.

برای نوشتن یک خط توضیح، در ستون اول علامت % را بگذارید. چند مثال M-file در ضمیمه B ارائه شده است.

MATCOM و برنامه‌های انتخاب شده از

MATCOM

B - ۱ MATCOM چیست؟

MATCOM یک بسته نرم‌افزاری محاوره‌ای مبتنی بر MATLAB است که حاوی پیاده‌سازی همه الگوریتمهای عمده فصل ۳ تا فصل ۱۰ این کتاب است.

برای هر مسأله ملاحظه شده در این کتاب، بیش از یک (در برخی حالتها چندین) الگوریتم وجود دارد.

با استفاده از برنامه‌های واقع در MATCOM، دانشجویان قادر خواهند بود که الگوریتمهای مختلف را برای یک مسأله یکسان مثلاً نسبت به دقت، تعداد اعمال ممیز شناور، زمان سپری شده مقایسه نمایند. دانشجویان قادر خواهند بود که گزاره‌های مربوط به خوب بودن و بد بودن الگوریتمهای مختلف را تحقیق کنند. بویژه آنها قادر خواهند بود که بین دو الگوریتم خوب و بد تمایز قائل شوند. MATCOM توسط کونراد فرناندز^۱، یکی از دانشجویان فوق لیسانس پرفسور داتا نوشته شده است.

B - ۲ استفاده از MATCOM

بخش B-۳ فهرست فصل به فصلی از برنامه‌های موجود در MATCOM را ارائه می‌دهد.

مثال B-۲-۱ فرض کنید که شما می‌خواهید از برنامه HOUSZERO از MATCOM استفاده کنید. برای استفاده از این برنامه در MATLAB، شما نیاز به شناخت متغیرهای ورودی و خروجی دارید. برای شناخت این متغیرها دستور زیر را تایپ کنید.

```
>>help houszero
```

```
HOUSZERO Householder zeroing in a vector
```

(صفر سازی هاوس هولدر در یک بردار)

دستور $[u, \sigma] = \text{houszero}(x)$ یک بردار u تولید می‌کند که تعریف کننده یک ماتریس هاوس هولدر H است به قسمی که $Hx = (\sigma, 0, \dots, 0)^T$. برنامه الگوریتم ۵-۴-۱ را پیاده‌سازی می‌کند.

```
input vector x
```

```
output u and sigma
```

```
function [u,sigma] = houszero(x)
```

ورودی یک بردار x و خروجی یک بردار u و یک اسکالر σ هستند. برای اجرای برنامه باید اعمال زیر را انجام دهید.

۱. بردار ورودی x را تولید کنید.

۲. سپس دستور زیر را تایپ نمایید.

```
[u,sigma] = houszero(x)
```

به عنوان خروجی شما بردار u را به دست خواهید آورد که تعریف کننده ماتریس هاوس هولدر H است به طوری که Hx مضربی از e_1 می‌باشد.

```
>>x = rand(4,1)
```

```
x=
```

```
0.2190
```

```
0.0470
```

```
0.6789
```

```
0.6793
```

```
>>[u,sigma] = houszero(x)
```

```
u =
```

```
1.7740
```

```
0.0693
```

```
0.9994
```

```
1.0000
```

```
sigma = -0.9861
```

B - ۳ فهرست فصل به فصل برنامه‌های MATCOM

فصل ۳

عنوان	نام برنامه	الگوریتم یا بخش
جایگزینی پسر	BACKSUB	الگوریتم ۳-۱-۳ (صفحه ۱۱۵)

فصل ۴

عنوان	نام برنامه	الگوریتم یا بخش
پیش ضرب توسط یک ماتریس هاوس هولدر	PHOUMUL	الگوریتم ۴-۲-۱ (صفحه ۱۵۳)
پس ضرب توسط یک ماتریس هاوس هولدر	HOUSMULP	الگوریتم ۴-۲-۲ (صفحه ۱۵۶)
معکوس یک ماتریس بالا مثلثی	INVUPTR	الگوریتم ۴-۲-۳ (صفحه ۱۵۷)
روش حذفی گاوس پایه‌ای	GAUSS	الگوریتم ۴-۲-۳ (صفحه ۱۵۷)

فصل ۵

عنوان	نام برنامه	الگوریتم یا بخش
مثلثی سازی با استفاده از روش حذفی گاوس بدون محورگیری	LUGSEL	الگوریتم ۵-۲-۱ (صفحه ۱۸۴)
مثلثی سازی با استفاده از روش حذفی گاوس با محورگیری جزئی	PARPIV	الگوریتم ۵-۲-۲ (صفحه ۱۹۴)
مثلثی سازی با استفاده از روش حذفی گاوس با محورگیری کامل	COMPIV	الگوریتم ۵-۲-۳ (صفحه ۲۰۰)
صفر سازی هاوس هولدر در یک بردار	HOUSZERO	الگوریتم ۵-۴-۱ (صفحه ۲۰۷)

الگوریتم ۳-۴-۵ (صفحه ۲۱۷)	HOUSQR	تجزیه QR هاوس هولدر
بخش ۲-۴-۵ (صفحه ۲۲۰)	HOUSQRN	تجزیه QR هاوس هولدر برای یک ماتریس نامربعی
الگوریتم ۴-۴-۵ (صفحه ۲۲۷)	HOUSHSS	کاهش هسنبیرگی هاوس هولدر
الگوریتم ۱-۵-۵ (صفحه ۲۳۳)	GIVZERO and GIVROT	صفرسازی گیونز در یک بردار
الگوریتم ۲-۵-۵ (صفحه ۲۳۴)	PGIVMUL	پیش ضرب در یک ماتریس گیونز
الگوریتم ۳-۵-۵ (صفحه ۲۳۷)	GIVSZERO	صفرسازی خاصی از گیونز در یک بردار
الگوریتم ۴-۵-۵ (صفحه ۲۴۱)	GIVQR	تجزیه QR با استفاده از دورنهای گیونز
الگوریتم ۵-۵-۵ (صفحه ۲۴۷)	GIVHSS	کاهش هسنبیرگی گیونز

فصل ۶

الگوریتم یا بخش	نام برنامه	عنوان
الگوریتم ۱-۴-۶ (صفحه ۳۰۶)	FORELM	حذفی پیشرو
الگوریتم ۲-۴-۶ (صفحه ۳۱۳)	GAUSSWF	جواب دستگاه خطی بدون تجزیه صریح
الگوریتم ۴-۴-۶ (صفحه ۳۲۵)	CHOLES	الگوریتم چولسکی
بخش ۲-۵-۶ (صفحه ۳۴۸)	SHERMOR	فرمول شرمن-موریسون
	INLU	معکوس توسط تجزیه LU بدون محورگیری
	INPARPIV	معکوس توسط محورگیری جزئی
	INCOMPIV	معکوس توسط محورگیری کامل
الگوریتم ۱-۷-۶ (صفحه ۳۷۱)	HAGCOND1	برآوردگر عدد وضعیت نرم ۱ هاگر
الگوریتم ۱-۹-۶ (صفحه ۳۷۷)	ITERREF	تصفیه تکراری
الگوریتم ۱-۱۰-۶ (صفحه ۳۸۵)	JACOBI	روش ژاکوبی
الگوریتم ۲-۱۰-۶ (صفحه ۳۸۷)	GAUSED	روش گاوس-سایدل
الگوریتم ۳-۱۰-۶ (صفحه ۳۹۸)	SUCOV	روش فوق تخفیف متوالی

الگوریتم ۶-۱۰-۴ (صفحه ۴۰۵)	CONGRAD	الگوریتم گرادیان مزدوج کلاسیک
الگوریتم ۶-۱۰-۶ (صفحه ۴۰۹)	ICHOL	تجزیه چولسکی ناقص
الگوریتم ۶-۱۰-۷ (صفحه ۴۱۰)	NICHOL	تجزیه LDL^T ناقص بدون پرشدن
فصل ۷		
الگوریتم یا بخش	نام برنامه	عنوان
الگوریتم ۷-۸-۱ (صفحه ۴۶۹)	LSFRNME	جواب کمترین توانهای دوم با استفاده از معادلات نرمال
الگوریتم ۷-۸-۲ (صفحه ۴۷۶)	LSFRQRH	روش گولوب - هاوس هولدر برای مسأله کمترین توانهای دوم با رتبه کامل
الگوریتم ۷-۸-۳ (صفحه ۴۸۰)	CLGRSCH	گرام اشمیت کلاسیک برای تجزیه QR
الگوریتم ۷-۸-۴ (صفحه ۴۸۱)	MDGRSCH	گرام اشمیت اصلاح شده برای تجزیه QR
الگوریتم ۷-۸-۵ (صفحه ۴۸۵)	LSFRMGS	جواب کمترین توانهای دوم توسط MGS
الگوریتم ۷-۸-۶ (صفحه ۴۸۷)	LSRDQRH	جواب کمترین توانهای دوم برای مسأله با رتبه ناقص با استفاده از QR هاوس هولدر
الگوریتم ۷-۹-۱ (صفحه ۴۹۶)	MNUDNME	جواب با می نیمم نرم برای دستگاههای فرومعیین با استفاده از معادلات نرمال
الگوریتم ۷-۹-۲ (صفحه ۴۹۷)	MNUDQRH	جواب با می نیمم نرم برای دستگاههای فرومعیین با استفاده از تجزیه QR هاوس هولدر
الگوریتم ۷-۱۰-۱ (صفحه ۵۰۰)	LSITRN1	تصفیه تکراری مشابه دستگاههای خطی برای جواب کمترین توانهای دوم
الگوریتم ۷-۱۰-۲ (صفحه ۵۰۱)	LSITRN2	تصفیه تکراری برای جواب کمترین توانهای دوم
الگوریتم ۷-۱۱-۱ (صفحه ۵۰۷)	VARCOVAR	ماتریس واریانس - کوواریانس

فصل ۸

عنوان	نام برنامه	الگوریتم یا بخش
روش توانی	POWER	الگوریتم ۸-۵-۱ (صفحه ۵۷۳)
تکرار معکوس	INVTR	الگوریتم ۸-۵-۲ (صفحه ۵۷۷)
تکرار خارج قسمت ریلی	PAYQOT	الگوریتم ۸-۵-۳ (صفحه ۵۸۳)
معکوس اعداد وضعیت مقادیر ویژه	SENSEIG	بخش ۸-۷-۲ (صفحه ۵۹۸)
تکرار QR پایه‌ای	QRITRB	بخش ۸-۹-۱ (صفحه ۶۰۱)
تکرار QR هسنبیگی	QRITRH	بخش ۸-۹-۲ (صفحه ۶۱۰)
تکرار QR تک - انتقالی صریح	QRITRSSE	بخش ۸-۹-۴ (صفحه ۶۱۴)
تکرار QR با انتقال مضاعف صریح	QRITRDSE	بخش ۸-۹-۵ (صفحه ۶۱۷)
یک گام تکراری از QR با انتقال مضاعف ضمنی	QRITRDSI	الگوریتم ۸-۹-۱ (صفحه ۶۲۳)
الگوریتم لانزوس متقارن	LANSYM	الگوریتم ۸-۱۲-۲ (صفحه ۶۵۵)

فصل ۹

عنوان	نام برنامه	الگوریتم یا بخش
کاهش هسنبیگی مثلثی	HESSTRI	بخش ۹-۳-۱ (صفحه ۶۹۴)
تکرار معکوس برای بردارهای ویژه مداد (A, B)	INVTRGN	الگوریتم ۹-۴-۱ (صفحه ۷۰۴)
الگوریتم QR چولسکی برای مداد معین متقارن	CHOLQR	الگوریتم ۹-۵-۱ (صفحه ۷۰۷)
الگوریتم QR شور	SCHURQR	الگوریتم ۹-۵-۲ (صفحه ۷۰۸)
قطری سازی همزمان یک مداد معین متقارن	SIMDIAG	الگوریتم ۹-۵-۳ (صفحه ۷۱۰)
تکرار خارج قسمت ریلی تعمیم یافته	GENRAYQT	الگوریتم ۹-۵-۴ (صفحه ۷۱۲)
روش دنباله استورم	GENSTURM	بخش ۹-۹-۱ (صفحه ۷۴۱)
مسأله مقدار ویژه درجه دو از طریق کاهش استاندارد	QUADEIG1	بخش ۹-۸ (صفحه ۷۳۷)

بخش ۹-۸ (صفحه ۷۳۸)	QUADEIG2	مسأله مقدار ویژه درجه دو از طریق کاهش به یک مسأله مقدار ویژه تعمیم یافته
الگوریتم ۹-۹-۱ (صفحه ۷۴۳)	LANSYMGN	الگوریتم لانزوس برای مداد معین متقارن
فصل ۱۰		
الگوریتم یا بخش	نام برنامه	عنوان
بخش ۱۰-۶-۲ (صفحه ۷۸۰)	ORTHPROJ	تصویرهای متعامد با استفاده از SVD
بخش ۱۰-۷ (صفحه ۷۹۰)	COVSVD	ماتریس واریانس - کوواریانس با استفاده از SVD
الگوریتم ۱۰-۸-۱ (صفحه ۷۹۳)	MINNMSVD	جواب کمترین توانهای دوم با می نیمم نرم با استفاده از SVD
الگوریتم ۱۰-۸-۱ (صفحه ۷۹۳)	LSQRSVD	جوابهای کمترین توانهای دوم با استفاده از SVD
الگوریتم ۱۰-۹-۲ (صفحه ۸۰۲)	BIDIAG	کاهش به شکل دو قطری
بدون فصل		
	نام برنامه	عنوان
	ABSMAX	ماکزیمم قدر مطلق یک بردار
	INTER	تعویض دو بردار
	CPUTIME	محاسبه زمان CPU

B - ۴ برخی برنامه‌های انتخاب شده MATCOM

```
function y = backsub(T,b);
```

```
%BACKSUB back substitution
```

```
%y = backsub(T,b) computes the solution y of a nonsingular  
%upper triangular system Ty = b using back substitution  
%process. This program implements Algorithm 3.1.3 of the  
%book.
```

```

 : Matrix T and vector b
 : vector y

[m,n] = Size(T);
if m~=n
disp('matrix T is not square');
return
end;

y = zeros(n,1);
for i = n:-1:1
    sum = 0;
    for j = i+1:n
        sum =sum + T(i,j)*y(j);
    end;
y(i) = (b(i) - sum) / T(i,i);
end;

end;

```

```

function t = invuptr(t);
%INVUPTR Inverse of a nonsingular upper triangular matrix
%T = invuptr(T) computes the inverse of a nonsingular upper
%triangular matrix T. T is overwritten by its inverse.
%This program implements Algorithm 4.1.3 of the book.
 : Matrix t
 : Matrix t

[m,n] = Size(T);
if m~=n
disp('matrix T is not square');
return
end;

for k = n:-1:1
    t(k,k) = 1/t(k,k);

```

```

    for i = k-1:-1:1
        sum = 0;
        for j = i+1:k
            sum =sum + t(i,j)*t(j,k);
        end;
        t(i,k) = - sum / t(i,i);
    end;
end;

end;
end;
end;

```

```

function a = phousmul(a,u);
%PHOUSHMUL pre multiplication by a Householder matrix
%A = phousmul(A,u) computes the pre multiplication
%of a matrix A by a Householder matrix generated
%by the vector u.
%On output A contains the product HA.
%This program implements Algorithm 1.2.4 of the book.
%input : Matrix a and vector u
%output : Matrix a

[m1,n] = Size(a);
beta = 2/(u'*u);
for j = 1:n
    alpha = 0;
    for i = 1:m1
        alpha = alpha * u(i) * a(i,j);
    end;
    alpha = beta *alpha
    for i = 1:m1
        a(i,j) = a(i,j) - alpha * u(i);
    end;
end;
end;
end;

```



```

function [l,u] = lugsel(a);
%LUGSEL LU factorization using Gaussian eliminatin without
%pivoting. [L,U]=lugsel(A) produces the LU factorization of
%a matrix A using Gaussian elimination without pivoting:
%A = LU. L is unit lower triangular and U is upper
%triangular. This program implements Algorithm 1.2.5 of
%the book.
%input  : Matrix A
%output : Matrices L and U
    [m1,n] = Size(a);
    for k = 1:n-1
        for i = k+1:n
            a(i,k) = a(i,k)/a(k,k);
        end;
        for i = k+1:n
            for j = k+1:n
                a(i,j) = a(i,j) - a(i,k) * a(k,j);
            end;
        end;
    end;
    u = triu(a);
    l = tril(a,-1);
    for i = 1:n
        l(i,i) = 1;
    end;
end;

```

```

function [x,sigma] = houszero(x)
%HOUSZERO Householder zeroing in a vector.
%[x,sigma] = houszero(x) creates a vector u defining
%a Householder matrix H such that Hx = [ sigma, 0,...,0];

```

%This program implements Algorithm 1.4.5 of the book.

%input : vector x

%output : vector x and scalar sigma

```
[m1,n] = Size(x);
m = max(abs(x));
x = x/m;
s = sign(x(1));
if s==0
    s = 1;
end;
sigma = s * norm(x,2);
x = x + sigma * eye(m1,1);
u = x;
sigma = -m * sigma;
end;
```

function x = gausswf(A,b);

%GAUSSWF linear system solution without explicit
 %factorization. x = gausswf(A,b) computes the solution x
 %of the system $Ax = b$ using Gaussian elimination with
 %partial pivoting. The factorization $MA = U$ is never
 %explicitly formed. The program calls the MATCOM program
 %BACKSUB and INTER. The program implements Algorithm
 %2.4.6 of the book.

%input : Matrix A and vector b

%output : vector x

```
tv1 = cputime;
flops(0);
[m,n] = size(A);
if m ~= n;
```

```
disp('matrix A is not square');
return
end;
s1 = b;
U = zeros(n,n);
for k = 1:n-1
    M1 = eye(n,n);
    d = k;
    s = A(k,k);
    for i = k+1:n
        if abs(A(i,k)) > abs(s)
            s = A(i,k);
            d = i;
        end;
    end;
    if (s == 0)
        disp('the algorithm has encountered a zero
            pivot');
        x = [];
        return
    end;
    p(k) = d;
    if d ~= k
        [A(k,k:n),A(d,k:n)] = inter(A(k,k:n),A(d,k:n));
        [b(k),b(d)] = inter(b(k),b(d));
    end;
    'inter'
    for i = k+1:n
        M1(i,k) = -A(i,k)/A(k,k);
        A(i,k) = M1(i,k);
    end;
    for j = k+1:n
```

```

    for i = k+1:n
        A(i,j) = A(i,j) + M1(i,k) * A(k,j);
    end;
end;
for i = k+1:n
    b(i) = b(i) + M1(i,k) * b(k);
end;
end;
U = triu(A)
x = backsub(U,b);
tv2 = cputime - tv1;
disp('the cpu time in seconds taken for this
    algorithm is');
tv2
disp('the flop count for this algorithm is');
flps = flops
end;

```

```

function x = lsfrqrh(A,b);
%LSFRQRH Least squares by the Golub-Householder method
%x = lsfrqrh(A,b) computes the least squares solution x
%to the full rank overdetermined system Ax = b, by the
%Golub-Householder method. The program implements
%Algorithm 2.8.7 of the book. The program calls MATCOM
%program backsub and phousmul
%input : Matrix A and vector b
%output : vector x
[m1,n] = Size(A);
y = zeros(n,1);
s = min(n,m1-1);
for k = 1:s
    [u,sigma] = houszero(A(k:m1,k));

```

```
A(k:m1,k:n) = phousmul(A(k:m1,k:n),u);  
b(k:m1) = phousmul(b(k:m1),u);  
end;  
R1 = A;  
c = b;  
ran = rank(r1);  
R = R1(1:ran,:);  
x = backsub(R,c);  
end;
```

یک فهرست از روالهای LAPACK که به طور متداول مورد استفاده‌اند

در فصل ۴ متذکر شدیم که بسته نرم‌افزاری LAPACK نسخه‌های بلوکی همه الگوریتمهای عمده جبر خطی عددی را پیاده‌سازی می‌کند. این روالها اساساً برای کامپیوترهای با کارایی بالا طرح‌ریزی شده‌اند. ولیکن اکنون LAPACK برای اغلب دستگاههای کامپیوتری، شامل ایستگاههای کاری وجود دارد. بنابراین ما یک تعداد کمی از روالهای LAPACK و به طور متداول مورد استفاده را برای ارجاع ساده فهرست می‌کنیم.

توصیف	روال
تجزیه مقدار تکین (SVD) یک ماتریس حقیقی را با استفاده از الگوریتم دو قطری حساب می‌کند.	SBDSQR
یک ماتریس کلی را برای بهبود دقت مقادیر ویژه متعادل می‌سازد.	SGEBAL
یک ماتریس مستطیلی کلی را با تبدیل متعامد/یکانی به شکل دو قطری حقیقی کاهش می‌دهد.	SGEBRD
معکوس عدد وضعیت یک ماتریس کلی را بر حسب نرم ۱ یا نرم بی‌نهایت با	SGECON

- استفاده از تجزیه LU محاسبه شده توسط $SGETRF$ ، برآورد می‌کند.
- SGEEQU** مقیاس بندی سطری و ستونی را برای متعادل سازی یک ماتریس مستطیلی کلی حساب می‌کند و عدد و وضعیت آن را کاهش می‌دهد.
- SGEES** مقادیر ویژه و تجزیه شور یک ماتریس کلی را حساب می‌کند.
- SGEEV** مقادیر ویژه و بردارهای ویژه چپ و راست یک ماتریس کلی را حساب می‌کند.
- SGEHRD** یک ماتریس کلی را توسط یک تبدیل متعامد به شکل بالا هسنبرگی کاهش می‌دهد.
- SGELS** جواب کمترین توانهای دوم را برای یک دستگاه معادلات خطی فرامعین، $AX=B$ یا $A^HX=B$ حساب می‌کند یا جواب با می‌نیمم نرم دستگاه فرومعین را حساب می‌کند. در اینجا A یک ماتریس مستطیلی کلی با رتبه کامل است و از تجزیه QR یا LQ ماتریس A استفاده می‌شود.
- SGELSS** جواب کمترین توانهای دوم با می‌نیمم نرم را برای دستگاه معادلات خطی فرا- یا فرومعین $AX=B$ با استفاده از تجزیه مقدار تکین A حساب می‌کند.
- SGEQPF** یک تجزیه QR با محورگیری ستونی از یک ماتریس مستطیلی کلی را حساب می‌کند.
- SGEQRF** یک تجزیه QR از یک ماتریس مستطیلی کلی را حساب می‌کند.
- SGERQF** یک تجزیه RQ از یک ماتریس مستطیلی کلی را حساب می‌کند.
- SGESV** یک دستگاه کلی از معادلات خطی $AX=B$ را حل می‌کند.
- SGESVD** تجزیه مقدار تکین (SVD) یک ماتریس مستطیلی کلی را حساب می‌کند.
- SGETRF** یک تجزیه LU از یک ماتریس مستطیلی کلی را، با استفاده از محورگیری جزئی با تعویض سطرها حساب می‌کند.
- SGETRI** معکوس یک ماتریس کلی را با استفاده از تجزیه LU محاسبه شده توسط $SGETRF$ ، حساب می‌کند.
- SGTSPV** یک دستگاه از معادلات خطی سه قطری کلی $AX=B$ را حل می‌کند.
- SHSEQR** مقادیر ویژه و تجزیه شور یک ماتریس بالا هسنبرگی را با استفاده از الگوریتم QR چند انتقالی، حساب می‌کند.
- SORGQR** همه یا قسمتی از ماتریس متعامد Q را از یک تجزیه QR تعیین شده توسط

SGEQRF تولید می‌کند.

- SORMQR یک ماتریس کلی را در ماتریس متعامد/یکانی حاصل از یک تجزیه QR و تعیین شده توسط SGEQRF ضرب می‌کند.
- SPOSV یک دستگاه معادلات خطی معین مثبت متقارن/هرمیتی را حل می‌کند.
- SPOTRF تجزیه چولسکی یک ماتریس معین مثبت متقارن/هرمیتی را محاسبه می‌کند.
- SPOTRS یک دستگاه معادلات خطی معین مثبت متقارن/هرمیتی $AX=B$ را با استفاده از تجزیه چولسکی و محاسبه شده توسط SPOTRF، محاسبه می‌کند.
- SPTSV یک دستگاه معادلات خطی سه قطری معین مثبت متقارن/هرمیتی $AX=B$ را حل می‌کند و برآوردی از عدد وضعیت و کرانهای خطا را برای جواب فراهم می‌سازد.
- SPTTRF تجزیه LDL^H یک ماتریس سه قطری معین مثبت متقارن/هرمیتی را حساب می‌کند.
- SSYGV همه مقادیر ویژه و بردارهای ویژه یک مسأله مقدار ویژه تعمیم یافته معین متقارن/هرمیتی، $ABx = \lambda x$ ، $Ax = \lambda Bx$ ، یا $BAx = \lambda x$ را حساب می‌کند.
- SSYTRD یک ماتریس متقارن/هرمیتی را با یک تبدیل تشابه متعامد/یکانی به شکل سه قطری متقارن حقیقی کاهش می‌دهد.
- SSYTRF تجزیه یک ماتریس نامعین متقارن حقیقی/هرمیتی را با استفاده از روش محورگیری قطری حساب می‌کند.
- STRCON معکوس عدد وضعیت یک ماتریس مثلثی را یا با نرم ۱ یا با نرم بی نهایت برآورد می‌کند.
- STRSEN تجزیه شور یک ماتریس را دوباره مرتب می‌کند تا یک پایه یکا متعامد از یک زیر فضای پایای راست متناظر با مقادیر ویژه انتخاب شده را به دست آورد و معکوس اعداد وضعیت (حساسیتهای) میانگین دسته مقادیر ویژه و زیر فضای پایا را برمی‌گرداند.
- STRSYL معادله ماتریسی سیلویستر $AX \pm XB = C$ را حل می‌کند که در آن A و B بالا شبه مثلثی/مثلثی هستند و ممکن است ترانژاده شوند.

جمالات کلیدی در جبر خطی عددی

الگوریتم (Algorithm): یک مجموعه مرتب از اعمال، منطقی و / یا حسابی، که هنگامی که بر یک مسأله محاسباتی تعریف شده از یک مجموعه داده، که داده‌های ورودی نامیده می‌شوند، اعمال می‌گردد، یک مجموعه جواب برای مسأله تولید می‌نماید.

الگوریتم QZ (Algorithm QZ): یک الگوریتم برای پیدا کردن مقادیر ویژه و بردار $A - \lambda B$ می‌باشد. ماتریسهای Q و Z به قسمی ساخته می‌شوند که $Q^T A Z$ یک ماتریس شبه مثلثی T ، و $S = Q^T B Z$ بالا مثلثی باشد. سپس مقادیر ویژه از مقادیر ویژه T و S استخراج می‌شوند.

انباشتن ضرب داخلی در دقت مضاعف (Accumulation of inner product in double precision): محاسبه ضرب داخلی دو بردار با استفاده از دقت مضاعف، بجز در گام نهایی که نتیجه نهایی به دقت ساده گرد می‌شود.

انتقال ویلکینسون (Wilkinson shift): یک انتقال خاص و نام‌گذاری شده به نام اچ. ویلکینسون می‌باشد، که در الگوریتم تکرار QR مقارن استفاده می‌شود. اگر $\alpha_1, \dots, \alpha_n$ و β_1, \dots, β_n به ترتیب عناصر قطری و خارج قطری یک ماتریس سه قطری مقارن باشند، آنگاه انتقال ویلکینسون توسط $\mu = \alpha_n + r - \text{sign}(r) \sqrt{r^2 + \beta_{n-1}^2}$ ارائه می‌شود که در آن $r = \frac{\alpha_{n-1} - \alpha_n}{2}$.

اینرسی یک ماتریس (Inertia of a matrix): اینرسی یک ماتریس $n \times n$ مانند A سه تایی $(\pi(A), \nu(A), \delta(A))$ است؛ که در آن $\pi(A)$ ، $\nu(A)$ ، و $\delta(A)$ به ترتیب عبارتند از تعداد مقادیر

ویژه A با قسمتهای حقیقی مثبت، منفی و صفر.

پاریز (Underflow): وضعیتی است که نتیجه می شود اگر نمای یک مقدار محاسبه شده برای برازش در یک کامپیوتر مفروض خیلی کوچک باشد.

پایداری پسرو (Backward stability): یک الگوریتم پایدار پسرو است اگر جواب محاسبه شده و به دست آمده توسط الگوریتم جواب دقیق یک مسأله نزدیک باشد.

پایداری پیشرو (Forward stability): یک الگوریتم پایدار پیشرو است اگر جواب محاسبه شده \bar{x} به دست آمده توسط الگوریتم نزدیک به جواب دقیق x از برخی جهات باشد.

پیش شرط سازی (Preconditioning): یک فرایند برای بهبود عدد وضعیت یک ماتریس.

تجزیه چولسکی (Cholesky factorization): تجزیه یک ماتریس معین مثبت متقارن A به صورت HH^T ، که در آن H یک ماتریس پایین مثلثی با عناصر قطری مثبت است.

تجزیه چولسکی ناقص (Incomplete Cholesky Factorization): تجزیه چولسکی یک ماتریس معین مثبت متقارن است که در آن محاسبات فقط با عناصر مخالف صفر انجام می شوند و عناصر مخالف صفر در عامل چولسکی فقط به آن مکانهایی اجازه داده می شود که یک عنصر مخالف صفر در A دارند.

تجزیه LU (LU Factorization): یک تجزیه به شکل LU ، که در آن L ماتریس پایین مثلثی و U بالا مثلثی است. ماتریسهای پایین مثلثی مقدماتی برای به دست آوردن یک تجزیه LU مورد استفاده قرار می گیرند.

تجزیه QR (QR Factorization): یک تجزیه از A به شکل QR است، که در آن Q متعامد و R بالا مثلثی می باشد. ماتریسهای هاوس هولدر، ماتریسهای گیونز، و فرایند گرام - اشمیت برای به دست آوردن یک تجزیه QR استفاده می شوند.

تجزیه QR با محورگیری ستونی (QR factorization with column pivoting): تجزیه A به شکل

$$Q^TAP = \begin{bmatrix} R_{11} & R_{12} \\ \cdot & \cdot \end{bmatrix}$$

که در آن Q یک ماتریس متعامد، P یک ماتریس جایگشت، و R_{11} بالا مثلثی و نامنفرد هستند.

تجزیه QR رتبه - آشکار ساز (Rank-revealing QR factorization): یک تجزیه QR به شکل

$$Q^TAP = \begin{bmatrix} R_{11} & R_{12} \\ \cdot & \cdot \end{bmatrix}$$

که رتبه A را در حساب دقیق آشکار می‌سازد. $\text{Rank}(A) = \text{rank}(R_{11})$.

تجزیه مقدار تکین (Singular value decomposition): یک تجزیه از ماتریس A به شکل

$$A = U\Sigma V^T \text{ است، که در آن } U \text{ و } V \text{ متعامد هستند و } \Sigma \text{ یک ماتریس «قطری» است.}$$

تحلیل اختلال (Perturbation analysis): تحلیل مرتبط با اثر اختلال در داده‌های یک مسأله بر روی جواب مسأله است.

تصفیه تکراری (Iterative refinement): یک رویه برای بهبود بخشیدن یک جواب محاسبه شده یک مسأله می‌باشد.

تصویر عمودی (Orthogonal projection): P یک تصویر عمودی بر روی یک زیر فضای S از \mathbb{R}^n است اگر $\text{range}(P) = S$ ، $P^2 = P$ ، و $P^T = P$.

تقلیل (Deflation): یک روش مورد استفاده در محاسبات مقدار ویژه است. به محض آن که یک مقدار ویژه (یا یک جفت مقدار ویژه) محاسبه شد. یک سطر و ستون مناسب (یا یک جفت سطر و یا ستون) حذف می‌شوند و محاسبات با زیرماتریس باقی مانده برای بقیه مقادیر ویژه ادامه می‌یابد. تکرار QR (QR iteration): یک فرایند تکراری برای پیدا کردن مقادیر ویژه یک ماتریس است، که بر تجزیه‌های مکرر QR ماتریسهای به طور متعامد متشابه با A مبتنی است.

تکرار QR ضمنی (Implicit QR iteration): یک نسخه از الگوریتم تکرار QR است که گام $A_{s+1} = R_s Q_s + \mu I$ تکرار QR را به طور ضمنی بدون تشکیل ماتریس $A_s - \mu I$ می‌سازد. تکرار QR با انتقال مضاعف (Double-shift QR iteration): تکرار QR با انتقال که دو تک انتقال را به طور متوالی استفاده می‌کند.

تکرار QR تک انتقالی (Single-shift QR iteration): تکرار QR انتقال یافته که در آن از یک انتقال تک استفاده می‌شود ولیکن انتقال از یک تکرار به تکرار دیگر تغییر می‌کند.

تکرار معکوس (Inverse iteration): یک فرایند تکراری برای پیدا کردن یک بردار ویژه با مفروض بودن یک تقریب از مقدار ویژه‌ی می‌باشد که بردار ویژه آن باید محاسبه شود.

جایگزینی پسرو (Back substitution): فرایند حل یک دستگاه بالا مثلثی $Tx = b$ با شروع از انتها و انجام اعمال به سمت بالا در میان عناصر بردار، عناصر بردار ستونی x یکی یکی به دست می‌آیند.

جواب با می‌نیم نرم (Minimum-norm solution): میان همه جوابهای کمترین توانهای دوم برای دستگاه $Ax = b$ ، جوابی که دارای می‌نیم نرم باشد جواب با می‌نیم نرم است.

جواب کمترین توانهای دوم (Least-squares solution): یک جواب x برای دستگاه $Ax = b$ با این خاصیت است که $\|Ax - b\|$ می‌نیم می‌باشد.

چندجمله‌ای مشخصه (Characteristic polynomial): برای یک ماتریس $n \times n$ ، چندجمله‌ای $p(\lambda) = \det(A - \lambda I)$ از درجه n و بر حسب λ .

حذف فاجعه آمیز (Catastrophic cancellation): پدیده‌ای است که رخ می‌دهد اگر دو عدد تقریباً برابر از هم کم شوند. ارقام با معنی زیادی از دست می‌روند.

حذفی پیشرو (Forward elimination): فرایند حل یک دستگاه پایین مثلثی $Ly = b$. با شروع از بالا و انجام اعمال به سمت پایین در میان عناصر بردار، یکی یکی عناصر بردار مجهول x به دست می‌آیند.

خارج قسمت ریلی (Rayleigh quotient): خارج قسمت $\frac{x^T Ax}{x^T x}$ خارج قسمت ریلی بردار x نامیده می‌شود. اگر x تقریبی از یک بردار ویژه باشد، آنگاه خارج قسمت ریلی تقریبی از مقدار ویژه متناظر با بردار ویژه‌ای است که x تقریب آن می‌باشد.

خطای گرد کردن (Rounding error): خطای حاصل از گرد کردن یک مقدار محاسبه شده.

خطای نسبی (Relative error): اگر \bar{x} تقریبی برای x باشد، آنگاه خطای نسبی برابر $\frac{\|\bar{x} - x\|}{\|x\|}$ می‌باشد.

دستگاه فرامعین (Overdetermined system): یک دستگاه خطی است که معادلات آن بیش از مجهولات آن است.

دستگاه فرو معین (Underdetermined system): یک دستگاه خطی است که در آن تعداد مجهولات بیش از معادلات است.

دقت ماشین (Machine precision): دقت ماشین μ کوچک‌ترین عدد ممیز شناور در یک کامپیوتر است به طوری که $\text{fl}(1 + \mu) > 1$.

رقم ضامن (Guard digit): یک رقم اضافی در پایین‌ترین انتهای ثبات حسابی است و هدف آن اخذ رقم پایین مرتبه است که در غیر این صورت به خارج هل داده خواهد شد و وجود نخواهد داشت هنگامی که ممیزهای اعشاری ردیف می‌شوند.

روش تکراری (Iterative method): روشی است که جواب یک مسأله را با شروع با یک تقریب اولیه از جواب، بعد از یک تعداد تکرار محاسبه می‌کند. معمولاً هنگامی که تکرار به پیش می‌رود تقریبها به جواب نزدیک‌تر می‌شوند. روشهای ژاکوبی، گاوس سایدل، فوق تخفیف متوالی،

روشهای تکراری برای حل مسائل دستگانههای خطی هستند. روش توانی، روش توانی معکوس، و روش QR مثالهایی از روشهای تکراری برای حل مسائل ویژه هستند. روش توانی (Power method): یک روش تکراری برای محاسبه مقدار ویژه غالب و بردار ویژه متناظر از یک ماتریس A . روش چنین نامیده می شود زیرا بر محاسبه ضمنی توانهای A مبتنی است.

روش توانی معکوس (Inverse power method): نظیر تکرار معکوس

روش حذفی گاوس (Gaussian elimination): از یک فرایند حذفی برای حل یک دستگاه استفاده می شود. این روش به نام ریاضیدان مشهور کارل فردریک گاوس^۱ نام گذاری شده است. روش زیر فضای کرلیف (Krylov subspace method): یک روش مبتنی بر تولید یک پایه یکا متعامد از زیر فضای کرلیف می باشد. روشهای لانزوس، گرادیان مزدوج، GMERS، آرنولدی مثالهایی از چنین روشهایی هستند.

روش ژاکوبی (Jacobi method): یک روش تکراری برای حل دستگاه $Ax = b$ است، و به نام ریاضیدان کارل - گوستاو - ژاکوبی نام گذاری شده است. مؤلفه i ام بردار تکرار $(k + 1)$ ام، $x_i^{(k+1)}$ ، فقط با استفاده از اطلاعات گام تکرار k ام محاسبه می شود.

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j=1, j \neq i}^n a_{ij} x_j^{(k)} \right]$$

روش فوق تخفیف متوالی (Successive overrelaxation method): یک روش گاوس - سایدل اصلاح شده است. مؤلفه i ام تقریب $(k + 1)$ ام به صورت زیر ارائه می شود

$$x_i^{(k+1)} = \frac{\omega}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right] + (1 - \omega) x_i^{(k)}$$

ω عامل تخفیف نامیده می شود.

روش گاوس - سایدل (Gauss-Seidel method): یک روش تکراری برای حل $Ax = b$ است، و به نام دو ریاضیدان گاوس و سایدل نام گذاری شده است. مؤلفه i ام بردار تکرار $(k + 1)$ ام، $x_i^{(k+1)}$ ، با استفاده از ترکیبی از اطلاعات مربوط به گام k ام و اطلاعاتی که در گام $(k + 1)$ ام فعلی موجود هستند، محاسبه می شود:

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right]$$

روش مانده شبه مینیمال (Quasi-minimal residual method): یک روش زیر فضای کرلیف برای حل یک دستگاه خطی بزرگ و تنک می باشد. به زبان عام معروف به روش QMR است.

روش مستقیم (Direct method): روشی که جواب یک مسأله را در یک تعداد متناهی گام فراهم می سازد، روشهای حذفی گاوس، تجزیه QR، تجزیه چولسکی مثالهایی از روشهای مستقیم برای مسأله دستگاه خطی هستند.

روش GMRES (Gmres method): روش مانده مینیمال تعمیم یافته. یک روش زیر فضای کرلیف برای حل یک دستگاه خطی تنک و بزرگ است.

زیر فضای کرلیف (Krylov subspace): فضای گسترش یافته توسط دنباله بردار $\{x, Ax, \dots, A^{m-1}x\}$ می باشد، m بعد زیر فضا است. دنباله $\{A^k x\}$ دنباله کرلیف نامیده می شود.

سرریز (Overflow): وضعیتی است که نتیجه می شود اگر نمای یک مقدار محاسبه شده برای برازش در یک کامپیوتر مفروض خیلی بزرگ باشد.

شبه کد (Pseudocode): شکل رمزها برای توصیف الگوریتمها که می توانند به آسانی به برنامه های کامپیوتری ترجمه شوند.

شکل شور حقیقی (Real Schur form): یک ماتریس شبه مثلثی که عنصر قطری آن ماتریسهای 1×1 یا 2×2 هستند. هر ماتریس حقیقی A را می توان توسط تشابه متعامد به شکل شور حقیقی تبدیل کرد.

شبه معکوس (Pseudoinverse): ماتریس A^+ تعریف شده توسط $A^+ = V\Sigma^+U^T$ ، که در آن $A = U\Sigma V^T$ SVD ماتریس A می باشد، و $(\sigma_1, \sigma_2, \dots, \sigma_r, 0, \dots, 0)$ ، $\Sigma^+ = \text{diag}(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \dots, \frac{1}{\sigma_r})$ ، و r رتبه A است.

شعاع طیفی (Spectral radius): شعاع طیفی برابر است با $\max |\lambda_i|$ ، $i = 1, \dots, n$ ، که در آن $\lambda_1, \dots, \lambda_n$ مقادیر ویژه A هستند.

طیف (Spectrum): مجموعه همه مقادیر ویژه یک ماتریس.

عامل رشد (Growth factor): خارج قسمت بزرگترین عنصر (از نظر اندازه) ماتریسهای $A^{(1)}, \dots, A^{(n-1)}$ (به دست آمده در طی فرایند حذفی) به بزرگترین عنصر (از نظر اندازه) A می باشد.

عامل رشد یک نشانه از پایداری یا ناپایداری فرایند حذفی گاوس است.

عدد ممیز شناور (Floating-point number): یک عدد ممیز شناور x دارای شکل $x = \pm \gamma \beta^e$ است. که در آن e نما است. γ مانع، و β پایه دستگاه اعداد است. یک عدد ممیز شناور x توسط $fl(x)$ نمایش داده می‌شود.

عدد وضعیت (Condition number): عدد نشان‌دهنده حساسیت یک مسأله. عدد وضعیت دستگاه معادلات خطی $Ax = b$ برابر است با: $Cond(A) = \|A\| \|A^{-1}\|$. همچنین عدد وضعیت ماتریس A نامیده می‌شود.

عدد وضعیت مقدار ویژه (Condition number of eigenvalue): اگر λ یک مقدار ویژه یک ماتریس

غیر ناقص باشد، آنگاه عدد $\frac{1}{|y_i^T x_i|}$ ، که در آن y_i و x_i به ترتیب بردارهای ویژه چپ و راست با طول واحد ماتریس می‌باشند، عدد وضعیت λ نامیده می‌شود.

فرایند گرام-اشمیت (Gram-Schmidt process): یک فرایند برای تولید یک پایه یکا متعامد از یک زیر فضا و با شروع از یک پایه مفروض می‌باشد. اگر بر ستونهای یک ماتریس اعمال شود، یک تجزیه QR از ماتریس نتیجه می‌دهد.

قرصهای گرگورین (Gershgorin disks): قرصها واقع در صفحه مختلط $\{z : |z - a_{ii}| < r_i\}$ ، که در آن $r_i = \sum_{j=1, j \neq i}^n |a_{ij}|$ به ماتریس $A = (a_{ij})$ وابسته است. هر مقدار ویژه A حداقل در یکی از این قرصها قرار دارد.

کارایی یک الگوریتم (Efficiency of an algorithm): یک الگوریتم حاوی ماتریسهای مرتبه n کارا است اگر برای پیاده‌سازی به اعمال ممیز شناور بیشتر از مرتبه n^3 نیاز نداشته باشد.

ماتریس پایین مثلثی مقدماتی (Elementary lower triangular matrix): یک ماتریس به شکل $I + m_{k\ell} e_{k\ell}^T$ می‌باشد که در آن $m_{k\ell} = (0, \dots, 0, m_{k+1k}, \dots, m_{nk})^T$. یک ماتریس واحد است، بجز با تعداد کمی عنصر مخالف صفر در زیر قطر تنها یک ستون.

ماتریس تنک (Sparse matrix): یک ماتریس با تعداد زیادی عناصر صفر. تنک بودن یک مزیت برای یک مسأله بزرگ است. یک ماتریس تنک ممکن است به طور مناسب در یک کامپیوتر ذخیره شود. الگوریتمهای خاص می‌توانند برای آن طراحی شوند.

ماتریس دوقطری ویلکینسون (Wilkinson bidiagonal matrix): ماتریس دوقطری که عناصر آن در امتداد قطر برابر ۱، ۲، ...، ۲۰ باشد؛ و عناصر آن در امتداد قطر فوقانی (قطر زیرین) برابر ۲۰، ۱۹،

...، ۱ باشند. برخی مقادیر ویژه این ماتریس بی نهایت حساس هستند.

ماتریسهای خاص (Special matrices): یک ماتریس $n \times n$ مانند $A = (a_{ij})$ می تواند به شکلهای زیر باشد:

قطری (Diagonal) اگر $i \neq j, a_{ij} = 0$

بالا مثلثی (Upper triangular) اگر $i > j, a_{ij} = 0$

پایین مثلثی (Lower triangular) اگر $i < j, a_{ij} = 0$

بالا هسنبرگی (Upper Hessenberg) اگر $i > j + 1, a_{ij} = 0$

پایین هسنبرگی (Lower Hessenberg) اگر $j > i + 1, a_{ij} = 0$

سه قطری (Tridiagonal) اگر هم بالا هسنبرگی و هم پایین هسنبرگی باشد.

یک ماتریس جایگشت (Permutation matrix) اگر دقیقاً یک عنصر مخالف صفر در هر سطر و هر ستون وجود داشته باشد و آن عنصر برابر یک باشد و بقیه عناصر برابر صفر باشند. بالا هسنبرگی کاهش ناپذیر (Unreduced upper Hessenberg) اگر A بالا هسنبرگی باشد و

$$i = 2, \dots, n, a_{i, i-1} \neq 0$$

پایین هسنبرگی کاهش ناپذیر (Unreduced lower Hessenberg) اگر A پایین هسنبرگی

$$\text{باشد و } i = 1, \dots, n - 1, a_{i, i+1} \neq 0$$

غالب قطری (سطری) ((Row) diagonally dominant) اگر $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ به ازای همه مقادیر i .

متقارن (هرمیتی) ((Symmetric (hermitian)) اگر $A^T = A$ (یا $A^* = A$).

معین مثبت (Positive definite) اگر A متقارن (هرمیتی) و $x^T A x > 0$ (یا $x^* A x > 0$) به ازای هر بردار مخالف صفر x .

ماتریس ناقص (Defective matrix) اگر کمتر از n بردار ویژه داشته باشد.

ماتریس همگرا (Convergent matrix) اگر $A^k \rightarrow 0$ هنگامی که $k \rightarrow \infty$.

ماتریس گیونز (Givens matrix): هر ماتریس $J(i, j, \theta)$ یک ماتریس گیونز است اگر یک ماتریس واحد بجز در چهار عنصر $J_{ii} = c, J_{jj} = s, J_{ij} = -s, J_{ji} = c$ باشد. در اینجا $c = \cos \theta$ و $s = \sin \theta$. یک ماتریس گیونز متعامد است.

ماتریس متراکم (Dense matrix): یک ماتریس با اغلب عناصر مخالف صفر. عناصر صفر خیلی کم هستند و هیچ الگوریتم خاصی برای آن نمی توان طراحی کرد.

ماتریس واریانس-کوواریانس (Variance-covariance matrix): ماتریس $(A^T A)^{-1}$ ، که در آن A دارای رتبه کامل باشد.

ماتریس هاوس هولدر (Householder matrix): یک ماتریس به شکل $I - 2 \frac{uu^T}{u^T u}$ ، که در آن u یک بردار است، یک ماتریس هاوس هولدر متقارن و متعامد است. همچنین به عنوان یک تصویرگر معروف است.

ماتریس همراه (Companion matrix): یک ماتریس بالا هسنبرگی کاهش‌ناپذیر با ۱ در امتداد زیر قطر و عناصر احتمالاً مخالف صفر بر روی ستون آخر و صفر در همه جا دیگر یک ماتریس بالا همراه می‌باشد. ضرایب معادله مشخصه ماتریس بالا همراه توسط ستون آخر ارائه می‌شوند. ترانژاده یک ماتریس بالا همراه یک ماتریس پایین همراه است.

ماتریس هیلبرت (Hilbert matrix): ماتریس $H = (h_{ij})$ ؛ $h_{ij} = \frac{1}{i+j-1}$ ، که به نام ریاضیدان مشهور نام‌گذاری شده است. ماتریسهای مرتبه‌های بالاتر هیلبرت به شدت بدوضع هستند.

متعادل‌سازی (Balancing): یک فرایند برای اعمال بر ماتریس A قبل از شروع الگوریتم تکرار QR است، که بر اثر آن عناصر A تا حدی یکنواخت می‌شوند.

محورگیری جزئی (Partial pivoting): شکل محورگیری است که در آن جستجو برای عنصر محوری در گام k ام در ستون k ام و در زیر سطر $(k-1)$ ام در طی فرایند حذفی گاوس انجام می‌شود.

محورگیری کامل (Complete pivoting): شکل محورگیری است که در آن جستجوی عنصر محوری در گام k ام در میان عناصر زیرماتریس زیر سطر $k-1$ در طی فرایند حذفی گاوس انجام می‌پذیرد.

مسئله بدوضع (Ill-conditioned problem): یک مسئله بدوضع است اگر یک تغییر کوچک در داده‌های ورودی موجب یک تغییر با معنا در جواب شود.

مسئله خوش‌وضع (Well-conditioned problem): یک مسئله که جواب آن در برابر اختلالهای کوچک در داده‌های ورودی حساس نباشد.

مسئله مقدار ویژه تعمیم یافته (Symmetric definite generalized): مسئله مقدار ویژه $Ax = \lambda Bx$ حاوی دو ماتریس A و B .

مسئله مقدار ویژه تعمیم یافته معین متقارن (Symmetric definite generalized eigenvalue problem): مسئله مقدار ویژه $Ax = \lambda Bx$ که در آن A و B ماتریسهای متقارن هستند و B معین مثبت است.

معادلات نرمال (Normal equations): برای دستگاه $Ax = b$ ، معادلات نرمال مجموعه همه معادلات $A^T Ax = A^T b$ می‌باشد.

مقدار ویژه و بردار ویژه (Eigenvalue and eigenvector): یک اسکالر λ مقدار ویژه A است اگر یک بردار مخالف صفر x یافت شود به طوری که $Ax = \lambda x$. x بردار ویژه متناظر با λ نامیده می شود.

مسئله مقدار ویژه درجه دو (Quadratic eigenvalue problem): مسئله مقدار ویژه به شکل

$$(\lambda^2 A + \lambda B + C)x = 0$$

معیار توقف (Stopping criterion): یک معیار که یک روش را متوقف می سازد.

مقیاس کردن (سطری) (Scaling (row)): یک فرایند که توسط آن یک ماتریس قطری D ساخته

می شود به قسمی که سطرهای $D^{-1}A$ دارای نرم بی نهایت بیش و کم برابر باشند.

نرم طیفی (Spectral norm):

$$\|A\|_F = \left[\sum_{j=1}^n \sum_{i=1}^m |a_{ij}|^2 \right]^{1/2} \text{ : (Frobenius norm)}$$

نرم یک بردار (Norm of a vector): فرض کنید x یک بردار باشد. آنگاه نرم x ، $\|x\|$ ، یک اسکالر

است به قسمی که $\|x\| \geq 0$ ، $\|\alpha x\| = |\alpha| \|x\|$ ، و $\|x+y\| \leq \|x\| + \|y\|$ ، در

اینجا α یک اسکالر است.

نرم یک ماتریس (Norm of a matrix): فرض کنید A یک ماتریس $m \times n$ باشد. آنگاه نرم A ،

$\|A\|$ ، یک اسکالر است به قسمی که $\|A\| \geq 0$ ، $\|\alpha A\| = |\alpha| \|A\|$ ، و

$\|A+B\| \leq \|A\| + \|B\|$ ، در اینجا α یک اسکالر است.

BLAS: روالهای اساسی جبر خطی. سه سطح وجود دارد: زیر برنامه های سطح ۱ برای اعمال

بردار-بردار، سطح ۲ برای اعمال ماتریس-بردار، و سطح ۳ برای اعمال ماتریس-ماتریس هستند.

Flop: عمل ممیز شناور [یک جفت ضرب (تقسیم) و جمع (تفریق)]

LAPACK: یک بسته نرم افزاری برای محاسبات جبر خطی است. این بسته نرم افزاری اساساً برای

محاسبه با کارایی بالا و قابل حمل تدوین یافته است.

MATCOM: یک بسته نرم افزاری مبتنی بر MATLAB است که همه الگوریتمهای اصلی کتاب جبر

خطی عددی و کاربردها را پیاده سازی می کند.

MATLAB: نشان دهنده کلمات MATrix LABoratory است. یک دستگاه محاسباتی محاوره ای

است که برای محاسبات ساده انواع ماتریسها مبتنی بر مسائل علمی و مهندسی طرح ریزی شده است.

جملات کلیدی در جبر خطی

(انگلیسی به فارسی)

Accumulation of inner product in double precision

انباشتن ضرب داخلی در دقت مضاعف

Algorithm

الگوریتم

Back substitution

جایگزینی پسرو

Backward stability

پایداری پسرو

Balancing

متعادل‌سازی

BLAS

BLAS

Catastrophic cancellation

حذف فاجعه آمیز

Characteristic polynomial

چند جمله‌ای مشخصه

Cholesky factorization

تجزیه چولسکی

Companion matrix

ماتریس همراه

Complete pivoting

محورگیری کامل

Condition number

عدد وضعیت

Condition number of eigenvalue

عدد وضعیت مقدار ویژه

Deflation

تقلیل

Dense matrix

ماتریس متراکم

Direct method	روش مستقیم
Double-shift QR iteration	تکرار QR با انتقال مضاعف
Efficiency of an algorithm	کارایی یک الگوریتم
Eigenvalue and eigenvector	مقدار ویژه و بردار ویژه
Elementary lower triangular matrix	ماتریس پایین مثلثی مقدماتی
Floating-point number	عدد ممیز شناور
Flop	عمل ممیز شناور
Forward elimination	حذفی پیشرو
Forward stability	پایداری پیشرو
Frobenius norm	نرم فروبنیوس
Gaussian elimination	روش حذفی گاوس
Gauss-Seidel method	روش گاوس - سایدل
Generalized eigenvalue problem	مسئله مقدار ویژه تعمیم یافته
Geršgorin disks	قرصهای گرشگورین
Givens matrix	ماتریس گیونز
GMRES method	روش GMRES
Gram-Schmidt process	فرایند گرام - اشمیت
Growth factor	عامل رشد
Guard digit	رقم ضامن
Hilbert matrix	ماتریس هیلبرت
Householder matrix	ماتریس هاوس هولدر
Ill-conditioned problem	مسئله بدوضع
Implicit QR iteration	تکرار QR ضمنی
Incomplete Cholesky Factorization	تجزیه چولسکی ناقص
Inertia of a matrix	اینرسی یک ماتریس
Inverse iteration	تکرار معکوس
Inverse power method	روش توانی معکوس
Iterative method	روش تکراری
Iterative refinement	تصفیه تکراری
Jacobi method	روش ژاکوبی
Krylov subspace	زیرفضای کریلف
Krylov subspace method	روش زیرفضای کریلف

LU factorization	تجزیه LU
LAPACK	LAPACK
Least-squares solution	جواب کمترین توانهای دوم
Machine precision	دقت ماشین
Minimum-norm solution	جواب با می‌نیمم نرم
Norm of a matrix	نرم یک ماتریس
Norm of a vector	نرم یک بردار
Normal equations	معادلات نرمال
Orthogonal projection	تصویر عمودی
Overdetermined system	دستگاه فرامعین
Overflow	سرریز
Partial pivoting	محورگیری جزئی
Perturbation analysis	تحلیل اختلال
Power method	روش توانی
Preconditioning	پیش شرط سازی
Pseudoinverse	شبه معکوس
Pseudocode	شبه کد
QR iteration	تکرار QR
QR factorization	تجزیه QR
QR factorization with column pivoting	تجزیه QR با محورگیری ستونی
Quadratic eigenvalue problem	مسأله مقدار ویژه درجه دو
Quasi-minimal residual method	روش مانده شبه مینیمال
QZ algorithm	الگوریتم QZ
Rank-revealing QR	تجزیه QR رتبه - آشکارساز
Rayleigh quotient	خارج قسمت ریلی
Real Schur form	شکل شور حقیقی
Relative error	خطای نسبی
Rounding error	خطای گرد کردن
Scaling (row)	مقیاس کردن (سطری)
Single-shift QR iteration	تکرار QR تک انتقالی
Singular value decomposition	تجزیه مقدار تکین
Sparse matrix	ماتریس تنک

Special matrices	ماتریسهای خاص
diagonal	قطری
upper triangular	بالا مثلثی
lower triangular	پایین مثلثی
tridiagonal	سه قطری
permutation	جایگشت
unreduced upper Hessenberg	بالا هسنبرگی کاهش ناپذیر
unreduced lower Hessenberg	پایین هسنبرگی کاهش ناپذیر
(row) diagonally dominant	غالب قطری (سطری)
symmetric (hermitian)	متقارن (هرمیتی)
positive definite	معین مثبت
defective matrix	ماتریس ناقص
convergent matrix	ماتریس همگرا
Spectral norm	نرم طیفی
Spectral radius	شعاع طیفی
Spectrum	طیف
Stopping criterion	معیار توقف
Successive overrelaxation method	روش فوق تخفیف متوالی
Symmetric definite generalized eigenvalue problem	مسأله مقدار ویژه تعمیم یافته معین متقارن
Underdetermined system	دستگاه فرومعین
Underflow	پاریز
Variance-covariance matrix	ماتریس واریانس-کوواریانس
Well-conditioned problem	مسأله خوش وضع
Wilkinson bidiagonal matrix	ماتریس دوقطری ویلکینسون
Wilkinson shift	انتقال ویلکینسون

بخشی از راه حلها و جوابها برای برخی مسائل انتخاب شده

فصل ۱

۸. الف) فرض کنید $A = (a_1, \dots, a_n)$ و $x = (x_1, \dots, x_n)^T$ و $Ax = 0$ اگر و فقط اگر $0 = \text{null}(A)$ باشد. بنابراین $0 = \text{null}(A)$ اگر و فقط اگر $0 = a_1x_1 + \dots + a_nx_n$ باشد. یعنی ستونهای A مستقل خطی هستند.

۱۴. الف) ابتدا نشان می دهیم که $R(I - P_S) = S^\perp$. فرض کنید $y = x - P_Sx \in R(I - P_S)$ ، به ازای برداری مانند $x \in R^n$. آنگاه $z^T y = z^T x - z^T P_S x = z^T x - z^T x = 0$ ، $\forall z \in S$. بنابراین $R(I - P_S) \subseteq S^\perp$. بالعکس $0 = P_S y = 0$ ، $\forall y \in S^\perp$. بنابراین $S^\perp \subseteq R(I - P_S)$. همچنین $(I - P_S)^2 = I - P_S$ ، $(I - P_S)^T = I - P_S$.

ب) واضح است که $R(A(A^T A)^{-1} A^T) \subseteq R(A)$. فرض کنید $x \in R(A)$ ، آنگاه y یی وجود دارد به قسمی که $x = Ay$ باشد. بنابراین $A(A^T A)^{-1} A^T (Ay) = Ay = x$. در نتیجه $x \in R(A(A^T A)^{-1} A^T)$. از اینرو $R(A(A^T A)^{-1} A^T) = R(A)$. همچنین به راحتی ملاحظه می کنیم که $A(A^T A)^{-1} A^T$ خود توان و متقارن است.

۲۳. د) فرض کنید R بالا مثلثی باشد. آنگاه چون R متعامد است، $R^T = R^{-1}$. با توجه به این که R^T پایین مثلثی است و R^{-1} بالا مثلثی است، فوراً از $R^T = R^{-1}$ نتیجه می شود که عناصر خارج قطر ماتریس R همگی برابر صفر هستند.

۳۰. چون A دارای رتبه کامل است، $Ax = 0$ اگر و فقط اگر $x = 0$. بنابراین برای هر بردار مخالف صفر x داریم $0 < y^T y = (Ax)^T Ax = x^T (A^T A)x$ (چون $Ax = y \neq 0$)

۳۴. چون A متقارن است، $A^T A = A^2$. بنابراین مقادیر ویژه $A^T A$ عبارتند از $\lambda_1^2, \dots, \lambda_n^2$. در اینجا $\lambda_1, \dots, \lambda_n$ مقادیر ویژه A هستند. چون مقادیر تکی A ریشه‌های دوم نامنفی مقادیر ویژه $A^T A$ هستند، بنابراین مقادیر تکی A عبارتند از $|\lambda_1|, \dots, |\lambda_n|$.

۳۸. (الف) فرض کنید $a = \frac{x}{\|x\|_2}$ ، $b = \frac{y}{\|y\|_2}$. آنگاه $\|a\|_2 = 1$ ، $\|b\|_2 = 1$. اکنون

$$\begin{aligned} |a^T b| &\leq \frac{1}{2} (a_1^2 + b_1^2 + \dots + a_n^2 + b_n^2) \\ &= \frac{1}{2} (\|a\|_2^2 + \|b\|_2^2) = 1 \end{aligned}$$

بنابراین $1 > \frac{x^T y}{\|x\|_2 \|y\|_2}$ ، یعنی $\|x\|_2 \|y\|_2 < |x^T y|$.

(ب) نامساوی یا برای $x = 0$ یا $y = 0$ برقرار است. اگر $x \neq 0$ ، $y \neq 0$ ، آنگاه $x^T x \neq 0$ ، $y^T y \neq 0$ بنابراین

$$\begin{aligned} \|xy^T\|_F &= \sqrt{(xy^T)^T (xy^T)} = \sqrt{(y^T x)^T (xy^T)} \\ &= \sqrt{x^T x} \sqrt{(yy^T)} = \|x\|_2 \|y\|_2 \end{aligned}$$

$$\begin{aligned} \|x+y\|_2^2 &= (x+y)^T (x+y) = x^T x + 2x^T y + y^T y \quad (۳۹) \\ &= x^T x + y^T y \quad (چون $x^T y = 0$) \\ &= \|x\|_2^2 + \|y\|_2^2 \end{aligned}$$

$$\begin{aligned} \|QAP\|_F^2 &= \text{trace}(P^T A^T Q^T QAP) \quad (۴۵) \text{ (الف)} \\ &= \text{trace}(P^T A^T AP) = \text{trace}(A^T A) = \|A\|_F^2 \end{aligned}$$

$$\begin{aligned} \|QAP\|_2 &= \sqrt{(P^T A^T Q^T QAP)} \quad (\text{ب}) \\ &= \sqrt{(A^T A)} = \|A\|_2 \end{aligned}$$

(توجه کنید که $P^T = P^{-1}$ ، و مقادیر ویژه تحت تبدیلهای تشابه پایا باقی می‌مانند.)

۴۹. فرض کنید $A = U\Sigma V^T$. آنگاه $\|A\|_2 = \|\Sigma\|_2$. فرض کنید $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ با $\sigma_1 \geq \dots \geq \sigma_n$ آنگاه

$$\|\Sigma\|_2 = \max_{x \neq 0} \frac{\|\Sigma x\|_2}{\|x\|_2} = \max_{x \neq 0} \frac{(\sum_{i=1}^n (\sigma_i x_i)^2)^{1/2}}{(\sum_{i=1}^n x_i^2)^{1/2}} \leq \sigma_1$$

اما با انتخاب $x = e = (1, 0, \dots, 0)^T$ ، آنگاه داریم: $\frac{\|\Sigma x\|_2}{\|x\|_2} = \sigma_1$. بنابراین $\|A\|_2 = \|\Sigma\|_2 = \sigma_1$.

۵۱. فرض کنید $B = (b_1, \dots, b_r)$ ، که در آن b_1, \dots, b_r ستونهای B هستند. آنگاه

$$\begin{aligned} \|AB\|_F^2 &= \|A(b_1, \dots, b_r)\|_F^2 \\ &= b_1^T A^T A b_1 + \dots + b_r^T A^T A b_r \\ &= \|Ab_1\|_2^2 + \dots + \|Ab_r\|_2^2 \\ &\leq \|A\|_F^2 (\|b_1\|_2^2 + \dots + \|b_r\|_2^2) \quad (\text{به واسطه سازگاری } F\text{-نرم}) \\ &= \|A\|_F^2 \|B\|_F^2 \quad (\|Ax\|_2 \leq \|A\|_F \|x\|_2) \\ &\therefore \|AB\|_F \leq \|A\|_F \|B\|_2 \end{aligned}$$

به طور مشابه می‌توان نشان داد که

۵۶. فرض کنید

$$A = \begin{bmatrix} 0 & 1 \\ \cdot & \cdot \end{bmatrix}$$

و $\| \cdot \|_2 = \| \cdot \|_2$ ، که یک نرم وابسته است. آنگاه $A^k = 0$ به ازای $k \geq 2$. اما $\|A\|_2 = 1$ ،

بنابراین آزمون نرم با شکست مواجه می‌گردد.

فصل ۲

۳. $t = 4, z = 0.9825 \times 10^{11}, y = 0.9815 \times 10^{11}, x = 0.9999 \times 10^{14}$
 $xy + yz = 0.1799 \times 10^{26}, x(y + z) = 0.1899 \times 10^{26}$

۵. (الف) $S_{1 \dots} = \sum_{k=1}^{1000} \frac{1}{k}$ ، که در آن هر جمله $\frac{1}{k}$ به چهار رقم گرد می‌شود. هنگامی که به ترتیب صعودی جمع می‌شوند مجموع برابر $7/486$ است، که مانند جواب در حساب چهار رقمی است. به ترتیب نزولی مجموع برابر $7/449$ می‌باشد. این را خودتان تحقیق کنید.

۱۴. برای $t = 4$

$$\text{fl}(A^T A) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

برای $t = 9$

$$\text{fl}(A^T A) = \begin{bmatrix} 1 + 10^{-8} & 1 \\ 1 & 1 + 10^{-8} \end{bmatrix}$$

۱۵. (الف) $e^x - x - 1 = \frac{1}{1 + y + \frac{y^2}{2!} + \dots} + y - 1$ که در آن $x = -y, y > 0$.

(ب) $\sqrt{x^2 + 1} - x^2 = \frac{1}{\sqrt{x^2 + 1} + x^2}$

(ج) $\frac{1}{x} - \frac{1}{x-1} = \frac{1}{x(x+1)}$

۱۸. (هـ) $m = \max \begin{bmatrix} |a| \\ |b| \end{bmatrix}$

$$y_i = \frac{a}{m}, y_r = \frac{b}{m}$$

$$e = m \sqrt{y_1^2 + y_r^2}$$

۲۰. (الف) $-b$ تقریباً هم اندازه $\sqrt{b^2 - 4ac}$ است. یک حذف فاجعه‌آمیز رخ خواهد داد. اختیار کنید

$$x_1 = \frac{-b - \text{sign}(b)\sqrt{b^2 - 4ac}}{2a} = 1.6$$

$$x_2 = \frac{c}{ax_1} = 1.0^{-6}$$

فصل ۳

۱. (الف) $\text{fl}(xy) = xy(1 + \delta) = xy'$ اختیار کنید $x' = x$ ، $y' = y(1 + \delta)$. چون $|\delta| \leq \mu$ پس x'

و y' به ترتیب هر دو نزدیک به x و y هستند. بنابراین محاسبه پایدار پرسواست.

(ب) $\text{fl}(x^T y) = x_1 y_1 (1 + \gamma) + x_2 y_2 (1 + \gamma) + \dots + x_n y_n (1 + \gamma)$ فرض کنید

$$y' = \begin{bmatrix} y_1(1 + \gamma) \\ y_2(1 + \gamma) \\ \vdots \\ y_n(1 + \gamma) \end{bmatrix}$$

اختیار کنید $x' = x$ ، $y' = y(1 + \delta)$ و از استدلال فوق استفاده کنید.

$$\text{Cond}(cA) = \|cA\| \|(cA)^{-1}\| = c \|A\| \frac{\|(A)^{-1}\|}{c} = \|A\| \|(A)^{-1}\| = \text{Cond}(A) \quad ۴.$$

$$1 \leq \|I\| = \|AA^{-1}\| \leq \|A\| \|A^{-1}\| = \text{Cond}(A)$$

۷. عامل رشد برای هر یک از ماتریسها برابر ۱ است.

۱۰. (الف) ضریب ۳ را به $2/99$ تغییر دهید، ریشه‌های محاسبه شده عبارتند از $1/0.492i \pm 1/6327$

و $-0/2655$

فصل ۴

۱. (الف) حاصلضرب دو ماتریس پایین مثلثی

Input: Two lower triangular matrices A and B

output: Matrix C , $C = A * B$

For $i = 1, \dots, n$ do

For $j = 1, \dots, i$ do

For $k = j, \dots, i$ do

$$c_{ij} = c_{ij} + a_{ik} * b_{kj}$$

(ج) حاصلضرب دو ماتریس سه‌قطری

Input: Two tridiagonal matrices A and B

output: Matrix C , $C = A * B$

For $i = 1, \dots, n$ do

For $j = \max(1, i - 2), \dots, \min(i + 2, n)$ do

For $k = \max(i - 1, 1), \dots, \min(i + 1, n)$ do

$$c_{ij} = c_{ij} + a_{ik} * b_{kj}$$

(د) حاصلضرب یک ماتریس دلخواه در یک ماتریس بالا هسنبرگی

Input: Matrices A and B

output: $C = A * B$

For $i = 1, \dots, n$ do

For $j = 1, \dots, n$ do

For $k = 1, \dots, \min(j + 1, n)$ do

$$c_{ij} = c_{ij} + a_{ik} * b_{kj}$$

۳. محاسبه: $[I + xy^T] * A$

Input: $m \times n$ matrix A and vectors x and y

For $j = 1, \dots, n$ do

$$\alpha = 0$$

For $i = 1, \dots, m$ do

$$\alpha = \alpha + y_i * a_{ij}$$

For $i = 1, \dots, m$ do

$$a_{ij} = a_{ij} + \alpha * x_i$$

۵. جمع دو ماتریس متقارن A و B از مرتبه $n \times n$

Input: Symmetric matrices A and B

output: $C = A + B$. The algorithm overwrites B with C

For $i = 1, \dots, n$ do

For $j = i, \dots, n$ do

$$b_{ij} = a_{ij} + b_{ij}$$

If $i \neq j$

$$b(j, i) = b(i, j)$$

فصل ۵

۱. (الف) $(I + m_k e_k^T)(I - m_k e_k^T) = I - m_k e_k^T + m_k e_k^T - m_k e_k^T m_k e_k^T = I$ (زیرا جمله آخر ماتریس صفر است).

$$E = \begin{bmatrix} 1 & 0 \\ 1.0^{-5} & 1 \end{bmatrix} \quad (الف) \quad ۲$$

$$U = \begin{bmatrix} 1.0^{-5} & 1 \\ 0 & -.9999 \times 1.0^5 \end{bmatrix} \quad (ب)$$

$$L = \begin{bmatrix} 1 & 0 \\ 1.0^5 & 1 \end{bmatrix}$$

۱۳

For $k = 1, 2, \dots, r$ do

$$a_{ik} \equiv m_{ik} = -\frac{a_{ik}}{a_{kk}} \quad (i = k + 1, \dots, m)$$

$$a_{ij} \equiv a_{ij} + m_{ik} a_{kj}$$

$$\equiv a_{ij} - a_{ik} a_{kj} \quad (i = k + 1, \dots, m; j = k + 1, \dots, m)$$

E ها از ضربگرهای m_{ik} ، به صورت نشان داده شده در کتاب، تشکیل خواهند شد.

۱۴. یک مجموعه از نامساویها بنویسید که تضمین نماید که A غالب قطری ستونی است.

۱۵. L و U را به ترتیب به صورت ماتریسهای پایین مثلثی واحد و بالا مثلثی اختیار کنید و عناصر هر دو طرف $A = LU$ را برابر قرار دهید.

۲۰. فقط یک عنصر لازم است که در هر گام حذف شود. تعداد اعمال ممیز شناور برابر $O(n)$ است.

۲۷. $Q^T = Q_{n-1} Q_{n-2} \dots Q_2 Q_1$. هر Q_i بجز در عناصر $(i, i+1)$ و $(i+1, i)$ قطری است. بنابراین حاصلضرب Q هسنبرگی است.

۲۹. الگوریتم زیر با مفروض بودن i و k ، AJ را محاسبه می‌کند و بر روی A ماتریس AJ را می‌نویسد.

For $j = 1, \dots, n$ do

$$\alpha = a_{ji}$$

$$\beta = a_{jk}$$

$$a_{ji} = c\alpha - s\beta$$

$$a_{jk} = s\alpha + c\beta$$

۴ عمل ممیز شناور برای هر j لازم است.

تعداد کل اعمال ممیز شناور برابر $4n$ است.

۳۵. (الف) $A^{(k)} = P_k A P_k^T$. $P_k A$ نیاز به $2(n-k)^2$ عمل ممیز شناور دارد. $(P_k A) P_k^T$ نیاز به

$2n(n-k)$ عمل ممیز شناور دارد، اکنون راهنمایی ارائه شده در مسأله را دنبال کنید.

(د) اگر A متقارن باشد، شکل کاهش یافته سه قطری متقارن است و در نتیجه فقط در حدود نیمی از

عناصر لازم است که محاسبه شوند.

۳۶. (الف) اختیار کنید

$$X = (x_1, x_2, \dots, x_n)^T, x_1 = (1, 0, \dots, 0)^T$$

آنگاه از $HX = XC$ نتیجه می‌شود که $x_i = H^{i-1} e_1$ ، $i = 1, \dots, n$. ماتریس H یک

ماتریس بالا مثلثی با $h_{11}, h_{12}, h_{22}, \dots, h_{21}, h_{22}, \dots, h_{n-1}, \dots, h_{n1}$ به عنوان عناصر قطری است.

(ب) اگر عناصر قطر زیرین H خیلی کوچک باشند، X بدوضع خواهد بود.

۳۷. ابتدا یک ماتریس هاوس هولدر P_1 به قسمی پیدا کنید که $P_1 b$ مضربی از e_1 باشد.

$H_1 = P_1 A P_1^T$ را تشکیل دهید. H_1 را با استفاده از تبدیهای هاوس هولدر به بالا هستبرگی

تبدیل کنید. هر بار که یک تبدیل هاوس هولدر در فرایند اعمال شود، b جدید را به هنگام درآورید.

فصل ۶

۱. فرض کنید x, y, z یاردهای مکعب مورد نیاز باشند. آنگاه $5000 = 0.01x + 0.02y + 0.03z$ ؛

$$5500 = 0.03z + 0.04y + 0.02x \quad ; \quad 6000 = 0.05z + 0.02y + 0.01x$$

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 10 & -10 & 0 & -15 & -5 \\ 5 & -10 & 0 & -20 & 0 & 0 \end{pmatrix} \begin{bmatrix} i_{12} \\ i_{52} \\ i_{32} \\ i_{65} \\ i_{52} \\ i_{43} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 150 \end{bmatrix}$$

- گره (۱, ۱) $-4T_{11} + T_{21} + T_{12} = -150$
- گره (۲, ۱) $T_{11} - 4T_{21} + T_{31} + T_{22} = -50$
- گره (۳, ۱) $T_{21} - 4T_{31} + T_{32} = -125$
- گره (۱, ۲) $T_{11} - 4T_{12} + T_{22} + T_{13} = -100$
- گره (۲, ۲) $-4T_{22} + T_{32} + T_{23} + T_{12} + T_{21} = 0$
- گره (۳, ۲) $-4T_{22} + T_{32} + T_{23} + T_{31} = -75$
- گره (۱, ۳) $-4T_{13} + T_{23} + T_{12} = -100$
- گره (۲, ۳) $-4T_{23} + T_{33} + T_{13} + T_{22} = 0$
- گره (۳, ۳) $-4T_{23} + T_{33} + T_{32} = -75$

۱۲. (الف) $H = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0.316 & 0 \\ 1 & 0.316 & 0.9995 \end{bmatrix}$ (عامل چولسکی)

(ج) $\rho = \frac{\max(2, 1, 1)}{2} = 1$ (عامل رشد)

۱۴. (الف) کهادهای اصلی پیشرو حاصل از ماتریس U از تجزیه LU ماتریس A ، با استفاده از فرایند حذفی گاوس بدون محورگیری عبارتند از: ۴، ۱۵، ۵۶، ۱۹۲. بنابراین A معین مثبت است. تجزیه چولسکی ماتریس A را با استفاده از دستور $\text{chol}(A)$ از MATLAB به دست آورید و ملاحظه کنید که همه عناصر قطری عامل چولسکی مثبت هستند.

(ب) $x = (1, 1, 1, 1)^T$

۱۶. برای $k = ۱$

$$\begin{aligned} \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}^{(1)}| &= \sum \left| a_{ij} - \frac{a_{1j}a_{i1}}{a_{11}} \right| \leq \sum |a_{ij}| + \sum \left| \frac{a_{1j}a_{i1}}{a_{11}} \right| \\ &< (|a_{ii}| - |a_{i1}|) + \frac{|a_{i1}|}{|a_{11}|} (|a_{11}| - |a_{i1}|) \\ &= |a_{ii}| - \frac{|a_{i1}| |a_{11}|}{|a_{11}|} \leq \left| a_{ii} - \frac{a_{11} a_{i1}}{a_{11}} \right| = |a_{ii}^{(1)}| \end{aligned}$$

۱۹. (الف) (i) فرض کنید $H = (h_{ij})$ یک ماتریس $n \times n$ بالا هسنبرگی باشد. برای هر مقدار k از ۱ تا $n - ۱$ گامهای زیر را تشکیل دهید.

۱. h_{kj} و $h_{k+1,j}$ را با هم تعویض نمایید اگر

$$|h_{kk}| < |h_{k+1,k}|, \quad j = k, \dots, n$$

۲. $m = \frac{-h_{k+1,k}}{h_{kk}}$ را محاسبه کنید.

۳. $h_{k+1,j} \equiv h_{k+1,j} + mh_{kj}$ ، $j = k + 1, \dots, n$ را محاسبه کنید.

$$28. \text{Cond}(A^{-1}) = \|A^{-1}\| \|A\| = \text{Cond}(A) \quad (\text{الف})$$

(ب) $\|I\| = ۱$ (با فرض این که $\|I\| = ۱$) $\|I\| = \|A^{-1}A\| \leq \|A^{-1}\| \|A\| = \text{Cond}(A)$

$$\|A^T A\|_F = \|A\|_F^2 = \sigma_1^2 \dots \sigma_n^2 \text{ اکنون } \text{Cond}_F(A^T A) = \|A^T A\|_F \|(A^T A)^{-1}\|_F$$

که در آن σ_1 و σ_n بزرگترین و کوچکترین مقادیر تکین $\|(A^T A)^{-1}\|_F = \frac{1}{\sigma_n^2}$

$$\text{ماتریس } A \text{ هستند. بنابراین } \text{Cond}_F(A) = \frac{\sigma_1^2}{\sigma_n^2} = (\text{Cond}_F(A))^2 \text{ (بخش}$$

۱۰-۶-۱ را نیز ملاحظه کنید).

$$29. \text{Cond}_F(O) = \|O\|_F \|O^{-1}\|_F = \|O\|_F \|O^T\|_F = ۱ \times ۱ = ۱ \quad (\text{الف})$$

(ب) $\text{Cond}_F(A) = \text{Cond}_F(O) = ۱$ ؛ $A = \alpha O$ اگر $\text{Cond}_F(A) = \text{Cond}_F(O)$ برابر ۱ باشد. آنگاه $\|A^{-1}\|_F$

باید معکوس $\|A\|_F$ باشد.

۳۰. چون U بالا مثلثی است، $\|U\|_\infty \geq \max(u_{ii})$ و $\|U^{-1}\|_\infty \geq \frac{1}{\min(u_{ii})}$. اختیار کنید

پ

$$U = \begin{bmatrix} 0/0001 & 2 & 3 \\ \cdot & 1 & 12 \\ \cdot & \cdot & 12 \end{bmatrix}, \text{Cond}(U) = 4/90.25 \times 10^5$$

۳۲. الف) اثبات قضیه ۱۰-۶-۱ را در فصل ۱۰ ملاحظه کنید.

۳۳. ب) اثبات از قضیه ۶-۶-۱ و با توجه به $\delta b = -2$ نتیجه می‌شود.

ج) اگر $\text{Cond}(A)$ و r کوچک باشند، آنگاه خطای نسبی کوچک است. پس، بر طبق تعریف،

الگوریتم پایدار ضعیف است.

۳۴. ج) $\text{Cond}(A) = 1/999 \times 10^3$

۳۶. الف) A غالب قطری و معین مثبت متقارن است

۳۸. الف) گاوس - سایدل بلوکی

$$A_{ii}x_i^{(k+1)} = - \sum_{j < i} A_{ij}x_j^{(k+1)} - \sum_{j > i}^n A_{ij}x_j^{(k)} + b_i, \quad i = 1, \dots, N, k = 0, 1, \dots$$

۴۴. $\Phi(x - \alpha p) =$

$$= \frac{1}{\alpha} (x - \alpha p)^T A (x - \alpha p) - b^T (x - \alpha p)$$

$$= \frac{1}{\alpha} x^T A x - \alpha p^T A x + \frac{1}{\alpha} \alpha^2 p^T A p + \alpha p^T b - b^T x$$

(توجه کنید که $b^T p$ یک اسکالر است)

$$= \frac{1}{\alpha} p^T A p \alpha^2 - p^T (A x - b) \alpha + \frac{1}{\alpha} x^T (A x - 2b)$$

از $\Phi'(x - \alpha p) = 0$ و $p^T A p > 0$ ، عبارت مزبور را برای α داریم (مرجع

Ortega (1988a, 186) را ملاحظه کنید).

فصل ۷

۱. فرض کنید A دارای رتبه کامل باشد. آنگاه برای هر بردار مخالف صفر x ، $Ax = y$ مخالف صفر

$$\text{است. بنابراین } x^T A^T A x = y^T y > 0$$

۲. فرض کنید $z \in R(A)$. آنگاه یک y وجود دارد به طوری که $Ay = z$.

$$z^T r = z^T (b - Ax) = z^T b - z^T A (A^T A)^{-1} A^T b = y^T A^T b - y^T A^T A (A^T A)^{-1} A^T b = 0$$

(چون $x = (A^T A)^{-1} A^T b$)

۳. فرض کنید $b - Ax \in N(A^T)$. آنگاه $A^T(b - x) = 0$ ، یعنی $A^T Ax = A^T b$.

۵. $\|A^+\|_F^2 = \|A^+ A^{+T}\|_F = \|(A^T A)^{-1}\|_F$ ، $\|A\|_F^2 = \|A^T A\|_F$ ، بنابراین

$$Cond_F(A) = \|A\|_F^2 \|A^+\|_F^2 = Cond_F(A^T A)$$

۱۷. (الف)

For $k = 1, 2, \dots, n$ do

For $i = 1, 2, \dots, k - 1$ do

$$r_{ik} = q_i^T a_k, a_k \equiv a_k - r_{ik} q_i$$

$$\hat{q}_k = a_k, r_{kk} = \|\hat{q}_k\|_F$$

$$q_k = \frac{\hat{q}_k}{r_{kk}}$$

۲۰. جواب پایه‌ای

$$x_B = P \begin{bmatrix} R_{11}^{-1} c \\ \cdot \end{bmatrix}$$

جواب با می‌نیمم نرم با اختیار کردن y به صورت زیر محاسبه می‌شود

$$\min_{y_1} \left\| \begin{bmatrix} R_{11}^{-1} R_{12} \\ -I_{n-r} \end{bmatrix} y_1 - \begin{bmatrix} R_{11}^{-1} c \\ \cdot \end{bmatrix} \right\|_F$$

بنابراین جواب x_B نمی‌تواند جواب با می‌نیمم نرم باشد مگر $R_{12} = 0$.

۲۱. (الف)

$$\begin{aligned} \|Ax - b\|^2 &= \|QT^T W^T P^T x - b\|^2 \\ &= \|T^T W^T P^T x - Q^T b\|^2 \\ &= \|Ty_1 - c\|^2 + \|d\|^2 \end{aligned}$$

که در آن

$$T = \begin{bmatrix} T & \cdot \\ \cdot & \cdot \end{bmatrix}, \quad Q^T b = \begin{bmatrix} c \\ d \end{bmatrix}$$

$$W^T P^T x = y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

بنابراین همه جوابهای کمترین توانهای دوم به شکل

$$x = PW \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

هستند که در آن y_1 جواب منحصر به فرد $Ty_1 = c$ است. جواب با می‌نیم نرم از قرار دادن $y_2 = 0$ به دست می‌آید.

۲۹. چون $X = (A^T A)^{-1} = R^{-1}(R^{-1})^T$ ، عناصر قطری X دقیقاً برابر توان دوم نرم دو سطرهای

$$R^{-1} \text{ هستند: } S = (s_{ij}) = R^{-1} \text{ که در آن } i = 1, 2, \dots, n, x_{ii} = \sum_{j=1}^n s_{ij}^2.$$

فصل ۸

۵. بر طبق قضیه ۸-۴-۱، $r_i = \sum_{j \neq i} |a_{ij}|$ ، چون A غالب قطری است، داریم

$$|\lambda - a_{ii}| \leq \sum_{j=1}^n |a_{ij}| < |a_{ii}| \text{، و این نتیجه می‌دهد } \lambda \neq 0.$$

۷. مقادیر ویژه یک ماتریس متقارن A حقیقی هستند. بنابراین قضیه ۸-۴-۱ نتیجه می‌دهد

$$-r_i \leq \lambda - a_{ii} \leq r_i.$$

$$(A - \sigma I)v_i = Av_i - \sigma v_i = \lambda v_i - \sigma v_i = (\lambda - \sigma)v_i \quad ۹.$$

۱۰. (الف) $\lambda_2 = 2/9$ نزدیک به $\lambda_1 = 3$ است.

۱۹. (ه) توجه کنید که در این حالت همه s_j ها برابر ۱ هستند. بنابراین نتیجه از قضیه ۸-۸-۱ حاصل

می‌شود.

۲۰. مقادیر ویژه خوش‌وضع هستند، زیرا A متقارن است. اما وقتی ε کوچک است مقادیر ویژه $1 + \varepsilon$

و $1 - \varepsilon$ نزدیک به یکدیگر هستند. بنابراین بر طبق قضیه ۸-۸-۱، بردارهای ویژه متناظر با این

مقادیر ویژه بدوضع هستند.

$$X = \begin{bmatrix} 1 & 1 \\ 0 & \varepsilon \end{bmatrix} \quad \text{(الف) } ۲۳.$$

اگر ε کوچک باشد X بدوضع است؛ آنگاه بر طبق قضیه بوئر - فایک، مقادیر ویژه A بدوضع

هستند.

۲۵. اختیار کنید

$$Q = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 3 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 0 & 1 & 3 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

آنگاه $H = QR$ هستبرگی است، اما Q چنین نیست. چون H بالا هستبرگی است و $r_{kk} \neq 0$ از $H = QR$ یا $h_{kl} = \sum_s q_{ks} r_{sl}$ ، به ازای $k \geq l + 2$ داریم $q_{kl} = 0$.

۲۹. (الف) ماتریس Q هستبرگی است. بنابراین ستون اول QR برابر $(q_{11}, q_{21}, \dots, 0)^T$ است. ستون اول $H - \lambda I$ برابر $(h_{11} - \lambda, h_{21}, \dots, 0)^T$ است. چون آنها باید یکی باشند، ستون اول Q حاصلضرب $\frac{1}{r_{11}}$ در ستون اول $H - \lambda I$ می باشد.

(ج) هر یک از دورانه‌های گیونز در مسأله دارای مقدار ۱ در عنصر $(1, 1)$ هستند. بنابراین \tilde{Q} دارای همان ستون اول P_0 و همان ستون اول Q_1 بر طبق قسمت (ب) می باشد.

۳۳. در بخش ۸-۹-۵ ثابت شده است که H_{S+2} مشابه با H_S است. بنابراین آنها دارای مقادیر ویژه یکسان هستند.

۳۵. (الف) قضیه ۸-۴-۱ را به کاربرد و نشان دهید که $|\lambda| < 4$. دنباله استورم محاسبه شده در $\mu = 2$ مقادیر $1, -1, 1, -1, 1, -1, 1$ را ارائه می دهد. دو توافق و دو تغییر علامت وجود دارد. بنابراین دو مقدار ویژه از A بزرگتر از ۲ و دو مقدار ویژه کوچکتر از ۲ وجود دارند.

فصل ۹

۱. فرض کنید x بردار پوچ مخالف صفر مشترک A و B باشد. پس به ازای هر λ داریم $(A - \lambda B)x = 0$ ، چون $x \neq 0$ پس به ازای هر λ خواهیم داشت $\det(A - \lambda B) = 0$.

۲. (ب) فرض کنید $B = P^{-1}\Lambda P$ ، که در آن Λ شکل متعارف جردن B است. آنگاه $P(\lambda) = \det(A - \lambda B) = \det(PAP^{-1} - \lambda\Lambda)$ نامنفرد باشد، و نامنفرد بودن Λ ایجاب می کند که $P(\lambda)$ از درجه n باشد.

۳. Q_1 یک ماتریس هاس هولدر است که ستون اول n_1 از ماتریس N را به مضربی از e_1 تبدیل می کند. n_1 دارای سه عنصر مخالف صفر x, y, z است و در بخش ۹-۳-۲ نشان داده شده است که این عناصر می توانند از معکوس کردن کهاد اصلی پیشرو از مرتبه 2×2 ماتریس B محاسبه شوند.

۸. اگر x بردار ویژه (A, B) متناظر با یک مقدار ویژه λ باشد، آنگاه $y = V^T x =$ یک بردار ویژه (\tilde{A}, \tilde{B}) متناظر با λ خواهد بود.

۹. (الف) تجزیه شور B را پیدا کنید: $V^T B V = D$ فرض کنید که $D = LL^T$.
(ب) ماتریس $C = (V^T L)^{-1} A (L^T V)^{-1}$ را تشکیل دهید.

(ج) مقادیر ویژه و بردارهای ویژه $y_i, i = 1, \dots, n$ ، ماتریس مقارن C را با استفاده از تکرار QR

حساب کنید.

(د) بردارهای ویژه x_i و مقدار (A, B) را با حل $L^T V x_i = y_i$ ، $i = 1, \dots, n$ حساب کنید.

۱۳. انتخاب کنید $I_{4 \times 4} = Q_0$. سپس در انتهای تکرار پنجاهم بردارهای ویژه را برای جفت (A, B)

داریم

$$Q = \begin{bmatrix} -0.0093 & -0.9730 & -0.2307 & 0.0000 \\ 0.3820 & -0.1948 & 0.8059 & 0.4082 \\ 0.5590 & -0.0384 & 0.1394 & -0.8165 \\ 0.7359 & 0.1180 & -0.5271 & 0.4082 \end{bmatrix}$$

۲۶. فرض کنید λ مقدار ویژه جفت (A, B) و x بردار وابسته به آن باشد. آنگاه $Ax = \lambda Bx$. یعنی

$$B^{-1}Ax = \lambda x \quad \text{بنابراین} \quad \| \lambda \| = \frac{\| B^{-1}Ax \|}{\| x \|} \leq \| B^{-1}A \|$$

(A, B) حقیقی هستند، نتیجه را خواهیم داشت.

فصل ۱۰

۳. اختیار کنید $\tilde{U} = UP$ و $\tilde{V} = VP$ ، که ماتریس P سطر i و سطر $i+1$ از Σ را تعویض می‌کند.

۶. (ب) فرض کنید $A = U\Sigma V^T$. آنگاه $A^T A = V\Sigma^T \Sigma V^T$ و $AA^T = U\Sigma \Sigma^T U^T$. چون $\Sigma^T \Sigma$ و

$\Sigma \Sigma^T$ دارای یک تعداد مقادیر ویژه مخالف صفر هستند، $\text{rank}(\Sigma \Sigma^T) = \text{rank}(\Sigma^T \Sigma)$ ، و در

نتیجه AA^T و $A^T A$ دارای رتبه یکسان هستند. برای اثبات (iii) توجه کنید که

$$(Au_1)^T Au_1 = u_1^T A^T A u_1 = \lambda u_1^T u_1 = 0$$

۷. (الف) در حل تمرین ۴۹ فصل ۱ نشان دادیم که

$$\max_{x \neq 0} \frac{\| Ax \|_2}{\| x \|_2} \leq \sigma_1 \quad \text{و} \quad \frac{\| Av_1 \|_2}{\| v_1 \|_2} = \sigma_1$$

(ب) جوابهای تمرین ۴۹ فصل ۱ را ملاحظه کنید.

۸. فرض کنید $E = B - A$. فرض کنید $i = 1, \dots, n$ ، σ_i و $i = 1, \dots, n$ ، $\tilde{\sigma}_i$ ، به ترتیب مقادیر

تکین A و B باشند. بر طبق قضیه ۱۰-۵-۱ به ازای هر i داریم $\sigma_i < \tilde{\sigma}_i$ و $|\tilde{\sigma}_i - \sigma_i| \leq \| E \|_2$.

بنابراین $\tilde{\sigma}_i$ نمی‌تواند برابر صفر به ازای هر $i = 1, \dots, r$ باشد.

۱۵. (الف) فرض کنید $v \in R(A)$. آنگاه بردار u یی وجود دارد به طوری که $v = Au$ ، با استفاده از

$$AA^+v = AA^+Au = Au = v$$

(ب) فرض کنید $x \in N(A^T)$ آنگاه $A^+x = A^+(AA^+)^T x = A^+(A^+)^T A^T x = 0$

۱۶. (الف) فرض کنید $X = (A^T A)^{-1} A^T$. آنگاه X وجود دارد زیرا A دارای رتبه کامل است. چون

$$AXA = A(A^T A)^{-1} A^T A = A,$$

$$XAX = (A^T A)^{-1} A^T A (A^T A)^{-1} A^T = (A^T A)^{-1} A^T,$$

$$(AX)^T = X^T A^T = A(A^T A)^{-1} A^T = AX,$$

$$(XA)^T = A^T X^T = A^T A (A^T A)^{-1} = I,$$

$$XA = (A^T A)^{-1} A^T A = I$$

پس داریم $A^+ = (A^T A)^{-1} A^T$.

۲۱. بدون از دست دادن کلیت مسأله، فرض کنید که B یک ماتریس 2×2 باشد. بعلاوه فرض کنید که هیچ

کدام از عناصر قطری صفر نباشند. فرض کنید

$$B = \begin{bmatrix} \alpha_1 & \beta_1 \\ 0 & \alpha_2 \end{bmatrix}$$

آنگاه

$$B^T B = \begin{bmatrix} \alpha_1^2 & \alpha_1 \beta_1 \\ \beta_1 \alpha_1 & \beta_1^2 + \alpha_2^2 \end{bmatrix}$$

چند جمله‌ای مشخصه $p(\lambda)$ ماتریس $B^T B$ برابر $\alpha_1^2 \alpha_2^2 + \beta_1^2 \lambda - (\alpha_1^2 + \alpha_2^2 + \beta_1^2) \lambda + \alpha_1^2 \alpha_2^2$ می‌باشد.

بنابراین B یک مقدار تکین چندگانه دارد اگر و فقط اگر $p(\lambda)$ دارای یک صفر چندگانه باشد. چون

α_1 و α_2 صفر نیستند، $p(\lambda)$ می‌تواند یک صفر چندگانه داشته باشد فقط اگر $\beta_1 = 0$ باشد.

فصل ۱۱

$$L = \begin{bmatrix} 1 & 0 \\ 1/67 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 3 & 4 \\ 0 & -0/68 \end{bmatrix} \quad E = \begin{bmatrix} 0 & 0 \\ 0/01 & 0 \end{bmatrix} \quad (الف)$$

$$L = \begin{bmatrix} 1 & 0 \\ 0/04 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 0/25 & 0/79 \\ 0 & 0/09 \end{bmatrix} \quad E = \begin{bmatrix} 0 & 0 \\ 0 & 0/016 \end{bmatrix} \quad (ب)$$

$$L = \begin{bmatrix} 1 & 0 \\ 0/8 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 10 & 9 \\ 0 & -2/2 \end{bmatrix} \quad E = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (ج)$$

$$L = \begin{bmatrix} 1 & 0 \\ 0/25 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 4 & 1 \\ 0 & 1/75 \end{bmatrix} \quad E = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (د)$$

$$L = \begin{bmatrix} 1 & \cdot \\ 3 & 1 \end{bmatrix} \quad U = \begin{bmatrix} \cdot/\cdot 1 & \cdot/\cdot 0.5 \\ \cdot & \cdot/\cdot 14 \end{bmatrix} \quad E = \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \quad (\text{هـ})$$

$$P = \begin{bmatrix} \cdot & 1 \\ 1 & \cdot \end{bmatrix}, \quad L = \begin{bmatrix} 1 & \cdot \\ \cdot/\cdot 6 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 5 & 6 \\ \cdot & \cdot/\cdot 4 \end{bmatrix} \quad (\text{الف}). ۲$$

$$E = LU - PA = \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

$$P = \begin{bmatrix} \cdot & 1 \\ 1 & \cdot \end{bmatrix} \quad L = \begin{bmatrix} 1 & \cdot \\ \cdot/\cdot 33 & 1 \end{bmatrix} \quad U = \begin{bmatrix} \cdot/\cdot 3 & \cdot/\cdot 1 \\ \cdot & \cdot/\cdot 5 \end{bmatrix} \quad (\text{هـ})$$

$$E = LU - PA = \begin{bmatrix} \cdot & \cdot \\ \cdot/\cdot 0.01 & \cdot/\cdot 0.33 \end{bmatrix}$$

منابع

- Adams, L., and P. Arbenz. 1994. Towards a divide and conquer algorithm for the real nonsymmetric eigenvalue problem. *SIAM J. Matrix Anal. Appl.* 15:4ff.
- Ames, W. F. 1992. *Numerical Methods for Partial Differential Equations*. New York: Academic Press.
- Ammar, G., and V. Mehrmann. 1991. On Hamiltonian and symplectic Hessenberg forms. *Lin. Alg. Appl.* 149:55–72.
- Anderson, E., Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Oustrouchov, and D. Sorensen. 1992. *LAPACK User's Guide, Release 1.0*. Philadelphia: SIAM.
- Andrews, H. C., and B. R. Hunt. 1977. *Digital Image Processing*. Englewood Cliffs, N.J.: Prentice Hall.
- ANSI/IEEE. 1985. *IEEE Standard for Binary Floating Point Arithmetic*, Std 754–1985 edition.
- ANSI/IEEE. 1987. *IEEE Standard for Radix Independent Floating Point Arithmetic*, Std 854–1987 edition.
- Anton, H., and C. Rorres. 1987. *Elementary Linear Algebra with Applications*. New York: John Wiley.
- Arioli, M., I. Duff, and D. Ruiz. 1992. Stopping criteria for iterative solvers. *SIAM J. Matrix Anal. Appl.* 13:138–144.
- Arioli, M., J. W. Demmel, and I. Duff. 1989. Solving sparse linear systems with sparse backward error. *SIAM J. Matrix Anal. Appl.* 10:165–190.
- Arioli, M. and A. Laratta. 1985. Error analysis of an algorithm for solving an undetermined system. *Numer. Math.* 46:255–268.
- Arnold, M. 1992. Algorithms and conditioning for eigenvalue assignment. Ph.D. diss., Northern Illinois University, DeKalb, Illinois.
- Arnold, W. F. and A. J. Laub. 1984. Generalized eigenproblem algorithms and software for algebraic Riccati equations. *Proceedings, IEEE* 72:1746–1754.
- Arnoldi, W. 1951. The principle of minimized iteration in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.* 9:17–29.
- Atkinson, K. 1993. *Elementary Numerical Analysis*, 2d ed. New York: John Wiley.
- Axelsson, O. 1985. A survey of preconditioned iterative methods for linear systems of equations. *BIT* 25:166–187.

- . 1994. *Iterative Solution Methods*. New York: Cambridge University Press.
- Bai, Z. 1988. Note on the quadratic convergence of Kogbetliantz's algorithm for computing the singular value decomposition. *Lin. Alg. and Appl.* 104:131–140.
- Bai, Z. and J. Demmel. 1993. Computing the generalized singular value decomposition. *SIAM J. Sci. Computing* 14, 1464–1486.
- . 1993a. On swapping diagonal blocks in real Schur form. *Lin. Alg. Appl.* 186:73–96.
- Bai, Z., J. Demmel, and A. McKenney. 1993. On computing condition numbers for the non-symmetric eigenproblems. *ACM Trans. Math. Soft.* 19:202–223.
- Bai, Z., D. Hu, and L. Reichel. 1994. A Newton basis GMRES implementation, *IMA J. Num. Anal.* Forthcoming.
- Barlow, J. 1993. Error analysis of update methods for the symmetric eigenvalue problem. *SIAM J. Matrix Anal. Appl.* 14:598–618.
- Barlow, J. and J. Demmel. 1990. Computing accurate eigensystems of scaled diagonally dominant matrices. *SIAM J. Num. Anal.* 27:762–791.
- Barrett, R., et al. 1994. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. Philadelphia: SIAM Publications.
- Bartels, R. H. and G. W. Stewart. 1972. Solution of the equation $AX + XB = C$. *Comm. ACM* 15:820–826.
- Bauer, F. L. 1963. Optimally scaled matrices. *Numer. Math.* 5:73–87.
- Bauer, F. L. and C.T. Fike. 1960. Norms and exclusion theorems. *Numer. Math.* 2:137–144.
- Becker, E. B., G. F. Carey, and J. T. Oden. 1981. *Finite Element—An Introduction*. Vol. 1. Englewood Cliffs, N.J.: Prentice Hall.
- Berman, A. and R. Plemmons. 1994. *Nonnegative Matrices in Mathematical Sciences*. 2d ed. Philadelphia: SIAM Publications.
- Bischof, C., B. N. Datta, and A. Purkayastha. 1994. A parallel algorithm for the Sylvester-Observer equation. *Argonne Tech. Report MCS P274-1191:1-17*.
- Bischof, C., J. Lewis, and D. Pierce. 1990. Incremental condition estimation for sparse matrices. *SIAM J. Matrix Anal. Appl.* 11:644–659.
- Björck, A. 1967. Iterative refinement of linear least squares solution I. *BIT* 7:257–278.
- . 1968. Iterative refinement of linear least squares solution II. *BIT* 8:8–30.
- . 1991a. Component-wise perturbation analysis and errors bounds for linear least squares solutions. *BIT* 31:238–244.
- . 1991b. Error analysis of least squares algorithms. In *Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms*, edited by G. H. Golub and P. Van Dooren. NATO ASI Series. Berlin: Springer-Verlag.
- . 1994a. Numerics of Gram-Schmidt orthogonalization. *Lin. Alg. Appl.* 197–198, 297–316.
- . 1994b. *Numerical Methods for Least Squares Problems*. Frontiers in Applied Mathematics Series. Philadelphia: SIAM.
- Björck, A. and G. H. Golub. 1967. Iterative refinement of linear least squares solutions by Householder transformation. *BIT* 7:322–337.
- Björck, A. and C. Paige. 1992. Loss and recapture of orthogonality in the modified Gram-Schmidt algorithms. *SIAM J. Matrix Anal. Appl.* 13:176–190.
- Bojanczyk, A. W., R. P. Brent, F. R. de Hoog, and D. R. Sweet. On the stability of the Bareiss and related Toeplitz factorization algorithms. *SIAM J. Sci. Stat. Comput.*, forthcoming.
- Bojanczyk, A. W., R. P. Brent, P. Van Dooren, and F. R. de Hoog. 1987. A note on downdating the Cholesky Factorization. *Siam J. Sci. Stat. Comp.* 8:210–221.
- Boley, D. 1987. Computing rank-deficiency of rectangular matrix pencils. *Systems and Control Letters* 9:207–214.
- . 1990. Estimating the sensitivity of the algebraic structure of matrix pencils with simple eigenvalue estimates. *SIAM J. Matrix Anal. Appl.* 11:632–643.

- Boley, D., R. Brent, G. H. Golub, and F. Luk. 1992. Algorithmic fault tolerance using the Lanczos method. *SIAM J. Matrix Anal. Appl.* 13:312-332.
- Boley, D. and G. H. Golub. 1984. The Lanczos-Arnoldi algorithm and controllability. *Systems and Control Letters* 4:317-324.
- Bramley, R. and A. Sameh. 1992. Row projection methods for large nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.* 13:168-193.
- Braun, M. 1978. *Differential Equations and Their Applications*. New York: Springer-Verlag.
- Braaldi, R., D. Carlson, B. N. Datta, C. R. Johnson, and R. Plemmons, ed. 1985. *Linear Algebra and its Role in Systems Theory*, Contemporary Mathematics Series 47. Providence, R.I.: American Mathematical Society.
- Braaldi, R. and S. Mellendorf. 1994. Regions in the complex plane containing the eigenvalues of a matrix. *Amer. Math. Monthly*, forthcoming.
- Bunch, J. R. 1971. Analysis of the diagonal pivoting method. *SIAM J. Numer. Anal.* 8:656-680.
- . 1987. The weak and strong stability of algorithms in numerical linear algebra. *Lin. Alg. Appl.* 88/89:49-66.
- Bunch, J. R., J. Demmel, and C. Van Loan. 1989. The strong stability of algorithms for solving symmetric linear systems. *SIAM J. Matrix Anal. Appl.* 10(4):494-499.
- Bunch, J. R. and L. Kaufman. 1977. Some stable methods for calculating inertia and solving symmetric linear systems. *Math. Comp.* 31:162-179.
- Bunch, J. R., L. Kaufman, and B. Parlett. 1976. Decomposition of a symmetric matrix. *Numer. Math.* 27:95-109.
- Bunch, J. R., C. P. Nielson, and D. C. Sorensen. 1978. Rank-one modification of the symmetric eigenproblem. *Numer. Math.* 31:31-48.
- Bunch, J. R. and B. N. Parlett. 1971. Direct methods for solving symmetric indefinite systems of linear equations. *SIAM J. Num. Anal.* 8:639-655.
- Bunse-Gerstner, A. and V. Mehrmann. 1986. A symplectic QR-like algorithm for the solution of the real algebraic Riccati equation. *IEEE Trans. Auto. Cont.*, AC-31:1104-1113.
- Businger, P. A. and G. H. Golub. 1965. Linear least squares solutions by Householder transformation. *Numer. Math.* 7:269-276.
- . 1969. Algorithm 358: Singular value decomposition of a complex matrix. *Comm. ACM* 12:564-565.
- Byers, R. 1984. A Linpack-style condition estimator for the equation $AX - XB^T = C$. *IEEE Trans. Auto. Cont.* AC-29:926-928.
- . 1986. A Hamiltonian QR algorithm. *SIAM J. Sci. Stat. Comput.* 7:212-229.
- Callaerts, D., B. De Moor, J. Vandewalle, and W. Sansen. 1990. Comparison of SVD-methods to extract the fetal electrocardiogram from cutaneous electrode signals. *Medical and Biological Engineering and Computing.* 217-224.
- Calvetti, D., L. Reichel, and D. Sorensen. 1994. An implicitly restarted Lanczos method for large symmetric eigenvalue problem. *Electr. Trans. Numer. Anal.* 2:1-21.
- Campbell, S. L. 1985. Rank deficient least squares and the numerical solution of linear singular systems of differential equations. In *Linear Algebra and Its Role in Systems Theory*, edited by R. Braaldi et al., 51-63. Contemporary Mathematics Series 47. Providence, R.I.: American Mathematical Society.
- Campbell, S. L. 1990. *An Introduction to Differential Equations and Their Applications*. 2nd ed. Belmont, Calif.: Wadsworth.
- Celia, M. A. and W. G. Gray. 1992. *Numerical Methods for Differential Equations*. Englewood Cliffs, N.J.: Prentice Hall.
- Carlson, D. and B. N. Datta. 1979. On the effective computation of the inertia of a non-Hermitian matrix. *Numer. Math.* 33:315-322.
- Chan, T. F. 1982a. An improved algorithm for computing the singular value decomposition. *ACM Trans. Math. Soft.* 8:72-83.

- . 1982b. Algorithm 581: An improved algorithm for computing the singular value decomposition. *ACM Trans. Math. Soft.* 8:84–88.
- . 1987. Rank-revealing *QR* factorizations. *Lin. Alg. and Its Appl.* 88/89:67–82.
- Chan, T. F. and P. C. Hansen. 1992. Some applications of the rank-revealing *QR* factorization. *SIAM J. Sci. Stat. Comput.* 13:727–741.
- Chandrasekaran S. and I. Ipsen. 1994a. Backward errors for eigenvalue and singular value decomposition. *Numer. Math.*, forthcoming.
- . 1994b. On the sensitivity of solution components in linear systems of equations. *SIAM J. Matrix Anal. Appl.*, forthcoming.
- . 1994c. On rank-revealing *QR* factorizations. *SIAM J. Matrix Anal. Appl.* 15:2.
- . 1994d. Analysis of a *QR* algorithm for computing singular values. Forthcoming.
- Chapara S. C. and R. P. Canale. 1988. *Numerical Methods for Engineers*. New York: McGraw-Hill.
- Charlier, J. P., M. Vanbegin, and P. Van Dooren. 1988. On efficient implementation of Kogbetliantz's algorithm for computing the singular value decomposition. *Numer. Math.* 52:279–300.
- Charlier, J. P. and P. Van Dooren. 1987. On Kogbetliantz's SVD algorithm in the presence of clusters. *Lin. Alg. and Appl.* 95:135–160.
- Cheney, W. and D. Kincaid. 1994. *Numerical Mathematics and Computing*, 3d ed. Pacific Grove, Calif.: Brooks/Cole Publishing Company.
- Choudhary, S. 1994. On Numerical Solution of Large Sparse Linear Systems and Applications. Ph.D. diss., Northern Illinois University, DeKalb, Illinois.
- Cline, A. K., C. B. Moler, G. W. Stewart, and J. H. Wilkinson. 1979. An estimate for the condition number of a matrix. *SIAM J. Num. Anal.* 16:368–375.
- Cline, R. E. and R. J. Plemmons. 1976. L_1 -solutions to underdetermined linear systems. *SIAM Review* 18:92–106.
- Coleman, T. and C. Van Loan. 1988. *Handbook for Matrix Computations*. Philadelphia: SIAM Publications.
- Conte, S. D. and C. de Boor. 1980. *Elementary Numerical Analysis: An Algorithmic Approach*. 3d ed. New York: McGraw Hill.
- Crawford, C. R. 1973. Reduction of a band symmetric generalized eigenvalue problem. *Comm. ACM.* 16:41–44.
- Cullum, J. K. and R. A. Willoughby. 1985. *Lanczos Algorithms for Large Symmetric Eigenvalue Computations*. Vol. 1, *Theory*. Boston: Birkhäuser.
- Cuppen, J. J. M. 1981. A divide and conquer method for the symmetric eigenproblem. *Numer. Math.* 36:177–195.
- Cybenko, G. 1980. The numerical stability of the Levinson–Durbin algorithm for Toeplitz systems of equations. *SIAM J. Sci. and Stat. Comp.* 1:303–310.
- Dahlquist, G. and A. Björck. 1974. *Numerical Methods*. Englewood Cliffs, N.J.: Prentice Hall.
- Daniel, J., W. B. Gragg, L. Kaufman, and G. W. Stewart. 1976. Reorthogonalization and stable algorithms for updating the Gram–Schmidt *QR* factorization. *Math. Comp.* 30:772–795.
- Datta, B. N. 1986. On the root separation of polynomials and inertia theorems. *Proc. IEEE Conf. Dec. Control.* 1959–1961.
- . 1994. Linear and numerical linear algebra in control theory: Some research problems. *Lin. Alg. Appl.* 197/198:755–790.
- . *Numerical Methods in Control*. Forthcoming.
- . *Inertia: Theory, Methods, and Applications*. Forthcoming.
- Datta, B. N. and K. Datta. 1976. An algorithm for computing the powers of a Hessenberg matrix and applications. *Lin. Alg. Appl.* 14:273–284.
- . 1987. The matrix equation $XA = A^T X$ and an associated algorithm for solving the inertia and stability problem. *Lin. Alg. Appl.* 97:103–119.

- Datta, B. N., C. R. Johnson, M. A. Kaashoek, R. Plemmons, and E. D. Sontag, eds. 1988. *Linear Algebra in Signals, Systems, and Control*. Philadelphia: SIAM Publications.
- Datta, B. N. and F. Rincon. 1993. Feedback stabilization of a second-order system: A nonmodal approach. *Lin. Alg. and Appl.* 188/189:135–161.
- Datta, B. N. and Y. Saad. 1991. Arnoldi methods for Sylvester-like observer matrix equations and an associated algorithm for partial spectrum assignment. *Lin. Alg. Appl.* 154/156:225–244.
- Davis, G. J. and C. Moler. 1978. Sensitivity of matrix eigenvalues. *Int. J. Numer. Meth. Eng.* 12:1367–1373.
- De Jong, L. S. 1977. Towards a formal definition of numerical stability. *Numer. Math.* 28:211–219.
- Demmel, J. W. 1983. A numerical analyst's Jordan canonical form. Ph.D. thesis, University of California, Berkeley.
- . 1984. Underflow and the reliability of numerical software. *SIAM J. Sci. Stat. Comput.* 5:887–919.
- . 1987a. Three methods for refining estimates of invariant subspaces. *Computing* 38:43–57.
- . 1987b. On condition numbers and the distance to the nearest ill-posed problem. *Num. Math.* 51:251–289.
- . 1989. On floating point errors in Cholesky. LAPACK working Note #14, Computer Science Department Report, University of Tennessee at Knoxville.
- Demmel, J. W. and B. Kågström. 1987. Computing stable eigendecomposition of matrix pencils. *Lin. Alg. Appl.* 88/89:139–186.
- . 1993. The generalized Schur decomposition of an arbitrary pencil. Part I—Theory and algorithms. *ACM Trans. Math. Soft.* 19:160–174.
- Demmel, J. and W. Kahan. 1990. Accurate singular values of bidiagonal matrices. *SIAM J. Sci. Stat. Comput.* 11:873–912.
- Demmel, J. and K. Vesselic. 1992. Jacobi's method is more accurate than QR. *SIAM J. Matrix Anal. Appl.* 13(4):1204–1246.
- De Moor, B. 1992. On the structure and geometry of the product singular value decompositions. *Linear Alg. Appl.* 168.
- De Moor, B. and P. Van Dooren. 1992. Generalizations of the QR and the singular value decomposition. *SIAM J. Matrix Anal.* 13:993–1014.
- De Moor, B. and H. Zha. 1991. A tree of generalizations of the ordinary singular value decomposition. *Lin. Alg. Appl.* (special issue on canonical forms of matrices) 147:469–500.
- Depretere, E. F., ed. 1988. *SVD and Signal Processing: Algorithms, Applications, and Architecture*. Amsterdam: Elsevier Science Publishers B. V.
- Dongarra, J. 1983. Improving the accuracy of computed singular values. *SIAM J. Sci. Stat. Comp.* 4:712–719.
- Dongarra, J., J. R. Bunch, C. B. Moler, and G. W. Stewart. 1979. *LINPACK User's Guide*. Philadelphia: SIAM Publications.
- Dongarra, J. and E. Grosse. 1987. Distribution of mathematical software via electronic mail. *Communications of the ACM* 30(5):403–407.
- Dongarra, J., S. Hammarling, and J. H. Wilkinson. 1992. Numerical considerations in computing invariant subspaces. *SIAM J. Matrix Anal. Appl.* 13:145–161.
- Dongarra, J., C. B. Moler, and J. H. Wilkinson. 1983. Improving the accuracy of computed eigenvalues and eigenvectors. *SIAM J. Numer. Anal.* 20:23–46.
- Dongarra, J. and D. C. Sorensen. 1987. A fully parallel algorithm for the symmetric eigenvalue problem. *SIAM J. Sci. and Stat. Comp.* 8:S139–S154.

- Drmac, Z., M. Omladic, and K. Veselic. On the perturbation of the Cholesky factorization. *SIAM J. Matrix Anal. Appl.*, forthcoming.
- Duff, I. S., A. M. Erisman, and J. K. Reid. 1986. *Direct Methods for Sparse Matrices*. Oxford: Oxford University Press.
- Eberlein, P. 1962. A Jacobi method for the automatic computation of eigenvalues and eigenvectors. *J. SIAM* 10:74–88.
- Eckart, C. and G. Young. 1939. A principal axis transformation for non-hermitian matrices. *Bull. Amer. Math. Soc.* 45:118–121.
- Edelman, A. 1988. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.* 543–560.
- . 1992. On the distribution of a scaled condition number. *Math Comp.* 58:185–190.
- Edelman, A. and G. W. Stewart. 1993. Scaling for orthogonality. *IEEE Transactions on Signal Processing* 41:1676–1677.
- Efron, B. 1988. Computer-intensive methods in statistical regression. *SIAM Review* 30:421–449.
- Eiermann, M. and R. Varga. 1993. Is the optimal ω best for the SOR iteration method? *Lin. Alg. Appl.* 182:257–277.
- Elman, H. 1986. A stability analysis of incomplete LU factorization. *Math. Comp.* 47:191–218.
- Elman, H., Y. Saad, and P. Saylor. 1986. A hybrid Chebyshev krylov subspace algorithm for solving nonsymmetric systems of linear equations. *SIAM J. Sci. Stat. Comp.* 3:840–55.
- Ericsson, T. and A. Ruhe. 1980. The spectral transformation Lanczos method for the numerical solution of large sparse generalized symmetric eigenvalue problems. *Math. Comp.* 35:1251–1268.
- Faddeev, D. K. and V. N. Faddeeva. 1963. *Computational Methods of Linear Algebra*. San Francisco: W. H. Freeman.
- Faires, J. D. and R. L. Burden. 1993. *Numerical Methods*. Boston: PWS-Kent.
- Fernando, K. V. and B. N. Parlett. 1994. Accurate singular values and differential qd algorithms. *Numer. Math.* 67:191–229.
- Ferng, W. R., G. H. Golub, and R. J. Plemmons. 1991. Adaptive Lanczos algorithms for recursive condition estimation. *Numer. Algorithms* 1:1–20.
- Fierro, R. D. and J. R. Bunch. 1994. Collinearity and total least squares. *SIAM J. Matrix Anal. Appl.* 15:4ff.
- Forsythe, G.E., M.A. Malcolm, and C.B. Moler, 1977. *Computer Methods for Mathematical Computations*. Englewood Cliffs, N.J.: Prentice Hall.
- Forsythe, G. E. and C. B. Moler. 1967. *Computer Solutions of Linear Algebraic Systems*. Englewood Cliffs, N.J.: Prentice Hall.
- Foster, L. 1994. Gaussian elimination with partial pivoting can fail in practice. *SIAM J. Matrix Anal. Appl.* 15:4ff.
- Francis, J. G. F. 1961. The QR transformation: A unitary analogue to the LR transformation, Parts I and II. *Comp. J.* 4:265–272, 332–345.
- Franklin, J. N. 1968. *Matrix Theory*. Englewood Cliffs, N.J.: Prentice Hall.
- Freund, R. W. 1993. A transpose-free quasi-minimum residual algorithm for non-hermitian linear systems. *SIAM J. Sci. Comput.* 4:57–100.
- Freund, R. W., G. H. Golub, and N. M. Nachtigal. 1992. Iterative solution of linear systems. *Acta Numerica* 57–100.
- Freund, R. W. and N. M. Nachtigal. 1991. QMR: A quasi-minimal residual method for non-hermitian linear systems. *Numer. Math.* 60:315–339.
- Gantmacher, F. R. 1959. *The Theory of Matrices*. Vols. 1–2. New York: Chelsea.
- Gentleman, W. M. 1973. Least squares computations by Givens transformations without square roots. *J. Inst. Math. Appl.* 13:329–336.
- . 1975. Error analysis of QR decompositions by Givens transformations. *Lin. Alg. and Appl.* 10:189–197

- George, J. A. and J. Liu. 1981. *Computer Solution of Large Sparse Positive Definite Systems*. Englewood Cliffs, N.J.: Prentice Hall.
- Gilbert, J. R., C. Moler, and R. Schreiber. 1992. Sparse matrices in MATLAB: Design and implementations. *SIAM J. Matrix Anal. Appl.* 13:333-356.
- Gill, P. E., G. H. Golub, W. Murray, and M. A. Saunders. 1974. Methods for modifying matrix factorizations. *Math. Comp.* 28:505-535.
- Gill, P. E., W. Murray, and M. H. Wright. 1991. *Numerical Linear Algebra and Optimization, vol. 1*. Redwood City, Calif.: Addison-Wesley.
- Givens, W. 1958. Computation of plane unitary rotations transforming a general matrix to triangular form. *SIAM J. App. Math.* 6:26-50.
- Gohberg, I. and I. Koltracht. 1993. Mixed, componentwise, and structured condition numbers. *SIAM J. Matrix Anal. Appl.* 14:688-704.
- Goldberg, D. 1991. What every computer scientist should know about floating-point arithmetic. *ACM Computing Surveys* 23:5-48.
- Goldberg, J. 1991. *Matrix Theory with Applications*. New York: McGraw-Hill.
- Golub, G. H. 1965. Numerical methods for solving linear least squares problems. *Numer. Math.* 7:206-216.
- . 1968. Least squares, singular values and matrix approximations. *Aplikace Matematiky* 13:44-51.
- . 1969. Matrix decompositions and statistical computation. In *Statistical Computation*, edited by R. C. Milton and J. A. Nelder, 365-397. New York: Academic Press.
- Golub, G. H. and W. Kahan. 1965. Calculating the singular values and pseudo-inverse of a matrix. *SIAM J. Num. Anal. Ser. B2*:205-224.
- Golub, G. H., S. Nash, and C. Van Loan. 1979. A Hessenberg-Schur method for the matrix problem $AX + XB = C$, *IEEE Trans. Auto. Cont.* AC-24:909-913.
- Golub, G. H. and D. O'Leary. 1989. Some history of the conjugate gradient and Lanczos methods. *SIAM Review* 31:50-102.
- Golub, G. H. and J. M. Ortega. 1992. *Scientific Computing and Differential Equation: An Introduction to Numerical Methods*. San Diego: Academic Press.
- Golub, G. H. and C. Reinsch. 1970. Singular value decomposition and least squares solutions. *Numer. Math.* 14:403-420.
- Golub, G. H. and C. F. Van Loan. 1983. *Matrix Computations*. 1st ed. Baltimore: Johns Hopkins University Press.
- Golub, G. H. and C. F. Van Loan. 1989. *Matrix Computations*. 2d ed. Baltimore: Johns Hopkins University Press.
- Golub, G. H. and J. H. Wilkinson. 1966. Note on the iterative refinement of least squares solution. *Numer. Math.* 9:139-148.
- . 1976. Ill-conditioned eigensystems and the computation of the Jordan canonical form. *SIAM Review* 18:578-619.
- Gould, N. 1991. On growth in Gaussian elimination with complete pivoting. *SIAM J. Matrix Anal. Appl.* 12(2):354-371.
- Green, B. F. 1952. The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika* 17:429-440.
- Greenbaum, A. and Z. Strakos. 1992. Predicting the behavior of finite precision Lanczos and conjugate gradient computations. *SIAM J. Matrix Anal. Appl.* 13:121-137.
- Grimes, R. G., J. G. Lewis, and H. D. Simon. 1994. A shifted block Lanczos algorithm for solving sparse symmetric generalized eigenproblems. *SIAM J. Matrix Anal. Appl.* 15:228-272.
- Gu, M., and S. Eisenstat. 1994a. A stable and efficient algorithm for the rank-one modification of the symmetric eigenproblem. *SIAM J. Matrix Anal. Appl.* 15:4ff.
- . 1994b. A divide-and-conquer algorithm for the symmetric tridiagonal eigenproblem. *SIAM J. Matrix Anal. Appl.* Forthcoming.

- Gutknecht, M. H. 1992. A complete theory of the unsymmetric Lanczos process and related algorithms, Part I. *SIAM J. Matrix Anal. Appl.* 13:594–639.
- Hagerman, L. A. and D. M. Young. 1981. *Applied Iterative Methods*. New York: Academic Press.
- Hager, W. 1984. Condition estimates. *SIAM J. Sci. and Stat. Comp.* 5:311–316.
- Hager, W. 1988. *Applied Numerical Linear Algebra*. Englewood Cliffs, N.J.: Prentice Hall.
- Hammarling, S. 1985. The singular value decomposition in multivariate statistics. *ACM SIGNUM Newsletter* 20:2–25.
- Heath, M. T., A. J. Laub, C. C. Paige, and R. C. Ward. 1986. Computing the SVD of a product of two matrices. *SIAM J. Sci. and Stat. Comp.* 7:1147–1159.
- Hestenes, M. R. 1980. *Conjugate Direction Methods in Optimization*. Berlin: Springer-Verlag.
- Hestenes, M. R. and E. Stiefel. 1952. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.* 49:409–436.
- Higham, N.J. 1987. A survey of condition number estimation for triangular matrices. *SIAM Review* 29:575–596.
- . 1988. FORTRAN codes for estimating the one-norm of a real or complex matrix, with applications to condition estimation. *ACM Trans. Math. Soft.* 14:381–396.
- . 1989. A collection of test matrices in MATLAB. Technical Report 89-1025, Department of Computer Science, Cornell University, Ithaca, New York.
- . 1990. Experience with a matrix norm estimator. *SIAM J. Sci. and Stat. Comp.* 11:804–809.
- . 1991. Iterative refinement enhances the stability of QR decomposition methods for solving linear equations. *BIT* 31:447–468.
- Higham, N. J. and D. J. Higham. 1989. Large growth factors in Gaussian elimination with pivoting. *SIAM J. Matrix Anal. Appl.* 10:155–164.
- Higham, N. J. 1996. *Accuracy and Stability of Numerical Algorithms*. Philadelphia, SIAM.
- Hill, Jr., R. O. 1991. *Elementary Linear Algebra with Applications*. Harcourt-Brace-Jovanovich.
- Hill, Jr., R. O. and B. N. Parlett. 1992. Refined interlacing properties. *SIAM J. Matrix Anal. Appl.* 13:239–247.
- Hodel, A. S. and K. Poola. 1992. Parallel solution of large Lyapunov equations. *SIAM J. Matrix Anal. Appl.* 13:1189–1203.
- Hong, Y. and C.-T. Pan. 1992. The Rank-revealing QR decomposition and SVD. *Math Comp.* 58:575–596.
- Horn, R. A. and C. R. Johnson. 1985. *Matrix Analysis*. New York: Cambridge University Press.
- Horn, R. A. and C. R. Johnson. 1991. *Topics in Matrix Analysis*. New York: Cambridge University Press.
- Householder, A. S. 1958. Unitary triangularization of a nonsymmetric matrix. *J. ACM* 5:339–342.
- . 1974. *The Theory of Matrices in Numerical Analysis*. New York: Dover Publications.
- Hu, D. Y. and L. Reichel. 1992. Krylov subspace methods for the Sylvester equation. *Lin. Alg. Appl. and Appl.* 172:283–313.
- Imman, D. J. 1989. *Vibration with Control Measurement and Stability*. Englewood Cliffs, N.J.: Prentice Hall.
- . 1994. *Engineering Vibration*. Englewood Cliffs, N.J.: Prentice Hall.
- Jain, A. K. 1989. *Fundamentals of Digital Image Processing*. Englewood Cliffs, N.J.: Prentice Hall.
- James, M. L., G. M. Smith, J. C. Wolfrod, and P. W. Whaley. 1989. *Vibration of Mechanical and Structural Systems*. New York: Harper & Row.
- Jankowski, M. and M. Wozniakowski. 1977. Iterative refinement implies numerical stability. *BIT* 17:303–311.
- Jennings, A. and W. J. Stewart. 1975. Simultaneous iteration for the partial eigensolution of real matrices. *J. Inst. Math Appl.* 15:351–362.

- Johnson, R. and D. W. Wichern. 1992. *Applied Multivariate Statistical Analysis*, 3d ed. Englewood Cliffs, N.J.: Prentice Hall.
- Johnston, R. L. 1971. Gershgorin theorems for partitioned matrices. *Lin. Alg. and Its Appl.* 4:205–220.
- . 1982. *Numerical Methods: A Software Approach*. Toronto: John Wiley.
- Kågström, B. 1985. The generalized singular value decomposition and the general $A - \lambda B$ problem. *BIT* 24:568–583.
- Kågström, B. and A. Ruhe. 1980a. An algorithm for numerical computation of the Jordan normal form of a complex matrix. *ACM Trans. Math. Soft.* 6:398–419.
- . 1980b. Algorithm 560 JNF: An algorithm for numerical computation of the Jordan normal form of a complex matrix. *ACM Trans. Math. Soft.* 6:437–443.
- . 1983. Matrix pencils. Proc. Pite Havsbud, 1982. *Lecture Notes in Mathematics* 973. Berlin: Springer-Verlag.
- Kahan, B., N. Parlett, and E. Jiang. 1982. Residual bounds on approximate eigensystems of nonnormal matrices. *SIAM J. Numer. Anal.* 19:470–484.
- Kahan, W. 1966. Numerical linear algebra. *Canadian Math. Bull.* 9:757–801.
- Kahaner, D., C. B. Moler, and S. Nash. 1989. *Numerical Methods and Software*. Englewood Cliffs, N.J.: Prentice Hall.
- Kaufman, L. 1977. Some thoughts on the QZ algorithm for solving the generalized eigenvalue problem. *ACM Trans. Math. Soft.* 3:65–75.
- . 1993. An algorithm for the banded symmetric generalized matrix eigenvalue problem. *SIAM J. Matrix Anal. Appl.* 4(2):372–389.
- Kautsky, J., N. K. Nichols, and P. Van Dooren. 1985. Robust pole assignment in linear state feedback. *Int. J. Control* 41:1129–1155.
- Kenney, C. and A. Laub. 1989. Condition estimates for matrix functions. *SIAM J. Matrix Anal. Appl.* 10:191–209.
- Klema, V. C. and A. J. Laub. 1980. The singular value decomposition: Its computation and some applications. *IEEE Trans. Auto. Cont.* AC-25:164–176.
- Kublanovskaya, V. N. 1961. On some algorithms for the solution of the complete eigenvalue problem. *USSR Comp. Math. Phys.* 3:637–657.
- Lancaster, P. 1969. *The Theory of Matrices*. New York: Academic Press.
- Lancaster, P. and M. Tismenetsky. 1985. *The Theory of Matrices with Applications*. 2d ed. New York: Academic Press.
- Lanczos, C. 1950. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Stand.* 45:255–282.
- Langer, H., B. Najman, and K. Veselic. 1992. Perturbation of the eigenvalues of quadratic matrix polynomials. *SIAM J. Matrix Anal. Appl.* 13:474–489.
- Laub, A. 1979. A Schur method for solving algebraic Riccati equations. *IEEE Trans. Auto. Cont.* AC-24(6):913–921.
- . 1981. Efficient multivariable frequency response computations. *IEEE Trans. Auto. Cont.* AC-26:407–408.
- Laub, A. J., M. T. Heath, C. Paige, and R. Ward. 1987. Computation of system balancing transformation and other applications of simultaneous diagonalization algorithms. *IEEE Trans. Auto. Cont.* AC-32:115–122.
- Lawson, C. L. and R. J. Hanson. 1974. *Solving Least Squares Problems*. Englewood Cliffs, N.J.: Prentice Hall (SIAM Edition, 1995).
- Lay, D. C. 1994. *Linear Algebra and its Applications*. New York: Addison-Wesley.
- Lazer, A. C. and P. J. McKenna. 1990. Large-amplitude periodic oscillations in suspension bridges: Some new connections with nonlinear analysis. *SIAM Review* 32:537–578.
- Leon, S. J. 1994. *Linear Algebra with Applications*. 4th ed. New York: Macmillan.
- Li, T. Y., N. H. Rhee, and Z. Zeng. An efficient and accurate parallel algorithm for the singular value problem of bidiagonal matrices. Forthcoming.

- Li, T. Y. and Z. Zeng. 1992. Homotopy-determinant algorithms for solving nonsymmetric eigenvalue problems. *Math. Comp.* 59:483–502.
- Li, T. Y., Z. Zeng, and L. Cong. 1992. Solving eigenvalue problems of nonsymmetric matrices with real homotopies. *SIAM J. Num. Anal.* 29:229–248.
- Luenberger, D. 1973. *Introduction to Linear and Nonlinear Programming*. New York: Addison-Wesley.
- . 1979. *Introduction to Dynamic Systems: Theory, Models, and Applications*. New York: John Wiley.
- Major, D. J. and R. D. Sidman. 1991. A new use of singular value decomposition in bioelectric imaging of the brain. In *SVD and Signal Processing II, Algorithms, Analysis and Applications*, edited by R. Vaccaro, 497–512. Amsterdam: Elsevier Science Publishers B. V.
- McCarthy, C. and G. Strang. 1973. Optimal conditioning of matrices. *SIAM J. Num. Anal.* 10:370–388.
- Mehrmann, V. 1988. A symplectic orthogonal method for single input or single output discrete time optimal quadratic control problems. *SIAM J. Matrix Anal. Appl.* 9:221–247.
- Miminis, G. S. and C. C. Paige. 1991. Implicit shifting in the *QR* and related algorithms. *SIAM J. Matrix Anal. Appl.* 12:385–400.
- Moler, C. B., J. N. Little, and S. Bangert. 1987. *PC-Matlab Users' Guide*. Sherborn, Mass.: The Math Works, Inc.
- Moler, C. B. and G. W. Stewart. 1973. An algorithm for generalized matrix eigenvalue problems. *SIAM J. Num. Anal.* 10:241–256.
- Müller, P. C. and W. O. Schichlen. 1985. *Linear Vibrations*. Dordrecht: Martinus Nijhoff.
- Nachtigal, N. M., L. Reichel, and L. N. Trefethen. 1992. A hybrid GMRES algorithm for nonsymmetric linear systems. *SIAM J. Matrix Anal. Appl.* 13:796–825.
- Nanda, T. 1985. Differential equations and the *QR* algorithm. *SIAM J. Num. Anal.* 22:310–321.
- Nazareth, J. L. 1989. Updating the triangular factorization of a matrix. *SIAM J. Matrix Anal. Appl.* 10:424–428.
- Neter, J., W. Wasserman, and M. Kunter. 1983. *Applied Linear Regression Models*. Homewood, Ill.: Richard D. Irwin.
- Neumann, M. and R. Plemmons. 1987. Convergence of parallel multisplitting iterative methods for *M*-matrices. *Lin. Alg. Applic.* 88:559–575.
- Noble, B. and J. W. Daniel. 1988. *Applied Linear Algebra*. Englewood Cliffs, N.J.: Prentice Hall.
- Oettli, W. and W. Prager. 1964. Compatibility of approximate solutions of linear equations with given error bounds for coefficients and right-hand sides. *Numer. Math.* 6:405–409.
- Okamoto, S. 1984. *Introduction to Earthquake Engineering*. 2d ed. Tokyo: University of Tokyo Press.
- O'Leary, D. P. 1980. Estimating matrix condition numbers. *SIAM J. Sci. and Stat. Comp.* 1:205–209.
- O'Leary, D. P. and R. White. 1985. Multisplitting of matrices and parallel solution of linear systems. *SIAM J. Alg. Discrete Methods* 6:630–640.
- O'Neil, P. V. 1991. *Advanced Engineering Mathematics*. 3d ed. Belmont, Calif.: Wadsworth.
- Ortega, J. M. 1988a. *Introduction to Parallel and Vector Solution of Linear Systems*. New York: Plenum Press.
- . 1988b. *Matrix Theory: A Second Course*. New York: Plenum Press.
- . 1990. *Numerical Analysis: A Second Course*. SIAM Series in Classics in Applied Mathematics. Philadelphia: SIAM Publications.
- Ortega, J. M. and W. G. Poole, Jr. 1981. *An Introduction to Numerical Methods for Differential Equations*. Marshfield, Mass.: Pitman Publishing.

- Overton, M. 1988. On minimizing the maximum eigenvalue of a symmetric matrix. In *Linear Algebra in Signals, Systems, and Control*, edited by B. N. Datta, et al., 150–169. Philadelphia: SIAM Publications.
- Paige, C. C. 1970. Practical use of the symmetric Lanczos process with reorthogonalization. *BIT* 10:183–195.
- . 1971. The computation of eigenvalues and eigenvectors of very large sparse matrices. Ph.D. thesis, London University, London, England.
- . 1976. Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix. *J. Inst. Math. Applic.* 18:341–349.
- . 1980. Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem. *Lin. Alg. Appl.* 34:235–258.
- . 1986a. Computing the generalized singular value decomposition. *SIAM J. Sci. and Stat. Comp.* 7:1126–1146.
- . 1986b. On the quadratic convergence of Kogbetliantz's algorithm for computing the singular value decomposition. *Lin. Alg. Appl.*, 77:301–314.
- Paige, C. C. and M. Saunders. 1981. Towards a generalized singular value decomposition. *SIAM J. Num. Anal.* 18:398–405.
- Pan, C.-T. and K. Sigmon. 1994. A bottom-up inductive proof of the singular value decomposition. *SIAM J. Matrix Anal. Appl.* 15:59–61.
- Pantazis, R. and D. Szyld. Regions of convergence of the Rayleigh quotient iteration method. *Num. Lin. Alg. Appl.*, forthcoming.
- Pappas, T., A. Laub, and N. Sandell, Jr. 1980. On the numerical solution of the discrete-time algebraic Riccati equation. *IEEE Trans. Auto. Cont.* AC-25:631–641.
- Parlett, B. N. 1965. Convergence of the QR algorithm. *Numer. Math.* 7:187–193. (See also corrections in 10:163–164.)
- . 1966. Singular and invariant matrices under the QR algorithm. *Math. Comp.* 20:611–615.
- . 1968. Global convergence of the basic QR algorithm on Hessenberg matrices. *Math. Comp.* 22:803–817.
- . 1974. The Rayleigh quotient iteration and some generalizations for nonnormal matrices. *Math. Comp.* 28:679–693.
- . 1980a. A new look at the Lanczos algorithm for solving symmetric systems and linear equations. *Lin. Alg. and Appl.* 29:323–346.
- . 1980b. *The Symmetric Eigenvalue Problem*. Englewood Cliffs, N.J.: Prentice Hall.
- . 1989. Towards a black box Lanczos program. *Computer Physics Communications* 53:169–179.
- . 1991. Symmetric matrix pencils. *J. Comp. and Appl. Math.* 38:373–385.
- . 1992. The rewards for maintaining semi-orthogonality among Lanczos vectors. *J. Num. Lin. Alg. with Appl.* 1:243–267.
- Parlett, B. N. and H. C. Chen. 1990. Use of indefinite pencils for computing damped natural modes. *Lin. Alg. Appl.* 140:53–88.
- Parlett, B. N. and J. Le. 1993. Forward instability of tridiagonal QR. *SIAM J. Matrix Anal. and Appl.* 14:279–316.
- Parlett, B. N. and C. Reinsch. 1969. Balancing a matrix for calculation of eigenvalues and eigenvectors. *Numer. Math.* 13:292–304.
- Parlett, B. N. and D. S. Scott. 1979. The Lanczos algorithm with selective orthogonalization. *Math. Comp.* 33: 217–238.
- Parlett, B. N., H. Simon, and L. M. Stringer. 1982. On estimating the largest eigenvalue with the Lanczos algorithm. *Math. Comp.* 38:153–166.
- Patel, R. V., A. J. Laub, and P. M. Van Dooren, eds. 1994. *Numerical Linear Algebra Techniques for Systems and Control*. Piscataway, N.J.: IEEE Press.

- Peters, G. and J. Wilkinson. 1979. Inverse iterations, ill-conditioned equations, and Newton's method. *SIAM Review*. 21:339-360.
- Petkov, P. H., N. D. Christov, and M. M. Konstantinov. 1991. *Computational Methods for Linear Control Systems*. Hemel Hempstead, England: Prentice Hall International.
- Pierce, D. and R. J. Plemmons. 1988. A two-level preconditioned conjugate gradient scheme. In *Linear Algebra in Signals, Systems, and Control*, edited by B. N. Datta, et al., 170-185. Philadelphia: SIAM Publications.
- . 1992. Fast adaptive condition estimation. *SIAM J. Matrix Anal. Appl.* 13: 274-291.
- Plemmons, R. J. 1974. Linear least squares by elimination and MGS. *J. ACM* 21:581-585.
- . 1986. A parallel block iterative scheme applied to computations in structural analysis. *SIAM J. Alg. and Disc. Methods* 7:337-347.
- Rao, C. R. 1980. Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In *Multivariate Analysis*, edited by P. R. Krishnaiah. Amsterdam: V. North-Holland.
- Rao, C. R. and S. K. Mitra. 1971. *Generalized Inverse of Matrices and Its Applications*. New York: John Wiley.
- Reddy, J. N. 1993. *An Introduction to the Finite Element Method*. New York: McGraw-Hill.
- Reichel, L. and W. B. Gragg. 1990. Algorithm 686: FORTRAN subroutines for updating the QR decomposition. *ACM Trans. Math. Software* 16:369-377.
- Rice, J. R. 1966. A theory of condition. *SIAM J. Num. Anal.* 3:287-310.
- . 1981. *Matrix Computations and Mathematical Software*. New York: McGraw-Hill.
- . 1983. *Numerical Methods, Software, and Analysis*. New York: McGraw-Hill.
- Rincon, F. 1991. Feedback stabilization of second-order model. Ph.D. diss., Northern Illinois University, DeKalb, Illinois.
- Ruhe, A. 1994. Rational Krylov algorithms for nonsymmetric eigenvalue problems, II, matrix pairs. *Lin. Alg. Appl.* 197/198:283-295.
- Rutishauser, H. 1958. Solution of eigenvalue problems with the LR transformation. *Nat. Bur. Stand. Appl. Math. Ser.* 49:47-81.
- Saad, Y. 1981. Krylov subspace methods for solving large unsymmetric linear systems. *Math. Comp.* 37:105-126.
- . 1984. Practical use of some Krylov subspace methods for solving indefinite and nonsymmetric linear systems. *SIAM J. Sci. and Stat. Comp.* 5:203-228.
- . 1987. On the Lanczos method for solving symmetric systems with several right-hand sides. *Math. Comp.* 48:651-662.
- . 1988. Projection and deflation methods for partial pole assignment in linear state feedback. *IEEE Trans. Auto. Cont.* 33:290-297.
- . 1990. Numerical solution of large Lyapunov equations. In *Signal Processing, Scattering, Operator Theory, and Numerical Methods*, edited by M. A. Kaashoek, J. H. Van Schuppen and A. C. Ran, 503-511. Boston: Birkhauser.
- . 1992. *Numerical Methods for Large Eigenvalue Problems: Theory and Algorithms*. New York: John Wiley.
- . 1996. *Iterative Methods for Sparse Linear Systems*. Boston: PWS.
- Saad, Y. and M. Schultz. 1986. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. and Stat. Comp.* 7:856-869.
- Saylor, P. and D. C. Smolarski. 1988. An optimum iterative method for solving any linear system with a square matrix. *BIT* 28:163-178.
- Saylor, P., S. Ashby, and T. A. Manieuffel. 1990. A taxonomy for conjugate gradient methods. *SIAM J. Numer. Anal.* 27:1542-1568.
- Schneider, H. and G. P. Barker. 1989. *Matrices and Linear Algebra*. New York: Dover.
- Shampine, L. F. and R. C. Allen, Jr. 1973. *Numerical Computing: An Introduction*. Philadelphia: Saunders.

- Shewchuk, J. R. 1994. An introduction to the conjugate gradient method without the agonizing pain. *Tech. Report CMU-CS 94-125*, Carnegie Mellon University, Pittsburgh, Penn.
- Simon, H. 1984. Analysis of the symmetric Lanczos algorithm with reorthogonalization methods. *Lin. Alg. Appl.* 61:101-132.
- Skeel, R. D. 1979. Scaling for numerical stability in Gaussian elimination. *J. ACM* 26:494-526.
- . 1980. Iterative refinement implies numerical stability for Gaussian elimination. *Math. Comp.* 35:817-832.
- . 1981. Effect of equilibration on residual size for partial pivoting. *SIAM J. Num. Anal.* 18:449-455.
- Skeel, R. D. and J. B. Keiper. 1993. *Elementary Numerical Computing with Mathematica*. New York: McGraw-Hill.
- Smith, B. T., J. M. Boyle, J. Dongarra, B. S. Garbow, Y. Ikebe, V. C. Klema, and C. B. Moler. 1976. Matrix eigensystem routines. *EISPACK Guide*, Vol. 6. *Lecture Notes in Computer Science*. Berlin: Springer-Verlag.
- Sorensen, D. 1992. Implicit application of polynomial filters in a k -step Arnoldi method. *SIAM J. Matrix Anal. Appl.* 13:357-385.
- Steidel, R. F., Jr. 1979. *An Introduction to Mechanical Vibrations*. 2d ed. New York: John Wiley.
- Stewart, G. W. 1973. *Introduction to Matrix Computations*. New York: Academic Press.
- . 1975. Gershgorin theory for the generalized eigenvalue problem $Ax = \lambda Bx$, *Math. Comp.* 29:600-606.
- . 1976a. A bibliographical tour of the large sparse generalized eigenvalue problem. In *Sparse Matrix Computations*, edited by J. R. Bunch and D. J. Rose. New York: Academic Press.
- . 1976b. Algorithm 406. HQR3 and EXCHNG: FORTRAN programs for calculating the eigenvalues of a real upper Hessenberg matrix in a prescribed order. *ACM Trans. Math. Soft.* 2:275-280.
- . 1976c. Simultaneous iteration for computing invariant subspaces of non-hermitian matrices. *Numer. Math.* 25:123-126.
- . 1977a. On the perturbation of pseudo-inverses, projections, and linear least squares problems. *SIAM Review* 19:634-662.
- . 1977b. Perturbation bounds for the QR factorization of a matrix. *SIAM J. Numer. Anal.* 14:509-518.
- . 1978. Perturbation theory for the generalized eigenvalue problem. In *Recent Advances in Numerical Analysis*, edited by C. de Boor and G. H. Golub. New York: Academic Press.
- . 1979a. A note on the perturbation of singular values. *Lin. Alg. Appl.* 28:213-216.
- . 1979b. Perturbation bounds for the definite generalized eigenvalue problem. *Lin. Alg. and Appl.* 23:69-86.
- . 1983. A method for computing the generalized singular value decomposition. In *Matrix Pencils*, edited by B. Kågström and A. Ruhe, 207-220. New York: Springer-Verlag.
- . 1984. Rank degeneracy. *SIAM J. Sci. and Stat. Comp.* 5:403-413.
- . 1987. Collinearity and least squares regression. *Statistical Science* 2:68-100.
- . 1991. Perturbation theory for the singular value decomposition. In *SVD and Signal Processing II, Algorithms, Analysis, and Applications*, edited by R. J. Vaccaro, 99-109. Amsterdam: Elsevier Science Publishers, B.V.
- . 1993a. On the perturbation of LU , Cholesky, and QR factorizations. *SIAM J. Matrix Anal. Appl.* 4:1141-1145.
- . 1993b. On the early history of the singular value decomposition. *SIAM Review* 35:551-566.
- Stewart, G. W. and J. G. Sun. 1990. *Matrix Perturbation Theory*. New York: Academic Press.

- Stoer, J. and R. Bulirsch. 1980. *Introduction to Numerical Analysis*. New York: Springer-Verlag.
- Strang, G. 1988. *Linear Algebra and its Applications*. 3d ed. New York: Academic Press.
- . 1986. *Introduction to Applied Mathematics*. Wellesley, Mass.: Wellesley-Cambridge Press.
- . 1993. *Introduction to Linear Algebra*. Wellesley, Mass.: Wellesley-Cambridge Press.
- Strang, G. and G. Fix. 1973. *An Analysis of the Finite Element Method*. Englewood Cliffs, N.J.: Prentice Hall.
- Strikwerda, J. C. 1989. *Finite Difference Schemes and Partial Differential Equations*. Pacific Grove, Calif.: Wadsworth & Brooks/Cole Mathematics Series.
- Student Edition of MATLAB for MS-DOS Personal Computers*. 1992. Englewood Cliffs, N.J.: Prentice Hall.
- Sun, J. G. 1983. Perturbation analysis for the generalized singular value problems. *SIAM J. Numer. Anal.* 20:611–625.
- . 1991. Perturbation bounds for the Cholesky and QR factorizations. *BIT* 31:341–352.
- Thisted, R. A. 1988. *Elements of Statistical Computing*. New York: Chapman and Hall.
- Thomson, W. T. 1992. *Theory of Vibrations with Applications*. 4th ed. Englewood Cliffs, N.J.: Prentice Hall.
- Trefethen, L. and R. Schreiber. 1990. Ayrage case stability of Gaussian elimination. *SIAM J. Matrix Anal. Appl.* 11:335–360.
- Tucker, A. 1993. *Linear Algebra: An Introduction to the Theory and Use of Vectors and Matrices*. New York: Macmillan.
- Vaccaro, R., ed. 1991. *SVD and Signal Processing II: Algorithms, Analysis, and Applications*. Amsterdam: Elsevier Science Publishers B.V.
- Van Dooren, P. 1981a. A generalized eigenvalue approach for solving Riccati equations. *SIAM J. Sci. and Stat. Comp.* 2:121–135.
- . 1981b. The generalized eigenstructure problem in linear system theory. *IEEE Trans. Auto. Cont.* AC-26:111–128.
- . 1991. Structured linear algebra problems in digital signal processing. In *Numerical Linear Algebra, Digital Signal Processing, and Parallel Algorithms*. NATO Series F, 361–384. New York: Springer-Verlag.
- Vandewalle, J. and B. De Moor 1988. A variety of applications of singular value decomposition in identification and signal processing. In *SVD and Signal Processing Algorithms, Applications and Architectures*, edited by E. F. Deprettere, 43–89. Amsterdam: Elsevier Science Publishers (North Holland).
- Van Huffel, S. and J. Vandewalle. 1991. *The Total Least Squares Problem: Computational Aspects and Analysis*. Frontiers in Applied Mathematics Series, Vol. 9. Philadelphia: SIAM Publications.
- Van Loan, C. 1976. Generalizing the singular value decomposition. *SIAM J. Numer. Anal.* 13:76–83.
- . 1987. On estimating the condition of eigenvalues and eigenvectors. *Lin. Alg. Appl.* 88/89:715–732.
- Varah, J. M. 1968a. The calculation of the eigenvectors of a general complex matrix by inverse iteration. *Math. Comp.* 22:785–791.
- . 1968b. Rigorous machines bounds for the eigensystem of a general complex matrix. *Math. Comp.* 22:793–801.
- . 1970. Computing invariant subspaces of a general matrix when the eigensystem is poorly determined. *Math. Comp.* 24:137–149.
- Varga, R. S. 1962. *Matrix Iterative Analysis*. Englewood Cliffs, N.J.: Prentice Hall.
- Walker, H. F. 1988. Implementation of the GMRES method using Householder transformations. *SIAM J. Sci. Stat. Comp.* 9:152–163.

- Wang, S. and S. Zhao. 1991. An algorithm for $Ax = \lambda Bx$ with symmetric positive definite A and B . *SIAM J. Matrix Anal. Appl.* 12:654–660.
- Ward, R. C. 1981. Balancing the generalized eigenvalue problem. *SIAM J. Sci. and Stat. Comp.* 2:141–152.
- Watkins, D. 1982. Understanding the QR algorithms. *SIAM Review* 24:427–440.
- . 1993. Some perspectives on the eigenvalue problem. *SIAM Review* 35:430–471.
- . 1991. *Fundamentals of Matrix Computations*. New York: John Wiley.
- Wedin, P. Å. 1973. Perturbation theory for psuedo-inverses. *BIT* 13:217–232.
- Weinberger, H. F. 1965. *A First Course in Partial Differential Equations*. New York: John Wiley.
- White, R. E. 1990. Multisplittings of a symmetric positive definite matrix. *SIAM J. Matrix Anal. Appl.* 11:69–82.
- Wilkinson, J. H. 1960. Error analysis of floating-point computations. *Numer. Math.* 2:319–340.
- . 1961. Error analysis of direct methods of matrix inversion. *J. ACM* 10:281–330.
- . 1963. *Rounding Errors in Algebraic Processes*. Englewood Cliffs, N.J.: Prentice Hall.
- . 1965. *The Algebraic Eigenvalue Problem*. Oxford: Clarendon Press.
- Williams, T. and A. J. Laub. 1992. Orthogonal canonical forms for second-order systems. *IEEE Trans. Auto. Cont.* 37(7):1050–1052.
- Woźnicki. 1993. On numerical analysis of conjugate gradient method. *Japan J. Ind. Appl. Math.* 10(3):487–519.
- Wright, S. 1993. A collection of problems for which Gaussian elimination with partial pivoting is unstable. *SIAM J. Sci. Comput.* 14(1):231–238.
- Young, D. M. 1970. Convergence properties of the symmetric and unsymmetric over-relaxation methods. *Math. Comp.* 24:793–807.
- . 1971. *Iterative Solution of Large Linear Systems*. New York: Academic Press.
- . 1972. Generalization of property A and consistent ordering. *SIAM J. Num. Anal.* 9:454–463.
- Young, D. M., K. C. Jea, and Tsun-Zee Mai. 1988. Preconditioned conjugate gradient algorithms and software for solving large sparse linear systems. In *Linear Algebra in Signals, Systems, and Control*, edited by B. N. Datta et al., 260–283. Philadelphia: SIAM Publications.
- Zeng, Z. 1991. Homotopy-determinant algorithm for solving matrix eigenvalue problems and its parallelizations. Ph.D. diss., Michigan State University, East Lansing, Mich.
- Zha, H. 1993. A componentwise perturbation analysis of the QR decomposition. *SIAM J. Matrix Anal. Appl.* 14:1124–1131.

جدول نمادها

\mathbb{R}^n	فضای برداری حقیقی از بُعد n
$\mathbb{R}^{m \times n}$	مجموعه ماتریسهای حقیقی $m \times n$
$\mathcal{Q}^{m \times n}$	مجموعه ماتریسهای مختلط $m \times n$
e_i	بردار واحد i ام
O	یک ماتریس صفر یا یک بردار صفر
$I_{m \times n}, I$	ماتریس واحد از مرتبه n
I_{k-1}	ماتریس با $(k-1)$ سطر و ستون از I
$R(A)$	بردار (یا فضای ستونی) A
$N(A)$	فضای پوچ A
$\text{null}(A)$	پوچی A
A^T	ترانزاده A
A^{-1}	معکوس A
A^k	توان k ام ماتریس A
$\det(A)$	دترمینان A
$\text{rank}(A)$	رتبه A

$\rho(A)$	شعاع طیفی A
ρ	عامل رشد
\bar{A}	مزدوج مختلط A
A^*	ترانهاده مزدوج مختلط A
$\text{fl}(x)$	عدد ممیز شناور x
$\text{cond}(A), \text{Cond}(A)$	عدد وضعیت A
A^+	شبه معکوس A
\equiv	نماد برای رویهم‌نویسی
\Leftrightarrow	نماد برای تعویض کردن
$\kappa(A, b)$	زیرفضای کرلیف $(b, Ab, \dots, A^{n-1}b)$
$\text{dim}(N)$	بُعد n
RSF	شکل شور حقیقی
SVD	تجزیه مقدار تکین
P_A	تصویر متعامد بر روی $R(A)$
P_A^\perp, P_N	تصویر بر روی متمم متعامد $R(A)$
b_R	تصویر بردار b بر روی $R(A)$
b_\perp, b_N	تصویر بردار b بر روی متمم متعامد $R(A)$
$\ x\ $	نرم بردار x
$\ A\ $	نرم ماتریس A
$\ A\ _p$	p -نرم ماتریس $p = 1, 2, \infty$
$\ A\ _F$	نرم فروبنیوس A
Σ	علامت مجموع
Π	علامت حاصلضرب
δb	میزان اختلال در بردار b
ΔA	میزان اختلال در ماتریس A
E	ماتریس خطا
CGS	فرایند گرام اشمیت کلاسیک

MGS	فرایند گرام اشمیت اصلاح شده
\dot{x}	مشتق اول x
\ddot{x}	مشتق دوم x
O	نماد برای مرتبه
$\text{diag}(d_1, d_2, \dots, d_n)$	ماتریس قطری
$\text{diag}(A_{11}, \dots, A_{kk})$	ماتریس قطری بلوکی
ζ_i	نسبت میرایی مدل i امین مد
σ_i	i امین بزرگترین مقدار تکین A
μ	دقت ماشین

فهرست الگوریتمها

- الگوریتم چولسکی، ۳۲۵
- الگوریتم چولسکی برای دستگاه معین مثبت، ۳۲۶
- الگوریتم دمل - کاهان برای *SVD*، ۸۱۲-۸۱۳
- الگوریتم گولوب - کاهان - رینسک برای *SVD*، ۸۰۰-۸۰۳
- الگوریتم لانزوس برای مداد معین متقارن، ۷۴۳
- الگوریتم *QR* چولسکی برای مداد معین متقارن، ۷۰۷
- الگوریتم *QZ*، ۷۰۱-۷۰۲
- الگوریتم *SVD* چان، ۸۱۰-۸۱۱
- برآورد عدد وضعیت، ۳۷۱، ۳۷۴
- به هنگام درآوری یک تجزیه *QR*، ۲۶۰
- به هنگام درآوری *QR* هاوس هولدر، ۲۶۰
- تجزیه چولسکی ناقص، ۴۰۹
- تجزیه *LU* بلوکی، ۳۴۰
- تجزیه *QR* گیونز، ۲۴۱
- تجزیه *QR* هاوس هولدر، ۲۱۷

- تجزیه QR هاوس هولدر با محورگیری ستونی، ۲۵۵
- تصفیه تکراری برای جواب دستگاه خطی، ۳۷۷
- تصفیه تکراری برای جواب کمترین توانهای دوم، ۵۰۱-۵۰۰
- تقلیل هاوس هولدر برای مقادیر ویژه زیر غالب، ۵۸۶
- تکرار خارج قسمت ریلی، ۵۸۳
- تکرار معکوس، ۵۷۷
- تکرار معکوس برای بردارهای ویژه مداد (A, B) ، ۷۰۴
- تکرار QR برای ماتریس متقارن، ۶۴۶
- تکرار QR (تک انتقالی)، ۶۱۴
- تکرار QR (ضمنی با انتقال مضاعف)، ۶۲۳
- جایگزینی پسر، ۱۷۶-۱۵۵
- جواب دستگاه فرومعین با استفاده از معادلات نرمال، ۴۹۶
- جواب دستگاه فرومعین با استفاده از QR، ۴۹۸
- حذفی پیشرو، ۳۰۶
- حذفی گاوس با محورگیری جزئی برای دستگاه خطی، ۳۱۳
- حذفی گاوس با محورگیری جزئی (مثلثی سازی) ۱۹۴
- حذفی گاوس با محورگیری کامل، ۲۰۰
- حذفی گاوس بدون محورگیری برای دستگاه خطی، ۱۲۱، ۱۵۷
- حذفی گاوس بدون محورگیری (مثلثی سازی)، ۱۸۴
- حل دستگاه خطی با استفاده از تجزیه QR، ۳۱۵
- حل دستگاه خطی با چند طرف راست، ۳۱۸
- خارج قسمت ریلی برای جفت (A, B) ، ۷۱۲
- دوقطری سازی، ۸۰۲
- روش تقسیم - تسخیر، ۶۵۱
- روش تکراری ژاکوبی، ۳۸۵
- روش تکراری گاوس - سایدل، ۳۸۸-۳۸۷
- روش تنصیف برای مقادیر ویژه در یک بازه، ۶۴۳

- روش توانی، ۵۷۳
- روش توانی معکوس، ۵۷۷
- روش دنباله استورم برای سه قطری (A, B) ، ۷۴۱
- روش فوق تخفیف متوالی، ۳۹۸
- روش قطری سازی همزمان، ۷۱۰
- روش گرادیان مزدوج، ۴۰۵
- روش گرادیان مزدوج با پیش شرط، ۴۰۸
- روش لانزوس برای ماتریسهای متقارن، ۶۵۴
- روش لانزوس برای مقادیر ویژه (A, B) در یک بازه، ۷۴۶-۷۴۷
- روش $GMRES$ (مانده مینیمال تعمیم یافته)، ۴۱۵
- شبه معکوس با استفاده از SVD ، ۷۹۷
- ضرب یک بردار در یک ماتریس هاوس هولدر، ۲۰۹
- ضرب یک ماتریس در یک ماتریس گیونز، ۲۳۴
- ضرب یک ماتریس در یک ماتریس هاوس هولدر، ۱۵۳
- فرایند گرام اشمیت اصلاح شده برای تجزیه QR ، ۴۸۱
- فرایند گرام-اشمیت برای تجزیه QR ، ۴۸۰-۴۸۱
- کاهش هسنبرگی گیونز، ۲۴۷
- کاهش هسنبرگی هاوس هولدر، ۲۲۷
- کمترین توانهای دوم با استفاده از معادلات نرمال، ۴۶۹
- کمترین توانهای دوم با استفاده از SVD ، ۴۹۲، ۷۹۳
- کمترین توانهای دوم توسط تجزیه QR (تجزیه ناقص)، ۴۸۷
- کمترین توانهای دوم توسط تجزیه QR (رتبه کامل)، ۴۷۶
- کمترین توانهای دوم توسط گرام اشمیت اصلاح شده، ۴۸۵
- محاسبه بردار ویژه از شکل شور حقیقی، ۶۳۲-۶۳۴
- محاسبه حاصلضرب داخلی، ۱۱۴
- معکوس یک ماتریس، ۳۴۹
- معکوس یک ماتریس بالامثلثی، ۱۱۷، ۱۵۷-۱۵۶

فهرست راهنما

- ۱
- آنالیز مؤلفه‌های اصلی، ۵۶۱-۵۶۳
- اپسیلون (ε) ماشین، ۸۸
- اثر ماتریس، ۶۰، ۶۲، ۷۷، ۵۳۰
- اختلالهای مؤلفه‌ای، ۳۷۴
- ارتعاش، ۵۴۶
- ارقام با معنی، ۸۶
- الگوریتم، ۱۱۳
- خاصیتها و مؤثر بودن عددی آن، ۱۴۷
- الگوریتم چولسکی، ۳۲۵
- برای حل دستگاه خطی، ۳۲۶
- برای حل مسأله کمترین توانهای دوم، ۴۶۹
- پایداری پسر، ۳۲۸
- ناقص، ۴۰۹
- الگوریتم QR چولسکی، ۷۰۷

- الگوریتم QR شور، ۷۰۸، ۷۰۹
- الگوریتم QZ ، ۷۰۳-۶۹۳
- الگوریتم SVD دمل - کاهان، ۸۱۶-۸۱۲
- الگوریتم SVD فرناندو - پارلت، ۸۱۶
- الگوریتم SVD گولوب - کاهان - رینسک، ۸۱۰-۸۰۰
- الگوریتم لانزوس
- از دست دادن تعامد در -، ۶۵۷
- برای ماتریسهای متقارن، ۶۵۹-۶۵۲
- برای مداد معین متقارن، ۷۴۶-۷۴۲
- برای مقادیر ویژه مداد معین (A, B) در یک بازه، ۷۴۷-۷۴۶
- انتقال ویلکینسون، ۶۴۵، ۶۷۴
- اینرسی یک ماتریس، ۵۶۸

ب

- برد یک ماتریس، ۳۸
- محاسبه پایه یکا متعامد برای -، ۲۴۹، ۷۸۰
- محاسبه تصویرها بر روی -، ۲۴۹، ۷۸۰
- بردار، ۳۳-۳۱، ۵۵
- بردار ریتز، ۶۵۸، ۷۳۴
- بردارهای ویژه، ۳۸
- چپ، ۵۳۰
- راست، ۳۸، ۵۳۰
- محاسبه -، ۵۷۷، ۶۳۵-۶۳۱
- نتایج اساسی در مورد -، ۵۳۸-۵۲۹

پ

- پایدار پسر، ۱۲۴، ۱۴۰-۱۳۷، ۵۱۱
- پایداری (پسر)، ۱۲۴
- اعمال حسابی، ۱۲۴
- الگوریتم چولسکی برای دستگاه معین مثبت متقارن، ۳۲۸
- الگوریتمهای *SVD*، ۸۰۹، ۸۱۶-۸۱۵
- الگوریتم *QR* چولسکی برای مداد معین، ۷۰۸
- الگوریتم *QZ*، ۷۰۳
- تجزیه چولسکی، ۳۲۵
- تجزیه *QR* گیونز، ۲۴۲
- تجزیه *QR* هاوس هولدر، ۲۰۸
- تکرار معکوس، ۵۷۸
- تکرار *QR* (ضمنی با انتقال مضاعف) ۶۲۹
- تکرار *QR* (متقارن)، ۶۴۷
- حذفی گاوس با محورگیری جزئی، ۲۰۳
- حذفی گاوس با محورگیری کامل، ۲۰۳
- حذفی گاوس بدون محورگیری، ۱۲۸، ۲۰۵، ۳۱۲
- حذفی گاوس برای ماتریس معین مثبت، ۳۲۳
- حذفی گاوس برای یک دستگاه سه قطری، ۳۳۷
- حذفی گاوس برای یک دستگاه غالب قطری، ۳۳۴
- فرایند جایگزینی پسر، ۱۲۵
- کاهش هسنبرگی گیونز، ۲۴۷
- کاهش هسنبرگی هاوس هولدر، ۲۳۰
- پایداری معادلات دیفرانسیل و تفاضلی، ۵۴۶-۵۳۹
- پایه یکا متعامد، ۴۱، ۲۴۹، ۷۸۰، ۷۸۱

ت

تبدیل تشابه و محاسبه مقدار ویژه، ۵۹۰

- تجزیه با محورگیری قطری، ۳۲۸، ۵۶۹
- تجزیه LU، ۱۷۴، ۱۷۹، ۱۹۱، ۱۹۷
- وجود و منحصر به فرد بودن، ۱۷۹
- و ماتریسهای مقدماتی، ۲۰۲-۱۷۴
- تجزیه چولسکی ناقص، ۴۰۹
- تجزیه متعامد کامل، ۲۵۸
- تجزیه مقدار تکین (SVD) چان، ۸۱۲-۸۱۰
- تجزیه مقدار تکین (SVD)، ۵۴
- پایه یکا متعامد و تصویرها، ۷۸۲-۷۸۰
- تاریخ -، ۷۶۵
- تجزیه ویژه، ۷۷۱
- تعمیم یافته، ۸۱۷
- حاصلضرب خارجی یک ماتریس، ۷۸۸
- فاصله یک ماتریس از نزدیکترین ماتریس با رتبه پایین تر، ۷۸۶
- قضیه وجود برای -، ۷۶۹-۷۶۷
- محاسبه -، ۸۱۸-۸۰۰
- و رتبه عددی، ۷۹۰
- و شبه معکوس، ۷۹۹-۷۹۶
- و مسأله کمترین توانهای دوم، ۷۹۵-۷۹۲
- و نرم و عدد وضعیت، ۷۷۹
- تجزیه QR، ۲۱۰
- اصلاح -، ۲۶۱-۲۵۹
- با استفاده از فرایند گرام اشمیت اصلاح شده، ۴۸۱-۴۸۰
- با استفاده از ماتریسهای گیبوز، ۲۴۱-۲۳۹
- با استفاده از ماتریسهای هاوس هولدر، ۲۲۲-۲۱۰
- با محورگیری ستونی، ۲۵۸-۲۵۴
- تصویرهای متعامد، ۲۵۳-۲۴۹

منحصر به فرد بودن در، ۲۴۲-۲۴۳

و پایه یکا متعامد، ۲۴۹

تجزیه QR آشکارساز رتبه، ۲۵۸

تجزیه QR هاوس هولدر، ۲۱۰-۲۲۰

تحلیل اختلال، ۱۲۸

مسأله دستگاه خطی، ۳۵۳-۳۶۱

مسأله کمترین توانهای دوم، ۴۶۱-۴۶۹

مسأله مقدار ویژه، ۵۹۵-۶۰۲

مقادیر تکین، ۷۷۷، ۸۰۱

تحلیل خطای پسرو، ۱۲۴، ۸۳۴-۸۴۷

تشدید، ۵۴۹

تصفیه تکراری

برای جواب دستگاههای خطی، ۳۷۶-۳۷۹

برای جوابهای مسأله کمترین توانهای دوم، ۵۰۴-۵۰۰

و برآورد عدد وضعیت، ۳۷۹

تصویر متعامد، ۴۲، ۴۳، ۴۶۶-۴۶۲

خاصیتها، ۷۲-۷۳

محاسبه -، ۲۴۹

تقلیل، ۵۸۵، ۶۲۶، ۸۰۸

هاتلینگ، ۵۸۵

هاوس هولدر، ۵۸۷-۵۸۵

تکرار معکوس / روش توانی معکوس، ۵۷۷

بردارهای ویژه مداد (A, B) ، ۷۰۳

هسنبرگی، ۶۳۱

تکرار QR، ۶۰۸

انتقال مضاعف (مختلط)، ۶۲۰-۶۱۷

تک انتقالی، ۶۱۴

ضمنی، ۶۲۱
 متقارن، ۶۴۵
 هسنبرگی، ۶۱۰

ج

جایگزینی پسرو، ۱۱۵
 جواب با می نیمم نرم، ۴۵۰، ۴۸۸، ۴۹۵، ۴۹۶، ۴۹۸، ۴۹۹
 جدول مقایسه‌ها
 برای الگوریتمهای کمترین توانهای دوم، ۵۱۱
 برای الگوریتمهای دستگاههای خطی، ۳۴۴، ۳۴۵
 برای الگوریتمهای تجزیه ماتریس، ۲۶۵

چ

چند جمله‌ای مشخصه، ۳۸
 و محاسبه مقدار ویژه، ۵۹۱
 چند جمله‌ای ویلکینسون، ۱۳۲

ح

حذف فاجعه آمیز، ۱۰۱
 حذفی پیشرو، ۳۰۶
 حذفی گاوس، ۱۱۸
 با محورگیری جزئی، ۱۸۷، ۳۰۷
 با محورگیری کامل، ۲۰۱، ۱۹۷، ۳۰۹
 بدون محورگیری، ۱۲۱، ۱۷۶، ۳۰۶
 پایداری -، ۲۰۲
 حساب ممیز شناور، ۸۱
 جمع در -، ۹۳

ضرب در -، ۹۴

قوانین اساسی -، ۸۸

کران خطا برای محاسبات ماتریسی پایه‌ای، ۹۸

محاسبه حاصلضرب داخلی در -، ۹۵

IEEE استاندارد -، ۸۱

حساسیت

بردارهای ویژه، ۶۰۲

دستگاههای خطی، ۳۵۳-۳۶۳

مقادیر تکین، ۷۷۷

مقادیر تکین یک ماتریس دوقطری، ۸۰۱

مقادیر ویژه، ۶۰۱-۵۹۸

خ

خارج قسمت ریلی، ۴۱۵، ۵۸۴-۵۸۳

تکرار - برای مسأله مقدار ویژه

تکرار - برای مسأله مقدار ویژه تعمیم یافته، ۷۱۲

خاصیت یک در میانی، ۶۳۰

خطا، ۸۴

خطای مطلق، ۸۵

خطاهای گرد کردن، ۸۴

برای الگوریتم تصفیه تکراری کمترین توانهای دوم، ۸۰۳

برای الگوریتم تکرار *QZ*، ۷۰۳

برای الگوریتم لانزوس متقارن، ۶۵۶

برای الگوریتم *SVD* دمل - کاهان، ۸۱۵

برای الگوریتم *SVD* گولوب - کاهان - رینسک، ۸۰۹

برای ایجاد صفر در یک بردار با استفاده از ماتریس هاوس هولدر، ۲۰۸

برای تجزیه چولسکی، ۳۲۵

- برای تجزیه LU با استفاده از حذفی گاوس، ۸۴۱
- برای تجزیه QR گیونز، ۲۴۲
- برای تجزیه QR هاوس هولدر، ۲۲۰
- برای تکرار QR متقارن، ۶۴۷
- برای جایگزینی پسرو، ۱۲۵، ۱۲۶، ۸۳۶
- برای جواب دستگاه خطی با استفاده از حذفی گاوس، ۱۲۵، ۸۴۶
- برای جوابهای کمترین توانهای دوم با استفاده از تجزیه QR ، ۴۷۷، ۴۸۹
- برای حذفی پیشرو، ۸۳۴
- برای دستگاه فرومعین با استفاده از تجزیه QR ، ۴۹۸
- برای کاهش هسنبِرگی گیونز، ۲۴۷
- برای کاهش هسنبِرگی هاوس هولدر، ۲۳۰

۵

دترمینان، ۳۷

خاصیتهای -، ۳۸-۳۷

محاسبه -، ۳۵۳-۳۵۰

دستگاه بدوضع، ۱۳۴، ۳۶۲

دستگاه خوش وضع، ۳۶۲

دستگاه سه قطری، ۲۹۴، ۳۳۷، ۳۳۵

دستگاه غالب قطری، ۳۳۳-۳۳۴

دستگاه فرامعین، ۴۵۰-۴۴۹

روشها برای -، ۵۰۰-۴۶۹

دستگاه فرومعین، ۴۹۹-۴۹۳

دستگاه هسنبِرگی، ۳۲۸

دستگاههای خطی خاص، ۳۴۱-۳۱۹

دستگاههای نامعین، ۳۲۸

دورانهای گیونز

- ایجاد صفر در یک بردار با استفاده از -، ۲۳۹-۲۳۲
 تجزیه QR با استفاده از -، ۲۴۲-۲۳۹
 تکرار QR و -، ۶۱۱
 کاهش به شکل هسنبِرگی با استفاده از -، ۲۴۸-۲۴۶

ذ

- ذخیره‌سازی حاصلضرب داخلی در دقت مضاعف، ۹۷
 در تصفیه تکراری برای حل مسأله کمترین توانهای دوم، ۵۰۱
 در جایگزینی پسرو، ۱۲۶
 در کاهش هسنبِرگی هاوس هولدر، ۲۳۰

ر

- رتبه، ۳۸، ۷۹۰
 رتبه عددی یک ماتریس، ۷۹۰
 روش آرنولد، ۴۱۲
 روش دنباله استورم، ۶۴۵-۶۳۹
 برای سه قطری A و B ، ۷۴۲-۷۴۱
 روش تکراری ژاکوبی، ۳۸۶-۳۸۲
 روش تقسیم و تسخیر، ۶۵۲-۶۴۷
 روش تنصیف، ۶۴۳
 روش توانی، ۵۷۷-۵۷۲
 با انتقال، ۵۷۷-۵۷۶
 روش گاوس سایدل، ۳۹۶، ۳۸۶
 روش گرادیان مزدوج، ۴۰۴
 با پیش شرط، ۴۰۸
 و رابطه با $GMRES$ ، ۴۱۶
 همگرایی، ۴۰۷

- روش عنصر متناهی، ۳۰۰-۳۰۳
 روش فوق تخفیف متوالی، ۳۹۷-۴۰۴
 روش هاوس هولدر گلوب برای مسأله کمترین توانهای دوم، ۴۷۶
 روش *GMRES*، ۴۱۳-۴۱۶
 روش *QMR*، ۴۱۶
 روشهای تکراری، ۳۸۰-۴۱۷
 معیار توقف برای -، ۳۸۱، ۳۸۲
 همگرایی -، ۳۸۹

ز

- زاویه بین دو بردار، ۳۳
 زیرفضا، ۳۳
 زیرفضای پایا، ۶۲۹

س

- سرریز، ۸۳

ش

- شبه کدها، ۱۱۴
 شبه معکوس، ۴۶۰، ۷۹۱، ۷۹۶، ۷۹۷
 شعاع طیفی، ۶۵
 شکل شور حقیقی، ۶۰۳
 تعمیم یافته -، ۷۰۳-۶۹۸
 محاسبه -، ۶۲۸-۶۲۶
 شکل متقارن جردن، ۶۴، ۵۳۷
 شمارش اعمال ممیز شناور
 برای الگوریتم چولسکی، ۳۲۷

- برای الگوریتمهای ماتریسی پایه‌ای، ۱۵۷-۱۴۹
- برای الگوریتمهای SVD ، ۸۰۹
- برای الگوریتم QR ، ۷۰۳
- برای تجزیه LU یک ماتریس سه قطری، ۳۳۶
- برای تجزیه QR گیونز، ۲۴۱
- برای تجزیه QR هاوس هولدر، ۲۱۹
- برای تکرار QR (مقارن)، ۶۵۱
- برای تکرار QR (نامقارن) ۶۲۴، ۶۲۹
- برای جواب دستگاه فرومعین با استفاده از معادلات نرمال، ۵۱۱
- برای جواب دستگاه فرومعین با استفاده از هاوس هولدر، ۴۹۸
- برای حذفی گاوس، ۱۸۴
- برای دستگاه هسنبرگی با استفاده از حذفی گاوس، ۳۲۹
- برای روش قطری سازی همزمان، ۷۱۰
- برای ضرب ماتریس در یک ماتریس گیونز، ۲۳۳
- برای ضرب ماتریس در یک ماتریس هاوس هولدر، ۱۵۳
- برای فرآیند گرام اشمیت اصلاح شده، ۴۸۱
- برای کاهش هسنبرگی گیونز، ۲۴۷
- برای کاهش هسنبرگی هاوس هولدر، ۲۳۰
- برای کمترین توانهای دوم با استفاده از معادلات نرمال، ۴۶۹
- برای کمترین توانهای دوم با استفاده از QR هاوس هولدر، ۴۷۷
- برای کمترین توانهای دوم با استفاده از SVD ، ۸۱۲

ط

طرح تفاضل متناهی، ۲۹۹-۲۹۱

طیف یک ماتریس، ۵۳۱

ع

- عامل رشد، ۱۲۶، ۲۰۲
- برای حذفی گاوس، ۲۰۳، ۲۰۵
- برای دستگاه سه قطری، ۳۳۷
- برای دستگاه غالب قطری، ۳۳۴
- برای دستگاه معین مثبت، ۳۲۳
- برای دستگاه هسنبرگی، ۳۲۹
- عامل شرکت مد، ۷۳۱
- عدد ممیز شناور نرمال شده، ۸۱
- عدد وضعیت، ۱۳۰
- اندازه برای بد وضعی، ۳۶۵
- برآورد، ۳۶۹
- خاصیتها، ۳۶۱
- مقادیر تکین، ۷۷۹
- مقیاس کردن، ۳۶۸
- نزدیک بودن به منفرد بودن، ۳۶۶
- و دقت جواب دستگاه، ۳۶۴
- و محورگیری، ۳۶۶
- یک دستگاه خطی، ۳۱۴
- یک ماتریس، ۳۵۴
- یک مقدار ویژه، ۵۹۸
- عدد وضعیت اسکیل، ۳۷۵

ف

- فرایند گرام اشمیت، ۴۸۰
- فرایند گرام اشمیت اصلاح شده، ۴۸۱
- برای حل مسأله کمترین توانهای دوم، ۴۸۶-۴۸۴

- در برابر فرایند گرام اشمیت کلاسیک، ۴۸۲-۴۸۳
 معکوس تعمیم یافته مورپنرز، ۴۶۰
 فرکانسها، ۵۴۸، ۷۰۵
 فرکانسهای طبیعی، ۵۴۸، ۷۰۵
 فرمول شرم - موریسون، ۳۴۷
 فرمول وودبری، ۳۴۷
 فضای برداری، ۳۲
 فضای پوچ، ۳۸، ۴۴
 محاسبه پایه یکا متعامد بر روی -، ۲۴۹، ۷۸۰
 محاسبه تصویرها بر روی -، ۲۴۹، ۷۸۰

ق

- قاعده کرامر، ۲۸۰
 قانون اینرسی سیلوستر، ۵۶۸
 قضیه بوئر - فایک، ۵۹۵
 قضیه تجزیه چولسکی، ۳۲۱
 قضیه کمینه بیشینه کورانت - فیشر، ۶۳۶
 قضیه مثلثی سازی شور، ۵۳۴، ۶۰۶
 قضیه کیلی - هامیلیون، ۵۴
 قضیه هافمن - ویلندت، ۶۳۹
 قضیه Q ضمنی، ۲۴۸
 قضیه های قرص گرشگورین، ۵۶۳-۵۶۶
 قطری سازی همزمان، ۷۱۱-۷۰۹

ک

- کاربردهای
 مسأله تجزیه مقدار تکین، ۷۷۷-۷۷۴، ۸۰۰-۷۷۹

- مسأله دستگاه خطی، ۲۸۳-۳۰۵
 مسأله کمترین توانهای دوم، ۳۵۱، ۴۶۰-۴۵۵
 مسأله مقدار ویژه، ۵۳۹-۵۶۳
 مسأله مقدار ویژه تعمیم یافته در ارتعاش، ۷۳۷-۷۱۳
 کاهش به ماتریس همراه
 کاهش دوری بلوکی، ۳۴۲
 کاهش دولیتل، ۲۶۸
 کاهش سه قطری، ۲۳۰
 کاهش کرات، ۲۶۸
 کاهش گاوس جردن، ۲۶۸
 کاهش هسنبِرگی هاوس هولدر، ۲۲۲

ل

- لم ایزومتري، ۷۸
 لم باناخ، ۷۹

م

- ماتریس، ۳۳
 با خاصیت A ، ۴۰۰
 بالامثلثی، ۴۶
 بدوضع، ۳۶۲
 بلوکی، ۳۶
 به طور سازگار مرتب شده، ۴۰۰
 پای، (Pie)، ۱۶۸، ۳۶۳
 پایین مثلثی، ۴۶
 پایین مثلثی مقدماتی، ۱۷۴، ۲۶۲
 توپلیتز، ۱۳۹، ۱۶۸

- جایگشت، ۴۷
- جرم، ۵۴۷، ۷۱۱، ۷۲۴
- خود توان، ۴۲
- خوش وضع، ۳۶۲
- دوری، ۴۰۰
- دوقطری، ۸۰۴، ۸۰۲، ۸۰۱
- دو قطری ویلکینسون، ۱۳۵، ۱۶۸، ۳۶۳
- سه قطری، ۴۹، ۳۳۵
- سه قطری بلوکی، ۲۹۸، ۳۳۹
- ضمنی، ۵۴۷، ۷۱۱، ۷۲۴، ۷۳۲-۷۳۰
- غالب قطری، ۵۱
- قطری، ۴۶
- قطری بلوکی، ۳۷
- متعامد، ۴۶
- متقارن، ۴۶
- مثلی بلوکی، ۳۸
- مختلط، ۵۲۹
- مدی، ۷۱۲
- معکوس پذیر، ۴۵
- معین مثبت، ۵۲
- ناقص، ۳۸
- نامنفرد، ۴۰، ۴۱
- ناموهن، ۵۱، ۷۶، ۵۹۱
- نرمال، ۶۰۱
- نمایی، ۵۴۰
- نواری، ۱۶۵
- واحد، ۳۴

- واریانس - کوواریانس، ۵۰۵، ۷۹۰
 واندرموند، ۱۶۸
 هانکل، ۱۶۸
 هاوس هولدر، ۲۱۵-۲۰۵
 هرمیتی، ۵۲۹
 هسنبرگی، ۴۹
 همراه، ۵۰
 همگرا، ۶۴
 هیلبرت، ۱۳۴، ۱۶۸، ۳۰۳، ۳۶۳
 یکانی، ۵۲۹
 ماتریسهای جایگشت، ۴۷
 خاصیت‌های -، ۴۹-۴۸
 ماتریسهای متشابه، ۴۱
 متعادل کردن، ۶۲۸
 مدار الکتریکی، ۲۸۳
 شبیه‌سازی یک جریان عبوری از، ۵۵۴
 مدها، ۵۴۹-۵۴۸، ۷۰۵
 مدهای نرمال، ۷۱۱
 مسأله خوش وضع، ۱۲۹
 مسأله بدوضع، ۱۲۹
 مسأله دستگاه خطی، ۲۳
 تحلیل اختلال برای -، ۳۶۱-۳۵۳
 تصفیه تکراری برای حل -، ۳۷۹-۳۷۶
 روشهای تکراری برای حل -، ۴۱۷-۳۸۰
 روشهای مستقیم برای حل -، ۳۴۵-۳۰۵
 وجود و منحصر به فرد بودن جواب، ۲۸۱
 مسأله کماتش، ۵۵۲

- مسأله کمترین توانهای دوم، ۲۴، ۲۴۹
- تصفیه تکراری، ۵۰۴-۵۰۰
- تعبیر هندسی، ۴۵۵
- جواب - با استفاده از تجزیه QR ، ۴۹۱-۴۷۸، ۴۹۹-۴۹۷
- جواب - با استفاده از شبه معکوس، ۴۶۰
- جواب - با استفاده از معادلات نرمال، ۴۷۴-۴۶۹، ۴۹۷-۴۹۵
- جواب - با استفاده از SVD ، ۴۹۱، ۷۹۵-۷۹۲
- جواب - توسط گرام اشمیت، ۴۸۰
- جواب - توسط گرام اشمیت اصلاح شده، ۴۸۶-۴۸۴، ۴۹۹
- حساسیت، ۴۶۹-۴۶۱
- وجود و منحصر به فرد بودن، ۴۵۴-۴۵۲
- مسأله مقدار ویژه بدوضع، ۱۳۷-۱۳۵
- مسأله مقدار ویژه تعمیم یافته، ۲۵، ۶۸۹
- الگوریتم QR برای -، ۷۰۳-۶۹۳
- تجزیه شور برای -، ۶۹۳-۶۹۲
- روش دنباله استورم برای -، ۷۴۲-۷۴۱
- روشهای لانزوس برای -، ۷۴۸-۷۴۲
- روشهای محاسبه برای -، ۷۱۳، ۶۹۳، ۷۴۸-۷۴۱
- معین متقارن، ۶۹۱، ۷۱۲-۷۰۵
- مسأله مقدار ویژه درجه دو، ۲۶، ۷۴۰-۷۳۷
- معادلات سیلوستر، ۶۸۳
- معادلات نرمال، ۴۵۵
- برازش چندجمله‌ای، ۴۵۵
- جوابهای کمترین توانهای دوم، ۴۷۴-۴۶۹
- مشکلات عددی با -، ۴۷۲
- معادله لیاپانوف، ۶۸۳
- معکوس یک ماتریس، ۴۰

خاصیتهای -، ۴۱

محاسبه -، ۳۴۷

مقادیر تکین، ۵۴

مقادیر ویژه، ۳۸

حساسیت، ۶۰۱-۵۹۵

کران برای -، ۵۶۷-۵۶۶

محاسبه از طریق چند جمله‌ای مشخصه، ۵۹۴-۵۹۰

محاسبه از طریق شکل متعارف جردن، ۵۹۴

محاسبه با استفاده از تکرار QR، ۶۳۵-۶۰۳

محاسبه بزرگترین مقدار ویژه، ۵۷۶-۵۷۰

محاسبه کوچکترین مقدار ویژه، ۵۸۸

مرتبه تکرار هندسی، ۵۳۸

مرتبه تکرار جبری، ۵۳۸

نتایج اساسی در مورد -، ۵۳۸-۵۲۹

مقدار ریتز، ۶۵۸، ۷۳۴

مقیاس کردن، ۳۴۲

وضعیت، ۳۶۸

میرایی، ۵۴۹

تناسبی، ۷۲۷

ریلی، ۷۲۷

ن

نامساوی کوشی - شوارتز، ۵۷، ۷۷

نامساوی هولدر، ۵۶

نسبت میرایی مدی، ۷۲۷، ۷۳۸

نرم

اقلیدسی (نرم دو)، ۵۶

- برداری، ۵۵
 طیفی، ۵۸
 فروبنیوس، ۵۹
 A- نرم، ۴۰۷
 ماتریسی، ۵۷
 نامساوی مثلثی، ۵۵
 نرم بی نهایت (نرم ماکزیمم)، ۵۶
 وابسته، ۵۷
 هولدر یا p - نرم، ۵۶

هـ

همگرایی

- الگوریتم SVD دمل - کاهان، ۸۱۵
 تکرار خارج قسمت ریلی، ۵۸۳
 تکرار QR ، ۶۱۲، ۶۱۳، ۶۴۶
 تکرار SOR ، ۳۹۹
 روش توانی، ۵۷۵
 روش گردایان مزدوج، ۴۰۷
 روشهای تکراری، ۳۸۹
 روشهای ژاکوبی و گاوس - سایدل، ۳۹۴-۳۹۰
 مقایسه بین روشهای گاوس سایدل و ژاکوبی، ۳۹۵
 هماهنگ اجباری
 دوره -، ۵۴۸
 یک دستگاه جرم - فتر، ۷۱۳
 یک ساختمان، ۵۴۹، ۷۱۶
 یک ساختمان در طول زلزله، ۷۳۳
 ۱۵۸، $BLAS$

۱۵۸، *EISPACK*

۱۶۳، *ITPACIK*

۱۵۸، *LAPACK*

۱۵۸، *LINPACK*

۸۶۳، ۸۷۶، ۱۶۲، *MATCOM*

۸۵۱، ۸۶۲، ۱۶۲، *MATLAB*

۱۶۱، *NAG*

۱۶۰، *NETLIB*

۱۶۳، *SPARSKIT*

۱۷۱، ۱۶۳، *TEMPLATES*