

آنالیز عددی

مترجمان:

علی اکبر عالمزاده

اسماعیل بابلیان

محمدرضا امیدوار

مؤلفان:

ریچارد ال. بوردن

ج. دوگلاس فیرز

آلبرت سی. رینولدز

آنالیز عددی

ریچارد ال. بوردن، ج. دوگلاس فیروز،
آلبرت سی. رینولدز

مترجمان:

علی اکبر عالم زاده، اسماعیل بابلیان، محمدرضا امیدوار



بیشگفتار مترجمان

آنالیز عددی علم و هنر محاسبه است. در این مبحث به تحلیل روشهایی می پردازیم که جواب عددی مسائل ریاضی را به ما می دهند. تاریخ قدمتش به دوران باستان بازمی گردد. حدود ۲۰۰۰ سال پیش از میلاد، بابلیها به تدوین جداول ریاضی مشغول بوده اند. یک لوح گلی بدست آمده که شامل مربعات اعداد صحیح از ۱ تا ۶۰ است. بابلیها که اجرام آسمانی را می پرستیدند از ستاره شناسی نیز اطلاع داشتند. ستاره شناس معروف اسکندری، کلودیوس بطلمیوس (حدود ۱۵۰ ب م) اتفاق کسوف و خسوف مربوط به ۷۴۷ ق م را خبر داده است.

در سال ۲۲۰ ق م، ارشمیدس از چند ضلعیهای منتظم به عنوان تقریبهایی به دایره استفاده کرد و نامساویهای $3\frac{1}{7} < \pi < 3\frac{1}{6}$ را نتیجه گرفت. تا قرن ۱۷ کارهای عددی عمدتاً " به تهیه جداول نجومی اختصاص داشت. پیدایش جبر در قرن ۱۶ تحرک جدیدی به همه جوانب ریاضی، بویژه آنالیز عددی، بخشید. در سال ۱۶۱۴، نیروالین جدول لگاریتم را منتشر کرد، و در سال ۱۶۲۰، لگاریتمهای توابع سینوس و تانژانت تا هفت رقم اعشاریه جدول در آمدند. تا سال ۱۶۲۸ جداول چهارده رقمی لگاریتم اعداد از ۱ تا ۱۰۰۰۰۰ محاسبه شده بودند.

در اواخر قرن ۱۷، محاسبات با سریهای نامتناهی، همراه با پیشرفت حساب دیفرانسیل و انتگرال، شروع به شکوفایی کرد. در اوایل قرن ۱۸، ژاکوب استرلینگ و بروک تیلور پایه های حساب تفاضلات متناهی را ریختند، که اینک نقش بسیار مهمی در آنالیز عددی دارد. با پیش بینی وجود و موضع سیاره نپتون توسط آدامز و لوریه در ۱۸۴۵، اهمیت علمی آنالیز عددی یکبار و برای همیشه به ثبوت رسید.

در اواخر قرن ۱۹، پیشرفت ماشینهای حساب موجب رشد بیشتر آنالیز عددی شد. این رشد پس از جنگ جهانی دوم به انفجار گشاده شد، و آن بخاطر ساخت ماشینهای کامپیوتر بود. ماشینهای جدید تحقیقات علمی مهم بسیاری را ممکن ساختند که پیشتر به نظر غیرقابل حصول می‌رسید، و اینک در آنالیز عددی نقش بسیار مهمی برعهده‌دارند. امروزه آنالیز عددی، یا به تعبیری ریاضیات مهندسی، شاخه‌ای از ریاضیات کاربردی است که تقریباً "همه رشته‌های مهندسی بدان نیاز دارند. ارج و منزلت رو به افزایش ما را بر آن داشت تا اثر بی‌ظنیر حاضر را ترجمه و تقدیم شیفتگان این علم نماییم. امید است این کتاب در شناساندن بیش از پیش این مبحث به طالبان فارسی زبان آن سهم موثری داشته باشد.

گروه آموزشی ریاضی

دانشگاه تربیت معلم

علی‌اکبر عالم‌زاده اسماعیل بابلیان

محمد رضا امیدوار

بیشگفتار مؤلفان

مطالب این کتاب در خلال یک دوره پنج ساله تهیه شده تا نیاز یک رشته درس در روشهای آنالیز عددی را برآورند. شاگردان این دوره عمدتاً دانشجویان سال سوم علوم و مهندسی اند که مبانی حساب دیفرانسیل و انتگرال را دیده و از یک زبان برنامه‌نویسی سطح بالا، عموماً "فرترن"، اطلاع دارند. دانستن مبانی نظریه ماتریسها و معادلات دیفرانسیل در بخشهای بعدی کتاب مفیدند، لیکن، برای بی‌نیازی از آنها، به عنوان پیشتاز مطالب مقدماتی به حد کافی گنجانده شده است.

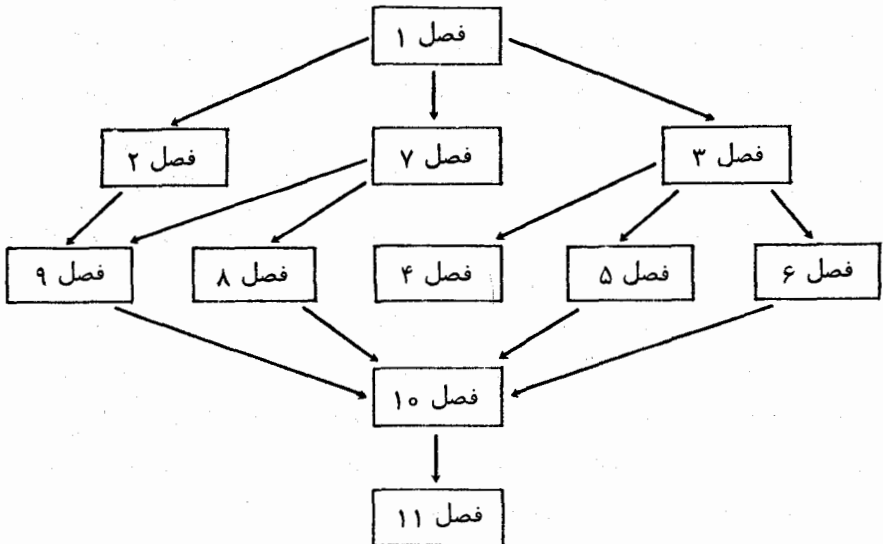
با آنکه مطالب کافی برای یک سال درس آنالیز عددی در سطح لیسانس گنجانده شده است، کتاب را می‌توان برای یک درس مقدماتی در روشهای عددی نیز بکار برد. در این درس، شاگردان انواع مسائل قابل حل بوسیله تکنیکهای عددی و مشکلات ناشی از این روشها را می‌آموزند؛ بقیه کتاب را می‌توان به عنوان یک مرجع برای مسائلی که در کارهای آینده پیش می‌آیند بکار برد. موارد زیر با فلسفه کتاب سازگارند: دادن درکی از روشهای عددی به دانشجو، چگونه و چرا این روشها کار می‌کنند، و حدود کارایی آنها، درکی که پایه استوار برای مطالعات آتی یا پذیرش تکنیکهای جدیدتری باشد که در آینده طرح می‌شوند.

برای هر روش مهم در متن، یک الگوریتم مشروح بدون ذکر برنامه‌ای خاص ارائه شده است. الگوریتمها را به نحوی عرضه کرده‌ایم که حتی یک شاگرد باتجربه محدود در برنامه نویسی می‌تواند آنها را به دستورات برنامه تبدیل کند. تأکید کنیم که گرچه الگوریتمها به شکل ذکر شده منجر به برنامه‌های منطقی صحیحی می‌شوند، لیکن همیشه نسبت به زمان یا ملزومات حافظه کامپیوتر کاراترین وضع را ندارند. وقتی شرح یک الگوریتم که به برنامه

بسیار مؤثری منجر می شود با سهولت طرز کار روش برخورد داشته است، قابل درک ترین شکل را اختیار کرده ایم. برنامه های عملی در متن گنجانده نشده اند چرا که، طبق تجربه و برخلاف مقاصد ما، شرح برنامه ها شاگرد را به تولید نتایج بدون روش مربوطه خواهد کشانید.

کتاب طوری تدوین شده که مدرس در مباحث و در سطوح دقت و کاربرد حق انتخاب دارد. در جهت این اهداف، برای تمام نتایج، نه فقط آنهایی که در متن توضیح داده شده اند بلکه برای تعمیمهای نظریه و کاربرد، مراجع مشروحاتی آورده ایم. این مراجع را حتی الامکان براین پایه اختیار کرده ایم که عمومی ترین منبع در کتابخانه های دانشکده ها بوده اند.

هر مفهوم آمده در متن را با مثال توضیح داده، و بیش از ۷۰۰ تمرین کلاسی گنجانده ایم؛ این تمرینات از کاربردهای ساده الگوریتمها تا تعمیمها و توسعه های پیچیده نظریه تغییر می کنند. بعلاوه، تعداد زیادی مسئله کاربردی از مباحث مختلف مهندسی و علوم فیزیک، بیولوژی، و اجتماعی ارائه شده اند. اکثر آنها از مباحثی گرفته شده اند که به تکنیک تقریب نیاز دارند. تمام کاربردهای انتخابی فشرده بوده و طرز اعمال روشها در "زندگی واقعی" را توضیح می دهند. شاگرد باید بداند مطالبی که اغلب در کلاس ساده می شوند در عمل کاملا پیچیده خواهند بود.



نمودار قبل پیشنهاد فصلها را نشان می‌دهد. اغلب رشته فصول تولید شده از این نمودار در دانشگاه ایالتی یانگز تاون با موفقیت به دانشجویان لیسانس تدریس شده‌اند. برای مدرسان این متن کتاب راهنمایی تدارک دیده‌ایم که شامل پیشنهادات درسی بسیار، برنامه‌های فرترن برای تمام الگوریتمها، و لیست کدهای متعارف برای روشهای آمده در متن است.

در تدوین این کتاب تنی چند شایسته^۱ سپاس‌اند. پیش از همه مایلم مراتب امتنان خود را نسبت به کتی اندرسون^۱ برای تایپ دستنویس و نیز دستیاران دانشجو در دانشگاه ایالتی یانگز تاون بخاطر نقش آنها در این پروژه ابراز داریم. همچنین، از منتقدان خود: رونالد گولداشتاین^۲، دانشگاه میامی؛ دیوید لوگان^۳، دانشگاه ایالتی کانزاس؛ فیلیپ اشمیت^۴، دانشگاه آکرون؛ شروددی. سیلمن^۵، دانشگاه ایالتی کلیولند؛ جیمزال. کورنت^۶، دانشگاه ایالتی آیووا؛ ریچارد اس. فالک^۷، دانشگاه روتگرز؛ دیوید فیندلی^۸، دانشگاه ایالتی کلیولند؛ لاورنسین^۹، دانشگاه دیتون؛ دونالد کالمی^{۱۰}، دانشگاه ایالتی اوهایو؛ شان کیوا^{۱۱}، دانشگاه نیوهمپشایر؛ جیمزال. فیلیپس^{۱۲}، دانشگاه ایالتی واشنگتن؛ و مری وینتر^{۱۳}، دانشگاه ایالتی میشیگان، که برای اصلاح کتاب پیشنهادات با ارزش بسیاری نموده‌اند سپاسگزاریم. بویژه، از ادوین پژاک^{۱۴} و لارن شرودر^{۱۵} بخاطر نظراتشان در تمرینهای کاربردی تشکر می‌نماییم.

در پایان مایلم از شربوردن^{۱۶}، که با مهارتهای دبیری رایگان خود به صورت یادداشتهای کوتاه و در لحظاتی حیاتی یاور ما بوده، و نیز از باربارا فیروز^{۱۷} که نسخ چاپی را غلطگیری کرده و چند پیشنهاد فنی با ارزش نموده است سپاسگزار می‌نماییم.

ریچارد ال. بوردن، ج. دوگلاس فیروز

آلبرت سی. رینولدز

- | | | |
|--------------------|------------------------|-----------------------|
| 1. Kathy Anderson | 2. Ronald Goldstein | 3. David Logan |
| 4. Phillip Schmidt | 5. Sherwood D. Sillman | 6. James L. Cornette |
| 7. Richard S. Falk | 8. David Findley | 9. Lawrence Jehn |
| 10. Donald Kalmey | 11. Shan Kuo | 12. James L. Phillips |
| 13. Mary Winter | 14. Edwin Pejack | 15. Lauren Schroeder |
| 16. Sher Burden | 17. Barabara Faires | |

فهرست مطالب

۱	پیشنمازهای ریاضی	۱
۳	۱۰۱ مروری بر حساب دیفرانسیل و انتگرال	۱۰۱
۱۲	۲۰۱ خطاهای گرد کردن و حساب کامپیوتر	۲۰۱
۲۸	۲ حل معادلات یک متغیره	۲
۳۰	۱۰۲ الگوریتم تنصیف	۱۰۲
۳۶	۲۰۲ نقاط ثابت توابع	۲۰۲
۴۱	۳۰۲ تکرار نقطه ثابت	۳۰۲
۵۲	۴۰۲ روش نیوتن - رفسون	۴۰۲
۶۸	۵۰۲ تحلیل خطا در روشهای تکراری و روشهای شتاب	۵۰۲
۸۳	۶۰۲ صفرهای چندجمله‌ایهای حقیقی	۶۰۲
۱۱۰	۳ درونیایی و تقریب چندجمله‌ای	۳
۱۱۲	۱۰۳ چندجمله‌ایهای تیلور	۱۰۳
۱۲۰	۲۰۳ درونیایی و چندجمله‌ای لاگرانژ	۲۰۳
۱۳۳	۳۰۳ درونیایی تکراری	۳۰۳
۱۴۱	۴۰۳ تفاضلهای منقسم	۴۰۳
۱۴۹	۵۰۳ درونیایی هرमित	۵۰۳
۱۵۹	۶۰۳ درونیایی اسپلاین مکعبی	۶۰۳

۱۸۰	۴ نظریه تقریب
۱۸۲	۱۰۴ آشنایی با تقریب و نرمها
۱۸۷	۲۰۴ تقریب کمترین مربعات گسسته
۲۰۳	۳۰۴ چندجمله‌ایهای متعامد و تقریب کمترین مربعات
۲۱۵	۴۰۴ تقریب کمترین مربعات وزندار
۲۲۷	۵۰۴ چندجمله‌ایهای چبیشف و با صرفه‌سازی سریهای توانی

۲۳۷	۵ مشتگیری و انتگرالگیری عددی
۲۳۹	۱۰۵ مشتگیری عددی
۲۵۴	۴۰۵ مشتقات بالاتر و برونمایی
۲۵۸	۳۰۵ مبانی انتگرالگیری عددی
۲۶۴	۴۰۵ فرمولهای نیوتن - کاتس
۲۷۳	۵۰۵ انتگرالگیری عددی مرکب
۲۸۴	۶۰۵ انتگرالگیری رامبرگ
۲۹۱	۷۰۵ انتگرالگیری گاوسی

۳۰۰	۶ مسائل مقدار اولیه در معادلات دیفرانسیل معمولی
۳۰۳	۱۰۶ نظریهٔ مقدماتی مسائل مقدار اولیه
۳۰۸	۲۰۶ روش اویلر
۳۲۱	۳۰۶ روشهای تیلور از مرتبهٔ بالاتر
۳۲۹	۴۰۶ روشهای رونگ - کوتا
۳۴۱	۵۰۶ کنترل خطا و روش رونگ - کوتا - فلبرگ
۳۵۳	۶۰۶ روشهای چندگامی
۳۷۳	۷۰۶ روشهای برونمایی
۳۸۲	۸۰۶ معادلات مرتبهٔ بالاتر و دستگاههای معادلات دیفرانسیل
۳۹۲	۹۰۶ پایداری و انتخاب روشها

۴۱۰	۷ روشهای مستقیم حل دستگاههای خطی
-----	----------------------------------

۴۱۲	دستگاههای معادلات خطی	۱۰۷
۴۲۰	حذف گاوسی و جایگذاری پسرو	۲۰۷
۴۳۱	جبر خطی و معکوس کردن ماتریس	۳۰۷
۴۶۱	استراتژیهای محورگیری	۴۰۷
۴۶۹	انواع ویژه ماتریسها	۵۰۷
۴۷۶	تجزیه مستقیم ماتریسها	۶۰۷

۴۹۵	روشهای تکراری در جبر ماتریسی	۸
۴۹۶	نرمهای بردارها و ماتریسها	۱۰۸
۵۱۷	تخمینهای خطا و تصفیه تکراری	۲۰۸
۵۲۹	روشهای تکراری برای حل دستگاههای خطی	۳۰۸
۵۴۵	مقادیر ویژه و بردارهای ویژه	۴۰۸
۵۶۶	الگوریتم QR	۵۰۸

۵۸۷	حل عددی دستگاههای معادلات غیرخطی	۹
۵۸۸	نقاط ثابت برای توابع چند متغیره	۱۰۹
۶۰۱	روش نیوتن	۲۰۹

۶۱۱	مسائل مقدار مرزی برای معادلات دیفرانسیل معمولی	۱۰
۶۱۳	روش تیراندازی خطی	۱۰۱۰
۶۲۳	روش تیراندازی برای مسائل غیرخطی	۲۰۱۰
۶۳۲	روشهای تفاضل متناهی برای مسائل مقدار مرزی	۳۰۱۰
۶۴۸	روش ریلی - ریتس	۴۰۱۰

۶۶۵	حل عددی معادلات دیفرانسیل جزئی	۱۱
۶۶۶	مسائل فیزیکی شامل معادلات دیفرانسیل جزئی	۱۰۱۱
۶۷۲	معادلات دیفرانسیل جزئی بیضوی	۲۰۱۱
۶۸۵	معادلات دیفرانسیل جزئی سهموی	۳۰۱۱

۷۰۴

۴۰۱۱ معادلات دیفرانسیل جزئی همدلولوی

۷۲۷

۵۰۱۱ آشنایی با روش عناصر متناهی

۷۴۲

کتابنامه

۷۴۶

جواب تمرینات منتخب

۷۶۷

واژه‌نامه فارسی به انگلیسی

۷۷۵

واژه‌نامه انگلیسی به فارسی

۷۸۸

فهرست راهنما

۸۰۳

فهرست نمادها

۸۰۵

فهرست الگوریتمها

آشنایی

آنالیز عددی در رابطه با طرح روشهایی است که برای تقریب جواب مسائل بیان شده به صورت ریاضی به روشی مؤثر بکار می‌روند. کارایی روش به دقت خواسته شده از آن و سهولت کاربردش بستگی دارد. در عمل، مسائل ریاضی از پدیده‌های فیزیکی ناشی شده که پذیرش چند فرض ساده‌ساز نمایش ریاضی مربوطه را ممکن می‌سازد. به‌طور کلی، انعطاف در فرضهای فیزیکی مدل ریاضی مناسبتری بدست می‌دهد، لیکن مدلی که حل صریح آن سخت‌تر و یا غیرممکن می‌باشد. چون مسئلهٔ ریاضی معمولاً "مسئلهٔ فیزیکی را دقیقاً" حل نمی‌کند، جواب تقریبی مدل ریاضی پیچیده‌تر اغلب دقیق‌تر از جواب کامل مدل ساده شده است. برای رسیدن به این تقریب، مدلی به نام الگوریتم طرح شده که معمولاً "از چند عمل حبری و منطقی تشکیل شده و تقریبی به مسئلهٔ ریاضی بدست می‌دهد، و امید است با خطایی قابل تحمل و یا دقتی مقرر تقریبی به مسئلهٔ فیزیکی نیز باشد.

از آنجا که کارایی یک روش به سهولت کاربردش بستگی دارد، انتخاب روش مناسب برای تقریب جواب یک مسئله تاحدودی تحت اثر دگرگونی تکنولوژی ماشینهای حساب و کامپیوتر می‌باشد. بیست و پنج سال قبل، پیش از رایج شدن کامپیوتر رقمی، روشهایی که نیاز به محاسبهٔ بسیار داشتند نمی‌توانستند به‌طور معقول بکار روند. ولی از آن زمان تاکنون پیشرفتهای کامپیوتر بعضی از این روشها را به طرز زایدالوصفی جذاب ساخته است. در حال حاضر، عامل محدودکننده معمولاً "گنجایش حافظهٔ کامپیوتر در رابطه با روش است، اگرچه البته عامل هزینهٔ مربوط به زمان زیاد محاسبه نیز اهمیت دارد. ماشینهای حساب نسبتاً کم‌خرج جدید که برنامه پذیرند نیز عامل مؤثری در انتخاب روش تقریب می‌باشند، چراکه این ماشینها در دسترس بوده و از آنها می‌توان برای حل بسیاری از مسائل نسبتاً ساده

استفاده کرد .

بهرحال ، ایده‌های اصلی اغلب تکنیکهای عددی جاری ، از جمله روشهای بکار رفته در پیش‌بینی کرانه‌های خطای ماکزیم ناشی از کاربرد روشها ، سالهاست شناخته شده‌اند . لذا ، به شیوهٔ رسیدن به این روشها و طرز تخمین خطای آنها توجهی خاص مبذول می‌شود ، زیرا از دگرگونی این تکنیکها ، صرف‌نظر از تکنولوژی ، در ارائه و اعمال روندهای عددی در آینده حتما " استفاده خواهد شد .

روشهای این کتاب روشهای معمول در زمان حاضر و روشهایی است که روشهای بهتر در آیندهٔ نزدیک احتمالا " بر آنها استوار خواهد بود .

بیشنیازهای ریاضی^۱

در درسهای مقدماتی شیمی، دانشجویان با رابطه

$$PV = NRT,$$

به نام قانون گازهای کامل، مواجه می شوند که فشار P ، حجم V ، دمای T ، و N ، تعداد مل^۱های یک گاز "کامل"، را بهم مربوط می کند. در این معادله R ثابتی است که فقط به دستگاه سنجش بکار رفته بستگی دارد.

فرض کنیم برای امتحان این قانون دو آزمایش، هر دو با یک گاز، صورت گرفته باشد. در آزمایش اول

$$P = 1.0 \text{ اتمسفر}، V = 10 \text{ مترمکعب}$$

$$R = 0.082، N = 0.0042 \text{ مل}$$

با استفاده از قانون گازهای کامل، پیش بینی می شود که دمای گاز چنین باشد:

$$T = \frac{PV}{NR} = \frac{(1.0)(10)}{(0.082)(0.0042)} = 290^\circ \text{ کلوین یا } 17^\circ \text{ سانتیگراد}$$

با اندازه گیری دمای گاز، ملاحظه می شود که دمای واقعی 15° سانتیگراد

است.

آزمایش را با همان مقادیر R و N تکرار می کنیم، ولی این بار فشار را چهار برابر کرده و حجم را به همین نسبت کاهش می دهیم. چون حاصل ضرب PV تغییر

۱) هر مل برابر $10^{23} \times 1/60$ است که تعداد مولکولهای یک ملکول گرم گاز

می باشد.

نکرده است، دمای پیش‌بینی شده همان 290° کلوین (یا 17° سانتیگراد) است، درحالی که دمای واقعی گاز در این حالت 32° سانتیگراد است.

واضح است که وقتی خطا تا این حد بزرگ باشد، قانون گازهای کامل مورد تردید واقع خواهد شد. اما، پیش از آنکه عدم اعتبار این قانون را نتیجه بگیریم، باید داده‌هایمان را بررسی و مشخص کنیم که آیا می‌توان خطا را به نتایج تجربی نسبت داد یا نه. اگر چنین باشد، تعیین اینکه نتایج تجربی چقدر باید دقیقتر باشند تا خطایی به این بزرگی روی ندهد سودمند خواهد بود.

این نوع تحلیل خطا در محاسبات، مبحث مهمی در آنالیز عددی است و در بخش دوم این فصل مطرح خواهد شد.

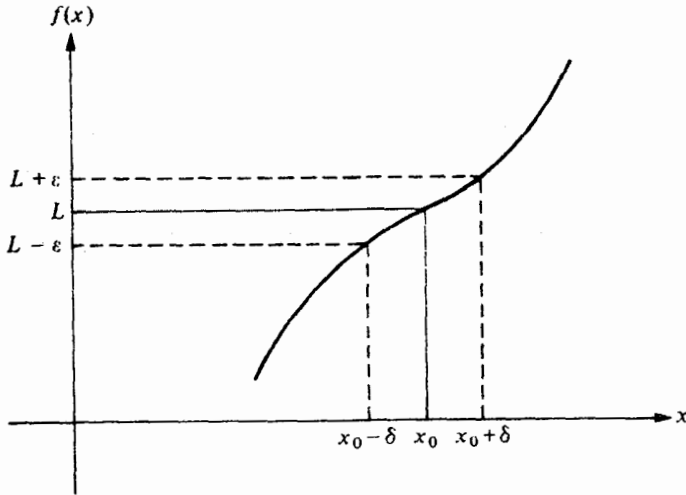
این فصل مرور کوتاهی است بر مطالب مورد نیاز از حساب دیفرانسیل و انتگرال یک متغیره^۶ مقدماتی در فصلهای آتی، و نیز آشنایی با اصطلاحاتی که در بحث همگرایی، تحلیل خطا، و نمایش ماشینی اعداد بکار می‌روند.

۱.۱ مروری بر حساب دیفرانسیل و انتگرال

در حساب دیفرانسیل و انتگرال، حد و پیوستگی یک تابع از مفاهیم بنیادی می‌باشند.

تعریف ۱.۱. فرض کنیم تابع f بر مجموعه^۶ X از اعداد حقیقی تعریف شده باشد؛ گوئیم f در x_0 دارای حد L است، که نوشته می‌شود $\lim_{x \rightarrow x_0} f(x) = L$ ، اگر به ازای هر عدد حقیقی $\varepsilon > 0$ عددی حقیقی مانند $\delta > 0$ وجود داشته باشد بطوری که $|f(x) - L| < \varepsilon$ به ازای هر $x \in X$ که $0 < |x - x_0| < \delta$ برقرار باشد. (ر.ک. شکل (۱.۱.۱))

تعریف ۲.۱. فرض کنیم تابع f بر مجموعه^۶ X از اعداد حقیقی تعریف شده باشد و $x_0 \in X$ ؛ گوئیم f در x_0 پیوسته است اگر $\lim_{x \rightarrow x_0} f(x) = f(x_0)$ و نیز، تابع f را بر X پیوسته گوئیم اگر در هر نقطه از X پیوسته باشد. مجموعه^۶ تمام توابع پیوسته بر X با $C(X)$ نشان داده می‌شود. وقتی X بازه‌ای از خط حقیقی باشد، پیرانتزها را در این نماد حذف می‌کنیم. مثلاً، مجموعه^۶ تمام توابع پیوسته بر بازه^۶ بسته^۶ $[a, b]$ را با $C[a, b]$ نشان خواهیم داد.



شکل ۱.۱

حد یک دنباله از اعداد حقیقی یا مختلط را می‌توان به همین ترتیب تعریف کرد.

تعریف ۳.۱. فرض کنیم $\{x_n\}_{n=1}^{\infty}$ دنباله‌ای نامتناهی از اعداد حقیقی یا مختلط باشد. گوییم این دنباله به عدد x (به نام حد دنباله) همگرا است اگر به ازای هر $\varepsilon > 0$ عدد صحیح مثبتی چون $N(\varepsilon)$ وجود داشته باشد بطوری که $n > N(\varepsilon)$ نامساوی $|x_n - x| < \varepsilon$ را ایجاب کند. نماد $\lim_{n \rightarrow \infty} x_n = x$ یا $x_n \rightarrow x$ وقتی $n \rightarrow \infty$ ، یعنی دنباله $\{x_n\}_{n=1}^{\infty}$ به x همگرا می‌باشد.

قضیه^۴ زیر مفاهیم همگرایی و پیوستگی را بهم ربط می‌دهد.

قضیه^{۴.۱}. هرگاه تابع f بر مجموعه X از اعداد حقیقی تعریف شده باشد و $x_0 \in X$ ،

آنگاه دو حکم زیر با هم معادل می‌باشند:

(A) f در x_0 پیوسته است؛

(B) اگر $\{x_n\}_{n=1}^{\infty}$ دنباله‌ای از نقاط در X و به x_0 همگرا باشد،

$$\lim_{n \rightarrow \infty} f(x_n) = f(x_0).$$

تعریف ۵.۱. هرگاه تابع f در نقطه x_0 تعریف شده باشد، گوئیم f در x_0 مشتقپذیر است اگر

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

موجود باشد. این حد را، در صورت وجود، با $f'(x_0)$ نشان داده و آن را مشتق f در x_0 می‌نامیم. تابعی که در هر نقطه از مجموعه X مشتق داشته باشد بر X مشتقپذیر خوانده می‌شود.

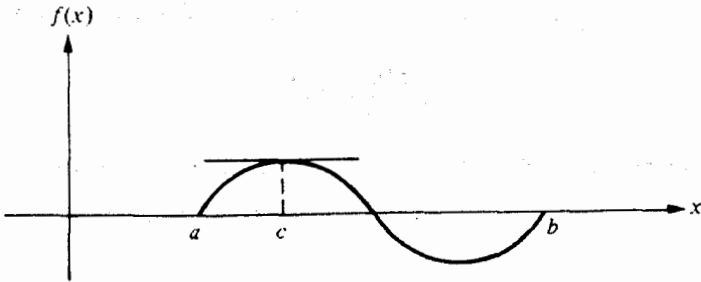
بازای مجموعه $X \subseteq \mathbb{R}$ ، مجموعه تمام توابعی که بر X ، n مشتق پیوسته دارند را با $C^n(X)$ ، و مجموعه تمام توابعی که در هر نقطه X از هر مرتبه مشتق دارند را با $C^\infty(X)$ نشان می‌دهیم. توابع چند جمله‌ای، گویا، مثلثاتی، نمایی، و لگاریتمی در رده $C^\infty(X)$ قرار دارند، که در آن X مجموعه تمام نقاطی است که این توابع در آنها تعریف شده‌اند. وقتی X بازه‌ای از خط حقیقی باشد، باز در این نماد پیرانتزها را حذف می‌کنیم. قضیه زیر نشان می‌دهد که این نوع توابع پیوسته نیز هستند.

قضیه ۶.۱. هرگاه تابع f در نقطه x_0 مشتقپذیر باشد، آنگاه f در x_0 پیوسته می‌باشد.

فضای زیر در بدست آوردن روشهای تخمین خطا اهمیتی بنیادی دارند. اثبات این قضایا و دیگر نتایج بی‌ارجاع در این بخش را می‌توان در هر کتاب حساب دیفرانسیل و انتگرال مقدماتی یافت (مثلاً، ر.ک. توماس^۱ [۶۴]).

قضیه ۷.۱ (قضیه رل^۲). فرض کنیم $f \in C[a, b]$ و f بر (a, b) مشتقپذیر باشد. هرگاه $f(a) = f(b) = 0$ ، آنگاه نقطه‌ای مانند c ، که $a < c < b$ ، وجود دارد بطوری که $f'(c) = 0$. (ر.ک. شکل (۰.۲.۱))

قضیه ۸.۱ (قضیه مقدار میانگین). هرگاه $f \in C[a, b]$ و f بر (a, b) مشتقپذیر باشد،



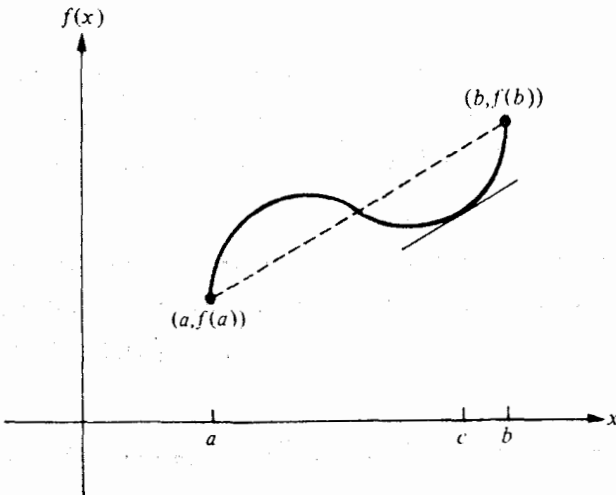
شکل ۲.۱

آنگاه نقطه‌ای مانند c ، که $a < c < b$ ، وجود دارد بطوری که

$$f'(c) = \frac{f(b) - f(a)}{b - a} \quad (\text{ر.م. ش. شکل ۳.۱})$$

قضیه ۹.۱ (قضیه مقدار اکستریم). هرگاه $f \in C[a, b]$ ، آنگاه نقاطی مانند

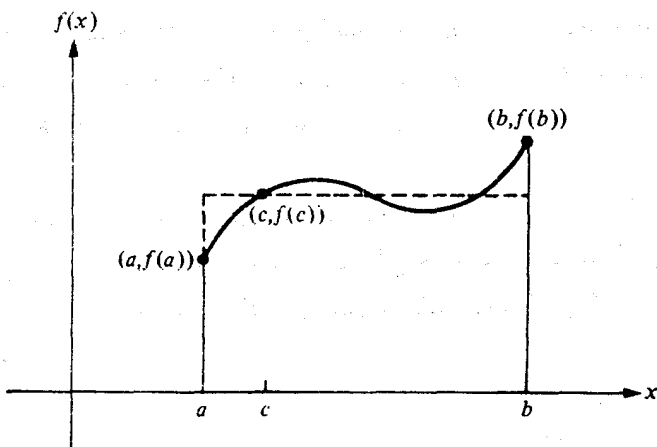
$f(c_1) \leq f(x) \leq f(c_2)$ ، $x \in [a, b]$ هر a, b وجود دارند بطوری که، $c_1, c_2 \in [a, b]$



شکل ۳.۱

بعلاوه، هرگاه f در هر نقطه از $[a, b]$ مشتقپذیر باشد، آنگاه، به ازای هر $i = 1, 2$ ،
 $f'(c_i) = 0$ یا $c_i = b$ یا $c_i = a$.

در بررسی روشهای عددی به دو نتیجه دیگر نیاز داریم. اولی تعمیمی از قضیه معمولی مقدار میانگین برای انتگرالها است. این قضیه، وقتی $g(x) \equiv 1$ ، مقدار متوسط تابع را روی بازه $[a, b]$ بدست می دهد (ر.ک. شکل ۴.۱).



شکل ۴.۱

برهان قضیه ۱۰.۱ عموماً "در یک درس حساب دیفرانسیل و انتگرال پایه نمی آید، اما می توان آن را در هر کتاب حساب دیفرانسیل و انتگرال پیشرفته متعارف یافت (مثلاً، ر.ک. فولکس^۱ [۲۹]، صفحه ۱۲۵).

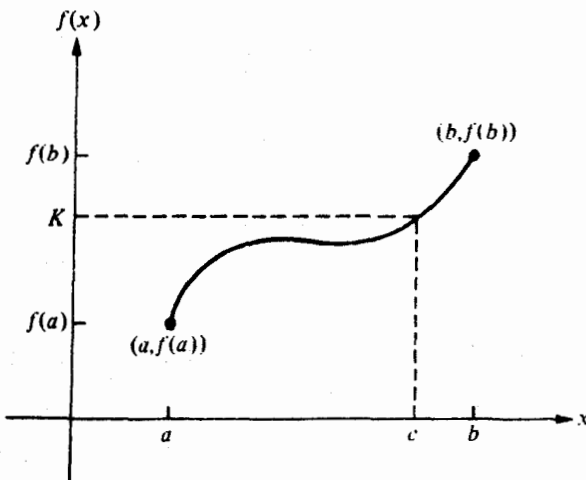
قضیه ۱۰.۱ (قضیه مقدار میانگین وزندار برای انتگرالها). هرگاه $f, g \in C[a, b]$ بر $[a, b]$ انتگرالپذیر باشد، و $g(x) \geq 0$ ، آنگاه عددی مانند c ، که $a < c < b$ وجود دارد بطوری که

$$\int_a^b f(x)g(x) dx = f(c) \int_a^b g(x) dx.$$

دیگر قضیه‌ای که مورد نیاز است، و عموماً "در یک درس حساب دیفرانسیل و انتگرال پایه ارائه نمی‌شود، از اعمال قضیه رل (قضیه ۷۰۱) بر f ، f' ، ...، و بالاخره بر $f^{(n-1)}$ بدست می‌آید.

قضیه ۱۱۰۱ (تعمیم قضیه رل). فرض کنیم $f \in C[a, b]$ بر (a, b) ، n بار مشتق‌پذیر باشد. هرگاه f در $n + 1$ نقطه متمایز x_0, \dots, x_n از $[a, b]$ صفر شود، آنگاه نقطه‌ای مانند c در (a, b) وجود دارد بطوری که $f^{(n)}(c) = 0$.

قضیه بعدی قضیه مقدار میانی است. با آنکه شهوداً "قضیه واضحی است، اثباتش از حوصله درس حساب دیفرانسیل و انتگرال معمولی خارج است. برهان آن را می‌توان در اغلب کتب حساب دیفرانسیل و انتگرال پیشرفته یافت (مثلاً، ر. ک. فولکس [۲۹]، صفحه ۵۹).



شکل ۵۰۱

قضیه ۱۲.۱ (قضیه مقدار میانی). هرگاه $f \in C[a, b]$ و K عددی بین $f(a)$ و $f(b)$ باشد، آنگاه نقطه‌ای مانند c در (a, b) وجود دارد که به‌ازای آن $f(c) = K$ (ر.ک. شکل (۵.۱)).

مثال ۱. برای اثبات اینکه $x^5 - 2x^3 + 3x^2 - 1 = 0$ بر بازه $[0, 1]$ جواب دارد، تابع $f(x) = x^5 - 2x^3 + 3x^2 - 1$ را در نظر می‌گیریم. واضح است که f بر $[0, 1]$ پیوسته است و $f(0) = -1$ و $f(1) = 1$ که $f(0) < 0 < f(1)$. چون f مقدار میانی را می‌گیرد، پس $f(c) = 0$ برای $c \in (0, 1)$ وجود داشته باشد که به‌ازای آن $x^5 - 2x^3 + 3x^2 - 1 = 0$.

همانطور که در این مثال ملاحظه شد، قضیه مقدار میانی از جهت کمک در تعیین اینکه چه وقت مسائل معینی جواب دارند مهم است. لکن، برای یافتن این جوابها راهی ارائه نمی‌دهد. این مطلب در فصل ۲ کاملتر بحث خواهد شد. آخرین قضیه این مرور از حساب دیفرانسیل و انتگرال بسط تیلور^۱ را توصیف می‌کند. از اهمیت بسط تیلور در آنالیز عددی هرچه بگوییم کم است، و نتیجه زیر بارها مورد استفاده ماقراری می‌گیرد. به‌اثبات این قضیه، در بخش ۲.۳، تمرین ۱۱، اشاره مختصری خواهد شد.

قضیه ۱۳.۱ (قضیه تیلور). فرض کنیم $f \in C^n[a, b]$ و $f^{(n+1)}$ بر $[a, b]$ موجود باشد. همچنین، $x_0 \in [a, b]$. در این صورت، به‌ازای هر $x \in [a, b]$ ، نقطه‌ای مانند $\xi(x) \in (a, b)$ وجود دارد بطوری که

$$f(x) = P_n(x) + R_n(x)$$

که در آن

$$P_n(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

$$= \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k,$$

9

$$R_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!} (x - x_0)^{n+1}.$$

در اینجا $P_n(x)$ چند جمله‌ای تیلور درجه n م f حول x_0 و $R_n(x)$ جمله باقیمانده (یا خطای برشی) وابسته به $P_n(x)$ نامیده می‌شود.

اصطلاح "خطای برشی" عموماً به خطای ناشی از بکارگیری یک جمع‌بندی متناهی یا بریده شده برای تقریب سازی مجموع یک سری نامتناهی اطلاق می‌شود. این اصطلاح در فصل‌های بعد مجدداً معرفی می‌شود.

مثال ۲. فرض کنیم $f(x) = \cos x$. چون $f \in C^\infty(\mathbb{R})$ ، می‌توان قضیه ۱۳.۱ را به ازای هر $n > 0$ بکار برد. به‌ازای $n = 2$ و $x_0 = 0$ ، قضیه ۱۳.۱ نتیجه می‌دهد که

$$\cos x = 1 - \frac{1}{2}x^2 + \frac{1}{6}x^3 \sin \xi(x).$$

که در آن $\xi(x)$ عددی بین 0 و x است.

به‌ازای $x = .001$ ، چند جمله‌ای تیلور نتیجه می‌دهد که

$$\begin{aligned} \cos .001 &= 1 - \frac{1}{2}(.001)^2 + \frac{1}{6}(.001)^3 \sin \xi(x) \\ &= .9999995 + (.16\bar{6}) \cdot 10^{-9} \sin \xi(x), \end{aligned}$$

که در آن $0.001 < \xi(x) < 0$. (تیره روی آخرین رقم در 166. نشان آن است که این رقم بی‌نهایت بار تکرار می‌شود.)

بنابراین، چون $|\sin \xi(x)| < 1$ ، می‌توان از 9999995 برای تقریب $\cos .001$ ، و با اطمینان از دقت دست‌کم تا نه رقم اعشار، استفاده کرد. با استفاده از جداول معمول، ملاحظه می‌شود که

$$\cos .001 = .9999995 \ 000 \ 000 \ 42.$$

و، در واقع، دقتی تا 13 رقم اعشار وجود دارد.

در این مثال، اگر چند جمله‌ای تیلور درجه سوم به‌ازای $x_0 = 0$ بکار رود، چون

$$f'''(0) = 0$$

$$\cos x = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 \cos \zeta(x),$$

که در آن $0 < \zeta(x) < .001$. چند جمله‌ای تقریب ساز همان چند جمله‌ای قبلی است، و تقریب همان 9999995 است، لیکن در اینجا نیز انتظار دقت تا 13 رقم اعشار می‌رود، چرا که

$$\left| \frac{1}{24}x^4 \cos \zeta(x) \right| \leq \frac{1}{24}(.001)^4(1) \approx 4.2 \times 10^{-14}.$$

این مقدار به مقدار واقعی نزدیکتر است.

مجموعه تمرینات ۱۰۶

۱. فرض کنید $f(x) = 1 - e^x + (e - 1)\sin(\pi/2)x$. نشان دهید که $f'(x)$ دست کم یک‌بار در $[0, 1]$ صفر می‌شود.

۲. نشان دهید که معادله $x = 3^{-x}$ در $[0, 1]$ جواب دارد.

۳. فرض کنید $f(x) = e^{-x}$. چند جمله‌ای تیلور درجه سوم f حول $x_0 = 1$ را بیابید. با استفاده از این چند جمله‌ای، $e^{-.99}$ را تقریب نمایید. انتظار دقت تا چند رقم اعشار می‌رود؟

۴. فرض کنید به ازای $x \in \mathbb{R}$ ، $f(x) = |x|$. نشان دهید که f در هر $x \neq 0$ مشتق‌پذیر است اما در $x = 0$ مشتق‌پذیر نیست.

۵. تعیین کنید که قضیه مقدار میانگین (قضیه ۸.۱) را در حالات زیر می‌توان بکار برد یا نه. اگر می‌توانید، نقطه c را که در این قضیه صدق می‌کند بیابید. در غیر این صورت، نشان دهید چرا چنین نقطه‌ای وجود ندارد:

$$f(x) = x^{2/3}, [a, b] = [0, 8] \quad (\text{ب}) \quad f(x) = x^{2/3}, [a, b] = [-1, 8] \quad (\text{آ})$$

$$f(x) = |x|, [a, b] = [0, 1] \quad (\text{ت}) \quad f(x) = |x|, [a, b] = [-1, 1] \quad (\text{پ})$$

۶. با استفاده از قضیه تیلور (قضیه ۱۳.۱)، به ازای $n = 2$ و $x_0 = 0$ ، تقریبی برای $\sin .001$ بدست آورید. آیا انتظار دارید دقت تقریب در این مسئله همان دقت حاصل برای $\cos .001$ در مثال ۲ باشد؟

۷. گوییم تابع $f: [a, b] \rightarrow \mathbb{R}$ بر $[a, b]$ در شرط لیب شیتس^۱، و با ثابت لیب شیتس L ، صدق می‌کند اگر، به ازای هر $x, y \in [a, b]$ ،

$$|f(x) - f(y)| \leq L|x - y|.$$

(T) نشان دهید که اگر f در شرط لیب شیتس، و با ثابت لیب شیتس L ، بر

بازه $[a, b]$ صدق کند، $f \in C[a, b]$.

(ب) نشان دهید که اگر مشتق f بر $[a, b]$ به وسیله L کراندار باشد، f بر $[a, b]$

در شرط لیب شیتس، و با ثابت لیب شیتس L ، صدق می کند.

(پ) تابعی مثال بزنید که بر یک بازه بسته پیوسته باشد ولی در شرط لیب شیتس

بر آن بازه صدق نکند.

۲۰۱ خطاهای گرد کردن و حساب کامپیوتر

وقتی در محاسبات عددی از ماشین حساب و یا کامپیوتر رقمی استفاده می شود، باید خطایی غیر قابل اجتناب، به نام خطای گرد کردن، در نظر گرفته شود. این خطا به این دلیل رخ می دهد که حساب ماشین با اعدادی صورت می گیرد که فقط تعدادی متناهی رقم دارند؛ و در نتیجه، خیلی از محاسبات با نمایشهای تقریبی از این اعداد انجام می شود. در یک کامپیوتر نوعی، برای نمایش همه اعداد حقیقی فقط زیر مجموعه "نسبتاً" کوچکی از دستگاه اعداد حقیقی بکار می رود. این زیر مجموعه فقط شامل اعداد گویا، هم مثبت و هم منفی، است، و یک جزء کسری، به نام مانتیس، همراه با یک جزء نمایی، به نام مشخصه، را ذخیره می کند. مثلاً، یک عدد با دقت معمولی و با ممیز سیار که در IBM سری 360 یا 370 بکار می رود از یک مانتیس 24 رقمی دودویی و یک نمای 7 رقمی دودویی با پایه 16 تشکیل شده است. چون 24 رقم دودویی متناظر 7 تا 8 رقم اعشاری است، می توان فرض کرد که IBM سری 360 یا 370 برای دستگاه اعداد با ممیز سیار دقتی دست کم تا هفت رقم اعشار دارد. نمای هفت رقم دودویی برد 0 تا 127 را بدست می دهد، اما، بخاطر پیشقدر نمایی، برد 64- تا 63+ است؛ یعنی، 64 خود بخود از نمای ثابت شده کم می شود. عدد ماشینی

0	1000010	101100110000010000000000
---	---------	--------------------------

دقیقاً "نمایش عدد اعشاری

$$+ \left(\left(\frac{1}{2} \right)^1 + \left(\frac{1}{2} \right)^3 + \left(\frac{1}{2} \right)^4 + \left(\frac{1}{2} \right)^7 + \left(\frac{1}{2} \right)^8 + \left(\frac{1}{2} \right)^{14} \right) \times 16^{66-64} = 179.015625$$

است، زیرا اولین رقم دودویی نمایش علامت است، 0 برای بعلاوه و 1 برای منهای هفت رقم دودویی بعدی نمایش نما است، و بیست و چهار رقم دودویی آخر نمایش مانتهیس است و در واقع برای نمایش هر عدد در بازه^۶

$$[179.01561737060546875, 179.01563262939453125]$$

بکار می‌رود.

در این نمایش، کوچکترین عدد مثبت قابل بیان $10^{-77} \approx 16^{-64}$ ، و بزرگترین عدد $10^{75} \approx 16^{63}$ است. چون دست‌کم یکی از سمت چپ‌ترین چهار رقم دودویی هر عدد ناصفر بزرگتر از 16^{-64} باید یک باشد، 15.2^{28} عدد به شکل

$$(1.1) \quad \pm .d_1 d_2 \dots d_{24} \times 16^{e_1 e_2 \dots e_7}$$

وجود دارد، که در این دستگاه برای نمایش جمیع اعداد حقیقی بکار می‌روند. در محاسبات، اعداد کوچکتر از 16^{-64} منتج به پاریز شده و اغلب آنها را صفر می‌گیرند، درحالی‌که اعداد بزرگتر از 16^{63} منتج به سرریز شده و باعث توقف محاسبات می‌گردند. دستگاه نمایش عددی فوق در همه^۷ ماشینهای کامپیوتر مرسوم نیست، لیکن مشکلات احتمالی را تا حدودی نشان می‌دهد. تا آخر این بحث، بخاطر سادگی، فرض می‌کنیم نمایش اعداد ماشینی به شکل اعشاری نرمال شده^۸

$$\pm .d_1 d_2 \dots d_k \times 10^n, \quad 1 \leq d_1 \leq 9, \quad 0 \leq d_i \leq 9,$$

به‌ازای هر $i = 2, \dots, k$ ، باشد، که، طبق آنچه هم‌اینک گذشت، در ماشینهای IBM تقریباً " $k = 7$ و $-77 \leq n \leq 75$ ".

نمایش یک عدد حقیقی دلخواه با ممیز سیار و به‌شکل (1.1) سودمند است. هر عدد حقیقی و مثبت y را، در صورتی که در برد عددی ماشین باشد، می‌توان به شکل نرمال شده^۹

$$y = .d_1 d_2 \dots d_k d_{k+1} d_{k+2} \dots \times 10^n$$

درآورد. شکل با ممیز سیار (1.1)، که با $f(y)$ نموده می‌شود، از مختوم کردن مانتهیس y به k رقم اعشار بدست می‌آید. برای انجام این مختومیت دوره وجود دارد. یک روش جدا کردن ارقام $d_{k+1} d_{k+2} \dots$ است که بدست می‌دهد

$$f(y) = .d_1 d_2 \dots d_k \times 10^n,$$

و دیگری افزودن $5 \times 10^{-(k+1)}$ به y و سپس جدا کردن تا بدست آید

$$f(y) = \delta_1 \delta_2 \dots \delta_k \times 10^n.$$

روش دوم را اغلب گرد کردن عدد می نامند، چراکه، اگر $d_{k+1} \geq 5$ ، به d_k یک می افزاییم تا $f(y)$ بدست آید و، اگر $d_{k+1} < 5$ ، همه ارقام جز k رقم اول را جدا می کنیم. مثلاً، اگر $k = 5$ و روش گرد کردن بکار رود، π و e به ترتیب با $10^1 \times 3.1416$ و $10^1 \times 2.7183$ نمایش داده می شوند. و اگر $k = 4$ و از روش جدا کردن استفاده شود، $\frac{1}{3}$ و $\frac{1}{4}$ به ترتیب با 3.333 و 1.428 ، نموده می شوند.

چون اعداد حقیقی همیشه دقیقاً در ماشین قابل نمایش نیستند، لازم است خطای ناشی از این تقریب متناهی رقم در نظر گرفته شود. تعریف زیر طرز اندازه گیری این خطا را مشخص می کند.

تعریف ۱۴.۱. هرگاه p^* تقریبی از p باشد، خطای مطلق مساوی $|p - p^*|$ و خطای نسبی برابر $(|p - p^*|/|p|)$ ، به شرط $p \neq 0$ ، است.

در مثال زیر، به خطاهای مطلق و نسبی ناشی از نمایش p با p^* توجه کنید.

مثال ۱

(آ) هرگاه $p = 3.000 \times 10$ و $p^* = 3.100 \times 10$ ، خطای مطلق ۱. و خطای نسبی $10^{-1} \times 3.333$ است.

(ب) هرگاه $p = 3.000 \times 10^{-3}$ و $p^* = 3.100 \times 10^{-3}$ ، خطای مطلق $10^{-4} \times 1$ و خطای نسبی $10^{-1} \times 3.333$ است.

(پ) هرگاه $p = 3.000 \times 10^4$ و $p^* = 3.100 \times 10^4$ ، خطای مطلق $10^3 \times 1$ و خطای نسبی $10^{-1} \times 3.333$ می باشد.

این مثال نشان می دهد که یک خطای نسبی، یعنی $10^{-1} \times 3.333$ ، برای چند خطای مطلق کاملاً متفاوت رخ می دهد. در نتیجه، خطای مطلق، به عنوان شاخصی از دقت، ممکن است گمراه کننده و خطای نسبی مناسبتر باشد همانطور که تعریف ذیل نشان می دهد، با استفاده از خطای نسبی، می توان درباب عده ارقام درست یک تقریب یا نمایش سخن گفت.

تعریف ۱۵.۱. گوئیم p^* ، p را تا t رقم (یا پیکر) با معنی تقریب می‌کند اگر t بزرگترین عدد صحیح نامنفی باشد که به‌ازای آن

$$\frac{|p - p^*|}{|p|} < 5 \times 10^{-t}.$$

دلیل استفاده از خطای نسبی در این تعریف بدست آوردن مفهومی پیوسته است. اعداد 1000، 5000، 9990، و 10,000 را در نظر می‌گیریم. برای آنکه p^* ، 1000 را طبق این تعریف تا چهار رقم با معنی تقریب کند، p^* باید در نامساوی

$$\left| \frac{p^* - 1000}{1000} \right| < 5 \times 10^{-4}.$$

صدق کند. در نتیجه، باید $999.5 < p^* < 1000.5$ ، که با تعریف شهودی ارقام با معنی مطابقت دارد. اگر همین مسئله را برای 5000 و 9990 در نظر بگیریم، p^* باید به ترتیب در $4997.5 < p^* < 4999.5$ و $9985.005 < p^* < 9994.995$ صدق کند تا اینکه تا چهار رقم با معنی دقیق باشد. این ممکن است با مفهوم شهودی ارقام با معنی مطابقت نداشته باشد. به‌رحال، توجه کنید که برای آنکه p^* یک تقریب 10,000 با چهار رقم با معنی باشد، p^* باید در $10005 < p^* < 9995$ صدق کند، که مجدداً با وضع شهودی مطابقت دارد. جدول زیر سرشت پیوسته این مفهوم را با ذکر کوچکترین کران بالایی $|p - p^*|$ ، به‌ازای مقادیر مختلف p ، یعنی $\max|p - p^*|$ ، وقتی p^* با p تا چهار رقم با معنی مطابقت دارد، نشان می‌دهد.

p	.1	.5	100	1000	5000	9990	10000
$\max p - p^* $.00005	.00025	.05	.5	2.5	4.995	5.

علاوه بر نمایش نادقیق اعداد، حساب معمول در یک کامپیوتر نیز دقیق نیست. این حساب معمولاً عبارت از دستکاری ارقام دودویی با انتقالهای متعدد یا اعمال منطقی است. چون مکانیک واقعی این اعمال ربطی به این نمایش ندارد، تقریب مطابق حساب کامپیوتر طرح ریزی می‌شود. اگرچه این حساب تصویر دقیق را بدست نمی‌دهد، اما باید برای توضیح مسائلی که پیش می‌آیند کافی باشد. (برای شرح اعمال لازم، خواننده

می‌تواند به کتب علوم کامپیوتر، نظیر چو [۲۰] *، مراجعه کند.

فرض کنیم نمایش یا ممیز سیار $f(y)$ ، به‌ازای عدد حقیقی

$$y = \pm .d_1 d_2 \cdots d_k d_{k+1} \cdots \times 10^n,$$

چنین باشد:

$$f(y) = \pm .d_1 d_2 \cdots d_k \times 10^n$$

و علامات $\oplus, \ominus, \otimes, \oslash$ به ترتیب نمایش اعمال جمع، تفریق، ضرب، و تقسیم ماشین باشند.

فرض کنیم حساب k رقمی عبارت باشد از

$$x \oplus y = f(f(x) + f(y)),$$

$$x \ominus y = f(f(x) - f(y)),$$

$$x \otimes y = f(f(x) \times f(y)),$$

$$x \oslash y = f(f(x) \div f(y)).$$

این حساب فرضی متناظر حساب دقیق بر نمایشهای با ممیز سیار x و y و برگرداندن نتیجه دقیق به نمایش با ممیز سیار آن است. مثلاً، اگر $x = \frac{1}{3}$ و $y = \frac{2}{7}$ ، و $k = 5$ ، با استفاده از روش جدا کردن،

$$f(y) = .71428 \quad \text{و} \quad f(x) = .33333$$

در جدول ۱۰۱ نتایج محاسبات مختلف در حساب کامپیوتری ثبت شده‌اند.

عمل	نتیجه	مقدار واقعی	خطای نسبی
$x \oplus y$	$.10476 \times 10^1$	$\frac{27}{21}$	$.182 \times 10^{-4}$
$y \ominus x$.38095	$\frac{8}{21}$	$.625 \times 10^{-5}$
$x \otimes y$.23809	$\frac{2}{21}$	$.220 \times 10^{-4}$
$y \oslash x$	$.21428 \times 10^1$	$\frac{4}{3}$	$.267 \times 10^{-4}$

جدول ۱۰۱

تا پایان بحث، این حساب فرضی را حساب k رقمی می‌نامیم، و تصریح می‌کنیم که روش گرد کردن بکار رفته است یا جدا کردن.

مسئله دیگر در ارتباط با استفاده از حساب متناهی رقم مسئله از دست دادن

دقت است. با استفاده از حساب چهار رقمی، مجموع $x = \frac{1}{3}$ و $y = 1111$ را در نظر می‌گیریم. چون $f(x) = .3333$ و $f(y) = .1111 \times 10^4$ ،
 $f(f(x) + f(y)) = .11113333 \times 10^4$ ، اما، چون $f(x) + f(y) = .1111 \times 10^4$ ، در مقایسه با y ، عملاً "صفر گرفته می‌شود. این مسئله می‌تواند با تقسیم بر عدد کوچکی نظیر $z = 1000 \times 10^{-4}$ حاد شود، چرا که، در این صورت، به جای مقدار درست 11,113,333، خواهیم داشت

$$\frac{x + y}{z} = 11,110,000,$$

اما خطای نسبی محاسبات برای این حساب در حد معقول و تقریباً " 3×10^{-4} " است. از دست دادن ارقام با معنی در بعضی از اعمال حسابی نیز رخ می‌دهد، زیرا، همانطور که مثال زیر نشان می‌دهد، اعمال با تعداد ثابتی رقم انجام می‌شوند. (همچنین، ر.ک. تمرین ۰۸)

مثال ۲. فرض کنیم $p = .54617$ و $q = .54601$. اگر $r = p - q$ ، مقدار دقیق r مساوی 00016 است. فرض کنیم تفریق با حساب چهار رقمی صورت گیرد. با گرد کردن p و q تا چهار رقم، به ترتیب $p^* = .5462$ و $q^* = .5460$ ، و $r^* = p^* - q^* = .0002$ تقریب چهار رقمی به r می‌باشد. چون

$$\frac{|r - r^*|}{|r|} = .25,$$

نتیجه دارای یک رقم با معنی است، حال آنکه p^* و q^* به ترتیب دارای چهار و پنج رقم دقیق بودند.

اگر در بدست آوردن این چهار رقم جدا کردن ساده‌ای انجام گیرد، تقریبهای

چهار رقمی به p ، q ، و r به ترتیب $p^* = .5461$ ، $q^* = .5460$ ، و $r^* = p^* - q^* = .0001$ می‌باشند. در نتیجه،

$$\frac{|r - r^*|}{|r|} = .375,$$

که هنوز یک رقم با معنی دقت دارد.

از دست دادن ارقام با معنی اغلب با تنظیم مجدد مسئله قابل اجتناب است.

مثال ۳. با استفاده از حساب سمرقمی، $f(x) = x^3 - 6x^2 + 3x - 149$ را در $x = 4.71$ محاسبه کنید.

	x	x^2	x^3	$6x^2$	$3x$
دقیق	4.71	22.1841	104.487111	133.1046	14.13
سمرقمی (جدا کردن)	4.71	22.1	104.	132.	14.1
سمرقمی (گرد کردن)	4.71	22.2	105.	133.	14.1

توجه کنید که اولین مقادیر که با حساب 3 رقمی بدست آمده‌اند فقط سه رقم پیشرو را حفظ می‌کنند، بدون هیچ گرد کردن، و با مقادیر گرد شده، سمرقمی تفاوت آشکار دارند.

$$\begin{aligned} \text{دقیق} \quad f(4.71) &= 104.487111 - 133.1046 + 14.13 - 149 \\ &= -14.636489; \end{aligned}$$

$$\begin{aligned} \text{سمرقمی} \quad f(4.71) &= 104. - 132. + 14.1 - 149 \\ \text{(جدا کردن)} \quad &= -14.0; \end{aligned}$$

$$\begin{aligned} \text{سمرقمی} \quad f(4.71) &= 105. - 133. + 14.1 - 149 \\ \text{(گرد کردن)} \quad &= -14.0. \end{aligned}$$

خطای نسبی در هر دو روش سه رقمی برابر است با

$$\left| \frac{-14.636489 + 14.0}{14.636489} \right| \approx .04.$$

به روش دیگر، $f(x)$ را می‌توان به صورت زیر نوشت:

$$f(x) = x^3 - 6x^2 + 3x - .149$$

$$= ((x - 6)x + 3)x - .149,$$

که نتیجه می‌دهد

$$f(4.71) = ((4.71 - 6)4.71 + 3)4.71 - .149 = -14.5,$$

(جداکردن)

و جواب گرد شده^۳ 3 رقمی 14.6 - است. خطاهای نسبی جدید عبارت خواهند بود از

$$\text{سهرقمی (جداکردن)} \quad \left| \frac{-14.636489 + 14.5}{-14.636489} \right| \approx .0093$$

$$\text{سهرقمی (گردکردن)} \quad \left| \frac{-14.636489 + 14.6}{-14.636489} \right| \approx .0025.$$

با اینکه هر دو تقریب سهرقمی تا دو رقم با معنی درست‌اند، خطای نسبی در روش دوم تقریباً تا یک‌چهارم روش اول تقلیل یافته است. در روش گرد کردن سهرقمی، یک رقم با معنی اضافی بدست آمده است.

کاهش خطا ناشی از این امر است که تعداد محاسبات از چهار ضرب و سه جمع به دو ضرب و سه جمع تقلیل یافته است. واضح است که یک راه تقلیل خطا تقلیل تعداد محاسبات خطاها است.

برای ملاحظه^۴ اثر خطای گردکردن بر محاسبات متوالی، شامل محاسبه^۵ تابعی، فرض کنیم p^* تقریب p باشد. همچنین، فرض کنیم محاسبه با $g(p)$ نموده شده باشد، که، بخاطر این بررسی، g را بر بازه‌ای شامل p و p^* دوبار مشتق‌پذیر می‌گیریم. بنابراین قضیه^۶ تیلور (قضیه^{۱۳۰۱})،

$$g(p) = g(p^*) + g'(p^*)(p - p^*) + \frac{g''(\xi)}{2}(p - p^*)^2,$$

که در آن ξ بین p و p^* قرار دارد. اگر $(p - p^*)^2$ در مقایسه با $|p - p^*|$ کوچک باشد،

$$|g(p) - g(p^*)| \approx |g'(p^*)||p - p^*|.$$

از اینرو، خطای اصلی $|p - p^*|$ در عامل $|g'(p^*)|$ ضرب شده است.

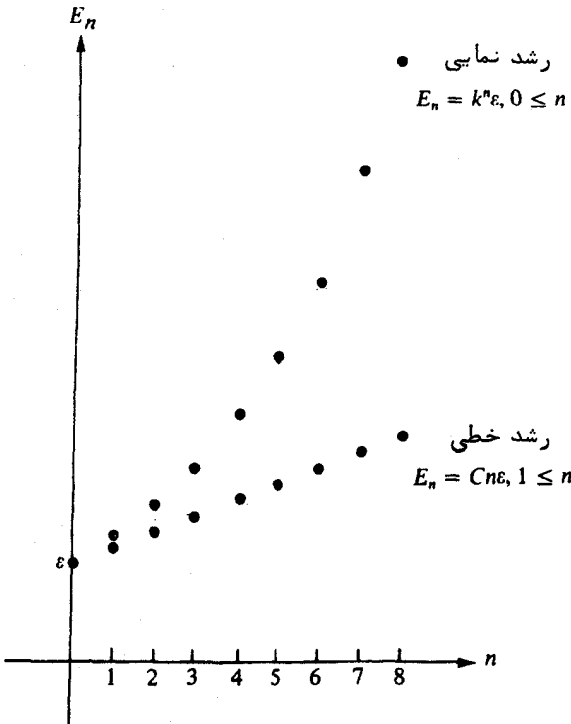
مثال ۴. فرض کنیم p^* یک تقریب p باشد و در محاسبه^۷ $g(p) = p^3 - 2$ بکار رفته باشد. طبق تذکرات فوق،

$$\frac{|g(p) - g(p^*)|}{|g(p)|} \approx \left| \left(\frac{3}{2} \right) (p^*)^{1/2} \right| \frac{|p - p^*|}{|p^{3/2}|} \approx \frac{3}{2} \frac{|p - p^*|}{|p|}$$

خطای نسبی در تقریب $p^{3/2}$ با $(p^*)^{3/2}$ حدوداً " $\frac{3}{2}$ خطای نسبی در تقریب p با p^* است .

محاسبات متوالی شامل تقریبها می‌توانند رشد یا انتشار خطاهای گرد کردن را نتیجه دهند . فرض کنیم خطا در مرحله‌ای از محاسبات ε بوده و پس از n عمل متوالی E_n باشد . در عمل اغلب دو حالت رخ می‌دهد ، که در زیر تعریف شده‌اند .

تعریف ۱۶.۱ . هرگاه $|E_n| \approx Cn\varepsilon$ ، که در آن C ثابتی مستقل از n است ، رشد خطا را خطی می‌نامیم . هرگاه ، به‌ازای $1 < k$ ای ، $|E_n| \approx k^n\varepsilon$ ، رشد خطا نمایی خوانده می‌شود . (ر.ک. شکل ۶.۱ .)



شکل ۶.۱

رشد خطی خطا، همانند در مثال ۴، معمولاً "غیرقابل اجتناب است؛ و، وقتی C و ε کوچک باشند، نتایج عموماً "قابل قبول می‌باشند. اما از رشد نمایی خطا بایستی پرهیز شود، چرا که جمله k^n ، حتی به‌ازای مقادیر نسبتاً "کوچک n ، بزرگ شده و، صرف‌نظر از اندازه ε ، به نادقتیهای غیرقابل قبول منجر می‌شود. به‌این دلیل، یک الگوریتم، و یا رشته‌ای از محاسبات، که رشد خطی خطا را نشان دهد **ناپایدار** نامیده می‌شود، درحالی که یک الگوریتم که رشد نمایی خطا را نشان دهد **ناپایدار** نام دارد.

مثال ۵. دنباله $\dots, \frac{1}{8}, \frac{1}{7}, \frac{1}{6}, \frac{1}{3}, 1$ از اعداد در نظر می‌گیریم. این دنباله را می‌توان با تعریف $p_0 = 1$ ، $p_1 = \frac{1}{3}$ ، و

$$(2.01) \quad p_n = \frac{10}{3} p_{n-1} - p_{n-2}$$

تولید کرد. اگر $\{p_n\}$ با حساب پنج رقمی و گرد کردن تولید شود، مقادیر جدول ۲.۱ بدست می‌آیند.

n	p_n محاسبه شده	مقدار دقیق p_n
0	1.0000	1.0000
1	.33333	.33333
2	.11110	.11111
3	$.37000 \times 10^{-1}$	$.37037 \times 10^{-1}$
4	$.12230 \times 10^{-1}$	$.12346 \times 10^{-1}$
5	$.37660 \times 10^{-2}$	$.41152 \times 10^{-2}$
6	$.32300 \times 10^{-3}$	$.13717 \times 10^{-2}$
7	$-.26893 \times 10^{-2}$	$.45725 \times 10^{-3}$
8	$-.92872 \times 10^{-2}$	$.15242 \times 10^{-3}$

جدول ۲.۱

$p_n = C_1(\frac{1}{3})^n + C_2(3)^n$ ، به‌ازای هر دو عدد حقیقی C_1 و C_2 ، در معادله (۲.۰۱) صدق می‌کند. برای آنکه $p_0 = 1$ و $p_1 = \frac{1}{3}$ ، باید $C_1 = 1$ و $C_2 = 0$ اختیار شوند. چون نمایش پنج رقمی عبارت است از $\hat{p}_0 = 1.0000$ و $\hat{p}_1 = .33333$ ، جواب تولید شده به وسیله معادله (۲.۰۱) عملاً "دارای $C_1 = 1.0000$ و $C_2 = -.12500 \times 10^{-5}$ است. بنابراین، خطای گرد کردن در نمایش $\frac{1}{3}$ مساوی $(-.125 \times 10^{-5})(3)^n$

است، که در هر مرحله به p_n افزوده می‌شود. از اینرو، این روند که با معادله (۲.۱) ارائه می‌شود رشد نمایی خطاهای گرد کردن را نشان می‌دهد، زیرا

$$|p_n - \hat{p}_n| = (3)(.125 \times 10^{-5})^n.$$

اگرچه تعیین خطای کل ناممکن است، می‌توان خطا را به‌طور آماری بررسی کرد. این بررسی از حوصله این کتاب خارج است، اما، در صورت امکان، خطای گرد کردن را تخمین خواهیم زد. می‌توان از حساب رقمی مرتبه بزرگ، نظیر حساب دو یا چند دقتی که در اغلب کامپیوترهای رقمی در دسترس است برای تقلیل اثرات خطای گرد کردن استفاده کرد. عیب استفاده از حساب دو دقتی این است که وقت بیشتری از ماشین را می‌گیرد. همچنین، رشد جدی خطای گرد کردن کاملاً "از بین نمی‌رود، بلکه، در صورت انجام تعداد زیادی محاسبه متوالی، فقط به تعویق می‌افتد.

راه دیگر تخمین خطای گرد کردن حساب بازه است؛ یعنی، بدست آوردن بزرگترین و کوچکترین مقدار ممکن در هر مرحله، بطوری که، در آخر بازه‌ای بدست می‌آید که شامل مقدار واقعی است. متأسفانه، جواب درست ممکن است نزدیک به مقادیر اکستریم این بازه باشد، و ممکن است برای اجرای مناسب این روش مجبور به تعیین بازه خیلی کوچکی شویم. ظاهراً "بررسی آماری بهترین روش است، اگرچه به تحلیل قابل ملاحظه‌ای نیاز دارد. هنریچی^۱ در [۳۴] بحثی از یک روش آماری برای تخمین خطای گرد کردن انباشته ارائه می‌دهد (ر.ک. هنریچی، صفحات ۳۰۵ تا ۳۰۹).

چون اغلب از روشهای تکراری شامل دنباله‌ها استفاده می‌شود، این بخش را با بحث کوتاهی از مفهوم مرتبه همگرایی به پایان می‌بریم.

تعریف ۱۷.۱. فرض کنیم $\{\alpha_n\}_{n=1}^{\infty}$ دنباله‌ای همگرا به عدد α باشد. گوئیم $\{\alpha_n\}_{n=1}^{\infty}$ به مرتبه همگرایی $O(\beta_n)$ به α همگرا است، که در آن $\{\beta_n\}_{n=1}^{\infty}$ دنباله‌ای دیگر با $\beta_n \neq 0$ به ازای هر n است، اگر

$$\frac{|\alpha - \alpha_n|}{|\beta_n|} \leq K, \text{ به ازای } n \text{ به قدر کافی بزرگ,}$$

که در آن K ثابتی است مستقل از n ؛ یا، معادلاً، اگر

$$\lim_{n \rightarrow \infty} \frac{|\alpha - \alpha_n|}{|\beta_n|} \leq K.$$

این حالت اغلب با نوشتن $\alpha_n = \alpha + O(\beta_n)$ نموده می‌شود.

مثال ۶. فرض کنیم به‌ازای $n \geq 1$ ،

$$\beta_n = \frac{1}{n} \quad \text{و} \quad \alpha_n = \frac{\sin n}{n}$$

در این صورت، $\{\alpha_n\}_{n=1}^{\infty} \rightarrow 0$ با مرتبه همگرایی $O(1/n)$ ، زیرا، به‌ازای هر n

$$\left| \frac{(\sin n)/n - 0}{1/n} \right| = |\sin n| \leq 1$$

این مفهوم را می‌توان به توابع نیز تعمیم داد:

تعریف ۱۸.۱. هرگاه $\lim_{h \rightarrow 0} F(h) = L$ ، گوئیم همگرایی $O(G(h))$ است اگر عددی مانند $K > 0$ ، مستقل از h ، وجود داشته باشد که به‌ازای آن

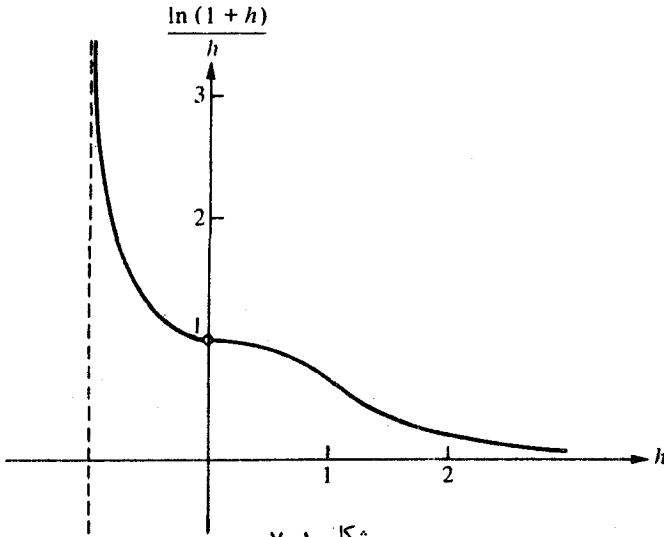
$$\lim_{h \rightarrow 0} \frac{|F(h) - L|}{|G(h)|} \leq K.$$

این مطلب اغلب با نوشتن $F(h) = L + O(G(h))$ نموده خواهد شد.

مثال ۷. بنا بر قاعده هوییتال^۱،

$$\lim_{h \rightarrow 0} \frac{\ln(1+h)}{h} = 1.$$

مرتبه همگرایی را می‌توان با استفاده از تعریف ۱۸.۱ و سری تیلور قضیه ۱۳.۱ معین کرد (ر. ک. شکل ۷.۱).



شکل ۷.۱

بمازای $f(x) = \ln(1+x)$

$$\begin{aligned} \ln(1+x) = f(x) &= f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \dots \\ &\quad + \frac{f^{(n)}(0)}{n!}x^n + \dots \\ &= x - \frac{x^2}{2} + \frac{x^3}{3} + \dots + (-1)^{n-1} \frac{x^n}{n} + \dots \end{aligned}$$

از اینرو،

$$\frac{\ln(1+h)}{h} = 1 - \frac{h}{2} + \frac{h^2}{3} + \dots + (-1)^{n-1} \frac{h^{n-1}}{n} + \dots$$

و چون

$$\lim_{h \rightarrow 0} \frac{|(\ln(1+h)/h) - 1|}{|h|} = \lim_{h \rightarrow 0} \left| -\frac{1}{2} + \frac{h}{3} + \dots + (-1)^{n-1} \frac{h^{n-2}}{n} + \dots \right| \leq \frac{1}{2},$$

داریم

$$\frac{\ln(1+h)}{h} = 1 + O(h)$$

با همگرایی $O(h)$ $\lim_{h \rightarrow 0} \frac{\ln(1+h)}{h} = 1$

قضیه ۱۹.۰۱. هرگاه $\alpha_n = \alpha + O(\beta_n)$ و $\alpha'_n = \alpha' + O(\beta'_n)$ ، که در آن $\beta'_n = O(\beta_n)$ ،
 آنگاه معادلات زیر برقرارند :

(۱) $\alpha_n \pm \alpha'_n = \alpha \pm \alpha' + O(\beta_n)$

(۲) $\alpha_n \alpha'_n = \alpha \alpha' + O(\beta_n)$

(۳) $\frac{\alpha_n}{\alpha'_n} = \frac{\alpha}{\alpha'} + O(\beta_n)$ ، مشروط بر اینکه $\alpha' \neq 0$.

برهان . ر.ک. تمرین ۵.

مجموعه تمرینات ۲.۰۱

۱. هرگاه x یک تقریب چهار رقمی به

π (۱) ؛ e (۲)

باشد ، با استفاده از جداول معمول و تعریف ۱۵.۰۱ ، کرانهای x را پیدا کنید .

۲. محاسبات زیر را (یک) دقیقاً " ؛ (دو) با استفاده از حساب سه رقمی (جدا کردن)

انجام دهید . بعد ، با این فرض که اعداد مفروض دقیق اند ، هرگونه از دست دادن ارقام با معنی را مشخص کنید :

(۱) $14.1 + .0981$ ؛ (۲) $.0218 \times 179$ ؛

(۳) $(164. + .913) - (143. + 21.0)$ ؛ (ت) $(913 - 21.0) + (164. - 143.)$.

۳. تمرین ۲ را با استفاده از حساب سه رقمی (گرد کردن) تکرار کنید .

۴. همه متعلمین حساب دیفرانسیل و انتگرال می دانند که

$$\lim_{h \rightarrow 0} \frac{\sin h}{h} = 1,$$

نشان دهید که

$$\frac{\sin h}{h} = 1 + O(h^2).$$

در نتیجه، همگرایی $O(h^2)$ است.

۵. قضیه ۱۹۰۱ را اثبات کنید.
۶. فرمول درجه دوم می گوید که ریشه های $ax^2 + bx + c = 0$ ، که در آن $a \neq 0$ ، عبارتند از

$$x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \quad \text{و} \quad x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

(T) نشان دهید که $x_1 + x_2 = -b/a$ و $x_1 \cdot x_2 = c/a$.

(ب) با استفاده از (T)، نشان دهید که شکل دیگر ریشه های $ax^2 + bx + c = 0$ عبارت است از

$$x_2 = \frac{-2c}{b - \sqrt{b^2 - 4ac}} \quad \text{و} \quad x_1 = \frac{-2c}{b + \sqrt{b^2 - 4ac}}$$

(پ) با استفاده از دو فرمول برای x_1 ، بزرگترین ریشه $x^2 + 62.10x + 1 = 0$ را به کمک حساب چهار رقمی (گرد کردن) محاسبه نمایید. مقدار صحیح x_1 تا هفت رقم با معنی برابر 0.01610723- است. تفاوت مقادیر x_1 را در این دو محاسبه چگونه توضیح می دهید؟

۷. چند جمله ای تیلور از درجه n برای $e^x = f(x)$ برابر $\sum_{i=0}^n (x^i/i!)$ است. با استفاده از چند جمله ای تیلور از درجه ۹، تقریبی برای e^{-5} به طرق زیر بیابید:

$$(T) \quad e^{-5} \approx \sum_{i=0}^9 \frac{(-5)^i}{i!} = \sum_{i=0}^9 \frac{(-1)^i 5^i}{i!}$$

$$(ب) \quad e^{-5} \approx 1/e^5 \approx \frac{1}{\sum_{i=0}^9 (5^i/i!)}$$

یک مقدار تقریبی e^{-5} که تا سه رقم درست است برابر 6.74×10^{-3} می باشد. کدام فرمول، (T) یا (ب)، بیشترین دقت را می دهد، و چرا؟

۸. معادلات زیر را در نظر بگیرید:

$$(T) \quad 31.69x + 14.31y = 45.00$$

$$(۲) \quad 13.11x + 5.89y = 19.00$$

جواب منحصر بفرد این دستگاه $x = 7.2$ و $y = -12.8$ است. روشی که در جبر مقدماتی برای حل این مسائل ارائه می شود اغلب این است که معادله (۱) را در

ضریب x از معادله^۲ و معادله^۲ را در ضریب x از معادله^۱ ضرب و، سپس، معادلات حاصل را از هم کم می‌کنند. در این مسئله، داریم

$$(3) \quad (13.11)(45.00) - (31.69)(19.00) = [(13.11)(14.31) - (31.69)(5.89)]y$$

(T) این عمل را با حساب چهاررقمی (جدا کردن) انجام داده، نتیجه را برای یافتن مقادیر چهاررقمی x و y بکار بندید.

(ب) توضیح دهید چرا جوابهای (T) با مقادیر واقعی x و y تفاوت قابل ملاحظه‌ای دارند.

۹. دنباله^۶ $1, \frac{1}{3}, \frac{1}{6}, \frac{1}{7}, \dots$ از اعداد، که در مثال ۵ در نظر گرفته شد، را می‌توان با تعریف $P_0 = 1$

$$P_n = \{P_{n-1}, n = 1, 2, \dots$$

تولید کرد.

(T) با استفاده از حساب پنج رقمی، P_n را، به‌ازای $n = 1, \dots, 8$ ، محاسبه کنید، و نتایج را با مقادیر حاصله در مثال ۵ مقایسه نمایید.

(ب) نشان دهید که این الگوریتم رشد خطی خطای گرد کردن را نشان می‌دهد.

حل معادلات یک متغیره^۲

رشد جمعیت‌های زیاد را می‌توان با این فرض که جمعیت به‌طور پیوسته با زمان و به میزانی متناسب با تعداد افراد در آن زمان رشد می‌کند، در فاصله زمانهای کوتاه مدلسازی کرد. اگر $N(t)$ تعداد افراد در لحظه t و λ ثابت میزان تولد باشد، جمعیت در معادله دیفرانسیل

$$\frac{dN(t)}{dt} = \lambda N(t)$$

صدق می‌کند. جواب این معادله $N(t) = N_0 e^{\lambda t}$ است، که در آن N_0 جمعیت اولیه می‌باشد.

این مدل فقط وقتی اعتبار دارد که جامعه از لحاظ مهاجرت ایزوله باشد؛ یعنی، هیچ مهاجرتی به آن صورت نگیرد. اگر مهاجرت به میزان ثابت v صورت گیرد، معادله دیفرانسیل این وضعیت به‌صورت

$$\frac{dN(t)}{dt} = \lambda N(t) + v$$

است، که جواب آن

$$N(t) = N_0 e^{\lambda t} + \frac{v}{\lambda} (e^{\lambda t} - 1).$$

می‌باشد.

فرض کنیم جمعیت اولیه جامعه‌ای یک میلیون باشد، در سال اول 435,000 نفر به آن مهاجرت کنند، و جمعیتش در آخر سال اول 1,564,000 باشد. برای تعیین

میزان تولد در این جامعه باید معادله

$$1,564,000 = 1,000,000 e^{\lambda} + \frac{435,000}{\lambda} (e^{\lambda} - 1).$$

را نسبت به λ حل کنیم. روشهای عددی این فصل برای یافتن تقریبهایی از جوابهای این نوع معادلات، وقتی جوابهای دقیق به روشهای جبری قابل حصول نیستند، بکار می‌روند. حل این مسئله خاص در تمرین ۱۸ در بخش ۴۰.۲ خواسته شده است.

در این فصل یکی از اساسی‌ترین مسائل آنالیز عددی مطرح می‌شود، و آن یافتن مقادیری از متغیر x است که در $f(x) = 0$ ، که در آن f یک تابع مفروض است، صدق می‌کنند. این مسئله یکی از قدیمی‌ترین مسائل تقریب عددی است، و هنوز تحقیق در این زمینه فعالانه ادامه دارد. روندهای مورد بحث از روش کلاسیک نیوتن ۱ - رفسون ۲، که اساساً "به‌وسیله" اسحق نیوتن حدود ۳۰۰ سال پیش ارائه شده، تا روش خارج قسمت - تفاضل برای توابع چندجمله‌ای که اخیراً "برای اولین بار چاپ شده‌اند، در تغییراند.

۱.۲ الگوریتم تنصیف

اولین روش، که بر قضیه مقسدار میانی (قضیه ۱۲.۱، ص ۹) استوار است، الگوریتم تنصیف، یا روش تحقیق دوپیی، نام دارد. فرض کنیم تابع پیوسته f بر بازه $[a, b]$ تعریف شده باشد، و $f(a)$ و $f(b)$ مختلف‌العلامه باشند. در این صورت، طبق قضیه ۱۲.۱، نقطه‌ای مانند p وجود دارد که $a < p < b$ و $f(p) = 0$. این روند برای حالت بیش از یک ریشه در بازه $[a, b]$ نیز کارا است، ولی برای سادگی فرض می‌کنیم ریشه موجود در این بازه منحصر بفرد باشد. این روش عبارت است از تکرار نصف کردن زیربازه‌های $[a, b]$ ، و در هر مرحله، تعیین "نیمه"‌ای که شامل p باشد. برای شروع، قرار می‌دهیم $a_1 = a$ و $b_1 = b$ ، و فرض می‌کنیم p_1 نقطه میانی $[a, b]$ باشد؛ یعنی،

$$p_1 = \frac{1}{2}(a_1 + b_1)$$

اگر $f(p_1) \neq 0$ ، $f(p_1)$ با $f(a_1)$ یا $f(b_1)$ همعلامت است . اگر $f(p_1)$ و $f(a_1)$ همعلامت باشند ، $p \in (p_1, b_1)$ و قرار می دهیم $a_2 = p_1$ و $b_2 = b_1$. اگر $f(p_1)$ و $f(b_1)$ همعلامت باشند ، $p \in (a_1, p_1)$ و قرار می دهیم $a_2 = a_1$ و $b_2 = p_1$. حال این عمل را درباره $[a_2, b_2]$ تکرار می کنیم . این کار روند زیر را نتیجه خواهد داد (ر.ک. شکل ۱۰۲) :

الگوریتم تنصیف ۱۰۲

تابع پیوسته f بر بازه $[a, b]$ مفروض است ، که در آن $f(a)$ و $f(b)$ مختلف‌العلامه‌اند . برای یافتن یک جواب $f(x) = 0$:

مرحله ۱ . قرار می دهیم $a_1 = a$ و $b_1 = b$.

مرحله ۲ . قرار می دهیم $i = 1$.

مرحله ۳ . قرار می دهیم $p_i = \frac{1}{2}(a_i + b_i)$.

مرحله ۴ . اگر p_i تقریب مناسبی باشد ، به مرحله ۱۰ می رویم .

اگر p_i تقریب مناسبی نباشد ، به مرحله ۵ می رویم .

مرحله ۵ . اگر $f(p_i)f(a_i) > 0$ ، به مرحله ۶ می رویم . اگر $f(p_i)f(a_i) < 0$ ، به مرحله ۸ می رویم .

مرحله ۶ . قرار می دهیم $a_{i+1} = p_i$ و $b_{i+1} = b_i$.

مرحله ۷ . به i یک می افزایشیم و به مرحله ۳ می رویم .

مرحله ۸ . قرار می دهیم $a_{i+1} = a_i$ و $b_{i+1} = p_i$.

مرحله ۹ . به i یک می افزایشیم و به مرحله ۳ می رویم .

مرحله ۱۰ . روند تمام خواهد بود .

حال به چند روند توقف که در مرحله ۴ الگوریتم اعمال می شوند و اغلبشان در

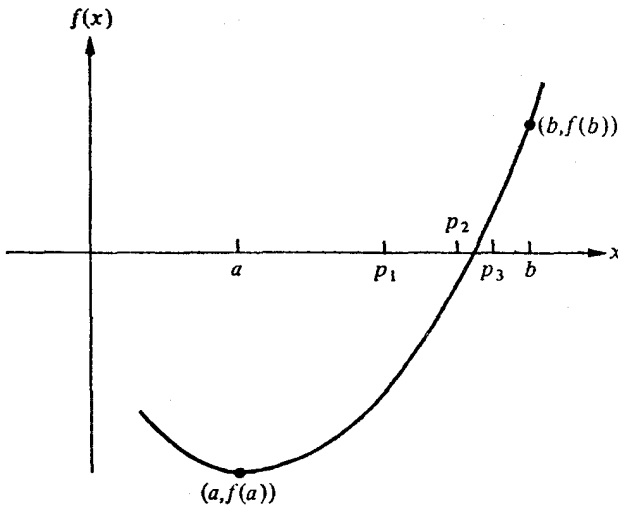
روشهای تکراری در این فصل بکار می روند مختصر اشاره ای می کنیم . $\epsilon > 0$ را

اختیار و p_1, \dots, p_N را تولید می کنیم تا یکی از شرایط زیر برقرار شود :

$$(۱۰۲) \quad |p_N - p_{N-1}| < \epsilon,$$

$$(۲۰۲) \quad \left| \frac{p_N - p_{N-1}}{p_N} \right| < \epsilon, \quad p_N \neq 0,$$

$$(۳۰۲) \quad |f(p_N)| < \epsilon.$$



شکل ۱۰۲

همچنین، در استفاده از یک کامپیوتر برای تولید تقریبها، افزودن شرطی که بر تعداد تکرارهای انجام شده کران بالایی قرار دهد کار خوبی است. این کار را می توان به آسانی با گذاردن کران اولیه‌ای چون N_0 و ختم روند در صورت

$$N > N_0$$

انجام داد.

توجه کنید که، برای شروع الگوریتم تنصیف، باید بازه‌ای مانند $[a, b]$ بیابیم که $f(a) \cdot f(b) < 0$. در هر مرحله از الگوریتم، طول بازه معلومی شامل صفر $f(x)$ نصف می شود؛ در نتیجه، اگر بازه اولیه $[a, b]$ ، شامل صفر، حتی المقدور کوچک اختیار شود مفید خواهد بود. به عنوان مثال، اگر $f(x) = 2x^3 - x^2 + x - 1$

$$f(0) \cdot f(1) < 0 \quad \text{و} \quad f(-4) \cdot f(4) < 0$$

در نتیجه، الگوریتم تنصیف را می توان بر هر یک از بازه‌های $[0, 1]$ و $[-4, 4]$ بکار برد. شروع الگوریتم تنصیف بر $[0, 1]$ به جای $[-4, 4]$ تعداد تکرارهای لازم برای دستیابی به هر دقت لازم را سمتا تقلیل می دهد.

برای توضیح الگوریتم تنصیف، مثال زیر را در نظر می گیریم. این مثال شامل روش

توقف خاصی نیست.

مثال ۱. تابع $f(x) = x^3 + 4x^2 - 10$ یک ریشه در $[1, 2]$ دارد، زیرا $f(1) = -5$ و $f(2) = 14$. به آسانی دیده می شود که فقط یک ریشه در $[1, 2]$ وجود دارد. الگوریتم تنصیف جدول ۱۰۲ را بدست می دهد.

n	a_n	b_n	p_n	$f(p_n)$
1	1.0	2.0	1.5	2.375
2	1.0	1.5	1.25	-1.79687
3	1.25	1.5	1.375	.16211
4	1.25	1.375	1.3125	-.84839
5	1.3125	1.375	1.34375	-.35098
6	1.34375	1.375	1.359375	-.09641
7	1.359375	1.375	1.3671875	.03236
8	1.359375	1.3671875	1.36328125	-.03215
9	1.36328125	1.3671875	1.365234375	.000072
10	1.36328125	1.365234375	1.364257813	-.01605
11	1.364257813	1.365234375	1.364746094	-.00799
12	1.364746094	1.365234375	1.364990235	-.00396

جدول ۱۰۲

پس از 12 تکرار، می بینیم که $p_{12} = 1.364990235$ ریشه p را با خطای

$$|p - p_{12}| \leq |b_{12} - a_{12}| = .000488281$$

تقریب می کند، و چون

$$\frac{|b_{12} - a_{12}|}{|a_{12}|} \approx 3.6 \times 10^{-4},$$

این تقریب دست کم تا چهار رقم با معنی درست است. مقدار درست p ، تا هشت رقم اعشار، $p = 1.36523001$ است. جالب است توجه شود که p_0 از تقریب نهایی p_{12} به p نزدیکتر است، اما راهی، جز دانستن جواب واقعی، برای تعیین آن وجود ندارد.^۱

همانطور که در مثال دیده شد، الگوریتم تنصیف، گرچه از حیث معنی روشن است، نواقص آشکاری دارد. یکی اینکه در همگرایی خیلی کند است (یعنی، پیش از آنکه $|p - p_N|$

۱. از اعداد زیر $f(p_n)$ معلوم می شود که این تقریب از تقریب نهایی بهتر است م.

به قدر کافی کوچک شود، N ممکن است کاملاً "بزرگ گردد"، و، دیگر اینکه، ممکن است یک تقریب میانی مناسب با بی توجهی حذف شود. بهر حال، این روش این خاصیت مهم را دارد که همواره به یک جواب همگراست، و به این دلیل اغلب به عنوان "چاشنی" روشهای کارتر که بعدها در این فصل ارائه می شوند بکار می رود.

قضیه ۱۰۲. فرض کنیم $f \in C[a, b]$ و $f(a) \cdot f(b) < 0$. روند تنصیف (الگوریتم ۱۰۲) دنباله $\{p_n\}$ از تقریبهای p را با خاصیت زیر تولید می کند:

$$(۴۰۲) \quad |p_n - p| \leq \frac{b-a}{2^n}, \quad n \geq 1.$$

برهان. واضح است که، به ازای هر $n \geq 1$ ، داریم

$$p \in (a_n, b_n) \quad \text{و} \quad b_n - a_n = \frac{1}{2^{n-1}}(b-a)$$

چون به ازای هر $n \geq 1$ ، $p_n = \frac{1}{2}(a_n + b_n)$ ، نتیجه می شود که

$$|p_n - p| \leq \frac{1}{2}(b_n - a_n) = 2^{-n}(b-a).$$

بنابراین تعریف ۱۷۰۱ (ص ۲۲)، نامساوی (۴۰۲) ایجاب می کند که $\{p_n\}_{n=1}^{\infty}$ ، با آهنگ همگرایی $O(2^{-n})$ ، به p همگرا باشد. به ازای $\varepsilon > 0$ معلوم، می توان p_n را به تکرار تعیین کرد تا اینکه

$$2^{-n}(b-a) < \varepsilon,$$

و، در این صورت، $|p_n - p| < \varepsilon$. به شکل دیگر، تعداد تکرارهای لازم N برای رسیدن به خطای کمتر از ε را می توان تخمین زد، یعنی، لازم است که

$$2^{-N} < \frac{\varepsilon}{(b-a)} \quad \text{یا} \quad 2^{-N}(b-a) < \varepsilon$$

این، با استفاده از لگاریتم، نتیجه می دهد که

$$-N \ln 2 < \ln \varepsilon - \ln(b-a),$$

یا

$$N > \frac{\ln(b-a) - \ln \varepsilon}{\ln 2}$$

بنابراین، در استفاده از قضیه ۱۰۲، تعداد $\lceil (\ln(b-a) - \ln \varepsilon) / \ln 2 \rceil$ تکرار لازم است

تا نامساوی $|p_n - p| < \varepsilon$ برقرار شود.

توجه کنید که هرچند در محاسبه این کران از لگاریتم طبیعی استفاده شده است، در عمل، همانطور که مثال بعدی روشن می‌کند، اغلب استفاده از لگاریتم در پایه ۱۰ در محاسبه راحتتر است. با استفاده از لگاریتم در پایه ۱۰، فرمول تعداد تکرارهای لازم برای حصول دقت به میزان ε عبارت است از

$$N > \frac{\log_{10}(b - a) - \log_{10} \varepsilon}{\log_{10} 2}$$

مثال ۲. تعداد تکرارهای لازم برای حل $f(x) = x^3 + 4x^2 - 10 = 0$ را با دقتی برابر $\varepsilon = 10^{-5}$ به ازای $a_1 = 1$ و $b_1 = 2$ به طور تقریبی معین می‌کنیم. برای این کار باید عدد صحیح N را طوری بیابیم که در روابط زیر صدق کند:

$$|p_N - p| \leq 2^{-N}(b - a) = 2^{-N} < 10^{-5}$$

با استفاده از لگاریتم در پایه ۱۰،

$$-N \log_{10} 2 < -5$$

یا

$$N > \frac{5}{\log_{10} 2} \approx 16.6$$

برای داشتن تقریبی که تا 10^{-5} دقیق باشد ۱۷ تکرار لازم است. به ازای $\varepsilon = 10^{-3}$ ، $N \geq 10$ تکرار لازم است و مقدار $p_9 = 1.36523475$ تا 10^{-4} دقیق می‌باشد. مهم است توجه شود که این روشها فقط یک کران برای تعداد تکرارهای لازم بدست می‌دهند، و در بسیاری از حالات این کران از عدد واقعی مطلوب خیلی بزرگتر است.

مجموعه تمرینات ۱.۲

۱. نشان دهید که $f(x) = x^3 - x - 1$ فقط یک ریشه در بازه $[1, 2]$ دارد.
۲. با استفاده از الگوریتم تنصیف، جوابهای مسائل زیر را با دقت 10^{-5} بیابید:

(ا) به ازای $0 \leq x \leq 1$ ، $f(x) = x - 2^{-x} = 0$ ؛

(ب) به ازای $1 \leq x \leq 2$ ، $f(x) = e^x + 2^{-x} + 2 \cos x - 6 = 0$ ؛

(پ) به ازای $0 \leq x \leq 1$ ، $f(x) = e^x - x^2 + 3x - 2 = 0$ ؛

- ۳ . با استفاده از الگوریتم تنصیف، برای $\sqrt[3]{25}$ تقریبی بیابید که تا دو رقم با معنی دقیق باشد.
- ۴ . با استفاده از الگوریتم تنصیف، برای $\sqrt{3}$ تقریبی بیابید که تا چهار رقم با معنی دقیق باشد.
- ۵ . با استفاده از قضیه ۱۰.۲، برای تعداد تکرارهای لازم جهت رسیدن به یک تقریب جواب $x^3 - x - 1 = 0$ در بازه $[1, 2]$ با دقت 10^{-4} یک کران بدست آورید. برای این ریشه، یک تقریب با این درجه از دقت پیدا نمایید.
- ۶ . با استفاده از قضیه ۱۰.۲، برای تعداد تکرارهای لازم جهت رسیدن به یک تقریب جواب $x^3 + x - 4 = 0$ در بازه $[1, 4]$ با دقت 10^{-3} یک کران بدست آورید. برای این ریشه، یک تقریب با این درجه از دقت پیدا نمایید.
- ۷ . با استفاده از بازه‌های $[1.2, 2.2]$ و $[1.5, 2.5]$ ، الگوریتم تنصیف را بر معادله

$$f(x) = \frac{4x - 7}{(x - 2)^2} = 0$$

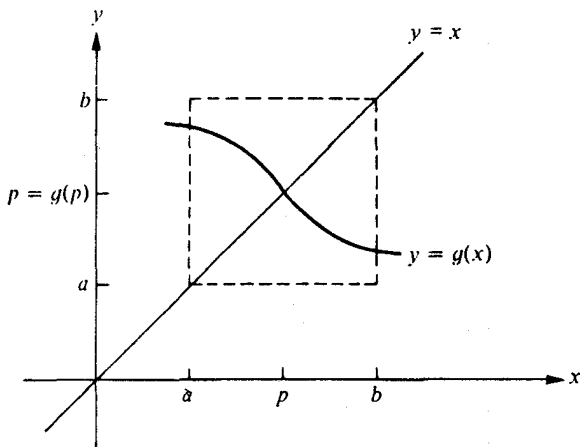
- اعمال کنید. نتایج را با شکل توضیح دهید.
- ۸ . تمام ریشه‌های $f(x) = x^2 + 10 \cos x$ را تا چهار رقم با معنی پیدا کنید. [راهنمایی. نمودار f را در نظر بگیرید.]

۲.۲ نقاط ثابت توابع

روشهای این بخش و کلاً "تا آخر این فصل بر تغییر معادله $f(x) = 0$ به $g(x) = x$ استوارند. برای این کار راههای متعددی وجود دارند. یک روند، تعریف $g(x)$ به صورت $x - f(x)$ است. در این حالت، \bar{x} در $f(\bar{x}) = 0$ صدق می‌کند اگر و فقط اگر $g(\bar{x}) = \bar{x}$. در نتیجه، اگر بتوان برای حل مسئله‌ای به شکل $g(x) = x$ یک روش کلی یافت، با تعریف تابع $g(x) = x - f(x)$ می‌توان جوابی از $f(x) = 0$ را بدست آورد. همانطور که بعدها در این فصل خواهید دید، با انتخاب g های دیگر تعیین جوابی برای $f(x) = 0$ امکان پذیر است، و بسیاری از آنها از انتخاب فوق مناسبتر می‌باشند.

تعریف ۲.۲ اگر g بر $[a, b]$ تعریف شده باشد و، به ازای $p \in [a, b]$ ، $g(p) = p$ ، گوئیم تابع g در $[a, b]$ دارای نقطه ثابت p است.

محک ساده‌ای برای وجود نقطه ثابت وجود دارد که بر قضیه مقدار میانی (قضیه ۱۲.۱، ص ۹) استوار است. (ر.ک. شکل ۲.۲).



شکل ۲.۲

لم ۳.۲. هرگاه $g \in C[a, b]$ و به ازای هر $x \in [a, b]$ ، $g(x) \in [a, b]$ ، آنگاه تابع g در $[a, b]$ یک نقطه ثابت خواهد داشت.

برهان. هرگاه $g(a) = a$ یا $g(b) = b$ ، وجود یک نقطه ثابت واضح است. فرض کنیم چنین نباشد؛ در این صورت، باید $g(a) > a$ و $g(b) < b$ ، تعریف می‌کنیم $h(x) = g(x) - x$ ؛ h بر $[a, b]$ پیوسته است، و، علاوه،

$$h(a) = g(a) - a > 0,$$

$$h(b) = g(b) - b < 0.$$

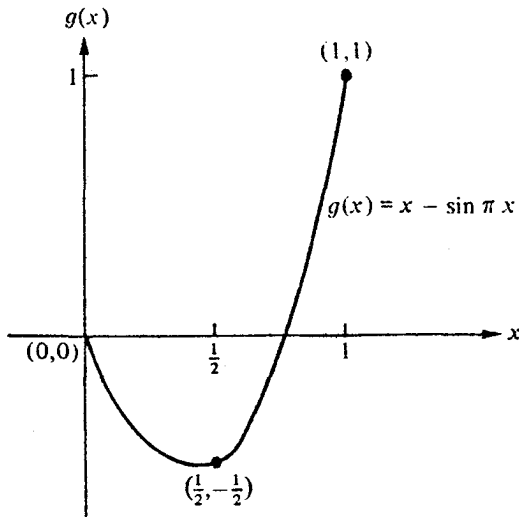
قضیه مقدار میانی ایجاب می‌کند که نقطه‌ای مانند $p \in (a, b)$ وجود داشته باشد که به ازای آن $h(p) = 0$. بنابراین، $g(p) - p = 0$ و p یک نقطه ثابت g می‌باشد.

مثال ۱

(آ) تابع $g(x) = x$ ، $0 \leq x \leq 1$ در هر نقطه x از $[0, 1]$ نقطه ثابت دارد.

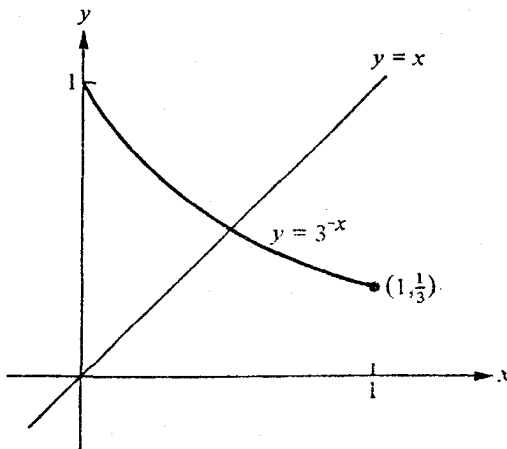
(ب) تابع $g(x) = x - \sin \pi x$ ، $0 \leq x \leq 1$ ، با اینکه در مفروضات لم ۳.۲ صدق نمی‌کند،

درست دو نقطه ثابت دارد، $x = 0$ و $x = 1$. (ر.ک. شکل ۳۰۲ .)



شکل ۳۰۲

(پ) تابع $g(x) = 3^{-x}$ ، $0 \leq x \leq 1$ ، که بوضوح پیوسته است، را در نظر می‌گیریم. چون $g'(x) = -3^{-x} \ln 3 < 0$ بر $[0, 1]$ ، g بر $[0, 1]$ نزولی است. از اینرو، به ازای $0 \leq x \leq 1$ ، در نتیجه، به ازای $x \in [0, 1]$ نیز $g(x) \in [0, 1]$ ، $g(1) = \frac{1}{3} \leq g(x) \leq 1 = g(0)$.



شکل ۴۰۲

(ر.ک. شکل ۴.۲) پس، بنا برلم ۳.۲، g باید در $[0, 1]$ یک نقطه ثابت داشته باشد.

جالب است بدانیم که نقطه ثابت چه وقت منحصر بفرد است، چرا که اگر g بر بازه $[a, b]$ بیش از یک نقطه ثابت داشته باشد، ممکن است بعضی روندها بین آنها نوسان کنند، یا، حتی اگر همگرایی دست دهد، ممکن است به مقدار مطلوب نباشد. قضیه زیر شرایط کافی برای وجود و یکتایی نقطه ثابت را بدست می دهد.

قضیه ۴.۲. فرض کنیم $g \in C[a, b]$ و، به ازای هر $x \in [a, b]$ ، $g(x) \in [a, b]$ ، همچنین، $g'(x)$ بر (a, b) موجود بوده و،

به ازای هر $x \in [a, b]$ ، $|g'(x)| \leq k < 1$ ، (۵.۲)
در این صورت، g در $[a, b]$ نقطه ثابت منحصر بفردی مانند p دارد.

برهان. بنا برلم ۳.۲، دست کم یک نقطه ثابت در $[a, b]$ وجود دارد. فرض کنیم p و q دو نقطه ثابت در $[a, b]$ بوده و $p \neq q$. طبق قضیه مقدار میانگین (قضیه ۸.۱، ص ۵)، نقطه‌ای مانند ξ در $[a, b]$ وجود دارد بطوری که

$$|p - q| = |g(p) - g(q)| = |g'(\xi)| |p - q| \leq k |p - q| < |p - q|,$$

که یک تناقض است. بنابراین، بنا بر این، $p \neq q$ و نقطه ثابت در $[a, b]$ منحصر بفرد است.

مثال ۲. فرض کنیم $g(x) = (x^2 - 1)/3$ بر $[-1, 1]$. با استفاده از قضیه مقدار اکستریم (قضیه ۹.۱، ص ۶)، به آسانی می توان نشان داد که مینیمم مطلق g در $x = 0$ روی می دهد و $g(0) = -\frac{1}{3}$. به همین نحو، ماکزیمم مطلق g در $x = \pm 1$ رخ می دهد که دارای مقدار $g(\pm 1) = 0$ است. بعلاوه، g پیوسته است و،

$$|g'(x)| = \left| \frac{2x}{3} \right| \leq \frac{2}{3}, \quad x \in [-1, 1]$$

در نتیجه، g در مفروضات قضیه ۴.۲ صدق می کند و در $[-1, 1]$ دارای نقطه ثابت منحصر بفرد است.

مثال ۳

(T) فرض کنیم p نقطه ثابت و منحصر بفرد مثال قبل در بازه $[-1, 1]$ باشد. در این صورت،

$$p^2 - 3p - 1 = 0 \quad \text{و} \quad p = g(p) = \frac{p^2 - 1}{3}$$

که، بنا بر فرمول معادله درجه دوم، نتیجه می دهد که

$$p = \frac{3 - \sqrt{13}}{2}$$

توجه کنید که g در بازه $[3, 4]$ نقطه ثابت منحصر بفرد دیگر $p = (3 + \sqrt{13})/2$ را دارد. اما $g(4) = 5$ و $g'(4) = \frac{8}{3} > 1$ ؛ در نتیجه، g در مفروضات قضیه ۴۰۲ صدق نمی کند. این امر نشان می دهد که مفروضات قضیه ۴۰۲ برای وجود نقطه ثابت منحصر بفرد کافی اند، ولی لازم نمی باشند.

(ب) فرض کنیم $g(x) = 3^{-x}$. بنا بر مثال ۱ (پ)، a در $[0, 1]$ یک نقطه ثابت دارد. اما

$$g'(0) = -\ln 3 = -1.098612289,$$

در نتیجه، $1 \not\leq |g'(x)|$ بر $[0, 1]$ ؛ و قضیه ۴۰۲ را نمی توان برای تعیین یکتایی بکاربرد. این مشکل را می توان با در نظر گرفتن بازه $[\frac{1}{3}, 1]$ به عنوان قلمرو g رفع کرد. چون g ، $[0, 1]$ را بتوی $[\frac{1}{3}, 1]$ می نگارد، واضح است که g ، $[\frac{1}{3}, 1]$ را بتوی خودش می نگارد و، بر $[\frac{1}{3}, 1]$

$$|g'(x)| \leq |g'(\frac{1}{3})| = 3^{-1/3} \ln 3 \leq .76174 < 1.$$

پس، بنا بر قضیه ۴۰۲، g بر $[\frac{1}{3}, 1]$ نقطه ثابت منحصر بفرد دارد.

مجموعه تمرینات ۲۰۲

۱. نشان دهید که $f(x) = 2^{-x}$ بر $[1, 1]$ نقطه ثابت منحصر بفرد دارد.
۲. فرض کنید $g(x) = \pi + .5 \sin x$. نشان دهید که g بر $[0, 2\pi]$ نقطه ثابت منحصر بفرد دارد.
۳. فرض کنید هدف یافتن یک جواب $f(x) = 0$ بر $[0, 2]$ باشد، که در آن $f(x) = x^2 - x - 1$ نشان دهید که یک نقطه ثابت هریک از توابع زیر یک جواب $f(x) = 0$ است:

(T) $g(x) = x^2 - 1$ ؛ (ب) $g(x) = (x + 1)^{1/2}$ ؛

$$p) \text{ به ازای هر ثابت } m, g(x) = x - \frac{x^2 - x - 1}{m}$$

به ازای چه g هایی می توان وجود نقطه ثابت منحصر بفرد در $[0, 2]$ و، در نتیجه، جواب منحصر بفرد برای $f(x) = 0$ بر $[0, 2]$ را تضمین کرد؟
۴. نشان دهید که

$$g(x) = \frac{2 - e^x + x^2}{3}$$

در $[0, 1]$ نقطه ثابت منحصر بفرد دارد.

۵. فرض برقراری معادله (5.2) در قضیه 4.2 را با این فرض که " g بر $[a, b]$ در شرط لیب شیتس، و بسا ثابت لیب شیتس $L < 1$ ، صدق می کند" عوض کنید. (ر.ک. تمرین ۷، بخش ۱.۱) نشان دهید که قضیه همچنان برقرار است.

۳.۲ تکرار نقطه ثابت

بخش پیش به بررسی وجود و یکتایی جوابهای مسائلی به شکل $g(x) = x$ ، که ممکن است تبدیلات معادله $f(x) = 0$ باشند، اختصاص یافته بود. در این بخش، روشهای تکراری برای حل معادله $g(x) = x$ را ارائه خواهیم داد.

برای حل $g(x) = x$ ، تقریب اولیه p_0 را اختیار و دنباله $\{p_n\}_{n=0}^{\infty}$ را با فرض $p_n = g(p_{n-1})$ به ازای هر $n \geq 1$ تولید می کنیم. اگر این دنباله به p همگرا بوده و g پیوسته باشد، بنا بر قضیه ۴.۱ (ص ۴)،

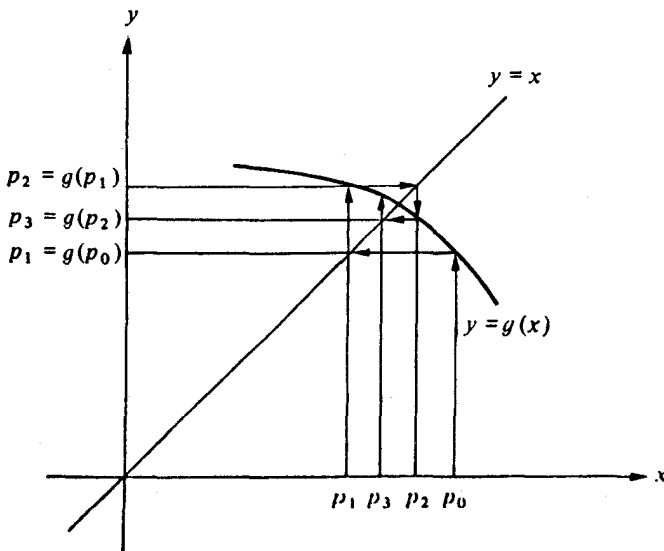
$$p = \lim_{n \rightarrow \infty} p_n = \lim_{n \rightarrow \infty} g(p_{n-1}) = g\left(\lim_{n \rightarrow \infty} p_{n-1}\right) = g(p).$$

و یک جواب $x = g(x)$ بدست می آید. این روش، روش تکرار نقطه ثابت، یا تکرار تابعی، نام دارد، و به طور صوری با الگوریتم زیر داده می شود. روند کار در شکل ۵.۲ شرح داده شده است.

الگوریتم نقطه ثابت ۲.۲

با معلوم بودن معادله $p = g(p)$ و مقدار اولیه p_0 ، مراحل زیر را انجام می دهیم:
مرحله ۱. قرار می دهیم $i = 1$.

- مرحله^۲ . قرار می‌دهیم $p_i = g(p_{i-1})$.
- مرحله^۳ . مشخص می‌کنیم که آیا روند باید ادامه یابد .
- اگر باید ، به مرحله^۴ می‌رویم .
- در غیر این صورت ، به مرحله^۵ می‌رویم .
- مرحله^۴ . به i عدد ۱ را افزوده به مرحله^۲ می‌رویم .
- مرحله^۵ . روند تمام خواهد بود .



شکل ۵-۲

عوامل تشخیص در مرحله^۳، درباب ادامه^۲ کار، ممکن است تابع تعداد تکرار i ، که قبلاً انجام شده، مقدار $|p_i - p_{i-1}|$ ، مقدار $|p_i - p_{i-1}|/|p_i|$ ، و یا ترکیبی از این عوامل باشد.

برای توضیح این الگوریتم، مثال زیر را در نظر می‌گیریم.

مثال ۱. معادله $x^3 + 4x^2 - 10 = 0$ در $[1, 2]$ ریشه^۲ منحصر بفرد دارد. برای تغییر این معادله به شکل $x = g(x)$ راههای متعددی وجود دارند، که از آنجمله‌اند:

$$x = g_1(x) = x - x^3 - 4x^2 - 10 \quad (T)$$

$$: x = g_2(x) = \left(\frac{10}{x} - 4x\right)^{1/2} \quad (ب)$$

$$: x = g_3(x) = \frac{1}{2}(10 - x^3)^{1/2} \quad (پ)$$

$$: x = g_4(x) = \left(\frac{10}{4+x}\right)^{1/2} \quad (ت)$$

$$: x = g_5(x) = x - \frac{x^3 + 4x^2 - 10}{3x^2 + 8x} \quad (ث)$$

$$: x = g_6(x) = x - \frac{x^3 + 4x^2 - 10}{16.51339907} \quad (ج)$$

نتایج روش تکرار نقطه ثابت، به ازای $p_0 = 1.5$ ، برای هر شش انتخاب g در جدول ۲۰۲ ثبت شده است.

n	(آ)	(ب)	(پ)	(ت)	(ث)	(ج)
0	1.5	1.5	1.5	1.5	1.5	1.5
1	- .875	.8165	1.28695377	1.34839973	1.37333333	1.3561774
2	6.732	2.9969	1.40254080	1.36737637	1.36526201	1.36518988
3	-469.7	$(-8.65)^{1/2}$	1.34545838	1.36495701	1.36523001	1.36523001
4	1.03×10^8		1.37517025	1.36526475		
5			1.36009419	1.36522559		
6			1.36784697	1.36523058		
7			1.36388700	1.36522994		
8			1.36591673	1.36523002		
9			1.36487822	1.36523001		
10			1.36541006			
15			1.36522368			
20			1.36523024			
23			1.36522998			
25			1.36523001			

جدول ۲۰۲

همانطور که در مثال ۱ از بخش ۱۰۲ ذکر شد، ریشه واقعی برابر 1.36523001 است. با مقایسه این نتایج با نتایج الگوریتم تنصیف در آن مثال، معلوم می شود که نتایج عالی در (پ)، (ت)، (ث) و (ج) بدست آمده اند، در حالی که روش تنصیف به 27 تکرار

برای این دقت نیاز دارد. جالب است توجه شود که انتخاب (T) به‌واگرایی، و انتخاب (ب) به ابهام منجر می‌شود، زیرا مستلزم ریشه دوم یک عدد منفی است.

مثال قبل‌نیاز به روندی را نشان می‌دهد که در آن g همگرایی جوابی از $x = g(x)$ را تضمین کند و نیز g به گونه‌ای اختیار شود که این همگرایی حتی‌المقدور سریع باشد. قضیه زیر نخستین گام در تعیین این روند است.

قضیه ۵.۲. فرض کنیم $g \in C[a, b]$ و، به‌ازای هر $x \in [a, b]$ ، $g(x) \in [a, b]$ ، بعلاوه، g' بر (a, b) وجود داشته باشد بطوری که،

$$(۶.۲) \quad |g'(x)| \leq k < 1, \quad x \in (a, b)$$

در این صورت، اگر p_0 نقطه‌ای در $[a, b]$ باشد، دنباله تعریف شده با

$$p_n = g(p_{n-1}), \quad n \geq 1,$$

به نقطه ثابت منحصربفردی چون p در $[a, b]$ همگرا می‌باشد.

برهان. بنا بر قضیه ۴.۲، نقطه ثابت منحصربفردی در $[a, b]$ وجود دارد. چون g ، $[a, b]$ را بتوی خودش می‌نگارد، دنباله $\{p_n\}_{n=0}^{\infty}$ به‌ازای هر $n \geq 0$ تعریف شده است و، به‌ازای هر n ، $p_n \in [a, b]$. بنا بر (۶.۲)، قضیه مقدار میانگین، و تعریف p_n ،

$$(۷.۲) \quad \begin{aligned} |p_n - p| &= |g(p_{n-1}) - g(p)| \\ &\leq |g'(\xi)| |p_{n-1} - p| \leq k |p_{n-1} - p|. \end{aligned}$$

که در آن $\xi \in [a, b]$. با استفاده از معادله (۷.۲) به استقرا، نتیجه می‌شود که

$$(۸.۲) \quad |p_n - p| \leq k |p_{n-1} - p| \leq k^2 |p_{n-2} - p| \leq \dots \leq k^n |p_0 - p|.$$

چون $k < 1$

$$\lim_{n \rightarrow \infty} |p_n - p| \leq \lim_{n \rightarrow \infty} k^n |p_0 - p| = 0$$

و $\{p_n\}_{n=0}^{\infty}$ به p همگرا می‌باشد.

در پرتو قضیه پیش، مثال ۱ را مجدداً " در نظر می‌گیریم :

مثال ۲

(آ) وقتی $g_1(x) = x - x^3 - 4x^2 + 10$ ، $g_1'(x) = 1 - 3x^2 - 8x$ ، بازه‌ای مانند $[a, b]$ ، شامل p ، وجود ندارد که در آن $|g_1'(x)| < 1$. با آنکه قضیه ۵.۲ عدم کارآیی روشن فوق را برای این g تضمین نمی‌کند ، دلیلی نیز برای تردید در همگرایی وجود ندارد .

(ب) به‌ازای $g_2(x) = ((10/x) - 4x)^{1/2}$ ، می‌توان دید که g_2 ، $[1, 2]$ را بتوی $[1, 2]$ نمی‌نگارد و دنباله $\{p_n\}_{n=0}^{\infty}$ به‌ازای $p_0 = 1.5$ تعریف نشده است . بعلاوه ، بازه‌ای شامل p وجود ندارد که در آن

$$|g_2'(x)| < 1 \text{ ، زیرا } |g_2'(p_0)| \approx 3.4$$

(پ) درمورد تابع $g_3(x) = \frac{1}{2}(10 - x^3)^{1/2}$ ،

$$g_3'(x) = -\frac{3}{4}x^2(10 - x^3)^{-1/2} < 0 \text{ بر } [1, 2] ؛$$

در نتیجه ، g بر $[1, 2]$ اکیدا "نزولی" است . اما $|g_3'(2)| \approx 2.12$ ؛ در نتیجه ، معادله (۶.۲) بر $[1, 2]$ برقرار نیست . بررسی دقیقتر دنباله $\{p_n\}_{n=0}^{\infty}$ که با $p_0 = 1.5$ شروع می‌شود نشان می‌دهد که کافی است به‌جای $[1, 2]$ بازه $[1, 1.5]$ را در نظر بگیریم . برای بازه هنوز داریم $g_3(x) < 0$ و g اکیدا "نزولی" است ، اما ، علاوه بر این ، به‌ازای هر $x \in [1, 1.5]$ ، $1 < 1.28 \approx g_3(1.5) \leq g_3(x) \leq g_3(1) = 1.5$ ، که نشان می‌دهد g_3 بازه $[1, 1.5]$ را بتوی خودش می‌نگارد . چون برای بازه نیز داریم $|g_3'(x)| \leq |g_3'(1.5)| \approx 0.66$ ، قضیه ۵.۲ همگرایی را تایید می‌کند که قبلاً "از آن مطلع بودیم . قسمتهای دیگر مثال ۱ را می‌توان به همین نحو بررسی کرد .

همانطور که نتایج زیر نشان می‌دهند ، با استفاده از قضیه ۵.۲ می‌توان تخمینهای دقیقی از خطاها بدست آورد .

نتیجه ۶.۲ . اگر g در مفروضات قضیه ۵.۲ صدق کند ، یک کران برای خطای حاصل از بگار بردن p_n به‌عنوان تقریب p از رابطه زیر بدست می‌آید :

$$|p_n - p| \leq k^n \max\{p_0 - a, b - p_0\} \text{ ، } n \geq 1 \text{ (۹.۲)}$$

برهان . از (۸.۲) نتیجه می‌شود که

$$|p_n - p| \leq k^n |p_0 - p| \leq k^n \max\{p_0 - a, b - p_0\},$$

زیرا $p \in [a, b]$.

نتیجه ۷.۲. هرگاه y در مفروضات قضیه ۵.۲ صدق کند، آنگاه

$$(10.2) \quad |p_n - p| \leq \frac{k^n}{1-k} |p_0 - p_1|, \quad n \geq 1$$

برهان. به ازای $n \geq 1$ ، روند بکار رفته در برهان قضیه ۵.۲ ایجاب می کند که

$$\begin{aligned} |p_{n+1} - p_n| &= |g(p_n) - g(p_{n-1})| \\ &\leq k |p_n - p_{n-1}| \\ &\vdots \\ &\leq k^n |p_1 - p_0|. \end{aligned}$$

بنابراین، به ازای $n \geq 1$ ،

$$\begin{aligned} |p_m - p_n| &= |p_m - p_{m-1} + p_{m-1} - \dots + p_{n+1} - p_n| \\ &\leq |p_m - p_{m-1}| + |p_{m-1} - p_{m-2}| + \dots + |p_{n+1} - p_n| \\ &\leq k^{m-1} |p_1 - p_0| + k^{m-2} |p_1 - p_0| + \dots + k^n |p_1 - p_0| \\ &= k^n (1 + k + k^2 + \dots + k^{m-n-1}) |p_1 - p_0|. \end{aligned}$$

بنابر قضیه ۵.۲، $\lim_{m \rightarrow \infty} p_m = p$ ، در نتیجه،

$$|p - p_n| = \lim_{m \rightarrow \infty} |p_m - p_n| \leq k^n |p_1 - p_0| \sum_{i=0}^{\infty} k^i = \frac{k^n}{1-k} |p_1 - p_0|.$$

این دو نتیجه آهنگ همگرایی را به کران k بر اولین مشتق y مربوط می کنند. واضح است که آهنگ همگرایی به عامل $k^n/(1-k)$ بستگی دارد و، هر قدر k کوچکتر شود، همگرایی سریعتر می شود. از اینرو، همگرایی در صورت نزدیک بودن k به 1 ممکن است خیلی کند باشد. برای توضیح کاربرد این نتایج در تخمین تعداد تکرارهای لازم برای یک دقت خاص، مثال زیر را در نظر می گیریم.

مثال ۳

(T) توابع مثال ۱ و مثال ۲ را مجدداً در نظر می گیریم. دیدیم که در مورد $g_3(x) = \frac{1}{2}(10 - x^3)^{1/2}$ ، هر وقت $x \in [1, 1.5]$ ، $|g'_3(x)| < .66$. اگر همین تحلیل در

$$g_4(x) = \left(\frac{10}{4+x} \right)^{1/2}$$

مذکور در مثال ۱ (ت) انجام شود، خواهیم دید که

$$|g'_4(x)| = \left| \frac{-5}{\sqrt{10}(4+x)^{3/2}} \right| \leq \frac{5}{\sqrt{10}(5)^{3/2}} < .15, \quad x \in [1, 2]$$

کران بزرگی $g'_4(x)$ خیلی از کران بزرگی $g'_3(x)$ کوچکتر است، که دلیل همگرایی سریعتر در استفاده از $g_4(x)$ می باشد.

(ب) در مثال ۲ از بخش ۲۰۲ معلوم شد که $g(x) = (x^2 - 1)/3$ بر بازه $[-1, 1]$ دارای نقطه ثابت $p = (3 - \sqrt{13})/2$ است. فرض کنیم تقریبی از p با دقت 00001 لازم باشد، و بخواهیم آن را از تکرار تابعی توصیف شده در الگوریتم ۲۰۲ بدست آوریم. اولین کار این است که تعداد تکرارهای لازم برای تضمین این درجه از دقت را تخمین بزنیم. با استفاده از کران خطا در نتیجه ۶۰۲ و مقدار $k = \frac{2}{3}$ مفروض در مثال ۲ از بخش ۲۰۲، تعداد تکرارهای لازم N باید در روابط زیر صدق کند:

$$(11.2) \quad |p_N - p| \leq \left(\frac{2}{3} \right)^N \max\{p_0 - (-1), 1 - p_0\} \leq .00001.$$

چون اطلاع دیگری در دست نیست، $p_0 = 0$ را اختیار می کنیم که $\max\{p_0 - (-1), 1 - p_0\} = 1$ را حتی الامکان کوچک می سازد. نامساوی (۱۱.۲) به صورت

$$\left(\frac{2}{3} \right)^N \leq 10^{-5}$$

درمی آید، که از آن نتیجه می شود

$$N(\log_{10} 2 - \log_{10} 3) \leq -5$$

یا

$$N \geq \frac{5}{\log_{10} 3 - \log_{10} 2} = 28.39.$$

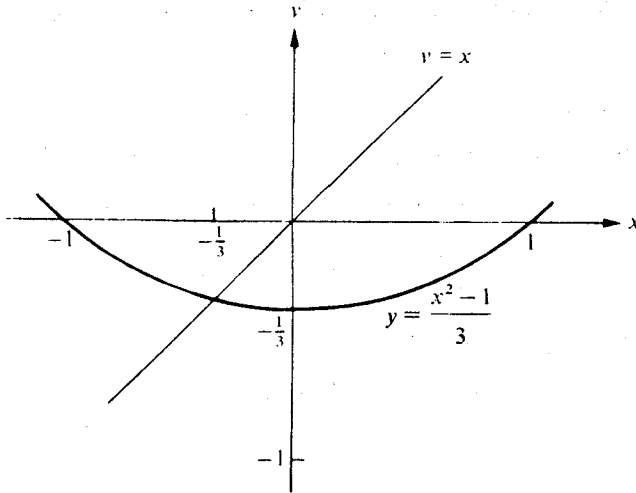
بنابراین، نتیجه ۶۰۲ ایجاب می کند که 29 یک کران تعداد تکرارهای لازم باشد. با استفاده از نتیجه ۷۰۲ به ازای $p_0 = 0$ و $p_1 = g(p_0) = -\frac{1}{3}$ ، از کران خطای

(۱۰.۲) معلوم می شود که

$$|p_N - p| \leq \frac{\left(\frac{2}{3}\right)^N}{1 - \frac{2}{3}} \left(\frac{1}{3}\right) = \left(\frac{2}{3}\right)^N.$$

این درست همان کران قبلی است، و مجدداً " 29 تکرار لازم می باشد.

به جای پرداختن به اصل مسئله، به تحلیل نموداری آن می پردازیم تا ببینیم آیا کاهش 29 تکرار امکان پذیر است یا نه. تعریف می کنیم $h(x) = g(x) - x$ ؛ در این صورت، $h(x) = 0$ اگر و فقط اگر x یک نقطه ثابت g باشد. چون $h(0) = -\frac{1}{3}$ و $h(-\frac{1}{3}) = \frac{1}{27}$ ، قضیه مقدار میانی ایجاب می کند که h یک ریشه در $[-\frac{1}{3}, 0]$ داشته باشد. (ر. ک. شکل (۰.۶.۲)



شکل ۰.۶.۲

با توجه به شکل ۰.۶.۲، می بینیم که ظاهراً " $-\frac{1}{3}$ " تقریب بهتری برای p است تا 0 چون بازه مورد نظر $[-\frac{1}{3}, 0]$ است، و g این بازه را بتوی خودش می نگارد، ممکن است برای این بازه مقدار کوچکتری از k پیدا شود تا برای بازه اصلی $[-1, 1]$. در واقع،

$$|g'(x)| = \left| \frac{2x}{3} \right| \leq \frac{2}{9} = k.$$

بمازی $-\frac{1}{3}$ و $p_0 = -\frac{1}{3}$ ، نتیجه ۰.۶.۲ ایجاب می کند که

$$|p_N - p| \leq \frac{\left(\frac{2}{9}\right)^N \cdot \frac{1}{27}}{1 - \frac{2}{9}} = \frac{1}{21} \left(\frac{2}{9}\right)^N \leq .00001$$

و

$$N \geq \frac{-\log_{10} .00001}{\log_{10} 9 - \log_{10} 2} \approx 5.6$$

حداکثر تکرارهای لازم است .

بمازای $\frac{1}{3}$ ، $p_0 = -\frac{1}{3}$ ، $k = \frac{2}{9}$ ، و $k = \frac{2}{9}$ ، $p_1 = [(-\frac{1}{3})^2 - 1]/3 = -\frac{8}{27}$ ، از نتیجه ۷.۲

حاصل می شود که

$$|p_N - p| \leq \left(\frac{2}{9}\right)^N \left(\frac{1}{3}\right) \leq .00001$$

و

$$N \geq \frac{-\log_{10} .00003}{\log_{10} 9 - \log_{10} 2} \approx 6.9$$

تکرار ، از اینرو ، نتیجه ۷.۲ تعداد تکرارهای لازم را بیشتر تقلیل می دهد .

در پایان این تحلیل ، باخوش بینی p_1, \dots, p_6 را به جای p_1, \dots, p_{29} حساب

می کنیم ، با این اطمینان که p_6 به قدر کافی دقیق می باشد . نتایج در جدول ۳.۲ ارائه

شده اند .

n	p_n	$p_n - p$
0	-.33333	.03056
1	-.29630	.00648
2	-.30407	.00129
3	-.30251	.00026
4	-.30283	.00005
5	-.30276	.00001
6	-.30278	.00000

جدول ۳.۲

مثال ۰۴ . برای حل معادله $x - 3^{-x} = 0$ ، فرض می کنیم $y(x) = 3^{-x}$ ، و تکرار تابعی را بر y اعمال می کنیم . طبق مثال ۳ از بخش ۲.۰۲ ، y در $[1, \frac{1}{3}]$ دارای نقطه ثابت

منحصر بفرد p است و $.762 \leq |g'(x)|$ بر $[\frac{1}{3}, 1]$.

برای تخمین تعداد تکرارهای لازم N در طرح تکرار تابعی جهت بدست آوردن

$$|p_N - p| < 10^{-5},$$

از نتیجه ۷۰۲ به ازای $p_0 = \frac{1}{3}$ استفاده می کنیم . چون

$$p_1 = g(p_0) = 3^{-1/3} \approx .693,$$

نامساوی (۱۰۰۲) نتیجه می دهد که

$$\begin{aligned} |p_N - p| &\leq \frac{(.762)^N}{1 - .762} |.333 - .693| \\ &\leq 1.513(.762)^N. \end{aligned}$$

برای اطمینان از اینکه $.00001 < |p_N - p|$ ، کافی است N را طوری حساب کنیم که

$$1.513(.762)^N < .00001.$$

بنابراین ،

$$N \geq \frac{\log_{10} .00001 - \log_{10} 1.513}{\log_{10} .762} \approx 43.9,$$

یا 44 یک کران برای تعداد تکرارهای مطلوب است .

نتایج محاسبات در جدول ۴۰۲ دیده می شوند .

n	p_n	n	p_n	n	p_n
0	.33333	8	.54386	16	.54774
1	.69336	9	.55019	17	.54785
2	.46686	10	.54638	18	.54778
3	.59876	11	.54867	19	.54782
4	.51799	12	.54729	20	.54780
5	.56605	13	.54812	21	.54781
6	.53694	14	.54762	22	.54781
7	.55439	15	.54792	23	.54781

جدول ۴۰۲

توجه کنید که در عمل فقط 21 تکرار برای دقت تا 10^{-5} لازم بود ، زیرا تکرارهای

بیشتر 54781 را بدست می دهند ، درحالی که کران لازم برای تعداد تکرارها 44 بود ،

رقمی که دارای بیش از 100 درصد خطاست. دلیل اصلی این است که مشتق بکار رفته در محاسبه کران N مساوی $g'(\frac{1}{3}) = 0.762$ بود، درحالی که $g'(0.54781) = 0.602$. اگر در این محاسبه 0.602. برای $g'(x)$ بکار می‌رفت، عدد حاصل 23 می‌شد، که به عدد واقعی مطلوب در این حالت خیلی نزدیکتر است.

اما واضح است که $g'(0.54781)$ نمی‌تواند پیش از انجام روند تکرار حساب شود و، وقتی روند با دقت کافی انجام شد، دیگر نیازی به مقدار مشتق نخواهد بود. در عمل، یک تقریب اولیه که حتی الامکان دقیق است بکار می‌رود و تکرارها تا جایی صورت می‌گیرند که دقت مطلوب حاصل شود یا تعداد تکرارها از عدد از پیش معلومی تجاوز کند، که عمدتاً بستگی به زمانی دارد که شخص می‌تواند در اجرای روند صرف نماید.

مجموعه تمرینات ۳.۲

۱. با استفاده از تکرار نقطه ثابت، یک جواب تقریبی $0 = 4^{-x} - x$ را که تا 10^{-2} دقیق باشد بیابید. با استفاده از نتیجه ۶.۲ یا ۷.۲، تعداد تکرارهای لازم برای حصول دقت 10^{-2} را تخمین بزنید، و این تخمین نظری را با عددی که واقعا لازم است مقایسه نمایید.
۲. با استفاده از تکرار نقطه ثابت، یک جواب تقریبی $0 = 2^{-x} - x$ را که تا 10^{-4} دقیق باشد بیابید. با استفاده از نتیجه ۶.۲ یا ۷.۲، تعداد تکرارهای لازم برای حصول دقت 10^{-4} را تخمین بزنید، و این تخمین نظری را با تعدادی که واقعا لازم است مقایسه نمایید.
۳. با استفاده از تکرار نقطه ثابت، $0 = x^3 - x - 1$ را در $[1, 2]$ حل کنید. برای این ریشه یک تقریب که تا 10^{-5} دقیق باشد بدست آورید.
۴. با استفاده از روند تکرار نقطه ثابت، تقریبی برای $\sqrt{3}$ بیابید که تا چهار رقم با معنی دقیق باشد. نتیجه و تعداد تکرارهای لازم را با جواب تمرین ۴ از بخش ۱.۲ مقایسه کنید.
۵. با استفاده از روند تکرار نقطه ثابت، تقریبی برای $\sqrt[3]{25}$ بیابید که تا دو رقم با معنی دقیق باشد. نتیجه و تعداد تکرارهای لازم را با جواب تمرین ۳ از بخش ۱.۲ مقایسه کنید.
۶. در هر یک از معادلات زیر، بازه $[a, b]$ را طوری تعیین کنید که تکرار نقطه ثابت

همگرا باشد. تعداد تکرارهای لازم برای حصول تقریبهای دقیق تا 10^{-5} را تخمین بزنید و محاسبات را انجام دهید:

$$(A) \quad x = \frac{2 - e^x + x^2}{3} \quad (A) \quad ; \quad x = 4^{-x} \quad (B) \quad ; \quad x = 5^{-x} \quad (P) \quad ; \quad x = 6^{-x} \quad (T)$$

$$(C) \quad x = 1.75 + \frac{4x - 7}{x - 2}$$

۷. درهریک از معادلات زیر، تابع $y(x)$ و بازه $[a, b]$ را طوری تعیین کنید که تکرار نقطه ثابت به یک جواب مثبت معادله همگرا باشد:

$$(A) \quad 3x^2 - e^x = 0 \quad (B) \quad x - \cos x = 0$$

جوابها را با تقریب 10^{-5} پیدا کنید.

۸. به روش تکرار نقطه ثابت و با یک تابع تکرار مناسب مانند y ، همه ریشههای

$$f(x) = x^2 + 10 \cos x$$

تکرارهای لازم را با تعداد تکرارهای تمرین ۸، بخش ۱۰.۲، مقایسه نمایید.

۹. ثابت کنید دنباله تعریف شده با

$$x_n = \frac{1}{2} \left(x_{n-1} + \frac{2}{x_{n-1}} \right), \quad n \geq 1$$

به ازای هر $x_0 > 0$ به $\sqrt{2}$ همگراست.

۱۰. فرض کنید g بر بازه (c, d) ، شامل نقطه ثابت ξ از g ، به طور پیوسته مشتقپذیر

باشد. نشان دهید که اگر $|g'(\xi)| < 1$ ، $\delta > 0$ وجود دارد بطوری که تکرار نقطه

ثابت به ازای هر تقریب اولیه p_0 که $|\rho_0 - \xi| \leq \delta$ همگرا می باشد.

۴.۲ روش نیوتن - رفسون

روش نیوتن - رفسون (یا فقط نیوتن) یکی از قویترین و معروفترین روشهای عددی برای

یافتن یک ریشه $f(x) = 0$ است. برای معرفی این روش دست کم سه راه وجود دارد.

معمولیترین آنها بررسی نموداری روش است (ر.ک. تمرین ۱). راه دیگر بدست آوردن

روش نیوتن به عنوان یک روش ساده برای دستیابی به همگرایی سریعتر از همگرایی حاصل

از انواع دیگر تکرار تابعی می باشد (ر.ک. بخش ۵.۲). راه سوم، که در زیر مطرح

می شود، یک طریقه شهودی است که بر چند جمله ای تیلور تعریف شده در قضیه ۱۳.۱

استوار است .

فرض کنیم تابع f بر بازه $[a, b]$ دوبار به طور پیوسته مشتقپذیر باشد ، یعنی ،
 $f \in C^2[a, b]$. فرض کنیم $\bar{x} \in [a, b]$ یک تقریب به p باشد بطوری که $|f'(\bar{x}) - p| \neq 0$
 "کوچک" باشد . چند جمله‌ای تیلور درجه دوم $f(x)$ را ، که حول \bar{x} بسط داده شده ،
 در نظر می‌گیریم :

$$(۱۲.۲) \quad f(x) = f(\bar{x}) + (x - \bar{x})f'(\bar{x}) + \frac{(x - \bar{x})^2}{2} f''(\xi(x)),$$

که در آن $\xi(x)$ بین x و \bar{x} قرار دارد . چون $f(p) = 0$ ، معادله (۱۲.۲) ، به ازای
 $x = p$ ، نتیجه می‌دهد که

$$0 = f(\bar{x}) + (p - \bar{x})f'(\bar{x}) + \frac{(p - \bar{x})^2}{2} f''(\xi(p)).$$

چون $|p - \bar{x}|$ کوچک فرض شده بود ، $|p - \bar{x}|^2$ از این هم کوچکتر است ، و اگر این
 مقدار را نادیده بگیریم ،

$$0 \approx f(\bar{x}) + (p - \bar{x})f'(\bar{x}).$$

با حل این معادله نسبت به p ، نتیجه می‌شود که

$$p \approx \bar{x} - \frac{f(\bar{x})}{f'(\bar{x})},$$

که باید تقریب بهتری از \bar{x} به p باشد . به این ترتیب ، طرح روش نیوتن - رفسون ریخته
 می‌شود ، که مستلزم تولید دنباله $\{p_n\}_{n=1}^{\infty}$ است که با

$$p_n = p_{n-1} - \frac{f(p_{n-1})}{f'(p_{n-1})}, \quad n \geq 1$$

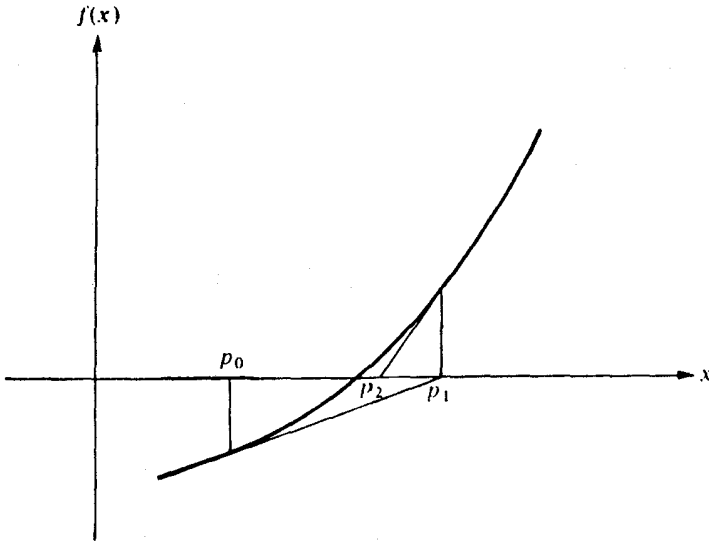
تعریف می‌شود . شکل ۷.۲ طرز بدست آوردن تقریبها را با استفاده از مماسهای متوالی
 نشان می‌دهد . (همچنین ، ر.ک. تمرین ۰۱)

الگوریتم نیوتن - رفسون ۳.۲

معادله $f(x) = 0$ و تقریب اولیه p_0 به جواب آن مفروضند .

مرحله ۱ . قرار می‌دهیم $i = 1$.

مرحله ۲ . قرار می‌دهیم



شکل ۷.۲

$$p_i = p_{i-1} - \frac{f(p_{i-1})}{f'(p_{i-1})}$$

- مرحله ۳. مشخص می‌کنیم که آیا روند باید ادامه یابد.
- اگر باید، به مرحله ۴ می‌رویم.
- در غیر این صورت، به مرحله ۵ می‌رویم.
- مرحله ۴. به i عدد ۱ را افزوده به مرحله ۲ می‌رویم.
- مرحله ۵. روند تمام خواهد بود.

روشهای توقف قبلی در روش نیوتن نیز قابل اعمالند؛ یعنی، $\epsilon > 0$ را اختیار

می‌کنیم و p_1, \dots, p_n را می‌سازیم تا

$$(۱۶.۲) \quad |p_n - p_{n-1}| < \epsilon,$$

$$(۱۷.۲) \quad \frac{|p_n - p_{n-1}|}{|p_n|} < \epsilon, \quad p_n \neq 0,$$

$$(۱۸.۰۲) \quad |f(p_n)| < \varepsilon.$$

توجه کنید که نامساوی (۱۸.۰۲) ممکن است اطلاع زیادی از خطای واقعی $|p_n - p|$ بدست ندهد.

روش نیوتن نمونه‌ای از تکرار تابعی $p_n = g(p_{n-1}), n \geq 1$ است، که در آن

$$g(p_{n-1}) = p_{n-1} - \frac{f(p_{n-1})}{f'(p_{n-1})}, \quad n \geq 1.$$

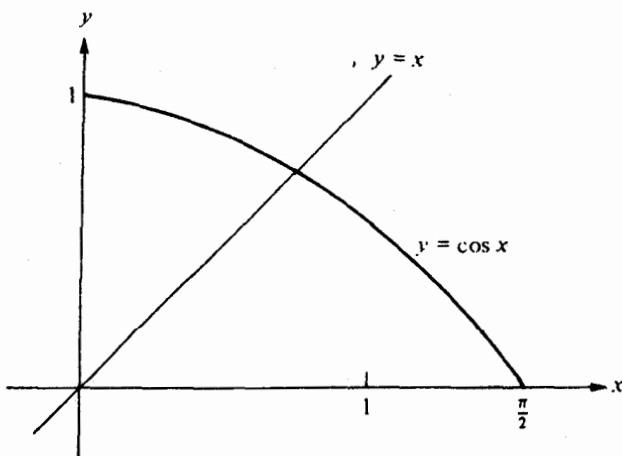
مسئله‌ای که وجود دارد این است که اگر برای $n \geq 1$ ی $f'(p_{n-1}) = 0$ ، p_n تعریف نشده است. بحث کلی روش نیوتن را بعد از مثال زیر ادامه می‌دهیم.

مثال ۱

(A) فرض کنیم تعیین جوابی از معادله $x = \cos x$ مورد نظر باشد. قرار می‌دهیم $f(x) = \cos x - x$ در این صورت،

$$f\left(\frac{\pi}{2}\right) = -\frac{\pi}{2} < 0 < 1 = f(0),$$

و، بنا بر قضیه مقدار میانی (قضیه ۱۲.۰۱، ص ۹)، ریشه‌ای از f در $[0, \pi/2]$ وجود دارد. نمودار معادلات $y = \cos x$ و $y = x$ را در شکل ۸.۰۲ می‌بینیم؛ اشتراک آنها نقطه ثابت $g(x) = \cos x$ می‌باشد.



شکل ۸.۰۲

از نمودار واضح است که $f(x) = 0$ در $[0, \pi/2]$ جواب منحصر بفرد دارد. چون $f'(x) = -\sin x - 1$ ، روش نیوتن شکل زیر را خواهد داشت:

$$p_n = p_{n-1} - \frac{(\cos p_{n-1} - p_{n-1})}{(-\sin p_{n-1} - 1)}, \quad n \geq 1,$$

که در آن p_0 هنوز باید انتخاب شود. در بعضی مسائل p_0 می تواند دلخواه باشد، در حالی که، در بقیه، انتخاب یک تقریب اولیه مناسب اهمیت خواهد داشت. در مسئله مورد نظر، نمودار شکل ۸.۲، $p_0 = \pi/4$ را به عنوان تقریب اولیه پیشنهاد می کند. به ازای $p_0 = \pi/4$ ، تقریبهای مندرج در جدول ۵.۲ تولید می شوند.

n	p_n	n	p_n
0	.7853981635	3	.7390851332
1	.7395361337	4	.7390851332
2	.7390851781	5	.7390851332

جدول ۵.۲

یک تقریب عالی به ازای $n = 3$ بدست آمده است.

(ب) برای یافتن جواب منحصر بفرد $0 = 10 - 4x^2 + x^3$ بر بازه $[1, 2]$ به روش نیوتن، دنباله $\{p_n\}_{n=1}^{\infty}$ را به وسیله

$$p_n = p_{n-1} - \frac{p_{n-1}^3 + 4p_{n-1}^2 - 10}{3p_{n-1}^2 + 8p_{n-1}}, \quad n \geq 1.$$

تولید می کنیم. با انتخاب $p_0 = 1.5$ ، نتایج مثال ۱ (ث) از بخش ۳.۲ بدست می آیند، که در آنها $p_3 = 1.36523001$ تا هشت رقم اعشار درست است. برای توضیح اهمیت یک تقریب اولیه خوب، روند فوق را به ازای $p_0 = -100$ بکار می بریم. نتایج این تقریب اولیه ضعیف در جدول ۶.۲ درج شده است.

همگرایی بکندی شروع می شود؛ در واقع، در $n = 18$ ، بنظر می رسد که اصولاً همگرایی وجود ندارد. بالاخره، وقتی $p_{22} \in [1, 2]$ ، همگرایی به جواب خیلی سریع خواهد شد. از این مثال معلوم می شود که، وقتی تقریب اولیه به ریشه واقعی نزدیک باشد، از روش نیوتن همگرایی سریعی انتظار می رود؛ اما، وقتی تقریب اولیه با جواب

n	p_n	n	p_n
0	-100	13	-2.0757
1	-67.1229	14	-2.5401
2	-45.2107	15	-3.1421
3	-30.6110	16	-2.8007
4	-20.8903	17	-2.2742
5	-14.4276	18	-2.6754
6	-10.1440	19	4.7255
7	-7.3217	20	2.9616
8	-5.4823	21	1.9405
9	-4.3043	22	1.4793
10	-3.5648	23	1.3711
11	-3.0995	24	1.36525
12	-2.7643	25	1.36523001

جدول ۶.۲

اختلاف فاحش دارد، روش همگرایی کندی خواهد داشت. در واقع، در بسیاری از مسائل، اگر تقریب اولیه به قدر کافی به ریشه واقعی نزدیک نباشد، روش نیوتن واگرایی به بار خواهد آورد.

قضیه زیر یک قضیه همگرایی برای روش نیوتن است، که اهمیت انتخاب p_0 را روشن می سازد.

قضیه ۸.۲. فرض کنیم $f \in C^2[a, b]$. هرگاه $p \in [a, b]$ چنان باشد که $f(p) = 0$ و $f'(p) \neq 0$ ، آنگاه $\delta > 0$ ای وجود دارد بطوری که روش نیوتن دنباله‌ای مانند $\{p_n\}_{n=1}^{\infty}$ تولید می کند که به ازای هر تقریب اولیه $p_0 \in [p - \delta, p + \delta]$ به p همگرا است.

برهان. برهان بر تحلیل روش نیوتن به عنوان یک طرح تکرار تابعی $p_n = g(p_{n-1})$ ، به ازای $n \geq 1$ ، استوار است، که در آن

$$g(x) = x - \frac{f(x)}{f'(x)}$$

روند کار یافتن بازه‌ای مانند $[p - \delta, p + \delta]$ است بطوری که g این بازه را بتوی خودش

بنگارد و، به ازای $|g'(x)| \leq k < 1$ ، $x \in [p - \delta, p + \delta]$.

چون $f'(p) \neq 0$ و f' پیوسته است، $\delta_1 > 0$ ی وجود دارد بطوری که به ازای $[a, b] \supset [p - \delta_1, p + \delta_1]$ ، $f'(x) \neq 0$ ، بنابراین، g بر $[p - \delta, p + \delta]$ تعریف شده و پیوسته است. همچنین، به ازای $[p - \delta_1, p + \delta_1]$ ،

$$g'(x) = 1 - \frac{f'(x)f''(x) - f(x)f'''(x)}{[f'(x)]^2} = \frac{f(x)f''(x)}{[f'(x)]^2}$$

و، چون $f \in C^2[a, b]$ ، پس $g \in C^1[p - \delta_1, p + \delta_1]$. طبق فرض $f(p) = 0$ ؛ در نتیجه،

$$(19.2) \quad g'(p) = \frac{f(p)f''(p)}{[f'(p)]^2} = 0.$$

چون g' پیوسته است، معادله (19.2) ایجاب می کند که δ ای با خاصیت $0 < \delta < \delta_1$ وجود داشته باشد و،

$$|g'(x)| \leq k < 1, \quad x \in [p - \delta, p + \delta]$$

باقی می ماند اثبات اینکه $[p - \delta, p + \delta] \rightarrow [p - \delta, p + \delta]$ گوییم اگر

$$x \in [p - \delta, p + \delta], \quad \text{به ازای } \xi \text{ ای بین } x \text{ و } p$$

$$|g(x) - p| = |g(x) - g(p)| = |g'(\xi)||x - p| \leq k|x - p| < |x - p|$$

چون $x \in [p - \delta, p + \delta]$ ، نتیجه می شود که $\delta < |x - p|$ و $|g(x) - p| < \delta$. این ایجاب می کند که $g: [p - \delta, p + \delta] \rightarrow [p - \delta, p + \delta]$.

حال جمیع مفروضات قضیه ۵.۲ (ص ۴۴) برای $g(x) = x - f(x)/f'(x)$ برقرارند؛

در نتیجه، دنباله $\{p_n\}_{n=1}^{\infty}$ تعریف شده با

$$p_n = g(p_{n-1}) \quad \text{به ازای } n = 1, 2, 3, \dots$$

به ازای هر $p_0 \in [p - \delta, p + \delta]$ به همگرا می باشد.

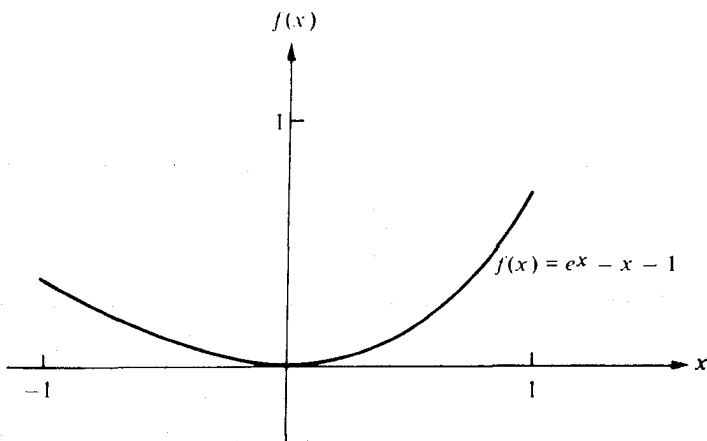
توجه کنید که قضیه ۸.۲ می گوید که، با مفروضاتی معقول، روش نیوتن همواره در صورتی که تقریب اولیه به قدر کافی دقیق باشد همگراست. گاهی، نظیر مثال ۱، روش نیوتن با تقریب اولیه بسیار ضعیفی نیز همگراست؛ اما در بسیاری حالات، مانند تمرین ۹، ضروری است که تقریب اولیه خوبی اختیار شود. در قضیه ۸.۲ فرض شده است که $f(p) = 0$ و $f'(p) \neq 0$. اگر $f'(p) = 0$ ، روش نیوتن ممکن است هنوز همگرا باشد، اگر چه همگرایی آن عموماً "کندتر است. در مثال زیر به این حالت می پردازیم. همگرایی در

این مثال رفتار غریبی دارد، که عمدتاً "بدلیل خطاهای گرد کردن است".

مثال ۰۲. فرض کنیم $f(x) = e^x - x - 1$ (ر.ک. شکل ۰۹۰۲). در این صورت،
 $f'(x) = e^x - 1$. توجه کنید که $f(x) = 0$ دارای ریشه منحصراً $p = 0$ است و
 $f'(0) = 0$. به ازای $p_0 = 1$ ، دنباله $\{p_n\}_{n=1}^{\infty}$:

$$p_n = p_{n-1} - \frac{e^{p_{n-1}} - p_{n-1} - 1}{e^{p_{n-1}} - 1}, \quad n \geq 1,$$

حاصل از روش نیوتن، در جدول ۰۷۰۲ ثبت شده است.



شکل ۰۹۰۲

قضیه دیگری که همگرایی همیشگی روش نیوتن را تضمین می‌کند و دارای این مزیت اضافی است که تقریب اولیه را می‌توان در آغاز برای تضمین همگرایی انتخاب کرد به صورت زیر است:

قضیه ۰۹۰۲. فرض کنیم $f \in C^1[a, b]$ ، $f''(x)$ در هر نقطه $x \in (a, b)$ وجود داشته باشد، و شرایط زیر برقرار باشند:

(یک) $f(a)f(b) < 0$ ؛

(دو) f' بر $[a, b]$ اکیدا "مثبت یا اکیدا" منفی باشد؛

n	p_n	n	p_n
0	1.0	9	2.7750×10^{-3}
1	.58198	10	1.3881×10^{-3}
2	.31906	11	6.9411×10^{-4}
3	.16800	12	3.4703×10^{-4}
4	.08635	13	1.7416×10^{-4}
5	.04380	14	8.8041×10^{-5}
6	.02206	15	4.2610×10^{-5}
7	.01107	16	1.9142×10^{-5}
8	.005545		

جدول ۷.۲

(سه) f'' بر (a, b) تغییر علامت ندهد؛

(چهار) هرگاه c نقطه انتهایی $[a, b]$ باشد که $|f'(x)|$ در آن مینیمم است، آنگاه

$$\left| \frac{f(c)}{f'(c)} \right| \leq b - a.$$

در این صورت، دنباله $\{p_n\}_{n=1}^{\infty}$ تولید شده به وسیله روش نیوتن به تنها صفر f در (a, b) همگراست مشروط بر اینکه $p_0 \in (a, b)$.

بعلاوه، $\{p_n\}_{n=1}^2$ یک دنباله صعودی است اگر به ازای هر $x \in (a, b)$ ، $f''(x)f'(x) \leq 0$ ؛

و یک دنباله نزولی است اگر به ازای هر $x \in (a, b)$ ، $f''(x)f'(x) \geq 0$.

شرطهای (یک) تا (چهار) را شرایط فوریه^۱ می نامند، و با آنکه از حیث نظری

مهمند، اعمال آنها در عمل عموماً مشکل است.

اثبات قضیه ۹.۲ کاملاً فنی است، و در اینجا حذف می شود. آن را می توان در

هنریچی [۳۴]، صفحات ۷۹ تا ۸۱، بویژه برای حالت $f(a) < 0 < f(b)$ و $f''(x) \leq 0$ یافت.

حالات دیگر را می توان به همین نحو اثبات و یا به این وضع درآورد.

مثال ۳. تابع $f(x) = x^2 - a$ را، که در آن $a > 0$ ، در نظر می گیریم. توجه کنید که

$x = \sqrt{a}$ تنها ریشه مثبت f است. فرض کنیم $[c, d]$ بازه‌ای با $c > 0$ بوده و چنان باشد که $f(c)f(d) < 0$ ، که ایجاب می‌کند $\sqrt{a} \in [c, d]$. چون به ازای هر $x \in [c, d]$ ، $f'(x) > 0$ و $f''(x) > 0$ ، سه شرط اول فوریه برقرارند. و چون مقدار مینیم f' بر $[c, d]$ مساوی $2c$ است، شرط (چهار) در صورت

$$\left| \frac{c^2 - a}{2c} \right| \leq d - c$$

برقرار است. از اینرو، شرط (چهار) در صورت $(a + c^2)/2c \leq d$ برقرار می‌باشد. بنابراین، به ازای هر چنین بازه $[c, d]$ و هر تقریب اولیه $p_0 \in [c, d]$ ، روش نیوتن به \sqrt{a} همگرا خواهد بود.

مشکل بزرگی که اغلب در روش نیوتن با آن مواجهیم این است که، ممکن است در عین اینکه $f(x)$ آسان حساب می‌شود، $f'(x)$ سخت بدست آید. مثلاً، اگر $f(x) = x^{23} \cos 2x$ ، $f'(x) = 2x^{23} \cos 2x + x^{23} (\cos 2x) \ln 3 - 2x^{23} \sin 2x$ ، محاسبه بسیار خسته کننده‌ای دارد، بخصوص اگر روش در یک ماشین حساب بدون برنامه پیاده شود. برای پیشگیری از این مشکل، می‌توان روش نیوتن را با مختصر تغییر بدست آورد. ابتدا $f(x)$ را با چند جمله‌ای تیلور درجه اول آن که حول p_{n-1} بسط داده شده تقریب می‌کنیم، داریم

$$f(x) \approx f(p_{n-1}) + f'(p_{n-1})(x - p_{n-1});$$

در این صورت،

$$f(p_{n-2}) \approx f(p_{n-1}) + (p_{n-2} - p_{n-1})f'(p_{n-1}),$$

و، با حل آن نسبت به $f'(p_{n-1})$ ، خواهیم داشت

$$p_n = p_{n-1} - \frac{f(p_{n-1})(p_{n-1} - p_{n-2})}{f(p_{n-1}) - f(p_{n-2})}.$$

با استفاده از این تقریب برای $f'(p_{n-1})$ در فرمول نیوتن، نتیجه می‌شود که

$$f'(p_{n-1}) \approx \frac{f(p_{n-1}) - f(p_{n-2})}{p_{n-1} - p_{n-2}}.$$

روشی که از این فرمول استفاده می‌کند روش خط قاطع نام دارد، و به صورت زیر می‌باشد:

الگوریتم خط قاطع ۴.۲

معادله $f(x) = 0$ و تقریبهای اولیه p_0 و p_1 ، که $p_0 \neq p_1$ ، مفروضند :

مرحله ۱. قرار می دهیم $i = 2$.

مرحله ۲. قرار می دهیم

$$p_i = p_{i-1} - \frac{f(p_{i-1})(p_{i-1} - p_{i-2})}{f(p_{i-1}) - f(p_{i-2})}$$

مرحله ۳. مشخص می کنیم که آیا روش باید ادامه یابد .

اگر باید ، به مرحله ۴ می رویم .

در غیر این صورت ، به مرحله ۵ می رویم .

مرحله ۴. به i عدد 1 را افزوده به مرحله ۲ می رویم .

مرحله ۵. روند تمام خواهد بود .

مثال زیر در ارتباط با مسئله‌ای است که در مثال ۱ (T) ، با استفاده از روش نیوتن

به صورت الگوریتم ۳.۲ ، با $p_0 = \pi/4$ ، در نظر گرفته شد .

مثال ۴. با استفاده از روش خط قاطع ، ریشه‌ای از $f(x) = \cos x - x$ را بدست آورید .

در مثال ۱ از تقریب اولیه $p_0 = \pi/4$ استفاده شد . در اینجا دو تقریب اولیه لازم است .

به ازای 5. $p_0 = \pi/4$ ، $p_1 = \pi/4$ ، و فرمول

$$p_n = p_{n-1} - \frac{(p_{n-1} - p_{n-2})(\cos p_{n-1} - p_{n-1})}{(\cos p_{n-1} - p_{n-1}) - (\cos p_{n-2} - p_{n-2})} \quad n \geq 2$$

حاصل از الگوریتم ۴.۲ ، جدول محاسبات ۸.۲ بدست خواهد آمد .

n	p_n
0	.5
1	.7853981635
2	.7363841390
3	.7390581394
4	.7390851492
5	.7390851334

با مقایسه این نتایج با نتایج مثال ۱، می‌توان دید که p_5 تا ده رقم اعشار دقیق است. جالب است توجه کنیم که همگرایی روش خط قاطع در این مثال از روش نیوتن، که این درجه از دقت در آن با p_3 بدست آمد، قدری کندتر است. همانطور که در بخش ۵.۲ خواهید دید، این نتیجه بطور کلی برقرار است.

روش نیوتن (یا روش خط قاطع) اغلب برای تهذیب جواب حاصل از روشهای دیگر بکار می‌رود. چون این روشها به تقریب اولیه مناسبی نیاز دارند، اما همگرایی بسیار سریعی بدست می‌دهند، این منظور را بخوبی برخواهند آورد.

مجموعه تمرینات ۴.۲

۱. در این تمرین، روش نیوتن تعبیر هندسی می‌شود: فرض کنید $f'(x)$ بر $[a, b]$ موجود بوده و $f'(x) \neq 0$ بر $[a, b]$. بعلاوه، $p \in [a, b]$ ای وجود داشته باشد بطوری که $f(p) = 0$. فرض کنید $p_0 \in [a, b]$ دلخواه باشد. همچنین، p_1 نقطه‌ای باشد که در آن خط مماس بر f در $(p_0, f(p_0))$ از محور x می‌گذرد. به‌ازای هر $n \geq 1$ ، p_n را قطع x خط مماس بر f در $(p_{n-1}, f(p_{n-1}))$ بگیرید. فرمول این روش را بدست آورید.

۲. به‌روش نیوتن و روش خط قاطع، ریشه $x^3 - x - 1 = 0$ در $[1, 2]$ را با دقت 10^{-5} بدست آورید.

۳. روش خط قاطع را تعبیر هندسی کنید. (راهنمایی. به تمرین ۱ و تعریف خط قاطع رجوع کنید.)

۴. به روش نیوتن، جوابهای معادلات زیر را تا 10^{-5} تقریب کنید:

$$(T) \quad x = \frac{2 - e^x + x^2}{3} \quad (ب) \quad 3x^2 - e^x = 0$$

$$(پ) \quad e^x + 2^{-x} + 2 \cos x - 6 = 0 \quad (ت) \quad x^2 + 10 \cos x = 0$$

۵. تمرین ۴ را با استفاده از روش خط قاطع حل کنید.

۶. به روشهای زیر، معادله $4 \cos x = e^x$ را با دقت 10^{-4} حل کنید:

$$(T) \quad \text{روش نیوتن به‌ازای } p_0 = 1$$

$$(ب) \quad \text{روش خط قاطع به‌ازای } p_0 = \pi/4 \text{ و } p_1 = \pi/2$$

۷. معادله زیر را به روش نیوتن حل کنید:

$$p_0 = \frac{\pi}{2} \quad \text{به ازای} \quad 0 = \left(\sin x - \frac{x}{2}\right)^2$$

روش نیوتن را آنقدر تکرار کنید تا ریشه^۲ $f(x) = (\sin x - x/2)^2$ با دقت 10^{-5} بدست آید. آیا این نتایج برای روش نیوتن غیرعادی است؟ توضیح دهید.

۸. (آ) به روش نیوتن و با فرض $p_0 = 2$ ، برای $\sqrt{3}$ تقریبی بیابید که تا چهار رقم دقیق باشد.

(ب) قسمت (آ) را با روش خط قاطع تکرار کنید، و سپس نتایج را با نتایج تمرین ۴ از بخش ۳۰۲ و تمرین ۴ از بخش ۱۰۲ مقایسه نمایید.

۹. تابع $f(x) = (4x - 7)/(x - 2)$ دارای ریشه^۳ $p = 1.75$ است. روش نیوتن را با تقریبهای اولیه^۴ زیر بکار برید:

$$(آ) p_0 = 1.625 \quad (ب) p_0 = 1.875 \quad (پ) p_0 = 1.5 \quad (ت) p_0 = 1.95$$

نتایج را تعبیر هندسی کنید.

۱۰. به ازای چه مقادیری از p_0 و p_1 می توان روش خط قاطع را برای حل معادله^۵

$$f(x) = \frac{4x - 7}{x - 2} = 0$$

بکار برد؟

۱۱. معادله^۶ تکرار برای روش خط قاطع را می توان به صورت زیر نوشت:

$$p_n = \frac{f(p_{n-1})p_{n-2} - f(p_{n-2})p_{n-1}}{f(p_{n-1}) - f(p_{n-2})}$$

می توانید بگویید چرا این معادله^۷ تکرار، در حالت کلی، از معادله^۸ تکرار الگوریتم ۴۰۲ کم اهمیت تر است؟

۱۲. با استفاده از مثال ۳، نشان دهید که روش نیوتن برای یافتن تقریبی از \sqrt{a} به ازای هر تقریب اولیه^۹ $p_0 > 0$ همگراست.

۱۳. (آ) فرض کنید $f \in C^1[a, b]$ و $\{p_n\}_{n=0}^{\infty}$ دنباله^{۱۰} تعریف شده به وسیله^{۱۱} روش نیوتن با $\lim_{n \rightarrow \infty} p_n = r$ باشد. نشان دهید که r باید یک جواب $f(x) = 0$ باشد.

(ب) فرض کنید $f \in C^1[a, b]$ ، و به ازای هر $x \in [a, b]$ ، $f'(x) \neq 0$. هرگاه $\{p_n\}_{n=0}^{\infty}$

دنباله ای باشد با این خاصیت که به ازای هر $n \geq 0$ ، $p_n \in (a, b)$ ، و $\lim_{n \rightarrow \infty} p_n = r$

نشان دهید که

$$|p_n - r| \leq \max_{x \in [a, b]} \frac{|f(p_n)|}{|f'(x)|}$$

۱۴. برای توابع زیر، بازه $[a, b]$ با خاصیت $a > 0$ را طوری بیابید که شرایط فوریه مذکور در قضیه ۹.۲ برقرار باشند:

(آ) $f(x) = x^3 - 2$ ؛ (ب) $f(x) = x^3 + 2x - 2$ ؛ (پ) $f(x) = x^4 + x - 19$ ؛

(ت) $f(x) = \frac{1}{x} + \frac{1}{2x^2} - 1$

۱۵. روشی دیگر برای یافتن ریشه معادله $f(x) = 0$ در بازه $[a, b]$ ، که از بعضی جهات با هر دو روش تنصیف و روش خط قاطع شباهت دارد، روش نابجایی است. این روش با روش تنصیف در این امر که بازه‌های $[a_i, b_i]$ که شامل یک ریشه اند تولید می‌شوند، و با روش خط قاطع در طرز بدست آمدن بازه‌های تقریبی جدید شباهت دارد. با این فرض که بازه $[a_i, b_i]$ شامل یک ریشه $f(x) = 0$ است، قطع x خط واصل بین نقاط $(a_i, f(a_i))$ و $(b_i, f(b_i))$ را حساب کرده آن را p_i بنامید. اگر $f(p_i)f(a_i) < 0$ ، تعریف کنید $a_{i+1} = p_i$ و $b_{i+1} = b_i$ ؛ در غیر این صورت، تعریف کنید $a_{i+1} = a_i$ و $b_{i+1} = p_i$. الگوریتمی شبیه به الگوریتمهای ۱.۲ و ۴.۲ بسازید که روش نابجایی را توصیف کند.

۱۶. به روش ساخته شده در تمرین ۱۵، یک ریشه تقریبی $\cos x - x = 0$ در بازه $[5, \pi/4]$ را بدست آورید.

۱۷. تمرین ۴ را با استفاده از روش نابجایی در تمرین ۱۵ حل کنید.

۱۸. به روش نیوتن، برای λ در معادله جمعیت

$$1,564,000 = 1,000,000 e^{\lambda} + \frac{435,000}{\lambda} (e^{\lambda} - 1),$$

مذکور در مقدمه این فصل، تقریبی بیابید که تا 10^{-4} دقیق باشد. با استفاده از این مقدار، و با این فرض که میزان مهاجرت در طول سال دوم در 435,000 نفر در سال ثابت است، جمعیت را در پایان این سال پیش‌بینی کنید.

۱۹. در طرح وسایل نقلیه زمینی باید به شکست وسیله در برخورد با دو نوع مانع توجه شود. یک نوع شکست، شکست تعلیق (HUF) نام دارد و نوعاً "وقتی رخ می‌دهد که وسیله می‌خواهد از مانعی بگذرد که باعث تماس ته آن با زمین (یا مانع) می‌شود.

نوع دیگر شکست دماغه (NIF) نام دارد و معمولاً "وقتی رخ می‌دهد که وسیله در یک گودال می‌افتد و جلوی آن با زمین تماس پیدا می‌کند.

شکل زیر، ماخوذ از بکر [۹]، مولفه‌های NIF یک وسیله نقلیه را نشان می‌دهد. در کتاب مرجع فوق نشان داده شده است که زاویه α زاویه ماکزیمم یک وسیله نقلیه در یک گودال، وقتی β زاویه ماکزیمم باشد که HUF در آن روی نمی‌دهد، در معادله

$$A \sin \alpha \cos \alpha + B \sin^2 \alpha - C \cos \alpha - E \sin \alpha = 0$$

صدق می‌کند، که در آن

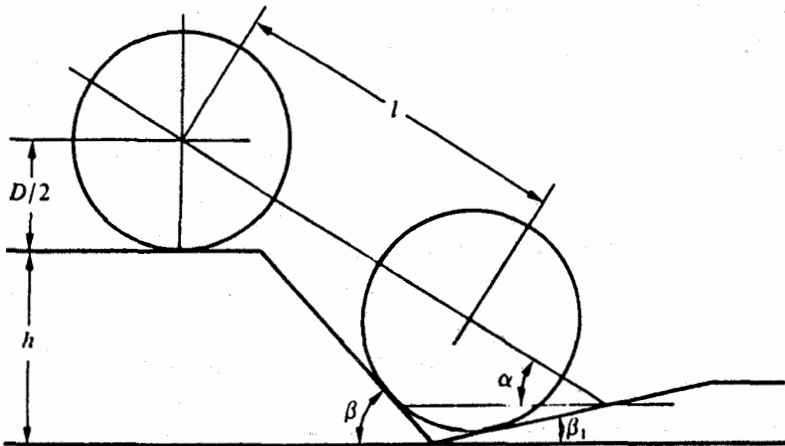
$$A = l \sin \beta_1,$$

$$B = l \cos \beta_1,$$

$$C = (h + .5D) \sin \beta_1 - .5D \tan \beta_1,$$

$$E = (h + .5D) \cos \beta_1 - .5D$$

(ر.ک. شکل ۱۰۰۲)



شکل ۱۰۰۲

(T) در کتاب مذکور آمده است که وقتی $l=222.5$ سانتیمتر، $h=122.5$ سانتیمتر،

حل معادلات یک متغیره ۶۷

$D=137.5$ سانتیمتر، و $\beta_1 = 11.5^\circ$ ، زاویه α تقریباً 33° است. این مطلب را تحقیق کنید.

(ب) α را درحالتی که h ، l و β_1 همان مقادیر قسمت (آ) باشند ولسی $D=75$ سانتیمتر بدست آورید.

۲۰. برای تعیین تقریبهایی برای کشش T ، و زاویه میل ϕ نسبت به افق در یک نقطه خاص یک کابل یا خط لوله کشیده شده در زیر آب، باید تعدادی معادله به شکل

$$(۱) \quad \phi = \tan^{-1} \left[\frac{T_0 \sin \phi_0 - F}{T_0 \cos \phi_0 - G} \right]$$

و

$$(۲) \quad T = [(T_0 \sin \phi_0 - F)^2 + (T_0 \cos \phi_0 - G)^2]^{1/2}$$

نسبت به ϕ و T حل شوند (ر.ک. کتاب مرجع [۶۶]). توابع F و G در (۱) و (۲) هر دو به ϕ بستگی دارند و، به ازای توابع "محموله‌ای" هیدرودینامیک ماسی و قائم مفروض f و g ، به شکلهای زیر می‌باشند:

$$F = \frac{s}{2} [-f(\phi_0) \sin \phi_0 - f(\phi) \sin \phi + g(\phi_0) \cos \phi_0 + g(\phi) \cos \phi] - ws$$

و

$$G = -\frac{s}{2} [f(\phi_0) \cos \phi_0 + f(\phi) \cos \phi + g(\phi_0) \sin \phi_0 + g(\phi) \sin \phi],$$

فرض کنید $w=1$ ، $T_0=2$ ، $s=.1$ ، $\phi_0 = \pi/2$ ، و توابع "محموله‌ای" f و g عبارت باشند از

$$f(\phi) = .2 \cos \phi$$

و

$$g(\phi) = .98 \sin^2 \phi + .02 \sin \phi.$$

(آ) با دو تقریب اولیه نزدیک به $\pi/2$ شروع کنید و به روش خط قاطع، تقریبی

از ϕ بیابید. چه نوع دقتی برای ϕ کافی است؟

(ب) چرا روش نیوتن برای این مسئله مناسب نیست؟

۲۱. بی.اچ. چن^۱ در مقاله‌ای تحت عنوان

“Holdup and Axial Mixing in Bubble Columns Containing Screen Cylinders”

[۱۸] ، با اولین تقریب کمیت

$$(1) \quad 2 \sum_{n=1}^{\infty} \frac{S_n \sin S_n}{S_n^2 + M^2 + 2M} \exp \left[M - \frac{S_n^2 + M^2}{2M} \cdot \frac{t}{\theta} \right]$$

که در آن t و M پارامترهای فیزیکی بوده و S_n ها کوچکترین مقادیری (از حیث اندازه) هستند که در

$$S_n \tan \left(\frac{S_n}{2} \right) = M$$

وقتی n فرد است ،

و در

$$S_n \cot \left(\frac{S_n}{2} \right) = -M$$

وقتی n زوج است

صدق می‌کند ، ارتفاع ایست بنزین در یک حباب استوانه‌ای را محاسبه کرده است .

(A) با فرض $M = 3.7$ ، S_1 ، S_2 ، S_3 ، و S_4 را بیابید .

(ب) با استفاده از نتایج قسمت (A) ، مجموع مذکور در معادله (۱) را ، وقتی $t = 0$ ، تقریب کنید .

۵.۲ تحلیل خطا در روشهای تکراری و روشهای شتاب

این بخش به بررسی مرتبه همگرایی روشهای تکراری تابعی اختصاص دارد و ، به عنوان ابزاری برای رسیدن به همگرایی سریع ، روش نیوتن را مجدداً " بدست می‌آوریم . همچنین ، راههای دیگر شتاب دادن به همگرایی را مورد توجه قرار می‌دهیم ، اما پیش از آنکه بتوان روشهایی برای شتاب دادن به همگرایی ارائه کرد لازم است روندی برای سنجش سرعت همگرایی تعریف شود .

تعریف ۱۰.۲ . فرض کنیم $\{p_n\}_{n=0}^{\infty}$ دنباله‌ای همگرا به p بوده و ، به ازای هر $n \geq 0$

$e_n = p_n - p$. هرگاه ثابتهای مثبتی چون λ و α باشند بطوری که

$$\lim_{n \rightarrow \infty} \frac{|e_{n+1}|}{|e_n|^\alpha} = \lambda,$$

گوییم $\{p_n\}_{n=0}^{\infty}$ همگرا به p از مرتبه α با ثابت خطای مجانبی λ است .

یک روش تکراری برای حل مسئله‌ای به شکل $x = g(x)$ را از مرتبه α گوئیم اگر، هر وقت روش برای دنباله $\{p_n\}_{n=0}^{\infty}$ ، که در آن به ازای $n \geq 1$ ، $p_n = g(p_{n-1})$ ، همگرایی بدست دهد، دنباله $\{p_n\}_{n=0}^{\infty}$ به جواب از مرتبه α همگرا باشد.

به دو حالت توجه خاص مبذول می‌شود. اگر $\alpha = 1$ ، روش خطی نامیده می‌شود. مثلاً، الگوریتم تنصیف ۱۰۲ به همگرایی خطی منجر می‌شود (ر.ک. تمرین ۹). یک روش مرتبه دو، یعنی $\alpha = 2$ ، روش درجه دو نام دارد، و همگرایی در چنین حالت همگرایی درجه دو نامیده می‌شود. مثال زیر نشان می‌دهد که تکرار نقطه ثابت عموماً خطی است، و نیز همگرایی خطی را با همگرایی درجه دو مقایسه می‌کند.

مثال ۱

(T) فرض کنید، با استفاده از تکرار نقطه ثابت $p_n = g(p_{n-1})$ به ازای $n \geq 1$ ، بخواهیم یک جواب تقریبی $g(x) = x$ را بیابیم.

با این فرض که g بازه $[a, b]$ را بتوی خودش می‌نگارد و عدد مثبتی چون k وجود دارد که به ازای هر $x \in [a, b]$ ، $|g'(x)| \leq k < 1$ ، بنا بر قضیه ۵۰۲ (ص ۴۴)، g نقطه ثابت منحصر بفردی مانند $p \in [a, b]$ دارد، و اگر $p_0 \in [a, b]$ ، دنباله نقطه ثابت $\{p_n\}_{n=0}^{\infty}$ همگرا به p می‌باشد. نشان می‌دهیم که همگرایی خطی است. گوئیم اگر n یک عدد صحیح مثبت باشد،

$$e_{n+1} = p_{n+1} - p = g(p_n) - g(p) = g'(\xi_n)(p_n - p) = g'(\xi_n)e_n,$$

که در آن ξ_n بین p_n و p است. چون $\{p_n\}_{n=0}^{\infty}$ همگرا به p است، $\{\xi_n\}_{n=0}^{\infty}$ نیز همگرا به p می‌باشد. با فرض پیوسته بودن g' بر $[a, b]$ ، داریم

$$\lim_{n \rightarrow \infty} g'(\xi_n) = g'(p).$$

از اینرو،

$$\lim_{n \rightarrow \infty} \frac{e_{n+1}}{e_n} = \lim_{n \rightarrow \infty} g'(\xi_n) = g'(p)$$

و

$$\lim_{n \rightarrow \infty} \frac{|e_{n+1}|}{|e_n|} = |g'(p)|.$$

بنابراین، تکرار نقطه ثابت عموماً "همگرایی خطی را نمایش می دهد .

(ب) برای مقایسه همگرایی خطی و درجه دو، فرض کنیم دو روش تکراری داریم که به صورت زیر توصیف می شوند:

$$(20.2) \quad \lim_{n \rightarrow \infty} \frac{|e_{n+1}|}{|e_n|} = \lambda \quad , \quad \text{که در آن } 0 < \lambda < 1$$

$$(21.2) \quad \lim_{n \rightarrow \infty} \frac{|\bar{e}_{n+1}|}{|\bar{e}_n|^2} = \bar{\lambda} \quad , \quad \text{که در آن } \bar{\lambda} > 0 \text{ و } |\bar{\lambda}\bar{e}_0| < 1$$

همچنین، فرض کنیم

$$|e_{n+1}| \leq \lambda |e_n|$$

و

$$|\bar{e}_{n+1}| \leq \bar{\lambda} |\bar{e}_n|^2 \quad , \quad n \geq 0$$

بنابراین

$$|e_{n+1}| \leq \lambda^{n+1} |e_0|$$

و، به ازای هر $n \geq 0$

$$|\bar{e}_{n+1}| \leq \bar{\lambda} |\bar{e}_n|^2 \leq \bar{\lambda} \bar{\lambda}^2 |\bar{e}_{n-1}|^4 \leq \dots \leq \bar{\lambda}^{(2^{n+1}-1)} |\bar{e}_0|^{2^{n+1}} \\ = (\bar{\lambda} |\bar{e}_0|)^{(2^{n+1}-1)} |\bar{e}_0|$$

فرض کنیم $\lambda = \bar{\lambda} = .75$ ، $|e_0| = |\bar{e}_0| = 1$ ، و بخواهیم تعداد تکرارهای لازم برای داشتن خطایی ناپیشتتر از 10^{-8} را بدانیم. در روش خطی، N ، یعنی تعداد تکرارهای لازم، باید دارای این خاصیت باشد که

$$(.75)^{N+1} \leq 10^{-8}$$

یا

$$N + 1 \geq \frac{-8}{\log_{10}.75} \approx 64.$$

بنابراین، روش خطی (20.2) حدود 63 تکرار لازم دارد.

به همین نحو، در روش درجه دو لازم است که

$$(.75)^{(2^{N+1}-1)} \leq 10^{-8}$$

یا

$$2^{n+1} \geq \frac{-8}{\log_{10} 0.75} + 1 \approx 65,$$

و روش درجه دو به بیش از پنج یا شش تکرار نیاز ندارد. در این وضع، روش به طور درجه دو همگرا برتری قابل ملاحظه‌ای خواهد داشت.

(پ) خاصیت جالب دیگر روشهای درجه دو را می‌توان به صورت زیر توضیح داد: فرض کنیم تکرارها (p_n ها) حدود 1 باشند، و $|e_n| = 10^{-b_n}$ ، $b_n > 0$ ، یعنی، اساساً "تعداد ارقام دقیق در تکرار n م باشد. چون $|e_{n+1}| \leq \lambda |e_n|^2$ ،

$$10^{-(b_{n+1})} \leq \lambda 10^{-2b_n}.$$

با حل آن نسبت به b_{n+1} ،

$$-b_{n+1} \leq \log_{10} \lambda - 2b_n,$$

در نتیجه،

$$b_{n+1} \geq 2b_n - \log_{10} \lambda,$$

و تعداد ارقام دقیق در هر مرحله از یک روش درجه دو تقریباً "دو برابر می‌شود.

حال که مطلوبیت روشهای به طور درجه دو همگرا بحث شد، قدم بعدی تعیین و توصیف آن روشهای تکراری تابعی است که درجه دو هستند.

قضیه ۱۱.۲. فرض کنیم p یک جواب $x = g(x)$ باشد. همچنین، $g'(p) = 0$ و g'' در بازه باز p پیوسته باشد. در این صورت، $\delta > 0$ ای وجود دارد بطوری که، به ازای $p_0 \in [p - \delta, p + \delta]$ ، تکرار نقطه ثابت یک روش مرتبه دو می‌شود. بعلاوه، هرگاه به ازای $p - \delta \leq x \leq p + \delta$ ، $|g''(x)|/2 \leq M$ ، و $|e_0| < 1$ ، آنگاه تکرار نقطه ثابت همگرا بوده و همگرایی درجه دو خواهد بود.

برهان. $\delta > 0$ را طوری اختیار می‌کنیم که بر بازه $[p - \delta, p + \delta]$ ، $|g'(x)| \leq k < 1$ و g'' پیوسته باشد. چون $|g'(x)| \leq k < 1$ ، جملات دنباله $\{p_n\}_{n=0}^{\infty}$ در $[p - \delta, p + \delta]$ قرار دارند. $g(x)$ را به ازای $x \in [p - \delta, p + \delta]$ به صورت یک چند جمله‌ای تیلور بسط می‌دهیم:

$$g(x) = g(p) + g'(p)(x - p) + \frac{g''(\xi)}{2}(x - p)^2,$$

که در آن ξ بین x و p قرار دارد. از مفروضات $g(p) = p$ و $g'(p) = 0$ نتیجه می‌گیریم که

$$g(x) = p + \frac{g''(\xi)}{2}(x-p)^2.$$

بخصوص، وقتی به‌ازای n ی $x = p_n$ ،

$$p_{n+1} = g(p_n) = p + \frac{g''(\xi_n)}{2}(p_n - p)^2$$

که در آن ξ_n بین p_n و p است. بنابراین ،

$$e_{n+1} = \frac{g''(\xi_n)}{2} e_n^2. \quad (22.2)$$

چون بر $[p - \delta, p + \delta]$ ، $|g'(x)| \leq k < 1$ ، از قضیه ۵.۲ (ص ۴۴) نتیجه می‌شود که $\{p_n\}_{n=0}^{\infty}$ همگرا به p است. چون ξ_n به‌ازای هر n بین p و p_n است ، $\{\xi_n\}_{n=0}^{\infty}$ نیز همگرا به p است ، و

$$\lim_{n \rightarrow \infty} \frac{|e_{n+1}|}{|e_n|^2} = \frac{|g''(p)|}{2}$$

این نشان می‌دهد که روش از مرتبه ۲ دو است. اگر به‌ازای هر $x \in [p - \delta, p + \delta]$ ، $|g''(x)|/2 \leq M$ معادله (۲۲.۲) ایجاب می‌کند که

$$|e_{n+1}| \leq M|e_n|^2 \leq M^3|e_{n-1}|^4 \leq \dots \leq (M|e_0|)^{(2^{n+1}-1)}|e_0|.$$

در نتیجه ، فرض $M|e_0| < 1$ برای آنکه روش به‌طور درجه ۲ دو همگرا باشد کافی است .

برای آنکه از قضیه ۱۱.۲ در حل معادله‌ای به‌شکل $f(x) = 0$ استفاده شود ، فرض می‌کنیم معادله $f(x) = 0$ جوابی چون p داشته باشد که در آن $f'(p) \neq 0$. روش نقطه ثابت

$$p_n = g(p_{n-1}), \quad n \geq 1$$

را در نظر می‌گیریم ، با g به‌شکل

$$g(x) = x - \phi(x)f(x),$$

که در آن $\phi(x)$ یک تابع دلخواه است که بعداً انتخاب می‌شود .

هرگاه $\phi(x)$ کراندار باشد ، $g(p) = p$ ، و ، برای آنکه روند تکراری ناشی از g به‌طور

درجه ۲ دو همگرا باشد ، کافی است داشته باشیم $g'(p) = 0$. اما

$$g'(x) = 1 - \phi'(x)f(x) - f'(x)\phi(x)$$

و

$$g'(p) = 1 - f'(p)\phi(p).$$

در نتیجه، $g'(p) = 0$ اگر و فقط اگر $\phi(p) = 1/f'(p)$

بخصوص، همگرایی درجه دو برای روش

$$(23.2) \quad p_n = g(p_{n-1}) = p_{n-1} - \frac{f(p_{n-1})}{f'(p)}$$

تحت شرایط مناسبی بر f برقرار است. بهر حال، چون p ، و در نتیجه $f(p)$ ، عموماً " مجهولند، بهترین کار این است که فرض کنیم $\phi(x) = 1/f'(x)$ ؛ در نتیجه، $\phi(p) = 1/f'(p)$. در این صورت، روند به شکل زیر درمی آید:

$$(24.2) \quad p_n = g(p_{n-1}) = p_{n-1} - \frac{f(p_{n-1})}{f'(p_{n-1})}$$

که روش نیوتن می باشد.

در بحث بالا لازم بود در جواب p ، $f'(p) \neq 0$. هر وقت p یک جواب معادله $f(x) = 0$ باشد، می توان تابع منحصر بفرد $q(x)$ و عدد صحیح و مثبت m را طوری یافت که

$$(25.2) \quad f(x) = (x - p)^m q(x).$$

که در آن p ریشه $q(x) = 0$ نباشد. عدد صحیح m مرتبه تکرار ریشه p نامیده می شود. می توان به آسانی نشان داد (ر. ک. تمرین ۱۴) که در یک جواب $f(x) = 0$ درست وقتی $f'(p) \neq 0$ که $m = 1$. در نتیجه، در بحث قبلی عملاً "فرض شده که p یک ریشه f از مرتبه تکرار یک است. در حالتی که p یک ریشه از مرتبه تکرار $m > 1$ باشد، اغلب می توان نشان داد که روش نیوتن به همگرایی منجر می شود ولی نه به صورت درجه دو. دلیلش این است که، اگر $f(x) = (x - p)^m q(x)$ ، که در آن $m > 1$ ، مرتبه تکرار p باشد،

$$\begin{aligned} g(x) &= x - \frac{f(x)}{f'(x)} \\ &= x - \frac{(x - p)^m q(x)}{(x - p)^m q'(x) + m(x - p)^{m-1} q(x)} \\ &= x - \frac{(x - p)q(x)}{(x - p)q'(x) + mq(x)} \end{aligned}$$

$$g'(x) = 1 - \frac{\{[q(x) + (x-p)q'(x)][(x-p)q'(x) + mq(x)] - [(x-p)q(x)] \times \{[q'(x)] + [(x-p)q''(x) + mq'(x)]\}}{[(x-p)q'(x) + mq(x)]^2}$$

بنابراین، $g'(p) = 1 - (1/m) \neq 0$ و قضیه ۱۱.۲ دیگر قابل اعمال نیست. بهر حال، با پیرایش جزئی روش نیوتن، می توان حالتی که مرتبه تکرار معلوم است را بررسی کرد. به جای روند متداول، فرض می کنیم

$$(26.02) \quad g(x) = x - \frac{mf(x)}{f'(x)}$$

از بحث بالا روشن است که برای این g ، $g'(p) = 1 - (m/m) = 0$ و همگرایی درجه دو بدست می آید.

این پیرایش روش نیوتن - رفسون ظاهراً "مشکل ریشه های مکرر را برطرف می کند، لیکن در عمل مرتبه تکرار در آغاز مجهول است؛ در نتیجه، مشکل به قوت خود باقی می ماند.

روش موفق تر دیگر برای پرداختن به این مسئله این است که تابع $\mu(x)$ را با

$$(27.02) \quad \mu(x) = \frac{f(x)}{f'(x)}$$

تعریف کنیم.

هرگاه p یک ریشه f از مرتبه تکرار $m \geq 1$ بوده و $f(x) = (x-p)^m q(x)$ ، آنگاه

$$\begin{aligned} \mu(x) &= \frac{(x-p)^m q(x)}{m(x-p)^{m-1} q(x) + (x-p)^m q'(x)} \\ &= \frac{(x-p)q(x)}{mq(x) + (x-p)q'(x)} \end{aligned}$$

نیز یک ریشه در p منتها از مرتبه تکرار یک دارد. در این صورت، روش نیوتن در مورد تابع μ قابل اجراست و نتیجه می دهد که

$$\begin{aligned} g(x) &= x - \frac{\mu(x)}{\mu'(x)} \\ &= x - \frac{f(x)/f'(x)}{\{[f'(x)]^2 - [f(x)][f''(x)]\}/[f'(x)]^2} \end{aligned}$$

$$(۲۸.۲) \quad g(x) = x - \frac{f(x)f'(x)}{[f'(x)]^2 - [f(x)][f''(x)]}$$

اگر g شرایط پیوستگی لازم را داشته باشد، تکرار تابعی اعمال شده بر g ، صرف نظر از مرتبه تکرار ریشه f ، به طور درجه دو همگرا می باشد. تنها نقص این روش محاسبه اضافی $f''(x)$ و محاسبه پر زحمت تر تکرارها است.

مثال ۲

(آ) در مثال ۱ از بخش ۳.۲، می بایست $f(x) = x^3 + 4x^2 - 10 = 0$ را نسبت به ریشه $p = 1.36523001$ حل می کردیم. در این مثال، مقدار ریشه را می دانیم و می توان به آسانی $f'(p)$ را حساب کرد و دید که مساوی 16.51339907 است. جهت مقایسه همگرایی برای یک ریشه از مرتبه تکرار یک با استفاده از معادله (۲۳.۲)، روش نیوتن، و روش پیراسته نیوتن مذکور در معادله (۲۸.۲)، قرار می دهیم

$$(یک) \quad p_n = p_{n-1} - \frac{p_{n-1}^3 + 4p_{n-1}^2 - 10}{16.51339907};$$

$$(دو) \quad p_n = p_{n-1} - \frac{p_{n-1}^3 + 4p_{n-1}^2 - 10}{3p_{n-1}^2 + 8p_{n-1}}; \text{ از روش نیوتن};$$

و، از معادله (۲۸.۲)،

$$p_n = p_{n-1} - \frac{(p_{n-1}^3 + 4p_{n-1}^2 - 10)(3p_{n-1}^2 + 8p_{n-1})}{(3p_{n-1}^2 + 8p_{n-1})^2 - (p_{n-1}^3 + 4p_{n-1}^2 - 10)(6p_{n-1} + 8)}. \text{ (سه)}$$

به ازای $p_0 = 1.5$ ، سه تکرار اول برای (یک)، (دو)، و (سه) عبارتند از

p_1	p_2	p_3	
1.3561774	1.36518988	1.36523001	(یک)
1.3733333	1.36526201	1.36523001	(دو)
1.3568990	1.36519585	1.36523001	(سه)

(ب) برای توضیح حالت یک ریشه مکرر، معادله $f(x) = x^4 - 4x^2 + 4 = 0$ را در نظر می گیریم که یک ریشه از مرتبه تکرار دو در $x = \sqrt{2}$ دارد. چون $f'(\sqrt{2}) = 0$ ، روش توصیف شده با معادله (۲۳.۲) را حتی نمی توان وقتی ریشه از قبل معلوم است بکاربرد.

اگر روش اول را با معادله^۲ (۲۶.۲) بهازای $m = 2$ عوض کنیم ،

$$(یک) \quad p_n = p_{n-1} - \frac{(p_{n-1}^2 - 2)}{2p_{n-1}} \quad ، \quad \text{از معادله}^۲ (۲۶.۲) :$$

$$(دو) \quad p_n = p_{n-1} - \frac{(p_{n-1}^2 - 2)}{4p_{n-1}} \quad ، \quad \text{از روش نیوتن} ؛$$

و

$$(سه) \quad p_n = p_{n-1} - \frac{(p_{n-1}^2 - 2)p_{n-1}}{(p_{n-1}^2 + 2)} \quad ، \quad \text{از معادله}^۲ (۲۸.۲) .$$

بهازای $p_0 = 1.5$ ، سه تکرار اول برای (یک) ، (دو) ، و (سه) عبارتند از

p_1	p_2	p_3	
1.416666667	1.414215686	1.414213562	(یک)
1.458333333	1.436607143	1.425497619	(دو)
1.411764706	1.414211438	1.414213562	(سه)

جواب واقعی دقیق تا 10^{-9} مقدار p_3 در هر یک از (یک) و (سه) است . برای آنکه به روش متعارف نیوتن - رفسون به این دقت برسیم 20 تکرار لازم خواهد بود .

حال یک روش درجه^۲ دو ارائه می دهیم که وابسته به f' نیست . این روش یک روش شتاب است به نام فرایند Δ^2 ایکن^۱ ، و می توان از آن برای شتاب دادن به همگرایی یک دنباله^۲ به طور خطی همگرا ، قطع نظر از ماهیتش ، استفاده کرد .

فرض کنیم $\{p_n\}_{n=0}^{\infty}$ یک دنباله^۲ به طور خطی همگرا با حد p باشد ؛ یعنی ، بهازای

$$e_n = p_n - p$$

$$0 < \lambda < 1 \quad \text{و} \quad \lim_{n \rightarrow \infty} \frac{|e_{n+1}|}{|e_n|} = \lambda$$

برای ساختن دنباله^۲ $\{\tilde{p}_n\}_{n=0}^{\infty}$ ، که سریعتر به p همگرا باشد ، فرض می کنیم حالت حدی بهازای هر عدد صحیح مثبت n رخ می دهد ، و همه^۲ e_n ها یک علامت داشته باشند . در این حالت ،

$$e_{n+1} = \lambda e_n, \quad n \geq 0$$

در نتیجه،

$$p_{n+2} = e_{n+2} + p = \lambda e_{n+1} + p,$$

یا

$$(29.2) \quad p_{n+2} = \lambda(p_{n+1} - p) + p, \quad n \geq 0$$

اگر در معادله (29.2) را با n عوض کنیم، خواهیم داشت

$$(30.2) \quad p_{n+1} = \lambda(p_n - p) + p;$$

و، حل معادلات (29.2) و (30.2) نسبت به p و حذف λ ، نتیجه می‌دهد که

$$\begin{aligned} p &= \frac{p_{n+2}p_n - p_{n+1}^2}{p_{n+2} - 2p_{n+1} + p_n} \\ &= \frac{p_n^2 + p_n p_{n+2} + 2p_n p_{n+1} - 2p_n p_{n+1} - p_n^2 - p_{n+1}^2}{p_{n+2} - 2p_{n+1} + p_n} \\ &= \frac{(p_n^2 + p_n p_{n+2} - 2p_n p_{n+1}) - (p_n^2 - 2p_n p_{n+1} + p_{n+1}^2)}{p_{n+2} - 2p_{n+1} + p_n} \\ &= p_n - \frac{(p_{n+1} - p_n)^2}{p_{n+2} - 2p_{n+1} + p_n} \end{aligned}$$

فرض اولیه $e_{n+1} = \lambda e_n$ عموماً "درست نیست؛ با اینحال، انتظار می‌رود که دنباله $\{\hat{p}_n\}_{n=0}^{\infty}$ ، تعریف شده با

$$(31.2) \quad \hat{p}_n = p_n - \frac{(p_{n+1} - p_n)^2}{p_{n+2} - 2p_{n+1} + p_n},$$

از دنباله اصلی $\{p_n\}_{n=0}^{\infty}$ سریعتر به p همگرا باشد.

مثال ۳. دنباله $\{p_n\}_{n=0}^{\infty}$ ، که در آن $p_n = n \ln(1 + (1/n))$ ، به‌طور خطی همگرا به p است. چند جمله اول دنباله $\{p_n\}_{n=1}^{\infty}$ و $\{\hat{p}_n\}_{n=1}^{\infty}$ در جدول ۹.۲ ثبت شده‌اند.

می‌بینیم که $\{\hat{p}_n\}_{n=1}^{\infty}$ سریعتر از $\{p_n\}_{n=1}^{\infty}$ به $p = 1$ همگراست.

این روش ساختن یک دنباله سریعتر همگرا، بخاطر تعریف زیر، روش Δ^2 اینتن

نام دارد.

n	p_n	\hat{p}_n
1	.6931471805	.9044075720
2	.8109302162	.9311744212
3	.8630462166	.9461287492
4	.8925742052	.9557185842
5	.9116077840	.9624007828
6	.9249040806	
7	.9347197496	

جدول ۹.۰۲

تعریف ۱۲.۰۲ . به ازای دنباله $\{p_n\}_{n=0}^{\infty}$ ، تفاضل پیشرو Δp_n به صورت زیر تعریف می شود :

$$\Delta p_n = p_{n+1} - p_n, \quad n \geq 0$$

توانهای بالاتر $\Delta^k p_n$ به طور بازگشتی به صورت زیر تعریف می شوند :

$$\Delta^k p_n = \Delta^{k-1}(\Delta p_n), \quad k \geq 2.$$

بخاطر تعریف ،

$$\begin{aligned} \Delta^2 p_n &= \Delta(p_{n+1} - p_n) \\ &= \Delta p_{n+1} - \Delta p_n \\ &= (p_{n+2} - p_{n+1}) - (p_{n+1} - p_n) \\ &= p_{n+2} - 2p_{n+1} + p_n. \end{aligned}$$

بنابراین ، فرمول مربوط به \hat{p}_n را می توان به صورت زیر نوشت :

$$\hat{p}_n = p_n - \frac{(\Delta p_n)^2}{\Delta^2 p_n}, \quad n \geq 0$$

تا اینجا در بحث روش Δ^2 ایتکن گفته ایم که دنباله $\{p_n\}_{n=0}^{\infty}$ از دنباله اصلی $\{p_n\}_{n=0}^{\infty}$ سریعتر به p همگراست ، اما دقیقاً " نگفته ایم که منظور ما از همگرایی " سریعتر " در این حالت چیست .

قضیه زیر این اصطلاح را توضیح داده و توجیه می کند .

قضیه ۱۳.۰۲ . فرض کنیم p_n دنباله ای همگرا به حد p بوده و $e_n = p_n - p$ ، که در آن

به ازای هر $n \geq 0$ ، $e_n \neq 0$ ، همچنین،

$$\lim_{n \rightarrow \infty} \frac{e_{n+1}}{e_n} = \lambda \quad \text{که در آن } 0 < \lambda < 1.$$

در این صورت، دنباله $\{\hat{p}_n\}_{n=0}^{\infty}$ سریعتر از $\{p_n\}_{n=0}^{\infty}$ همگرا به p است به این معنی که

$$\lim_{n \rightarrow \infty} \frac{\hat{p}_n - p}{p_n - p} = 0.$$

برهان. ر.ک. تمرین ۱۲.

روش Δ^2 ایتن همگرایی یک دنباله به طور خطی همگرا را، قطع نظر از منشاء آن، شتاب می‌دهد. با اعمال روش Δ^2 ایتن بر تکرار نقطه ثابت روندی به نام روش استفنسن^۱ بدست می‌آید.

الگوریتم استفنسن ۵.۲

به فرض آنکه معادله $g(x) = x$ و تقریب اولیه $p_0^{(0)}$ داده شده باشند:

- مرحله ۱. قرار می‌دهیم $i = 0$.
- مرحله ۲. قرار می‌دهیم $p_1^{(i)} = g(p_0^{(i)})$.
- مرحله ۳. قرار می‌دهیم $p_2^{(i)} = g(p_1^{(i)})$.
- مرحله ۴. قرار می‌دهیم

$$p_0^{(i+1)} = p_0^{(i)} - \frac{(\Delta p_0^{(i)})^2}{\Delta^2 p_0^{(i)}} = p_0^{(i)} - \frac{[p_1^{(i)} - p_0^{(i)}]^2}{[p_2^{(i)} - 2p_1^{(i)} + p_0^{(i)}]}$$

- مرحله ۵. مشخص می‌کنیم که آیا روش باید ادامه یابد.
- اگر باید، به \bar{i} یک می‌افزاییم و به مرحله ۲ می‌رویم.
- در غیر این صورت، به مرحله ۶ می‌رویم.
- مرحله ۶. روند تمام خواهد بود.

مثال ۰۴. برای حل $x^3 + 4x^2 - 10 = 0$ به روش استفنسن، قرار می‌دهیم $x^3 + 4x^2 = 10$ و آن را با تقسیم بر $x + 4$ نسبت به x حل می‌کنیم. لذا، اگر

$$g(x) = \left(\frac{10}{x+4} \right)^{1/2},$$

$x = g(x)$ ایجاب می‌کند که $x^3 + 4x - 10 = 0$.

به ازای $p_0 = 1.5$ ، روند استفنسن نتیجه می‌دهد که

k	$p_0^{(k)}$	$p_1^{(k)}$	$p_2^{(k)}$
0	1.5	1.348399725	1.367376372
1	1.365265224	1.365225534	1.365230583
2	1.365230013		

تکرار $p_0^{(2)} = 1.365230013$ تا نه رقم اعشار دقیق است. جالب است توجه شود که، در این مثال، روش استفنسن تقریباً "همان مرتبه" همگرایی روش نیوتن را بدست می‌دهد (ر. ک. مثال ۲).

توجه کنید که $\Delta^2 p_n$ ممکن است صفر باشد. در این حالت، دنباله را ختم و $p_2^{(n-1)}$ را جواب تقریبی می‌گیریم، چرا که، در غیر این صورت، در مخرج تکرار بعدی صفر ایجاد می‌شود.

از آخرین مثال دیده می‌شود که روش استفنسن همگرایی درجه دو را بدون محاسبه مشتق بدست می‌دهد. قضیه‌ای با این نتیجه در زیر بیان شده است. برهان این قضیه در اینجا عرضه نمی‌شود، لیکن می‌توان آن را در هنریچی [۳۴]، صفحات ۹۰ تا ۹۲، یا در ایزاکسون و کلر [۳۹]، صفحات ۱۰۳ تا ۱۰۷، یافت.

قضیه ۱۴۰۲. فرض کنیم $x = g(x)$ دارای جواب p باشد و $g'(p) \neq 1$. هرگاه $\delta > 0$ باشد بطوری‌که $g \in C^3[p - \delta, p + \delta]$ ، آنگاه روش استفنسن به ازای هر $p_0 \in [p - \delta, p + \delta]$ همگرایی درجه دو بدست می‌دهد.

مجموعه تمرینات ۵.۲

۱. با استفاده از روش استفسن و تمرین ۳ از بخش ۳.۰۲، ریشه $x^3 - x - 1 = 0$ را در $[1, 2]$ با دقت 10^{-4} بدست آورید.
۲. با استفاده از روش استفسن، ریشه $x - 2^{-x} = 0$ در $[0, 1]$ را با دقت 10^{-4} بدست آورده و با نتایج تمرین ۲ از بخش ۳.۰۲ مقایسه کنید.
۳. با استفاده از روش استفسن و $p_0 = 2$ ، $\sqrt{3}$ را تا چهار رقم با معنی حساب کنید. این نتیجه را با نتایج تمرین ۸ از بخش ۴.۰۲، تمرین ۴ از بخش ۳.۰۲، و تمرین ۴ از بخش ۱.۰۲ مقایسه نمایید.
۴. با استفاده از روش استفسن، جوابهای تقریبی معادلات زیر را با تقریب 10^{-5} بدست آورید:

(ت) $x = \frac{2 - e^x + x^2}{3}$ ، که در آن g تابع تمرین ۶ (ت) از بخش ۳.۰۲ است؛

(ب) $3x^2 - e^x = 0$ ، که در آن g تابع تمرین ۷ (ت) از بخش ۳.۰۲ است؛

(پ) $x - \cos x = 0$ ، که در آن g تابع تمرین ۷ (ب) از بخش ۳.۰۲ است.

۵. با استفاده از معادله (۲۶.۰۲) به ازای $m = 2$ و تکرار تابعی، و با شروع از $p_0 = \pi/2$ ، یک ریشه $f(x) = (\sin x - (x/2))^2$ را با دقت 10^{-5} پیدا کنید. نتایج خود را با نتایج تمرین ۷ در بخش ۴.۰۲ مقایسه کنید.

۶. با استفاده از معادله (۲۸.۰۲) و تکرار تابعی، و با شروع از $p_0 = \pi/2$ ، یک ریشه $f(x) = (\sin x - (x/2))^2$ را با دقت 10^{-5} پیدا کنید. نتایج خود را با نتایج تمرین ۵ در فوق و تمرین ۷ در بخش ۶.۰۲ مقایسه کنید.

۷. با استفاده از روند توصیف شده در معادله (۲۶.۰۲) به ازای $m = 1, 2$ ، و با شروع از $p_0 = 0$ و انجام ۱۰ تکرار در هر حالت، یک ریشه تقریبی معادله

$$f(x) = x^2 + 2xe^x + e^{2x}$$

را پیدا کنید. آیا می‌توانید مرتبه تکرار ریشه بدست آمده را مشخص کنید؟ اگر با کامپیوتر کار می‌کنید، محاسبات خود را با استفاده از دقت مضاعف مجدداً انجام داده، اثرات خطای گرد کردن را بر محاسبات با دقت معمولی بررسی کنید. آیا می‌توانید بگویید چرا خطای گرد کردن این قدر زیاد است؟

۸. با استفاده از روش پیراسته نیوتن-رفسون در (۲۸.۰۲) ، و با شروع از $p_0 = 0$ و

انجام 10 تکرار، ریشه تقریبی

$$f(x) = x^2 + 2xe^x + e^{2x}$$

را پیدا کنید. اگر با کامپیوتر کار می‌کنید، محاسبات را با استفاده از دقت مضاعف

نجدداً انجام داده و نتایج خود را با نتایج مسئله ۷ مقایسه کنید.

۹. نشان دهید که الگوریتم تنصیف ۱۰.۲ همگرایی خطی بدست می‌دهد.

۱۰. نشان دهید که روش تکراری حل $f(x) = 0$ ، که با

$$p_n = p_{n-1} - \frac{f(p_{n-1})}{f'(p_{n-1})} - \frac{f''(p_{n-1})}{2f'(p_{n-1})} \left[\frac{f(p_{n-1})}{f'(p_{n-1})} \right]^2, \quad n = 1, 2, 3, \dots,$$

داده شده، عموماً "همگرایی درجه سه ($\alpha = 3$) بدست می‌دهد. با تحلیل مثال

۱ (ب)، همگرایی درجه دو و سه را با هم مقایسه کنید.

۱۱. فرض کنید روش خط قاطع به ریشه p از $f(x)$ همگرا باشد، f دوبار به‌طور پیوسته

مشقی‌پذیر باشد، و $f'(p) \neq 0$ نشان دهید که

$$\lim_{n \rightarrow \infty} \frac{e_{n+1}}{e_n e_{n-1}} = \frac{f''(p)}{2f'(p)}$$

و این ایجاب می‌کند که همگرایی روش خط قاطع بیش از خطی است اما کاملاً "درجه"

دو نیست.

۱۲. قضیه ۱۳.۲ را ثابت کنید.

۱۳. مفهوم مذکور در تعریف ۱۰.۲ را می‌شد طوری تعریف کرد که دنباله‌هایی که همگرا

نباشند رانیز شامل شود. گاهی دنباله $\{p_n\}_{n=0}^{\infty}$ را نسبت به عدد p از مرتبه α

گویند اگر

$$\lim_{n \rightarrow \infty} \frac{|p_{n+1} - p|}{|p_n - p|^\alpha} = \lambda > 0 \quad \text{به‌ازای } \lambda \text{ ای}$$

(آ) فرض کنید $\{p_n\}_{n=0}^{\infty}$ نسبت به p از مرتبه $\alpha = 1$ باشد.

(یک) نشان دهید که اگر $\lambda < 1$ ، $\{p_n\}_{n=0}^{\infty}$ همگرا به p است.

(دو) نشان دهید که اگر $\lambda > 1$ ، $\{p_n\}_{n=0}^{\infty}$ همگرا به p نیست.

(سه) درحالی که $\lambda = 1$ چه می‌توان نتیجه گرفت؟

(ب) فرض کنید $\{p_n\}_{n=0}^{\infty}$ نسبت به p از مرتبه $\alpha > 1$ باشد. نشان دهید که اگر

$$|\lambda(p - p_0)| < 1 \quad \{p_n\}_{n=0}^{\infty} \text{ همگرا به } p \text{ است.}$$

۱۴. نشان دهید p یک ریشه از مرتبه p تکرار یک تابع مشتقپذیر f است اگر و فقط اگر

$$f(p) = 0 \text{ و } f'(p) \neq 0$$

۶.۲ صفرهای چند جمله‌ایهای حقیقی

یک تابع به شکل

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0,$$

که در آن a_i ها ثابت بوده و $a_n \neq 0$ ، یک چند جمله‌ای از درجه n نام دارد. همه چند جمله‌ایهای این بخش حقیقی فرض می‌شوند؛ یعنی، همه a_i ها اعداد حقیقی می‌باشند. تابع صفر، یعنی $P(x) = 0$ ، بازای هر مقدار از x ، را معمولا "یک چند جمله‌ای می‌گیرند اما به آن درجه‌ای نسبت نمی‌دهند.

قضیه ۱۵.۲ (قضیه اساسی جبر). هرگاه $P(x)$ یک چند جمله‌ای از درجه $n \geq 1$ باشد، آنگاه $P(x) = 0$ (لااقل یک ریشه (یا صفر) دارد؛ یعنی، لااقل یک عدد مانند \hat{x} ، احتمالا "مختلط، هست که $P(\hat{x}) = 0$.

با آنکه قضیه ۱۵.۲ در بررسی توابع مقدماتی اساسی است، برهانش نیاز به روشهایی از نظریه توابع مختلط دارد. برای بیان اصولی مطالب لازم در اثبات قضیه ۱۵.۲ خواننده را به دتمن^۱ [۲۲]، صفحه ۱۱۶، ارجاع می‌دهیم. یک نتیجه مهم، قضیه ۱۵.۲ عبارت است از

نتیجه ۱۶.۲. هرگاه $P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ یک چند جمله‌ای از درجه $n \geq 1$ باشد، آنگاه ثابتهایی منحصر بفرد چون x_1, x_2, \dots, x_k ، احتمالا "مختلط، و اعدادی مثبت چون m_1, m_2, \dots, m_k وجود دارند بطوری که $\sum_{i=1}^k m_i = n$ و

$$P(x) = a_n (x - x_1)^{m_1} (x - x_2)^{m_2} \dots (x - x_k)^{m_k}.$$

برهان. ر.ک. تمرین ۳ (آ).

اگر $(x - x_i)^m$ یک عامل $P(x)$ باشد، x_i را یک ریشه (یا صفر) از مرتبه m_i تکرار می‌گوییم. نتیجه ۱۶.۰۲ می‌گوید که ریشه‌های یک چند جمله‌ای منحصر بفردهستند، و اگر هر ریشه x_i به تعداد مرتبه m_i تکرارش حساب شود، یک چند جمله‌ای درجه n درست n ریشه خواهد داشت.

نتیجه زیر از قضیه اساسی جبر اغلب در این بخش و در فصول آتی بکار می‌رود. برهان این نتیجه در تمرین ۳ (ب) خواسته می‌شود.

نتیجه ۱۷.۰۲. فرض کنیم P و Q چند جمله‌ایهایی از درجه حداکثر n باشند. هرگاه $k > n$ ، x_1, x_2, \dots, x_k نقاط متمایزی باشند بطوری که به ازای $i = 1, 2, \dots, k$ ، $P(x_i) = Q(x_i)$ ، آنگاه، به ازای هر مقدار x ، $P(x) = Q(x)$.

قضیه ۱۸.۰۲. هرگاه $z = a + bi$ یک صفر مختلط چند جمله‌ای P از مرتبه m تکرار باشد، آنگاه $\bar{z} = a - bi$ نیز یک صفر چند جمله‌ای P از مرتبه m تکرار است و $(x^2 - 2ax + a^2 + b^2)^m$ یک عامل P می‌باشد.

برهان. ر.ک. تمرین ۴.

برای آنکه از روند نیوتن-رفسون در تعیین ریشه‌های تقریبی چند جمله‌ای P استفاده کنیم، باید P و مشتق آن را در مقادیر خاص حساب کنیم. اگر بخواهیم چند جمله‌ای درجه n $P(x) = a_n x^n + \dots + a_1 x + a_0$ را در $x = x_0$ محاسبه کنیم، یک روند "طبیعی" این است که ابتدا $x_0^0, x_0^1, \dots, x_0^n$ را حساب کنیم، روندی که به $(n-1)$ ضرب نیاز دارد، بعد $a_i x_0^i$ را به ازای $i = 1, 2, \dots, n$ حساب می‌کنیم، که به n ضرب دیگر نیاز دارد، و بالاخره با انجام n جمع مقدار $P(x_0) = a_0 + a_1 x_0 + a_2 x_0^2 + \dots + a_n x_0^n$ را بدست می‌آوریم. بنابراین، برای محاسبه $P(x_0)$ به این طریق $(2n-1)$ ضرب و n جمع لازم است. قضیه بعد الگوریتمی به نام روش هورنر^۱ یا، به شکل فشرده‌اش، تقسیم ترکیبی بدست می‌دهد. روش هورنر برای محاسبه یک چند جمله‌ای در یک نقطه فقط به n ضرب

n جمع نیاز دارد و، در نتیجه، از حیث زمان لازم برای محاسبه یک چند جمله‌ای از روند "طبیعی" بالا کاراتر است. چون روش هورنر اعمال ریاضی کمتری از روند "طبیعی" می‌طلبد، احتمالاً از نظر تقلیل اثر خطای گرد کردن، مورد بحث در فصل ۱، نیز از آن بهتر است.

قضیه ۱۹.۲ (روش هورنر). فرض کنیم

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 \quad \text{و} \quad b_n = a_n$$

هرگاه $b_k = a_k + b_{k+1} x_0$ ، $k = n-1, n-2, \dots, 1, 0$ ، آنگاه

$$b_0 = P(x_0) \quad \text{بعلاوه، هرگاه}$$

$$Q(x) = b_n x^{n-1} + b_{n-1} x^{n-2} + \dots + b_2 x + b_1,$$

آنگاه

$$(32.2) \quad P(x) = (x - x_0)Q(x) + b_0.$$

برهان. طبق تعریف $Q(x)$ ،

$$\begin{aligned} (x - x_0)Q(x) + b_0 &= (x - x_0)(b_n x^{n-1} + \dots + b_2 x + b_1) + b_0 \\ &= (b_n x^n + b_{n-1} x^{n-1} + \dots + b_2 x^2 + b_1 x) \\ &\quad - (b_n x_0 x^{n-1} + \dots + b_2 x_0 x + b_1 x_0) + b_0 \\ &= b_n x^n + (b_{n-1} - b_n x_0) x^{n-1} + \dots + (b_1 - b_2 x_0) x \\ &\quad + (b_0 - b_1 x_0). \end{aligned}$$

بنابر مفروضات قضیه، $b_n = a_n$ و $b_k - b_{k+1} x_0 = a_k$ ؛ در نتیجه، $(x - x_0)Q(x) + b_0 = P(x)$

$$\text{و} \quad b_0 = P(x_0).$$

برای آنکه روش هورنر آسان اجرا شود، الگوریتم زیر را بیان می‌کنیم.

الگوریتم هورنر ۶.۲

به فرض معلوم بودن چند جمله‌ای

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0,$$

مقدار $P(x_0)$ را پیدا می‌کنیم:

مرحله ۱. قرار می‌دهیم $b_n = a_n$.

مرحله ۲. قرار می‌دهیم $k = n - 1$.

مرحله ۳. قرار می‌دهیم $b_k = a_k + b_{k+1}x_0$.

مرحله ۴. اگر $k \neq 0$ ، به مرحله ۵ می‌رویم.

اگر $k = 0$ ، به مرحله ۶ می‌رویم.

مرحله ۵. ۱ را از k کم کرده به مرحله ۳ می‌رویم.

مرحله ۶. روند تمام خواهد بود.

$$P(x) = (b_n x^{n-1} + b_{n-1} x^{n-2} + \dots + b_2 x + b_1)(x - x_0) + b_0$$

و

$$P(x_0) = b_0.$$

مثال ۱. با استفاده از روش هورنر، $P(x) = 2x^4 - 3x^2 + 3x - 4$ را در $x_0 = -2$ حساب کنید. با استفاده از قضیه ۱۹.۲،

$$b_4 = 2, \quad b_3 = 2(-2) + 0 = -4.$$

$$b_2 = (-4)(-2) - 3 = 5, \quad b_1 = 5(-2) + 3 = -7$$

و، بالاخره،

$$P(-2) = b_0 = (-7)(-2) - 4 = 10.$$

بعلاوه، قضیه ۱۹.۲ نتیجه می‌دهد که

$$P(x) = (x + 2)(2x^3 - 4x^2 + 5x - 7) + 10.$$

وقتی در روش هورنر محاسبات را با دست انجام می‌دهیم، ابتدا یک جدول می‌سازیم، که نام "تقسیم ترکیبی" که اغلب به روش هورنر اطلاق می‌شود از آن ناشی شده است. در مثال قبل، این جدول به صورت زیر است.

	(ضریب x^4)	(ضریب x^3)	(ضریب x^2)	(ضریب x)	(جمله ثابت)
$x_0 = -2$	$a_4 = 2$	$a_3 = 0$	$a_2 = -3$	$a_1 = 3$	$a_0 = -4$
	$b_4 x_0 = -4$	$b_3 x_0 = 8$	$b_2 x_0 = -10$	$b_1 x_0 = 14$	
	$b_4 = 2$	$b_3 = -4$	$b_2 = 5$	$b_1 = -7$	$b_0 = 10$

مزیت دیگر استفاده از روند هورنر (یا تقسیم ترکیبی) این است که، چون

$$P(x) = (x - x_0)Q(x) + b_0,$$

که در آن

$$Q(x) = b_n x^{n-1} + b_{n-1} x^{n-2} + \dots + b_2 x + b_1,$$

مشتگیری نسبت به x نتیجه می دهد که

$$(۳۳.۲) \quad P'(x) = Q(x) + (x - x_0)Q'(x)$$

و

$$(۳۴.۲) \quad P'(x_0) = Q(x_0).$$

از اینرو، وقتی برای یافتن یک ریشه تقریبی چند جمله ای P روش نیوتن - رفسون بکار رود، می توان هر دوی P و P' را به یک طریق محاسبه کرد.

مثال ۲. با استفاده از روند نیوتن - رفسون، یک ریشه تقریبی

$$P(x) = 2x^4 - 3x^2 + 3x - 4$$

را بیابید، و، با استفاده از تقسیم ترکیبی، $P(x_n)$ و $P'(x_n)$ را به ازای هر تکرار x_n محاسبه کنید. با استفاده از $x_0 = -2$ به عنوان یک تقریب اولیه، $P(-2)$ در مثال ۱ را به صورت زیر بدست می آوریم:

$$x_0 = -2 \quad \left| \begin{array}{cccc} 2 & 0 & -3 & 3 & -4 \\ & -4 & 8 & -10 & 14 \\ \hline & 2 & -4 & 5 & -7 & 10 = P(-2) \end{array} \right.$$

با استفاده از قضیه ۱۹.۲ و معادله (۳۴.۲)، داریم

$$P'(-2) = Q(-2) \quad \text{و} \quad Q(x) = 2x^3 - 4x^2 + 5x - 7$$

در نتیجه، $P'(-2)$ را می توان با محاسبه $Q(-2)$ به همین طریق پیدا کرد.

$$x_0 = -2 \quad \left| \begin{array}{ccc} 2 & -4 & 5 & -7 \\ & -4 & 16 & -42 \\ \hline & 2 & -8 & 21 & -49 = Q(-2) = P'(-2) \end{array} \right.$$

و

$$x_1 = x_0 - \frac{P(x_0)}{P'(x_0)} = -2 - \frac{10}{-49} \approx -1.796.$$

با تکرار روند برای یافتن x_2 ، خواهیم داشت

-1.796	2	0	-3	3	-4
		-3.592	6.451	-6.197	5.742
	2	-3.592	3.451	-3.197	1.742 = $P(x_1)$
		-3.592	12.902	-29.368	
	2	-7.184	16.353	-32.565	$Q(x_1) = P'(x_1)$

در نتیجه ، $P(-1.796) = 1.742$ ، $P'(-1.796) = -32.565$ ، و

$$x_2 = -1.796 - \frac{(1.742)}{-32.565} \approx -1.7425.$$

یک ریشه واقعی تا پنج رقم اعشار عبارت است از -1.74259

توجه کنید که چندجمله‌ای $Q(x)$ به تقریب بکار رفته و تغییرات از یک تکرار به تکرار دیگر بستگی دارد .

هرگاه تکرار N م ، یعنی x_N ، در روند نیوتن-رفسون یک ریشه تقریبی $P(x)$ باشد ،

آنگاه

$$P(x) = (x - x_N)Q(x) + h_0 = (x - x_N)Q(x) + P(x_N) \approx (x - x_N)Q(x);$$

در نتیجه ، $x - x_N$ یک عامل تقریبی $P(x)$ است . با این فرض که $\hat{x}_1 = x_N$ ریشه تقریبی $P(x)$ بوده و $Q_1(x)$ عامل تقریبی است ،

$$(۳۵.۲) \quad P(x) \approx (x - \hat{x}_1)Q_1(x).$$

می توان یک ریشه تقریبی دوم $P(x)$ را با اعمال روند نیوتن-رفسون بر $Q_1(x)$ بدست آورد . حال ، با یافتن ریشه تقریبی \hat{x}_2 از $Q_1(x)$ و تجزیه تقریبی $Q_1(x) \approx (x - \hat{x}_2)Q_2(x)$ ، روند نیوتن-رفسون را بر $Q_2(x)$ اعمال کرده یک ریشه تقریبی سوم $P(x)$ را پیدایم کنیم ، زیرا

$$P(x) \approx (x - \hat{x}_1)Q_1(x) \approx (x - \hat{x}_1)(x - \hat{x}_2)Q_2(x).$$

هرگاه $P(x)$ یک چندجمله‌ای درجه n م با n ریشه حقیقی باشد ، تکرار این روند سرانجام $(n - 2)$ ریشه تقریبی از $P(x)$ و یک عامل تقریبی درجه دو مانند $Q_{n-2}(x)$ بدست می‌دهد . در این مرحله ، $Q_{n-2}(x) = 0$ را می توان با فرمول درجه دو حل کرد و دو ریشه تقریبی

آخر $P(x)$ را بدست آورد .

با آنکه می‌توان از روش فوق برای دستیابی به ریشه‌های تقریبی چندجمله‌ایهای زیادی استفاده کرد، این روش به استفاده مکرر از تقریبها بستگی دارد و گاهی به ریشه‌های تقریبی بسیار نادقیق منجر می‌شود. بویژه، اگر P_1 و P_2 دو چندجمله‌ای درجه n باشند که ضرایبشان تقریباً مساوی‌اند، ریشه‌های P_1 ممکن است خیلی شبیه ریشه‌های P_2 نباشد. به عنوان یک مثال بدیهی، چندجمله‌ایهای $P_1(x) = x^2 - 4x + 4$ و $P_2(x) = x^2 - 4x + 4.1$ را در نظر می‌گیریم. چندجمله‌ای P_1 دارای ریشه مکرر در $x = 2$ است، حال آنکه ریشه‌های P_2 هر دو مختلط می‌باشند.

با مرور روندی که توصیف شد، ملاحظه می‌شود که مشکل بخاطر آن است که، وقتی ریشه‌های تقریبی P بدست آمد، روند نیوتن - رفسون بر چند جمله‌ای تقلیل یافته Q_k اعمال می‌شود، یعنی، بر چند جمله‌ای با این خاصیت که

$$P(x) \approx (x - \hat{x}_1)(x - \hat{x}_2) \cdots (x - \hat{x}_k)Q_k(x).$$

یک ریشه تقریبی مانند \hat{x}_{k+1} از Q_k تعیین می‌شود و از چند جمله‌ای تقلیل یافته جدید $Q_{k+1}(x)$ برای یافتن ریشه تقریبی بعدی استفاده می‌شود، که

$$Q_k(x) \approx (x - \hat{x}_{k+1})Q_{k+1}(x).$$

یک راه برای رفع این مشکل استفاده از معادلات تقلیل یافته، یعنی عاملهای تقریبی چندجمله‌ای اصلی P ، برای یافتن تقریبهای $\hat{x}_2, \hat{x}_3, \dots, \hat{x}_k$ به ریشه‌های P ، و سپس اصلاح این تقریبها با اعمال روند نیوتن - رفسون بر چند جمله‌ای اصلی P است. این روند بعد از بحث زیر در نحوه یافتن تقریبهای اولیه به ریشه‌های چندجمله‌ایها توضیح داده خواهد شد.

قبلاً متذکر شدیم که موفقیت روش نیوتن اغلب به یافتن یک تقریب اولیه مناسب بستگی دارد. چون چندجمله‌ایها بر خط حقیقی پیوسته‌اند، یک روش اساسی برای یافتن یک تقریب به ریشه‌های از یک چندجمله‌ای از قضیه مقدار میانی (ص ۹) بدست می‌آید. در مثال ساده

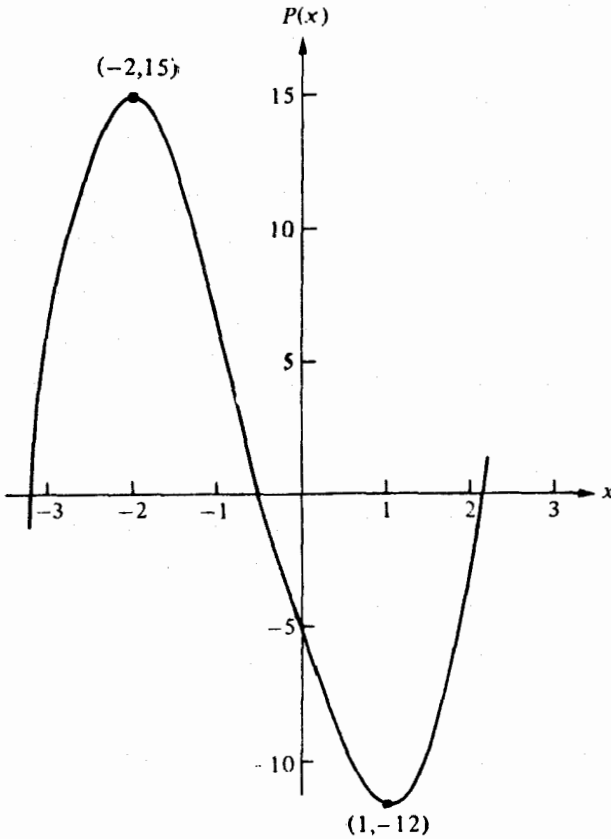
$$P(x) = 2x^3 + 3x^2 - 12x - 5,$$

یک ماکزیمم نسبی در $x = -2$ و یک مینیمم نسبی در $x = 1$ وجود دارد. بعلاوه،

$$P(1) = -12 < 0, \quad P(-2) = 15 > 0,$$

$$\lim_{x \rightarrow \infty} P(x) = +\infty \quad \text{و} \quad \lim_{x \rightarrow -\infty} P(x) = -\infty,$$

در نتیجه، نمودار P به شکل ۱۱.۲ می‌باشد.



شکل ۱۱.۲

چون $P(1)P(-2) < 0$ ، یک ریشه P در بازه $(-2, 1)$ قرار دارد. یک تقریب اولیه مناسب به ریشه واقع در $(-2, 1)$ ، $x_0 = 0$ است. برای یافتن بازه‌ای که شامل بزرگترین ریشه P باشد، P را در $x = 2, 3, \dots$ حساب می‌کنیم تا عدد صحیحی چون k بدست آید با این خاصیت که $P(k) > 0$ و $P(k-1) < 0$ ، چرا که در این صورت P باید یک ریشه در بازه $(k-1, k)$ داشته باشد. در این مثال، $P(2) = -1$ و $P(3) = 40$ و یک تقریب اولیه مناسب به این ریشه $x_0 = 2$ است. یک تقریب اولیه برای ریشه‌ای که کوچکتر

از 2- است را می توان به همین ترتیب بدست آورد .

با آنکه این روند هندسی برای چند جمله ایهای با درجه کوچک بخوبی کاراست ، برای چند جمله ایهای درجات بالا مفید نیست ، زیرا ، برای دستیابی به نمودار تقریبی $P(x)$ ، معمولا " باید مقادیر ماکزیم و مینیم P را بدست آورد ؛ و در نتیجه ، باید $P'(x) = 0$ را حل کرد . اگر $P(x)$ از درجه n باشد ، باید صفرهای $P'(x)$ ، که یک چند جمله ای از درجه $n - 1$ است ، را پیدا کرد که مسئله ای مشابه و تقریبا " به همان سختی مسئله اصلی است .

ایده اصلی یافتن ریشه های تقریبی $P(x)$ به صورت زیر است : $P(x)$ را در نقاط x_i به ازای $i = 1, 2, \dots, k$ حساب می کنیم . هرگاه $P(x_i)P(x_{i+1}) < 0$ ، یک صفر بین x_i و x_{i+1} دارد . مسئله به صورت انتخاب x_i ها درمی آید بطوری که شانس از دست رفتن تغییر علامت حداقل بوده و درعین حال تعداد x_i ها به میزان معقولی کوچک باشد .

برای توضیح مشکل احتمالی این مسئله ، چند جمله ای

$$P(x) = 16x^4 - 40x^3 + 5x^2 + 20x + 6.$$

را در نظر می گیریم . اگر x_i عددی صحیح باشد ، می توان نشان داد که $P(x_i) > 0$. در واقع ، اگر به ازای هر عدد صحیح i ، $x_i = i/4$ ، باز هم $P(x_i) > 0$ برقرار است ؛ لذا ، محاسبه $P(x)$ در این بی نهایت x_i ها هیچ بازه ای چون (x_i, x_{i+1}) شامل ریشه های $P(x)$ را مشخص نمی کند . در حالی که ، در مثال ۳ ملاحظه می شود که $P(x)$ ریشه های حقیقی دارد . در اینجا انتخاب x_i های فوق برای این چند جمله ای ، بدلیل نزدیک بودن ریشه ها ، نامناسب بود .

سوال اصلی این است : چطور می توان x_i هایی انتخاب کرد که مناسب باشند ؟ معقول به نظر می رسد که ، اگر $P(x)$ ریشه های حقیقی داشته و از x_i ها به قدر کافی انتخاب شود ، سرانجام بازه هایی که شامل صفرهای $P(x)$ باشند مشخص می شوند ، اما ، واضح است که این روند می تواند فوق العاده وقت گیر باشد .

روش پیشنهادی برای یافتن بازه های شامل ریشه های حقیقی $P(x)$ به صورت زیر است : برای یافتن یک تقریب اولیه به یک ریشه حقیقی

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0,$$

ابتدا توجه می کنیم که دقیقا " وقتی $y \in (0, 1]$ که $1/y \in [1, \infty)$ ، و دقیقا " وقتی $y \in [-1, 0)$ که $1/y \in (-\infty, -1]$. همچنین ،

$$P\left(\frac{1}{y}\right) = a_n\left(\frac{1}{y}\right)^n + a_{n-1}\left(\frac{1}{y}\right)^{n-1} + \dots + a_1\left(\frac{1}{y}\right) + a_0$$

$$= \frac{1}{y^n} (a_n + a_{n-1}y + \dots + a_1y^{n-1} + a_0y^n).$$

بنابراین، اگر \hat{P} با

$$\hat{P}(y) = a_0y^n + a_1y^{n-1} + \dots + a_{n-1}y + a_n,$$

تعریف شده باشد، $\hat{P} \neq 0$ ریشه \hat{P} است اگر و فقط اگر $1/\hat{y}$ ریشه P باشد.

برای یافتن تقریبها به همه ریشه‌های حقیقی P ، کافی است بازهای $[-1, 0]$ و

$[0, 1]$ را بگردیم و P و \hat{P} را در هر x_i حساب کنیم تا تغییر علامات مشخص شوند.

بنابر قضیه مقدار میانی، اگر $P(x_i)P(x_{i+1}) < 0$ ، P ریشه‌ای بین x_i و x_{i+1} دارد. (اگر

$P(x_i) = 0$ ، از خوش شانسی، یک ریشه P مشخص شده است.) اگر $\hat{P}(x_i)\hat{P}(x_{i+1}) < 0$ ،

\hat{P} ریشه‌ای بین x_i و x_{i+1} و P ریشه‌ای بین $1/x_i$ و $1/x_{i+1}$ دارد. بازه‌های $[-1, 0]$ و $[0, 1]$

را جداگانه در نظر می‌گیریم که ابهام حالتی که در آن به‌ازای x_i در $[-1, 0]$ و x_{i+1} در

$(0, 1]$ ، $\hat{P}(x_i)\hat{P}(x_{i+1}) < 0$ از بین برود.

راه‌های زیادی برای انتخاب x_i وجود دارند؛ اما، در اکثر مسائل عملی، انتخاب

x_i ها به‌صورت زیر کافی خواهد بود: عدد زوج N را طوری اختیار می‌کنیم که یک کران

برای تعداد نقاطی که باید محاسبه شوند باشد. فرض کنیم $-1 = x_0$ ، و به‌ازای

$x_i = x_{i-1} + 2/N$ ، $i = 1, \dots, N$ این انتخاب به‌استفاده از $(N+1)$ نقطه $-1 = x_0$

و $x_{N/2} = 1$ ، $x_N = 1$ منجر می‌شود.

مثال زیر کاربرد عملی این روند را نشان می‌دهد.

مثال ۳. برای یافتن ریشه‌های

$$P(x) = 16x^4 - 40x^3 + 5x^2 + 20x + 6,$$

قرار می‌دهیم $y = 1/x$ در این صورت،

$$P\left(\frac{1}{y}\right) = 16\left(\frac{1}{y}\right)^4 - 40\left(\frac{1}{y}\right)^3 + 5\left(\frac{1}{y}\right)^2 + 20\left(\frac{1}{y}\right) + 6$$

$$= \frac{1}{y^4} (16 - 40y + 5y^2 + 20y^3 + 6y^4),$$

در نتیجه،

$$\hat{P}(y) = 6y^4 + 20y^3 + 5y^2 - 40y + 16$$

یا

$$\hat{P}(x) = 6x^4 + 20x^3 + 5x^2 - 40x + 16.$$

ریشه‌های \hat{P} معکوس ریشه‌های ناصفر $P(x)$ اند. حال برای یافتن بازه‌های (x_i, x_j) که شامل ریشه‌های P یا ریشه‌های \hat{P} اند باید $[-1, 1]$ را برگردیم. تقسیم ترکیبی (روش هورنر) را برای محاسبه $P(x_i)$ و $\hat{P}(x_i)$ بکار می‌بریم. مثلاً، محاسبه $P(-1)$ و $\hat{P}(-1)$ به صورت زیر انجام می‌شود:

	P					
$x = -1$		6	-40	5	20	6
		-16	56	-61	41	
		16	-56	61	-41	<u>47</u> = $P(-1)$
	\hat{P}					
$x = -1$		6	20	5	-40	16
		-6	-14	9	31	
		6	14	-9	-31	<u>47</u> = $\hat{P}(-1)$

در جدول ۱۰.۲ نقاط بکاررفته در جستجوی اصولی $[-1, 1]$ و مقادیر P و \hat{P} در این نقاط درج شده‌اند.

x_i	$P(x_i)$	$\hat{P}(x_i)$	x_i	$P(x_i)$	$\hat{P}(x_i)$
-1.00	47.000	47.000	.25	10.75	6.6484
-.75	15.750	42.273	.50	13.25	.1250
-.50	3.250	35.125	.75	12.00	-.8516
-.25	2.000	26.023	1.00	7.00	7.0000
.00	6.000	16.000			

در این مثال، تعداد نقاطی که باید محاسبه شوند نه است.

چون $\hat{P}(5) = .125$ ، $\hat{P}(75) = -.8516$ ، و $\hat{P}(1) = 7$ ، \hat{P} یک ریشه در $(.5, .75)$ و ریشه دیگری در $(.75, 1)$ دارد. در نتیجه، P یک ریشه در $(4/3, 2) = (1.33, 2)$ و ریشه دیگری در $(1, 4/3)$ دارد. چون $|\hat{P}(5)|$ از هر یک از $|\hat{P}(1)|$ و $|\hat{P}(75)|$ به صفر نزدیکتر است، احتمالاً "5. از هر یک از 1 و 75 به یک ریشه \hat{P} نزدیکتر است؛ و حدس اینکه 2 از هر یک از 1 یا $\frac{4}{3}$ به یک ریشه P نزدیکتر است معقول می باشد. با استفاده از روش نیوتن با تقریب اولیه $x_0 = 2$ ، اولین ریشه تقریبی P را مشخص می کنیم. با آزایی $x_0 = 2$ و با استفاده از روند مثال ۲ برای یافتن $x_1 = x_0 - (P(x_0)/P'(x_0))$ ، نتیجه می شود که

2	16	-40	5	20	6
		32	-16	-22	-4
	16	-8	-11	-2	<u>2</u> = P(2)
		32	48	74	
	16	24	37	<u>72</u> = P'(2)	

و $x_1 = 2 - 2/72 \approx 1.97$ با تکرار این روند، خواهیم داشت

1.97	16	-40	5	20	6
		31.52	-16.7056	-23.0600	-6.0283
	16	-8.48	-11.7056	-3.0600	<u>-.0283</u> = P(1.97)
		31.52	45.3888	66.3559	
	16	23.04	33.6832	<u>63.2959</u> = P'(1.97)	

و

$$x_2 = x_1 - \frac{P(x_1)}{P'(x_1)} \approx 1.9705.$$

به یاد می آوریم که روند تقسیم ترکیبی یک تجزیه تقریبی P را نیز بدست می دهد. در این مثال، تجزیه تقریبی عبارت است از

$$P(x) \approx (x - 1.97)Q(x),$$

که در آن

$$Q(x) = 16x^3 - 8.48x^2 - 11.7056x - 3.0600.$$

حال می‌توان چند جمله‌ای Q را برای ریشه‌های تقریبی دیگر P بکار برد. چون فقط یک عامل تقریبی P است، صفرهای Q فقط ریشه‌های تقریبی P می‌باشند؛ با اینحال، چون Q یک چند جمله‌ای از درجه ۶ پایین‌تر از P است، یافتن ریشه‌های تقریبی Q به محاسبات کمتری نیاز دارد. همانطور که قبلاً بحث شد، روند پیشنهادی این است که اول از روش نیوتن - رفسون برای یافتن ریشه‌های تقریبی Q استفاده شود و سپس به وسیله تکرار روی چند جمله‌ای اصلی P دقت را بهتر کرد.

مثال ۳ را ادامه داده، فرض می‌کنیم $x_0 = 1.33 \approx \frac{4}{3}$ و روند تقسیم ترکیبی را بر Q اعمال می‌کنیم تا بدست آید

1.33	16	-8.48	-11.7056	-3.0600
		21.28	17.024	7.0735
	16	12.8	5.3184	<u>4.0135</u> = $Q(1.33)$
		21.28	45.3264	
	16	34.08	<u>50.6448</u> = $Q'(1.33)$	

لذا، بنا بر روش نیوتن،

$$x_1 = 1.33 - \frac{Q(1.33)}{Q'(1.33)} \approx 1.251$$

یک ریشه تقریبی Q و، در نتیجه، یک ریشه تقریبی P است. برای بهبود تقریب، یک تکرار روش نیوتن را با P انجام می‌دهیم. از این کار نتیجه می‌شود که

1.251	16	-40	5	20	6
		20.016	-25.000	-25.020	-6.280
	16	-19.984	-20.000	-5.020	<u>-.280</u> = $P(1.251)$
		20.016	.040	-24.970	
	16	.032	-19.960	<u>-29.990</u> = $P'(1.251)$	

$$x_2 = 1.251 - \frac{(-.280)}{-29.990} \approx 1.2417.$$

حال که ریشه‌های تقریبی $\hat{x}_1 = 1.9705$ و $\hat{x}_2 = 1.2417$ می‌توان بار دیگر تقسیم ترکیبی را جهت تجزیه $P(x)$ اعمال کرد و بدست آورد که

$$P(x) \approx (x - \hat{x}_1)(x - \hat{x}_2)\hat{Q}(x).$$

روند به صورت زیر می‌باشد :

1.2417	16	-40	5	20	6
		19.8672	-24.9989	-24.8326	-6.0007
1.9705	16	-20.1328	-19.9989	-4.8326	-.0007
		31.528	22.4542	4.8382	
	16	11.3952	2.4553	.0056	

بنابراین ،

$$(۳۶.۰۲) \quad P(x) \approx (x - 1.2417)(x - 1.9705)(16x^2 + 11.3952x + 2.4553),$$

که در آن باقیمانده‌های $-.0007$ و $.0056$ حذف شده‌اند. حال تقریبهای \hat{x}_3 و \hat{x}_4 برای دو ریشه دیگر P را می‌توان با حل

$$16x^2 + 11.3952x + 2.4553 = 0$$

بدست آورد، که نتیجه می‌شود $\hat{x}_3 = -.3561 + .16325i$ و $\hat{x}_4 = -.3561 - .16325i$. برای بهبود تقریبهای \hat{x}_3 و \hat{x}_4 ، روشهای قبلی نیاز به کاربردی دیگر از روش نیوتن دارند، که در آن با $\hat{x}_3 = -.3561 + .16325i$ شروع و از چند جمله‌ای اصلی

$$P(x) = 16x^4 - 40x^3 + 5x^2 + 20x + 16$$

استفاده می‌شود.

چون این روند به محاسبات زیادی با اعداد مختلط نیاز دارد، به جای آن یک روش عمومی برای یافتن ریشه‌های مختلط یک چند جمله‌ای بدست می‌آوریم. ایده اصلی در این روند یافتن عاملهای درجه دو متناظر با ریشه‌های مختلط و سپس یافتن ریشه‌های این چند جمله‌ایهای درجه دو به وسیله فرمول درجه دو است. اگر چند جمله‌ای

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

دارای ریشه مختلط $x = \alpha + \beta i$ باشد، قضیه ۱۸.۲ ایجاب می کند که $\bar{x} = \alpha - \beta i$ نیز یک ریشه P و $P(x) = x^2 - 2\alpha x + \alpha^2 + \beta^2$ یک عامل P باشد. کار با عامل درجه دو $R(x)$ آسانتر از کار با خود ریشه هاست، زیرا $R(x)$ تنها شامل اعداد حقیقی می باشد. در حالت کلی، اگر $R(x)$ یک چند جمله ای درجه دو باشد، با تقسیم $P(x)$ بر $R(x)$ خواهیم داشت

$$P(x) = R(x)Q(x) + Ax + B,$$

که در آن A و B ثابت بوده و Q یک چند جمله ای است که درجه اش دو تا کمتر از درجه P است. ذیلاً "یک روند تقسیم ترکیبی برای تقسیم چند جمله ای P بر یک چند جمله ای درجه دو به شکل $x^2 + b_1x + b_0$ را توضیح می دهیم. با آنکه این روند تا حدودی پیچیده تر از تقسیم ترکیبی معمولی است، روند را می توان به آسانی اجرا کرد و، مانند روش هورنر، این مزیت را دارد که تعداد محاسبات آن مینیمم است. هرگاه R و P عبارت باشند از

$$R(x) = x^2 + b_1x + b_0,$$

$$P(x) = a_nx^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0,$$

آنگاه روش تقسیم P بر R به وسیله تقسیم ترکیبی به صورت زیر است:

$-b_1, -b_0$	a_n	a_{n-1}	a_{n-2}	a_{n-3}	\dots	a_1	a_0
	0	0	$-b_0a_n$	$-b_0(a_{n-1} - b_1a_n)$	\dots		
	0	$-b_1a_n$	$-b_1(a_{n-1} - b_1a_n)$	\dots			0
	a_n	$a_{n-1} - b_1a_n$	\dots	\dots		A	B

ابتدا صفرها را جایی قرار می دهیم که با مقادیر حساب شده پرنمی شوند. به طور مشخص، درایه های ستون یک سطرهای دو و سه، درایه های ستون دو سطر دو، و ستون آخر سطر سه صفرند. درایه های سطر اول ضرایب P اند. درایه های سطر دوم $-b_0$ برابر مجموع درایه های دو ستون به طرف چپ اند، یا به طور روشن تر، درایه ستون z از سطر دو مساوی $-b_0$ برابر مجموع درایه های ستون $z - 2$ است، که ستونها با اعداد 1 تا $n + 1$ از چپ به راست شماره گذاری شده اند. به ازای $z = 1, 2, \dots, n$ درایه ستون z از سطر سوم جدول تقسیم ترکیبی $-b_1$ برابر مجموع درایه های ستون $z - 1$ است. درایه آخرین ستون سطر سه همواره صفر است. مجموع درایه های ستون یکی به آخر مقدار A در

جمله باقیمانده $Ax + B$ ، و مجموع درایه‌های ستون آخر B را بدست می‌دهد.

مثال ۴. با استفاده از روند تقسیم ترکیبی درجه دو،

$$x^6 - 2x^5 + 7x^4 - 4x^3 + 11x^2 - 2x + 5$$

را بر $x^2 - 2x + 1$ تقسیم کنید.

$$\begin{array}{r|rrrrrrr} 2, -1 & 1 & -2 & 7 & -4 & 11 & -2 & 5 \\ & 0 & 0 & -1 & 0 & -6 & -8 & -21 \\ & 0 & 2 & 0 & 12 & 16 & 42 & 0 \\ \hline & 1 & 0 & 6 & 8 & 21 & \underline{32} & \underline{-16} \end{array}$$

از این جدول نتیجه می‌شود که

$$\begin{aligned} x^6 - 2x^5 + 7x^4 - 4x^3 + 11x^2 - 2x + 5 \\ = (x^2 - 2x + 1)(x^4 + 6x^2 + 8x + 21) + 32x - 16. \end{aligned}$$

برای سهولت در اجرای روند تقسیم ترکیبی درجه دو، الگوریتم زیر را بیان می‌کنیم.

الگوریتم تقسیم ترکیبی درجه دو ۷.۲

فرض کنید بخواهیم چند جمله‌ای

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

را بر چند جمله‌ای $x^2 + b_1 x + b_0$ تقسیم کنیم:

مرحله ۱. قرار می‌دهیم $z_n = a_n$ و $c_n = a_{n-1}$.

مرحله ۲. قرار می‌دهیم $k = n - 1$.

مرحله ۳. قرار می‌دهیم $c_k = -b_0 z_{k+1} + a_{k-1}$ ، $z_k = -b_1 z_{k+1} + c_{k+1}$.

مرحله ۴. اگر $k \neq 1$ ، به مرحله ۵ می‌رویم.

اگر $k = 1$ ، به مرحله ۶ می‌رویم.

مرحله ۵. k را از ۱ کم کرده به مرحله ۳ می‌رویم.

مرحله ۶. روند تمام خواهد بود.

$$\begin{aligned} P(x) = (z_n x^{n-2} + z_{n-1} x^{n-3} + \dots + z_3 x + z_2)(x^2 + b_1 x + b_0) \\ + z_1 x + c_1. \end{aligned}$$

توجه کنید که $c_1 x + c_0$ جمله باقیمانده در فرایند تقسیم است؛ در نتیجه، اگر $c_1 = c_0 = 0$ ، همچنین، توجه کنید که برای بکار بردن الگوریتم ۷.۲ ضریب x^2 باید یک باشد.

مثال ۵. با استفاده از الگوریتم ۷.۲، $x^6 - 2x^5 + 7x^4 - 4x^3 + 11x^2 - 2x + 5$ را بر $x^2 - 2x + 1$ تقسیم کنید.

در این مثال، $b_1 = -2$ و $b_0 = 1$ ، و مراحل الگوریتم به صورت زیرند:

$z_6 = 1$	$c_6 = -2$	مرحله ۱
$z_5 = (2)(1) + (-2) = 0$	$c_5 = (-1)(1) + 7 = 6$	مرحله ۳
$z_4 = (2)(0) + 6 = 6$	$c_4 = (-1)(0) + (-4) = -4$	
$z_3 = (2)(6) + (-4) = 8$	$c_3 = (-1)(6) + 11 = 5$	
$z_2 = (2)(8) + 5 = 21$	$c_2 = (-1)(8) - 2 = -10$	
$z_1 = (2)(21) + (-10) = 32$	$c_1 = (-1)(21) + 5 = -16$	
$x^6 - 2x^5 + 7x^4 - 4x^3 + 11x^2 - 2x + 5$		مرحله ۶
$= (x^4 + 6x^2 + 8x + 21)(x^2 + 2x - 1) + 32x - 16$		

توجه کنید که این همان تجزیه بدست آمده از مثال ۴ است.

به محض یافتن یک عامل درجه دو تقریبی نظیر به ریشه‌های مختلط چندجمله‌ای P ، می‌توان با استفاده از فرمول درجه دو ریشه‌های مختلط تقریبی را بدست آورد. یادآور می‌شویم که ریشه‌های مختلط تقریبی $\alpha \pm \beta i$ نظیر به عامل درجه دو $x^2 - 2\alpha x + (\alpha^2 + \beta^2)$ هستند. به محض یافتن ریشه‌های درجه دو تقریبی، می‌توان روش نیوتن را بکار برد و، با استفاده از حساب مختلط، دقت این ریشه‌ها را بهتر کرد. ضمن استفاده از روش نیوتن در این وضع، درخواست می‌شود که نتایج الگوریتم ۷.۲ مفیدند. اگر $x_0 = \alpha + \beta i$ یک ریشه تقریبی $P(x)$ باشد، و بخواهیم این تقریب را بهتر کنیم،

$$x_1 = x_0 - \frac{P(x_0)}{P'(x_0)}$$

را محاسبه می‌کنیم. چون $x_0 = \alpha + \beta i$ نظیر یک عامل درجه دو است،

$$R(x_0) = 0 \quad \text{و} \quad R(x) = x^2 - 2\alpha x + (\alpha^2 + \beta^2)$$

اگر برای تقسیم $P(x)$ بر $R(x)$ از الگوریتم ۷.۲ استفاده شود، حاصل می‌شود

$$P(x) = R(x)Q(x) + Ax + B$$

و

$$P(x_0) = Ax_0 + B.$$

بعلاوه،

$$P'(x) = R(x)Q'(x) + R'(x)Q(x) + A;$$

در نتیجه،

$$P'(x_0) = R(x_0)Q'(x_0) + (2x_0 - 2x)Q(x_0) + A = 2\beta iQ(x_0) + A.$$

و $P'(x_0)$ را می‌توان ابتدا با محاسبه $Q(x_0)$ به وسیله الگوریتم ۷.۲ و سپس محاسبه

$$P'(x_0) = 2\beta iQ(x_0) + A.$$

بدست آورد.

مثال ۶. در مثال ۳، $\hat{x}_3 = -3561 + .16325i$ به عنوان یک ریشه تقریبی

$$P(x) = 16x^4 - 40x^3 + 5x^2 + 20x + 6$$

بدست آمد. با یک تکرار روش نیوتن، دقت این ریشه را بالا می‌بریم. تقریب \hat{x}_3 نظیر به

عامل درجه دو $.1535x + .7122x^2 + R(x)$ است. حال، با استفاده از روند تقسیم ترکیبی

نظیر به الگوریتم ۷.۲، داریم

- .7122, - .1535	16	-40	5	20	6
	0	0	-2.456	7.8892	-6.0092
	0	-11.3952	36.6037	27.8810	0
	16	-51.3952	39.1477	<u>.0082</u>	<u>-.0092</u>
	0	0	-2.456		
	0	-11.3952	0		
	16	<u>-62.7904</u>	<u>36.6917</u>		

بنابراین،

$$\begin{aligned} P(\hat{x}_3) &= P(-.3561 + .16325i) \\ &= (.0082)(-.3561 + .16325i) - .0092 \\ &= -.01212 + .00134i, \end{aligned}$$

$$\begin{aligned}
 P'(\hat{x}_3) &= 2(.16325i)Q(\hat{x}_3) + .0082 \\
 &= 2(.16325i)[(-62.7904)(-.3561 + .16325i) + 36.6917] \\
 &\quad + .0082 \\
 &= 3.35500 + 19.2803i.
 \end{aligned}$$

بمازای

$$x_0 = \hat{x}_3 = -.3561 + .16325i,$$

$$x_1 = x_0 - \frac{P(x_0)}{P'(x_0)} = -.356061 + .162628i$$

تا چهار رقم اعشار دقیق می باشد .

ضعف روند ریشه یابی مختلط در الگوریتم ۷.۲ این است که برای عامل درجه دو یک تقریب اولیه لازم است . به یاد می آوریم که روش بدست آوردن تقریب اولیه برای یک ریشه حقیقی نسبتاً پیچیده بود ، و نباید از اهمیت این ضعف تعجب کرد . یک روش برای یافتن ریشه های چند جمله ایها ، که اخیراً " بخاطر عدم نیاز به تقریب اولیه شهرت یافته ، بر الگوریتم خارج قسمت - تفاضل (QD) که به وسیله اچ . روتیشاوزر^۱ عرضه شده استوار است . نظریه اساسی این روند و تحلیل شرایط لازم برای توفیق آن نسبتاً پیچیده است ، و برای جزئیات آن خواننده را به هنریچی [۳۴] یا به مقاله ای که توسط هنریچی در رالستون^۲ و ویلف^۳ [۵۰] نوشته شده ارجاع می دهیم .

همگرایی روش QD وقتی روی دهد در حالت کلی نسبتاً کند و بسیار مستعد برای خطای گرد کردن است ، اما می تواند در بدست آوردن تقریب های اولیه ، هم به ریشه های حقیقی و هم به عامل های درجه دو مربوط به ریشه های مختلط ، که در روش های همگرای سریعتر مورد نیازند خیلی مفید باشد .

روش QD برای تقریب ریشه های چند جمله ای

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

عبارت است از ساختن دنباله های زیر از اعداد حقیقی :

$k = 1, 2, \dots, (n + 1)$ هر بازارای هر $\{e_i^{(k)}\}_{i=1}^{\infty}$

9

$k = 1, 2, \dots, n$ هر بازارای هر $\{q_i^{(k)}\}_{i=1}^{\infty}$

که در آنها

بازارای هر $e_i^{(1)} = 0$ ، $i = 1, 2, \dots$

بازارای هر $e_i^{(n+1)} = 0$ ، $i = 1, 2, \dots$

بازارای هر $e_1^{(k)} = \frac{a_n - k}{a_{n-k+1}}$ ، $k = 2, 3, \dots, n$

$$q_1^{(1)} = -\frac{a_{n-1}}{a_n}$$

بازارای هر $q_1^{(k)} = 0$ ، $k = 2, 3, \dots, n$

بازارای هر $i = 1, 2, \dots$ و $k = 1, 2, 3, \dots$

$$q_{i+1}^{(k)} = e_i^{(k+1)} + q_i^{(k)} - e_i^{(k)}$$

و، بالاخره،

بازارای هر $i = 1, 2, \dots$ و $k = 2, 3, \dots, n$

$$e_{i+1}^{(k)} = (q_{i+1}^{(k)} e_i^{(k)}) / q_{i+1}^{(k-1)}$$

با آنکه ساختن این دنباله ظاهراً " پیچیده است ، در عمل می توان آن را خیلی ساده

به کمک یک جدول انجام داد ، و این جدول را ابتدا با وارد کردن تمام مقادیر برای $q_i^{(k)}$ ،

$e_i^{(1)}$ ، $e_i^{(n)}$ و $e_i^{(n+1)}$ می سازیم (جدول ۱۱.۲) .

i	$e_i^{(1)}$	$q_i^{(1)}$	$e_i^{(2)}$	$q_i^{(2)}$	$e_i^{(3)}$	$q_i^{(3)}$...	$e_i^{(n)}$	$q_i^{(n)}$	$e_i^{(n+1)}$
1	0	$\left(\frac{-a_{n-1}}{a_n}\right)$	$\left(\frac{a_{n-2}}{a_{n-1}}\right)$	0	$\left(\frac{a_{n-3}}{a_{n-2}}\right)$	0		$\left(\frac{a_0}{a_1}\right)$	0	0
2	0									0
3	0									0
⋮	⋮									⋮

جدول ۱۱.۲

مرحله بعد ساختن $q_i^{(k)}$ ها در سطر دوم است به این ترتیب که یکی به راست و بالای آن، یعنی $e_i^{(k+1)}$ ، را گرفته به آن عنصر بلافاصله بالای آن، یعنی $q_i^{(k)}$ ، را می افزاییم، و یکی به چپ و بالای آن، یعنی $e_i^{(k)}$ ، را کم می کنیم تا جدول ۱۲.۲ بدست آید.

i	$e_i^{(1)}$	$q_i^{(1)}$	$e_i^{(2)}$	$q_i^{(2)}$	$e_i^{(3)}$	$q_i^{(3)}$...	$e_i^{(n)}$	$q_i^{(n)}$	$e_i^{(n+1)}$
1	0	$\left(\frac{-a_{n-1}}{a_n}\right)$	$\left(\frac{a_{n-2}}{a_{n-1}}\right)$	$\left(\frac{a_{n-3}}{a_{n-2}}\right)$	0	$\left(\frac{a_0}{a_1}\right)$		0	0	
2	0	$q_2^{(1)}$		$q_2^{(2)}$		$q_2^{(3)}$...		$q_2^{(n)}$	0
3	0									0
\vdots	\vdots									\vdots

جدول ۱۲.۲

حال $e_2^{(k)}$ ها را می توان با اختیار عنصر راست آن، یعنی $q_2^{(k)}$ ، و ضرب آن در درایهء بالای آن، یعنی $e_1^{(k)}$ ، و تقسیم برد درایهء چپ آن، یعنی $q_2^{(k-1)}$ ، بدست آورد (جدول ۱۳.۲).

i	$e_i^{(1)}$	$q_i^{(1)}$	$e_i^{(2)}$	$q_i^{(2)}$	$e_i^{(3)}$	$q_i^{(3)}$...	$e_i^{(n)}$	$q_i^{(n)}$	$e_i^{(n+1)}$
1	0	$\left(\frac{-a_{n-1}}{a_n}\right)$	$\left(\frac{a_{n-2}}{a_{n-1}}\right)$	0	$\left(\frac{a_{n-3}}{a_{n-2}}\right)$	0	...	$\left(\frac{a_0}{a_1}\right)$	0	0
2	0	$q_2^{(1)}$	$e_2^{(2)}$	$q_2^{(2)}$	$e_2^{(2)}$	$q_2^{(3)}$...	$e_2^{(n)}$	$q_2^{(n)}$	0
3	0									0
\vdots	\vdots									\vdots

جدول ۱۳.۲

اگر به ازای هر $k = 1, 2, \dots, n$ ، $\lim_{i \rightarrow \infty} e_i^{(k)} = \lim_{i \rightarrow \infty} e_i^{(k+1)} = 0$ ، به آسانی معلوم می شود که $\lim_{i \rightarrow \infty} q_i^{(k)}$ وجود دارد (ر.ک. تمرین ۱۳). بهر حال، نکتهء مهم آن است که این حد یک ریشهء چند جمله ای P است. بعلاوه، اگر $\{e_i^{(k)}\}_{i=1}^{\infty}$ به ازای k ای به صفر همگرا نباشد، دنباله های $\{r_i\}_{i=1}^{\infty}$ و $\{s_i\}_{i=1}^{\infty}$ ، که در آنها

$$r_i = q_i^{(k-1)} + q_i^{(k)}, \quad i = 1, 2, \dots$$

و

$$s_i = q_{i-1}^{(k-1)} q_i^{(k)}, \quad i = 2, 3, \dots$$

همگرا به اعداد r و s اند بطوری که $x^2 - rx + s$ یک عامل درجه دوی $P(x)$ متناظر با یک جفت از ریشه‌های مزدوج مختلط است.

همانطور که قبلاً گفتیم، این روند عموماً " برای یافتن ریشه‌های واقعی چند جمله‌ای بکار نمی‌رود، بلکه برای تعیین تقریبهای اولیه رضایتبخش بکار می‌رود.

الگوریتم خارج قسمت - نفاضل (QD) ۸.۲

چند جمله‌ای $P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ مفروض است: مرحله ۱. قرار می‌دهیم

$$e_1^{(1)} = 0$$

$$e_1^{(n+1)} = 0,$$

$$q_1^{(1)} = -\frac{a_{n-1}}{a_n},$$

به‌ازای هر $k = 2, 3, \dots, n$ $q_1^{(k)} = 0$

$$e_1^{(k)} = \frac{a_{n-k}}{a_{n-k+1}}, \quad k = 2, 3, \dots, n$$

مرحله ۲. قرار می‌دهیم $i = 2$.

مرحله ۳. قرار می‌دهیم

$$e_i^{(1)} = 0,$$

$$e_i^{(n+1)} = 0,$$

به‌ازای هر $k = 1, 2, \dots, n$ $q_i^{(k)} = e_{i-1}^{(k+1)} + q_{i-1}^{(k)} - e_{i-1}^{(k)}$

به‌ازای هر $k = 2, 3, \dots, n$ $e_i^{(k)} = (q_i^{(k)} e_{i-1}^{(k)}) / q_i^{(k-1)}$

مرحله ۴. مشخص می‌کنیم که آیا می‌توان در باب همگرایی دنباله‌های $\{e_j^{(k)}\}_{j=1}^r$ به صفر تصمیم گرفت.

اگر بتوان تصمیم گرفت، به مرحله ۶ می‌رویم.

اگر نتوان تصمیم گرفت، به مرحله ۵ می‌رویم.

مرحله ۵. ۱ را به i افزوده به مرحله ۳ می‌رویم .

مرحله ۶. روند تمام است. اگر مشخص شده باشد که $\{e_j^{(k)}\}_{j=1}^{\infty}$ و $\{e_j^{(k+1)}\}_{j=1}^{\infty}$ هر دو به ازای

k ای همگرا به صفرند، $q_i^{(k)}$ را یک تقریب به ریشه‌ای از P می‌گیریم .

اگر مشخص شده باشد که $\{e_j^{(k)}\}_{j=1}^{\infty}$ به ازای k ای همگرا نیست،

$$x^2 - (q_i^{(k-1)} + q_i^{(k)})x + q_i^{(k-1)}q_i^{(k)}$$

را یک تقریب به عامل درجه ۲ دویی از P مربوط به ریشه‌های مزدوج مختلط می‌گیریم .

مثال ۷. چند جمله‌ای $P(x) = 16x^4 - 40x^3 + 5x^2 + 20x + 6$ را، که در امثله

۳ و ۶ بررسی شد، در نظر می‌گیریم . الگوریتم QD مقادیر نموده شده در جدول ۱۴.۲ را

بدست می‌دهد .

i	$e_i^{(1)}$	$q_i^{(1)}$	$e_i^{(2)}$	$q_i^{(2)}$	$e_i^{(3)}$	$q_i^{(3)}$	$e_i^{(4)}$	$q_i^{(4)}$	$e_i^{(5)}$
1	0	2.5	-.125	0	4	0	.3	0	0
2	0	2.375	-.217105	4.125	-3.587879	-3.7	.024324	-.3	0
3	0	2.157895	-.075883	.754226	.417653	-.087797	.089855	-.324324	0
4	0	2.08201	-.045477	1.247762	-.139108	-.415595	.089549	-.414179	0
5	0	2.036535	-.025772	1.154130	.022532	-.186937	.241301	-.503728	0
6	0	2.010763	-.015412	1.202434	.000596	.031832	-5.647705	-.745029	0
7	0	1.995351	-.009411	1.218442	-.002749	-5.616470	4.929942	4.902676	0
8	0	1.985940	-.005806	1.225104	.001535	-.683778	.196581	-.027266	0
9	0	1.980135	-.003613	1.232444	-.000609	-.488732	.090037	-.223846	0
10	0	1.976521	-.002259	1.235449	.000196	-.398087	.070992	-.313883	0
11	0	1.974263	-.001416	1.237904	-.000052	-.327391	.083481	-.384875	0
12	0	1.972846	-.000890	1.239268	.000010	-.243757	.160405	-.468358	0
13	0	1.971957	-.000560	1.240168	-.000001	-.083362	1.209859	-.628762	0
14	0	1.971397	-.000352	1.240727	-.000001	1.126498	1.974681	-1.838622	0

جدول ۱۴.۲

از این تحلیل ظاهراً " روشن است که

$$\lim_{i \rightarrow \infty} e_i^{(3)} = 0 \quad \text{و} \quad \lim_{i \rightarrow \infty} e_i^{(2)} = 0$$

$$q_{14}^{(2)} = 1.240727 \text{ و } q_{14}^{(1)} = 1.971397$$

ریشه‌های تقریبی باشند.

با آنکه درآیه‌های $e_9^{(4)}$ ، $e_{10}^{(4)}$ ، و $e_{11}^{(4)}$ کوچکند، ساختن درایه‌های بعدی $e_{12}^{(4)}$ ، $e_{13}^{(4)}$ ، و $e_{14}^{(4)}$ ظاهراً "ایجاب می‌کند که $\{e_i^{(4)}\}_{i=1}^{\infty}$ همگرا به صفر نباشد. اگر مقادیر $(q_i^{(3)} + q_i^{(4)})$ و $q_{i-1}^{(3)}q_i^{(4)}$ ، همانطور که در جدول ۱۵.۲ نموده شده، حساب شوند، می‌توان دید که $x^2 + .712124x + .153271$ یک تقریب برای عامل درجه دو است، که دارای ریشه‌های $-.356062 \pm .162760i$ - می‌باشد. مقادیر واقعی ریشه‌های چند جمله‌ای $16x^4 - 40x^3 - 5x^2 + 20x + 6$ عبارتند از $1.241677, 1.970446, -.356062 \pm .162758i$ که همه تا شش رقم اعشار دقیق‌اند.

i	$r_i = q_i^{(3)} + q_i^{(4)}$	$s_i = q_{i-1}^{(3)}q_i^{(4)}$
1	.000000	—
2	-4.000000	.000000
3	-.412121	1.200000
4	-.829774	.036364
5	-.690665	.209347
6	-.713197	.139273
7	-.713794	.156062
8	-.711044	.153139
9	-.712578	.153061
10	-.711970	.153405
11	-.712266	.153214
12	-.712115	.153336
13	-.712124	.153265
14	-.712124	.153271

جدول ۱۵.۲

برای یافتن تقریبهای بهتر به این ریشه‌ها، می‌شود از روندهای الگوریتم ۶.۲ و ۷.۲، همراه با روش نیوتن، استفاده کرد. با اینحال، روشی که عموماً "با آن تمام ریشه‌های یک چند جمله‌ای بدست آید وجود ندارد. بخصوص، ریشه‌های مکرر و ریشه‌های مختلط که دارای یک اندازه هستند مشکلات زیادی به بار می‌آورند.

مجموعه تمرینات ۶.۲

۱. با استفاده از قضیه ۱۹.۲،

$$P(x) = 5x^4 - 2x^2 + 3x + 4$$

را در $x_0 = 3$ حساب کنید. همچنین، به وسیله تقسیم ترکیبی روش هورنر، P را در $x_0 = 3$ محاسبه نمایید.

۲. ریشه‌های تقریبی هریک از چندجمله‌ایهای زیر را به این ترتیب پیدا کنید که ابتدا ریشه‌های حقیقی آنها را بیابید و سپس، برای تعیین ریشه‌های مختلط، چندجمله‌ایها را به درجات پایین‌تر تقلیل دهید:

(آ) $P(x) = x^4 + 5x^3 - 9x^2 - 85x - 136$

(ب) $P(x) = x^4 - 2x^3 - 12x^2 + 16x - 40$

(پ) $P(x) = x^4 + x^3 + 3x^2 + 2x + 2$

(ت) $P(x) = x^5 + 11x^4 - 21x^3 - 10x^2 - 21x - 5$

۳. (آ) نتیجه ۱۶.۲ را ثابت کنید.

(ب) نتیجه ۱۷.۲ را ثابت کنید.

۴. قضیه ۱۸.۲ را ثابت کنید. (راهنمایی. ابتدا حالت $m = 1$ را در نظر بگیرید

و ببینید اگر $u - bi$ ریشه $P(x)$ نباشد چه روی می‌دهد.)

۵. قضیه زیر را ثابت کنید.

قضیه. فرض کنید $P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ یک چندجمله‌ای از درجه n بوده، و x_0 عدد حقیقی مثبتی با خاصیت $P(x_0) > 0$ باشد. هرگاه $Q(x)$ چند

جمله‌ای ساخته شده در الگوریتم ۶.۲ باشد، یعنی

$$P(x) = (x - x_0)Q(x) + P(x_0) \\ = (x - x_0)(b_n x^{n-1} + \dots + b_2 x + b_1) + P(x_0)$$

و به ازای $i = 1, 2, \dots, n$ ، $b_i > 0$ ، آنگاه همه ریشه‌های P نابیشتر از x_0 اند.

(توجه کنید که این قضیه را گاهی می‌توان برای تقلیل کار لازم برای جستجو در

$[-1, 1]$ برای صفرهای P بکار برد.)

۶. فرض کنید $P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$. نشان دهید که اگر به ازای

$a_i > 0$ ، $i = 1, 2, \dots, n$ ، ریشه نامنفی ندارد (اگر P به این شکل باشد، برای

ریشه‌های P فقط در $[-1, 0]$ جستجو کنید.)

۷. (آ) نشان دهید که اگر x_0 یک ریشه مضاعف $P(x) = a_n x^n + \dots + a_1 x + a_0$ باشد، x_0 یک ریشه $P'(x)$ است. نمودار $P(x)$ نزدیک $x = x_0$ به چه شکل است؟
 (ب) چند جمله‌ای

$$P(x) = x^3 - 4.7x^2 + 2.9225x - .49.$$

را در نظر بگیرید. با محاسبه $P(x)$ و $\hat{P}(x) = -.49x^3 + 2.9225x^2 - 4.7x + 1$ و بازای $x_j = -1 + .1j$ ، $j = 0, 1, \dots, 20$ ، در $[-1, 1]$ بدنیال ریشه‌های P بگردید. چون P یک ریشه مضاعف دارد، فقط یک ریشه حقیقی را مشخص می‌کنید. با توجه به نتایج حاصل، آیا علائمی برای وجود یک ریشه مضاعف برای P وجود دارد؟

۸. $P(x) = 10x^3 - 8.3x^2 + 2.295x - 21141$ دارای یک ریشه در $x = .29$ است. با استفاده از روش نیوتن به ازای تقریب اولیه $x_0 = .26$ ، سعی کنید این ریشه را بیابید. چه اتفاقی می‌افتد؟ فرض کنید تنها ریشه‌ای که می‌خواهید $x = .29$ باشد؛ چگونه می‌توان تقریب اولیه به قدر کافی مناسبی بدست آورد بطوری که روش نیوتن به $x = .29$ همگرا شود؟

۹. تمرین ۲ را تکرار کنید، ابتدا، با استفاده از الگوریتم QD 8.02 ، تقریبهای اولیه‌ای به ریشه‌ها بیابید.
 ۱۰. همه ریشه‌های

$$P(x) = 16x^4 + 88x^3 + 159x^2 + 76x - 240$$

را، ابتدا با تقریب ریشه‌ها با استفاده از الگوریتم 8.02 و سپس با استفاده از روش نیوتن برای بهبود تقریبها، تا شش رقم اعشار بدست آورید.

۱۱. وقتی یکی از ضرایب a_0, a_1, \dots, a_{n-1} از چند جمله‌ای

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

صفر باشد، الگوریتم QD را نمی‌توان برای بدست آوردن ریشه‌های آن بکار برد. (آ) توضیح دهید چرا مطلب فوق صحت دارد.

(ب) نشان دهید که همیشه تغییر متغیری به شکل $z = x - a$ وجود دارد که می‌توان با آن چند جمله‌ای P را به چند جمله‌ای با همان درجه تغییر داد که همه ضرایب آن ناصفر باشند.

(پ) رابطه بین ریشه‌های P و ریشه‌های چند جمله‌ای Q ، که از P با استفاده از تغییر متغیر $z = x - a$ بدست می‌آید، را مشخص کنید.

۱۲. با استفاده از روند تمرین ۱۱ و الگوریتم QD، ۸۰۲، برای ریشه‌های

$$P(x) = x^4 - 4x^2 - 3x + 5$$

تقریبهای اولیهای بدست آورید، و سپس تقریبهایی از این ریشه‌ها را که تا شش رقم اعشار دقیق باشند پیدا کنید.

۱۳. با استفاده از نماد گذاری الگوریتم ۸۰۲، نشان دهید هرگاه

$$\lim_{i \rightarrow \infty} q_i^{(k)} = 0 \text{ و } \lim_{i \rightarrow \infty} e_i^{(k+1)} = 0 \text{ و } \lim_{i \rightarrow \infty} e_i^{(k)} = 0$$

۱۴. در توصیف یک مدل ریاضی برای نامنظم کار کردن ابزار ماشین در [۴۸]، مولفان

به‌ازای مقادیر مختلف n و ثابتهای $\phi_1, \phi_2, \dots, \phi_{2n}$ به ریشه‌های یک چندجمله‌ای

به‌شکل

$$P_{2n}(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_{2n} B^{2n}$$

نیاز دارند. مقدار n به میزان داده‌های تجربی بکاررفته بستگی دارد، و با استفاده

از این داده‌ها و یک روش کمترین مربعات خطی، که در بخش ۲۰۴ مطرح می‌شود،

تقریبهایی به ثابتهای ϕ_i ، $i = 1, 2, \dots, 2n$ ، بدست می‌آیند.

(T) در یک وضعیت تجربی معلوم شده است که، به‌ازای $n = 2$ ، تقریبهایی معقول

برای ϕ_i ها عبارتند از

$$\phi_1 = 1.8310, \quad \phi_2 = -.5218$$

$$\phi_3 = -.4754, \quad \phi_4 = .1595.$$

با استفاده از این مقادیر، ریشه‌های $P_4(B)$ را بیابید.

(ب) در همین آزمایش، به‌ازای $n = 3$ ، مقادیر زیر برای ϕ_i بدست آمده‌اند:

$$\phi_1 = 1.742, \quad \phi_2 = -.0385, \quad \phi_3 = -.8133$$

$$\phi_4 = -.1061, \quad \phi_5 = .2019, \quad \phi_6 = .0383.$$

با استفاده از این مقادیر، ریشه‌های $P_6(B)$ را پیدا کنید.

درونیابی و تقریب چند جمله‌ای^۳

در ایالات متحده، هر ۱۰ سال یکبار سرشماری می‌شود. در جدول زیر، تعداد جمعیت این کشور از ۱۹۲۰ تا ۱۹۷۰، به هزار، درج شده است.

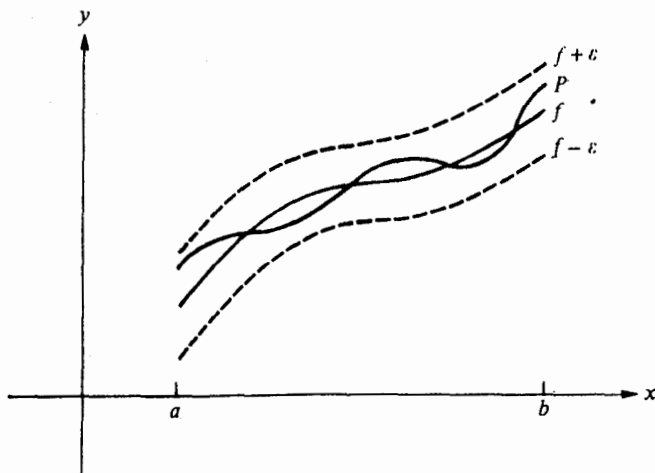
سال	1920	1930	1940	1950	1960	1970
جمعیت (به‌هزار)	105,711	123,203	131,669	150,697	179,323	203,212

با توجه‌بما این داده‌ها، می‌پرسیم آیا می‌توان با استفاده از آن جمعیت را، مثلاً در ۱۹۶۵ یا حتی در ۲۰۰۰، به‌طور معقول تخمین زد. این نوع پیشگویی را می‌توان با استفاده از تابعی که مناسب این داده‌هاست بدست آورد. این مبحث را درونیابی می‌نامند و موضوع این فصل می‌باشد.

یکی از مفیدترین و معروفترین رده‌های توابع که خط حقیقی را بتوی خود می‌نگارد رده چند جمله‌ایهای جبری است؛ یعنی، مجموعه توابعی به شکل

$$P_n(x) = a_0 + a_1x + \dots + a_nx^n,$$

که در آن n یک عدد صحیح نامنفی است و a_0, \dots, a_n ثابتهایی حقیقی می‌باشند. یک دلیل عمده بر اهمیت آنها این است که توابع پیوسته را به‌طور یکنواخت تقریب می‌کنند؛ یعنی، به‌ازای هر تابع تعریف شده و پیوسته بر یک بازه بسته، چند جمله‌ای وجود دارد



شکل ۱.۳

که هر قدر بخواهیم به تابع مفروض "نزدیک" است. این نتیجه در قضیه زیر دقیقترین شده است. (ر.ک. شکل ۱۰۳).

قضیه ۱۰۳ (قضیه تقریب وایراشتراس^۱). هرگاه f بر $[a, b]$ تعریف شده و پیوسته باشد، و $\varepsilon > 0$ نیز مفروض باشد، آنگاه چندجمله‌ایی چون P ، که بر $[a, b]$ تعریف شده است، وجود دارد بطوری که

$$|f(x) - P(x)| < \varepsilon, \quad x \in [a, b]$$

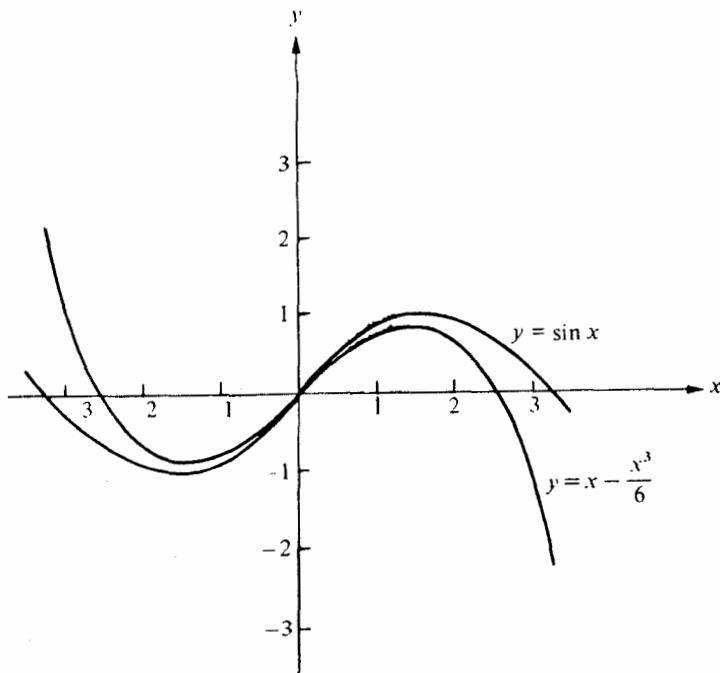
اثبات این قضیه از حوصله این درس خارج است، و می‌توان آن را در هر کتاب مقدماتی آنالیز حقیقی یافت. (مثلاً، ر.ک. بارتل^۲ [۸]، صفحات ۱۶۵ تا ۱۷۲).

خاصیت مهم دیگر رده چندجمله‌ایها در تقریب توابع این است که مشتق و انتگرال نامعین هر چندجمله‌ای به آسانی حساب می‌شوند و هر کدام یک چندجمله‌ای اند. به این دلایل، رده چندجمله‌ایها اغلب برای تقریب توابع دیگر که یا پیوسته‌اند و یا پیوسته فرض می‌شوند بکار می‌رود.

۱۰۳ چندجمله‌ایهای تبلور

قضیه وایراشتراس از دیدگاه نظری بسیار مفید است، لیکن نمی‌توان از آن برای مقاصد محاسبه‌ای به‌طور موثر استفاده کرد. به‌جای یافتن یک چندجمله‌ای که یک تابع را بر کل یک بازه به‌طور یکنواخت تقریب کند، اغلب بهتر است چندجمله‌ایی بیابیم که در شرایطی که برای مسئله مورد نظر مفیدند صدق کند و درعین حال بنوعی به آن تابع "نزدیک" باشد.

مثال ۱. یک چندجمله‌ای از درجه ۳ یا کمتر بیابید که $f(x) = \sin x$ را در مجاورت $x_0 = 0$ تقریب کند، و، با استفاده از این چندجمله‌ای، \sin را تقریب نمایید. (ر.ک. شکل ۲۰۳).



شکل ۲.۳

مثال فوق ممکن است به نظر مبهم بیاید، چرا که هیچ توضیحی برای یافتن چند جمله‌ای تقریب ساز داده نشده است، و نیز عبارت "در مجاورت x_0 " بوضوح تعریف نشده است. یک شرط منطقی برای یک چندجمله‌ای تقریب ساز می‌تواند این باشد که با تابع مفروض و حداکثر مشتقات آن در نقطه x_0 یکی باشد، که اثرش این است که "شکل" چند جمله‌ای حتی الامکان به شکل تابع مفروض در مجاورت x_0 نزدیک است.

برای انجام کار به این طریق، قرار می‌دهیم

$$P(x) = a_0 + a_1x + a_2x^2 + a_3x^3.$$

و a_0, a_1, a_2, a_3 را طوری تعیین می‌کنیم که $P(0) = f(0)$ ، $P'(0) = f'(0)$ ، و غیره. با مشتقگیری از P و محاسبه در $x = 0$ ، داریم

$$P(x) = a_0 + a_1x + a_2x^2 + a_3x^3 \quad ; \quad \text{در نتیجه، } P(0) = a_0,$$

$$P'(x) = a_1 + 2a_2x + 3a_3x^2 \quad ; \quad \text{در نتیجه، } P'(0) = a_1,$$

، $P''(0) = 2a_2$ ، در نتیجه ، $P''(x) = 2a_2 + 6a_3x$

، $P'''(0) = 6a_3$ ، در نتیجه ، $P'''(x) = 6a_3$

چون $f(0) = \sin 0 = 0$ ، $f'(0) = \cos 0 = 1$ ، $f''(0) = -\sin 0 = 0$ ، و $f'''(0) = -\cos 0 = -1$ نتیجه می شود که

، $6a_3 = -1$ و $2a_2 = 0$ ، $a_1 = 1$ ، $a_0 = 0$

چندجمله‌ای تقریب ساز از درجه ۳ یا کمتر عبارت است از

$$P(x) = x - \frac{1}{6}x^3$$

و

$$\sin .1 = f(.1) \approx P(.1) = .1 - \frac{1}{6}(.001) = .09983333.$$

با مراجعه به جداول موجود ، نتیجه می شود که این تقریب و $\sin .1$ با خطایی حدود 10^{-7} یکی اند .

باید توجه داشت که روند مثال ۱ دقیقاً " یافتن چندجمله‌ای تیلور درجه ۳ است ، که در قضیه ۱۳۰۱ ، صفحه ۹ ، برای تابع f ، که حول نقطه x_0 بسط داده شده ، مطرح گردید . در نتیجه ، چندجمله‌ایهای تیلور را می توان به عنوان چندجمله‌ایهای تقریب ساز " در مجاورت " یک نقطه " مفروض بکار برد .

مثال ۲

(T) چندجمله‌ای تیلور درجه ۳ را برای

$$f(x) = (1+x)^{1/2} \text{ در مجاورت } x_0 = 0$$

حساب کنید .

(ب) با استفاده از چندجمله‌ای قسمت (T) ، $\sqrt{1.1}$ را تقریب کنید ، و برای خطای حاصل کرانی بیابید .

(پ) با استفاده از چندجمله‌ای قسمت (T) ، $\int_0^1 (1+x)^{1/2} dx$ را تقریب کنید ، و برای خطای این تقریب کرانی بیابید .

با مشتقگیریهای لازم ، داریم

، $f(0) = 1$ ، در نتیجه ، $f(x) = (1+x)^{1/2}$

، $f'(0) = \frac{1}{2}$ ، در نتیجه ، $f'(x) = \frac{1}{2}(1+x)^{-1/2}$

$$, \quad \underline{f''(0) = -\frac{1}{4}}, \quad \text{در نتیجه, } f''(x) = -\frac{1}{4}(1+x)^{-3/2}$$

$$, \quad \underline{f'''(0) = \frac{3}{8}}, \quad \text{در نتیجه, } f'''(x) = \frac{3}{8}(1+x)^{-5/2}$$

$$, \quad \underline{f^{(iv)}(\xi) = -\frac{15}{16}(1+\xi)^{-7/2}}, \quad \text{در نتیجه, } f^{(iv)}(x) = -\frac{15}{16}(1+x)^{-7/2}$$

که در آن $0 < \xi < x$. طبق قضیهٔ تیلور ،

$$(۱.۳) \quad P_3(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3$$

$$= 1 + \frac{1}{2}x - \frac{1}{4} \cdot \frac{1}{2!}x^2 + \frac{3}{8} \cdot \frac{1}{3!}x^3$$

$$= 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3,$$

چندجمله‌ای تیلور درجهٔ سه مطلوب در (T) است .

برای پاسخ دادن به قسمت (ب) ، داریم

$$\sqrt{1.1} = f(1) \approx P_3(1) = 1 + \frac{1}{2}(1) - \frac{1}{8}(1)^2 + \frac{1}{16}(1)^3 = 1.0488125,$$

خطای حاصل درقضیهٔ ۱۳.۰۱ با $R_3(1)$ داده شده است ، و یک کران برای آن عبارت است

از

$$\begin{aligned} |R_3(1)| &= \frac{|-\frac{15}{16}(1+\xi)^{-7/2}|}{4!}(1)^4 \\ &\leq \frac{15}{(16)(24)}(1)^4 \max_{\xi \in [0,1]} (1+\xi)^{-7/2} \\ &= \frac{.0005}{128}(1) \\ &\leq 3.91 \times 10^{-6}. \end{aligned}$$

چون مقدار دقیق $\sqrt{1.1}$ (تا ۸ رقم) ۱.۰۴۸۸۰۸۸ است ، خطای واقعی حدود 3.7×10^{-6} می‌باشد .

محاسبات قسمت (پ) به صورت زیر است :

$$\int_0^1 (1+x)^{1/2} dx \approx \int_0^1 P_3(x) dx$$

$$\begin{aligned}
 &= \int_0^1 \left(1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} \right) dx \\
 &= \left(x + \frac{x^2}{4} - \frac{x^3}{24} + \frac{x^4}{64} \right) \Big|_0^1 \\
 &= .1024598958,
 \end{aligned}$$

با باقیمانده $\int_0^1 R_3(x) dx$. با استفاده از روشهایی شبیه آنچه که در قسمت (ب) بکار رفت، کران زیر برای خطا بدست می‌آید:

$$\begin{aligned}
 \left| \int_0^1 R_3(x) dx \right| &= \frac{15}{(16)4!} \int_0^1 (1 + \xi)^{-7/2} x^4 dx \\
 &\leq \frac{5}{128} \int_0^1 x^4 dx \\
 &= \frac{5}{128} \cdot \frac{x^5}{5} \Big|_0^1 \\
 &\leq 7.82 \times 10^{-8}
 \end{aligned}$$

توجه کنید که جمله $\int_0^1 R_3(x) dx$ واقعی، یعنی $\int_0^1 R_3(x) dx$ منفی است؛ در نتیجه، انتگرال مطلوب نمی‌تواند از تقریب بیشتر شود. بنابراین،

$$\int_0^1 P_3(x) dx - \int_0^1 |R_3(x)| dx \leq \int_0^1 (1+x)^{1/2} dx \leq \int_0^1 P_3(x) dx$$

یا

$$.1024598176 \leq \int_0^1 (1+x)^{1/2} dx \leq .1024598958.$$

چون مقدار واقعی $\int_0^1 (1+x)^{1/2} dx$ مساوی $.102459822$ است، خطای دقیق حدود 7.4×10^{-8} می‌باشد.

در این مرحله ممکن است خواننده احساس کند که چند جمله‌ای تیلور همیشه کارا است. مثال زیر نیاز به روشهای تقریب ساز دیگر را روشن می‌سازد.

مثال ۳. شکل ۲.۳ پیشنهاد می‌کند که چند جمله‌ای تیلور درجه سه برای تابع سینوس،

که حول $x_0 = 0$ بسط داده شده، وقتی x از صفر دور شود، نتایجی با دقت نزولی بدست می‌دهد. در جدول زیر مقادیر چندجمله‌ای تیلور درجه سه (۱.۳) برای تابع $f(x) = \sqrt{1+x}$ که در مثال پیش در نظر گرفته شد، و خطای ناشی از استفاده از این چندجمله‌ای به‌ازای مقادیر مختلف x درج شده‌اند.

x	.1	.5	1	2	10
$P_3(x)$	1.048813	1.2266	1.438	2.00	56.00
$f(x)$	1.048809	1.2247	1.414	1.73	3.32
$ P_3(x) - f(x) $	4×10^{-6}	2×10^{-3}	2×10^{-2}	3×10^{-1}	53

در بعضی مواقع، اگر از چندجمله‌ایهای تیلور درجات بالاتر استفاده کنیم، تقریبهای بهتری بدست می‌آیند، اما همیشه چنین نیست. به‌عنوان مثال، مسئله استفاده از چندجمله‌ایهای تیلور با درجات مختلف برای $f(x) = 1/x$ ، که حول $x_0 = 1$ بسط داده شده‌اند، را برای تقریب $\frac{1}{3} = f(3)$ در نظر می‌گیریم. چون $f(x) = 1/x$ ، به‌آسانی معلوم می‌شود که چندجمله‌ای تیلور به‌ازای $n \geq 1$ عبارت است از

$$P_n(x) = \sum_{k=0}^n \frac{f^{(k)}(1)}{k!} (x-1)^k = \sum_{k=0}^n (-1)^k (x-1)^k.$$

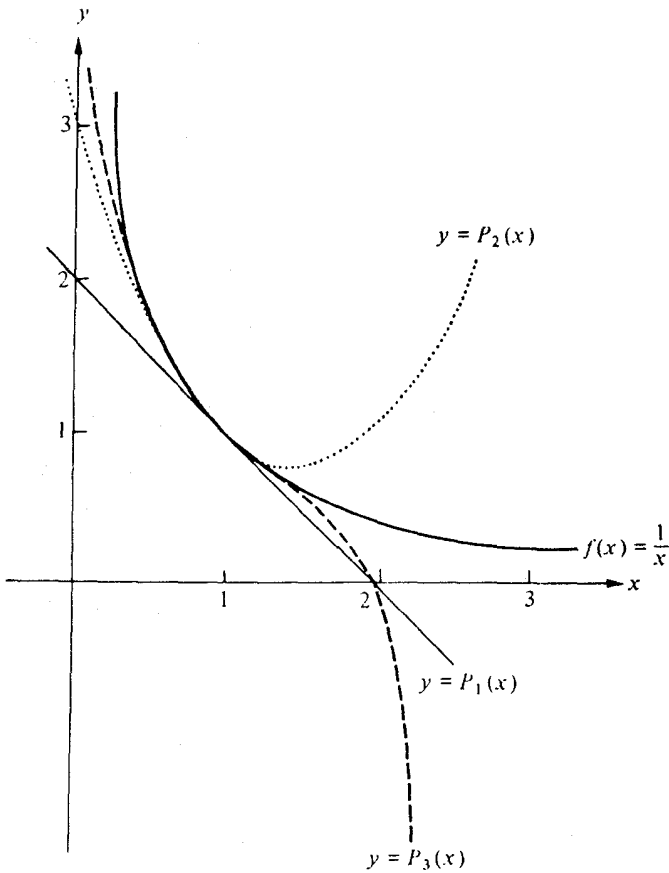
برای تقریب $f(3) = \frac{1}{3}$ به‌وسیله $P_n(3)$ ، به‌ازای مقادیر صعودی n داریم

n	0	1	2	3	4	5	6	7
$P_n(3)$	1	-1	3	-5	11	-21	43	-85

دلیل آنکه این روش تقریب ساز کارا نیست این است که قدر مطلق جمله خطا، یعنی

$$R_n(x) = \frac{(-1)^{n+1}(x-1)^{n+1}}{\xi^{n+2}}, \quad \text{که در آن } 1 < \xi < x$$

با افزایش n رشد می‌کند (ر.ک. شکل ۳.۳). علت رشد خطا این است که $x = 3$ "به قدر کافی نزدیک" $x_0 = 1$ نیست. برای بحث کاملتری از مشکلات مربوط به سریهای توانی، به کتابی در باب نظریه توابع تحلیلی، مثلاً "آلفرس" [۳]، مراجعه کنید.



شکل ۳.۳

چون چند جمله‌ایهای تیلور دارای این خاصیت‌اند که همهٔ اطلاعات بکار رفته در تقریب در یک نقطه، یعنی در x_0 ، متمرکز شده‌اند، نوع مشکلی که در مثال ۳ یا آن مواجه بودیم کاملاً معمولی است. این امر استفاده از چند جمله‌ایهای تیلور را به حالتی محدود می‌کند که در آن الزاماً "تقریبها در نقاطی خیلی نزدیک به x_0 باشند. برای مقاصد محاسبه‌ای معمولی، استفاده از روشهایی که شامل اطلاعات در نقاط مختلف‌اند موثرتر است، و ساختن این نوع چند جمله‌ای تا پایان این فصل مورد نظر خواهد بود.

مجموعه تمرینات ۱.۳

۱. چندجمله‌ای تیلور درجه ۲ برای $f(x) = x^2 - 3$ که حول نقطه $x_0 = 1$ (آ) بسط داده شده است را بیابید. (ب) $x_0 = 0$ بسط داده شده است را بیابید.
۲. چندجمله‌ای تیلور درجه ۳ برای $f(x) = (1+x)^{-2}$ که حول $x_0 = 0$ بسط داده شده است را بیابید، و، با استفاده از آن، $f(0.05)$ را تقریب کنید. برای خطای این تقریب کرانی بیابید و نتیجه خود را با مقدار دقیق $f(0.05)$ مقایسه کنید.
۳. با استفاده از چندجمله‌ای تمرین ۲، $\int_0^{0.05} (1+x)^{-2} dx$ را تقریب کنید. برای خطای این تقریب کرانی بیابید، و نتیجه خود را با مقدار واقعی این انتگرال مقایسه نمایید.
۴. با استفاده از چندجمله‌ای تیلور درجه n حول $x_0 = 1$ ، کوچکترین عدد صحیح n را برای تقریب $f(x) = 1/x$ در $x = 1.25$ با دقت 10^{-8} بیابید.
۵. خطای ناشی از بکار بردن $\sin x \approx x$ در تقریب $\sin 1^\circ$ چقدر است؟
۶. فرض کنید $f(x) = \ln(1+x)$. چندجمله‌ای تیلور درجه ۴ چهارم برای f که حول $x_0 = 0$ بسط داده شده است را بیابید، و، با استفاده از آن، $\ln(1.1)$ را تقریب نمایید. برای خطای ناشی از این تقریب کرانی بیابید.
۷. فرض کنید $F(x) = \int_0^x (1+t)^{-1} dt$. با استفاده از چندجمله‌ای تیلور درجه ۳ برای $f(x) = (1+x)^{-1}$ که حول $x_0 = 0$ بسط داده شده است، $F(1)$ را تقریب کنید. نتایج خود را با نتایج تمرین ۶ مقایسه نمایید.

چندجمله‌ایهای برنشتاین^۱. اگر f تابعی بر $[0, 1]$ باشد، چندجمله‌ای برنشتاین از درجه n برای f عبارت است از

$$B_n(x) = \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k},$$

که در آن $\binom{n}{k}$ مساوی $\frac{n!}{k!(n-k)!}$ است. می‌توان نشان داد که، اگر f بر $[0, 1]$ پیوسته بوده و $x_0 \in [0, 1]$ ،

$$\lim_{n \rightarrow \infty} B_n(x_0) = f(x_0).$$

از این چند جمله‌ایها می‌توان در یک برهان ساختنی قضیهٔ وایراشتراس استفاده کرد (ر. ک. بارتل [۸]).

تمرینهای زیر مربوط به چند جمله‌ایهای برنشتاین‌اند.

۸. $B_3(x)$ را برای توابع زیر بیابید:

$$f(x) = x, x \in [0, 1] \quad (\Gamma) \quad ; \quad f(x) = 1, x \in [0, 1] \quad (\beta)$$

۹. نشان دهید که به‌ازای هر $k \leq n$,

$$\binom{n-1}{k-1} = \binom{k}{n} \binom{n}{k}$$

۱۰. با استفاده از تمرین ۹ و این مطلب که

$$1 = \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k}, \quad \text{به‌ازای هر } n$$

نشان دهید که، به‌ازای $f(x) = x^2$,

$$B_n(x) = \left(\frac{n-1}{n}\right)x^2 + \frac{1}{n}x.$$

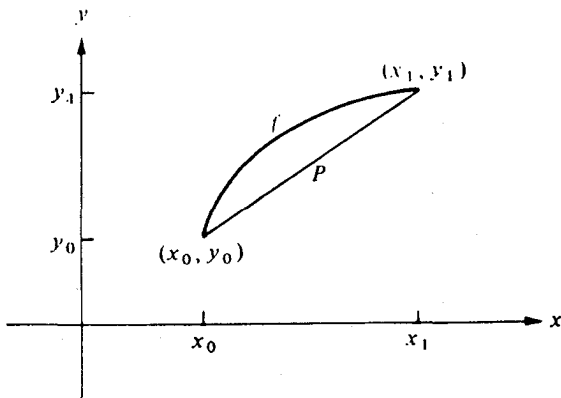
۱۱. با استفاده از تمرین ۱۰، اندازهٔ n لازم برای آنکه $|B_n(x) - x^2| \leq 10^{-6}$ به‌ازای هر x در $[0, 1]$ برقرار باشد را تخمین بزنید. به‌نظر شما چند جمله‌ایهای برنشتاین ابزارهایی عملی برای تقریب سازی توابع پیوسته‌اند؟

۲.۳ درونیایی و چند جمله‌ای لاگرانژ^۱

در بخش قبل، یک چند جمله‌ای تقریب ساز که با یک تابع و بعضی از مشتقاتش در یک نقطه یکی باشد مورد بحث قرار گرفت. این چند جمله‌ای برای توابعی که مشتقات آنها موجود و به‌آسانی قابل محاسبه‌اند روی بازه‌های کوچک کاملاً مفید است، اما واضح است که همیشه وضع به‌این صورت نیست. در نتیجه، چند جمله‌ای تیلور اغلب مفید نیست، و باید روشهای دیگری برای تقریب جستجو شود. مطالب این بخش مربوطند به یافتن چند جمله‌ایهای تقریب ساز که بتوان آنها را فقط با تعیین چند نقطه در صفحه که باید از آنها بگذرند مشخص کرد.

مسئلهٔ تعیین یک چند جمله‌ای از درجهٔ l که از نقاط متمایز (x_0, y_0) و (x_1, y_1)

می‌گذرد را در نظر می‌گیریم. این مسئله عبارت است از تقریب سازی یک تابع مانند f ، که $f(x_0) = y_0$ و $f(x_1) = y_1$ ، به وسیله یک چندجمله‌ای درجه اول که با مقادیر f در نقاط مفروض حالت درونیابی دارد یا با آنها یکی است. (ر.ک. شکل ۴.۳)



شکل ۴.۳

اگر

(۲.۳)

$$P(x) = a_0 + a_1x$$

این چندجمله‌ای باشد، a_0 و a_1 باید در روابط زیر صدق کنند:

$$y_0 = P(x_0) = a_0 + a_1x_0,$$

$$y_1 = P(x_1) = a_0 + a_1x_1.$$

با حل این معادلات نسبت به a_0 و a_1 ، داریم

$$a_1 = \frac{y_0 - y_1}{x_0 - x_1}$$

و

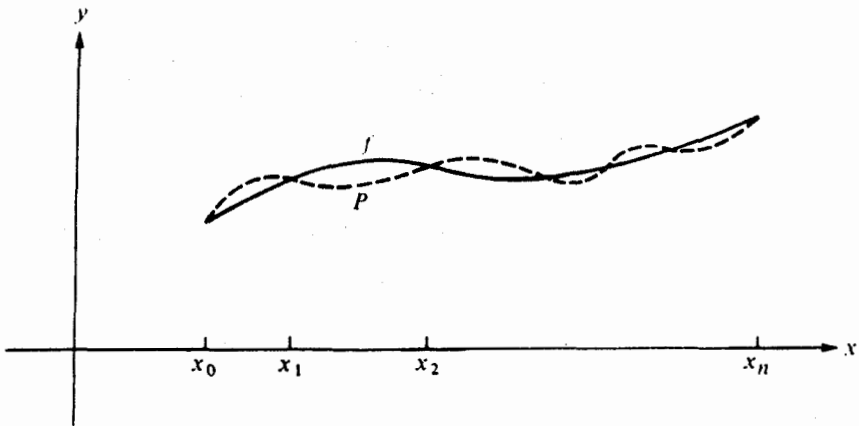
$$a_0 = y_1 - a_1x_1 = y_1 - \left(\frac{y_0 - y_1}{x_0 - x_1}\right)x_1.$$

با جایگذاری مقادیر a_0 و a_1 در معادله (۲.۳)، نتیجه می‌شود

$$\begin{aligned}
 P(x) &= y_1 - \left(\frac{y_0 - y_1}{x_0 - x_1} \right) x_1 + \left(\frac{y_0 - y_1}{x_0 - x_1} \right) x \\
 &= \frac{y_1(x_0 - x_1) - x_1(y_0 - y_1) + x(y_0 - y_1)}{x_0 - x_1} \\
 &= \frac{y_1(x_0 - x_1 + x_1 - x)}{x_0 - x_1} + \frac{y_0(-x_1 + x)}{x_0 - x_1} \\
 &= \frac{(x - x_1)}{(x_0 - x_1)} y_0 + \frac{(x - x_0)}{(x_1 - x_0)} y_1.
 \end{aligned}$$

واضح است که با نوشتن $P(x)$ به شکل اخیر، $P(x_0) = y_0$ و $P(x_1) = y_1$ ، همچنین، با این روش می‌توان مقادیر یک تابع را بین دو مقدار ثبت شده تقریب کرد، یعنی، این یک روش "درونیابی" است که اغلب در جداول مثلثاتی یا لگاریتمی بکار می‌رود. آنچه روشن نیست این است که P تنها چند جمله‌ای از درجه ۱ یا کمتر با خاصیت درونیابی است. بهر حال، این مطلب فوراً از نتیجه ۱۷.۲، ص ۸۴، بدست می‌آید.

برای تعمیم مفهوم درونیابی خطی، یک چند جمله‌ای از درجه حداکثر n می‌یابیم که از $(n + 1)$ نقطه معلوم بگذرد. این کار را می‌توان یک روش تقریبی دانست که در آن، به ازای تابع معلوم f ، چند جمله‌ایی مانند P می‌یابیم که با مقادیر تابع در نقاطی معلوم



شکل ۵.۳

یکی بوده، و سپس، چندجمله‌ای P را برای تقریب f در نقاط دیگر بکار می‌بریم. روند درونیابی که هم‌اکنون به اختصار توضیح داده شد به تفصیل در قضیه بعد توصیف می‌شود. این چندجمله‌ای شکل لاگرانژ چندجمله‌ای درونیاب نامیده می‌شود. (ر. ک. شکل ۰.۵.۳)

قضیه ۲.۳. هرگاه $(n+1)$ ، x_0, x_1, \dots, x_n نقطه متمایز بوده و f تابعی با مقادیر معلوم در این نقاط باشد، آنگاه چندجمله‌ای منحصر بفردی مانند P ، از درجه حداکثر n ، وجود دارد با این خاصیت که،

$$f(x_k) = P(x_k) \quad , \quad k = 0, 1, \dots, n$$

به ازای هر

این چندجمله‌ای با رابطه زیر داده می‌شود:

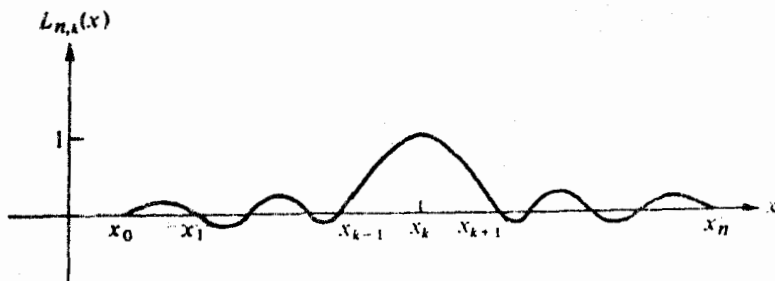
$$(۳.۳) \quad P(x) = f(x_0)L_{n,0}(x) + \dots + f(x_n)L_{n,n}(x) \\ = \sum_{k=0}^n f(x_k)L_{n,k}(x),$$

که در آن،

به ازای $k = 0, 1, \dots, n$

$$(۴.۳) \quad L_{n,k}(x) = \frac{(x-x_0)(x-x_1)\dots(x-x_{k-1})(x-x_{k+1})\dots(x-x_n)}{(x_k-x_0)(x_k-x_1)\dots(x_k-x_{k-1})(x_k-x_{k+1})\dots(x_k-x_n)} \\ = \prod_{\substack{i=0 \\ i \neq k}}^n \frac{(x-x_i)}{(x_k-x_i)}$$

وقتی در مورد درجه اشتباهی رخ ندهد، به جای $L_{n,k}(x)$ فقط می‌نویسیم $l_k(x)$.



شکل ۶.۳

برهان. توجه کنید که چندجمله‌ای L_k به‌ازای هر $k = 0, 1, \dots, n$ از درجه n است و،
بعلاوه، به‌ازای هر $i = 0, 1, \dots, n$ (ر.ک. شکل ۵.۳)،

$$(۵.۳) \quad L_k(x_i) = \begin{cases} 0 & , \quad k \neq i \\ 1 & , \quad k = i \end{cases}$$

از اینرو، به‌ازای $i = 0, 1, \dots, n$ ، چون $P(x_i) = \sum_{k=0}^n f(x_k)L_k(x_i) = f(x_i)L_i(x_i) = f(x_i)$ ، هر L_k یک چندجمله‌ای درجه n است، P یک چندجمله‌ای از درجه حداکثر n است و وجود P ای صادق در $P(x_k) = f(x_k)$ ثابت می‌شود. یکتایی آن فوراً از نتیجه ۱۷.۲ حاصل خواهد شد.

مثال ۱. با استفاده از نقاط، یا گره‌های $x_0 = 2$ ، $x_1 = 2.5$ ، و $x_2 = 4$ ، چندجمله‌ای درونیاب درجه ۲ دورا برای $f(x) = 1/x$ تعیین و برای تقریب $\frac{1}{3} = f(3)$ ابتدا چندجمله‌ایهای ضریب یعنی L_0 ، L_1 ، و L_2 را به‌صورت زیر معین می‌کنیم:

$$L_0(x) = \frac{(x - 2.5)(x - 4)}{(2 - 2.5)(2 - 4)} = x^2 - 6.5x + 10,$$

$$L_1(x) = \frac{(x - 2)(x - 4)}{(2.5 - 2)(2.5 - 4)} \approx -1.333x^2 + 8x - 10.667,$$

$$L_2(x) = \frac{(x - 2)(x - 2.5)}{(4 - 2)(4 - 2.5)} \approx .333x^2 - 1.5x + 1.667.$$

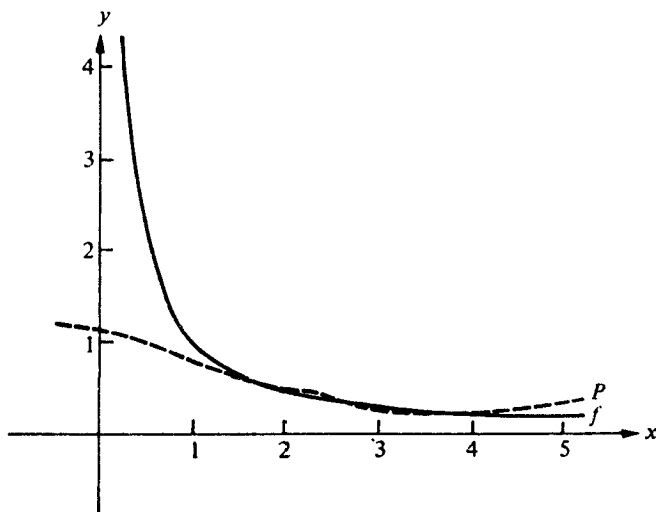
چون $f(x_0) = f(2) = .5$ ، $f(x_1) = f(2.5) = .4$ ، و $f(x_2) = f(4) = .25$ ،

$$\begin{aligned} P(x) &= \sum_{k=0}^2 f(x_k)L_k(x) \\ &\approx .5(x^2 - 6.5x + 10) + .4(-1.333x^2 + 8x - 10.667) \\ &\quad + .25(.333x^2 - 1.5x + 1.667) \\ &\approx .05x^2 - .425x + 1.15 \\ &\approx (.05x - .425)x + 1.15. \end{aligned}$$

بنابراین،

$$\frac{1}{3} = f(3) \approx P(3) \approx .325.$$

از مقایسه این مثال با مثال ۳ از بخش ۱.۳، که در آن از هیچ چندجمله‌ای تیلور (بسط داده شده حول $x_0 = 1$) نمی‌توان برای تقریب معقول $f(3) = \frac{1}{3}$ استفاده کرد، معلوم می‌شود که چندجمله‌ای درونیاب می‌تواند تقریبهای مفیدی بدست دهد. (ر.ک. شکل ۷.۳)



شکل ۷.۳

قدم بعدی محاسبه جمله باقیمانده یا کران خطای ناشی از تقریب سازی یک تابع با یک چندجمله‌ای درونیاب است. این کار در قضیه زیر انجام می‌شود.

قضیه ۳.۳. هرگاه x_0, x_1, \dots, x_n نقاط متمایزی در بازه $[a, b]$ بوده و $f \in C^{n+1}[a, b]$ ، آنگاه، به ازای هر x در $[a, b]$ ، نقطه‌ای مانند $\xi(x)$ در (a, b) وجود دارد با این خاصیت که

$$(۶.۳) \quad f(x) = P(x) + \frac{f^{(n+1)}(\xi(x))}{(n+1)!} (x - x_0)(x - x_1) \cdots (x - x_n),$$

که در آن P چندجمله‌ای درونیاب است که با معادله (۳.۳) داده می‌شود.

برهان. ابتدا توجه می‌کنیم که، هرگاه به‌ازای $x = x_k$ ، $k = 0, 1, \dots, n$ ، آنگاه $f(x_k) = P(x_k)$ ، و $\xi(x_k)$ را نقطه دلخواهی در (a, b) بگیریم، معادله (۶.۳) نتیجه خواهد شد.

به‌ازای $x \neq x_k$ ($k = 0, 1, \dots, n$)، تابع g را به‌ازای هر t در $[a, b]$ با

$$g(t) = f(t) - P(t) - [f(x) - P(x)] \frac{(t - x_0)(t - x_1) \cdots (t - x_n)}{(x - x_0)(x - x_1) \cdots (x - x_n)}$$

$$= f(t) - P(t) - [f(x) - P(x)] \prod_{i=0}^n \frac{(t - x_i)}{(x - x_i)}$$

تعریف می‌کنیم. چون $f \in C^{n+1}[a, b]$ ، $P \in C^\infty[a, b]$ ، و به‌ازای هر k ، $x \neq x_k$ ، نتیجه می‌شود که $g \in C^{n+1}[a, b]$ ، و به‌ازای $t = x_k$

$$g(x_k) = f(x_k) - P(x_k) - [f(x) - P(x)] \prod_{i=0}^n \frac{(x_k - x_i)}{(x - x_i)}$$

$$= 0 - [f(x) - P(x)] \cdot 0$$

$$= 0.$$

بعلاوه،

$$g(x) = f(x) - P(x) - [f(x) - P(x)] \prod_{i=0}^n \frac{(x - x_i)}{(x - x_i)}$$

$$= f(x) - P(x) - [f(x) - P(x)]$$

$$= 0.$$

بنابراین، $g \in C^{n+1}[a, b]$ و g در $n + 2$ نقطه متمایز x, x_0, x_1, \dots, x_n صفر می‌شود. طبق تعمیم قضیه رل (قضیه ۱۱.۱ ص ۸)، نقطه‌ای مانند $\xi \equiv \xi(x)$ در (a, b) وجود دارد که به‌ازای آن $g^{(n+1)}(\xi) = 0$. با محاسبه $g^{(n+1)}$ در نقطه ξ ، نتیجه می‌شود که

$$0 = g^{(n+1)}(\xi) = f^{(n+1)}(\xi) - P^{(n+1)}(\xi)$$

$$(۷.۳) \quad - [f(x) - P(x)] \frac{d^{n+1}}{dt^{n+1}} \left(\prod_{i=0}^n \frac{(t - x_i)}{(x - x_i)} \right) \Big|_{t=\xi}$$

چون P یک چندجمله‌ای از درجه حداکثر n است، مشتق $(n + 1)$ م، یعنی $P^{(n+1)}$ ، باید متحده صفر باشد. همچنین، $\prod_{i=0}^n [(t - x_i)/(x - x_i)]$ یک چندجمله‌ای از درجه $(n + 1)$ است، در نتیجه،

$$\prod_{i=0}^n \frac{(t - x_i)}{(x - x_i)} = \left(\frac{1}{\prod_{i=0}^n (x - x_i)} \right) t^{n+1} + (\text{جملاتی با درجه کمتر از } t)$$

$$\frac{d^{n+1}}{dt^{n+1}} \prod_{i=0}^n (t - x_i) = \frac{(n+1)!}{\prod_{i=0}^n (x - x_i)}$$

حال معادله (۷.۳) به صورت زیر درمی‌آید .

$$0 = f^{(n+1)}(\xi) - 0 - [f(x) - P(x)] \frac{(n+1)!}{\prod_{i=0}^n (x - x_i)}$$

و، با حل آن نسبت به $f(x)$ ، نتیجه می‌شود

$$f(x) = P(x) + \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^n (x - x_i)$$

فرمول خطای (۶.۳) از جنبه نظری مهم است، اما استفاده عملی آن به توابعی محدود می‌شود که مشتقات آنها کرانهای معلومی دارند. این اغلب حالت توابع جدولی نظیر توابع مثلثاتی یا لگاریتمی است.

مثال ۲. فرض کنیم قرار است برای تابع $f(x) = e^x, 0 \leq x \leq 1$ جدولی تهیه شود. همچنین، تعداد ارقام اعشاری هردراییه d باشد؛ مثلاً، اگر $d = 5$ ، $f(1) = 2.71828$ ، و تفاضل بین مقادیر مجاور x ، یعنی طول گام h ، باشد.

(آ) با فرض $d \geq 6$ ، h چقدر باید باشد تا درونیابی خطی (یعنی، چندجمله‌ای لاگرانژ از درجه ۱) خطایی با قدر مطلق حداکثر 10^{-6} بدست دهد؟

(ب) اگر $d < 6$ ، h چقدر باید باشد تا درونیابی خطی خطایی با قدر مطلق حداکثر 10^{-6} بدست بدهد؟

فرض کنیم $x \in [0, 1]$ و z چنان باشد که $x_j \leq x \leq x_{j+1}$. بنابراین معادله (۶.۳) خطا در درونیابی خطی عبارت است از

$$f(x) - P(x) = \frac{f^{(2)}(\xi)}{2!} (x - x_j)(x - x_{j+1})$$

چون طول گام h است، نتحه می‌شود که $x_j = jh$ ، $x_{j+1} = (j+1)h$ ،

$$|f(x) - P(x)| \leq \frac{|f^{(2)}(\xi)|}{2!} |(x - jh)(x - (j + 1)h)|.$$

بنابراین ،

$$\begin{aligned} |f(x) - P(x)| &= \frac{1}{2} \max_{\xi \in [0, 1]} |f^{(2)}(\xi)| \max_{x_j \leq x \leq x_{j+1}} |(x - jh)(x - (j + 1)h)| \\ &= \frac{1}{2} \max_{\xi \in [0, 1]} e^{\xi} \max_{x_j \leq x \leq x_{j+1}} |(x - jh)(x - (j + 1)h)| \\ &\leq \frac{1}{2} e \max_{x_j \leq x \leq x_{j+1}} |(x - jh)(x - (j + 1)h)|. \end{aligned}$$

با توجه به $g(x) = (x - jh)(x - (j + 1)h)$ به بازای $jh \leq x \leq (j + 1)h$ و استفاده از روشهای حساب دیفرانسیل و انتگرال (ر.ک. تمرین ۱۲) ، می توان نشان داد که

$$(۸.۳) \quad \max_{x_j \leq x \leq x_{j+1}} |g(x)| = |g((j + \frac{1}{2})h)| = \frac{h^2}{4}.$$

در نتیجه ، خطا در درونیابی خطی به وسیله

$$|f(x) - P(x)| \leq \frac{eh^2}{8}$$

کراندار است .

برای پاسخ دادن به قسمت (آ) ، کافی است h طوری اختیار شود که

$$\frac{eh^2}{8} \leq 10^{-6},$$

$$h^2 \leq \frac{8}{e} \cdot 10^{-6},$$

$$h^2 < 2.944 \times 10^{-6},$$

یا

$$h < 1.72 \times 10^{-3}.$$

$h = .001$ یک انتخاب منطقی برای طول گام می باشد .

دلیل فرق گذاشتن بین قسمتهای (آ) و (ب) این است که می خواهیم تاکید کنیم که ، اگر جدولها فقط تا پنج رقم اعشار دقیق باشند ، نمی توان از طریق درونیابی مقادیری تا شش رقم دقیق بدست آورد . خطاهایی که تا پنج رقم اعشار گرد شده اند باقی خواهند

ماند، در نتیجه، برای قسمت (ب) جوابی وجود نخواهد داشت.

آخرین مثال این بخش روشهای درونیابی را در وضعی که بخش خطای معادله (۶.۳) قابل استفاده نیست توضیح می‌دهد. این مثال همچنین می‌گوید که برای رسیدن به تقریب از طریق درونیابی باید به جستجوی راه موثری بپردازیم.

مثال ۳. در جدول ۱.۳ مقادیر یک تابع (تابع بسل^۱ نوع اول و از مرتبه صفر) در نقاط مختلف درج شده‌اند. تقریبهای $f(1.5)$ که با چندجمله‌ایهای لژرانژ متفاوت بدست آمده‌اند را با هم مقایسه می‌کنیم.

x	$f(x)$
1.0	.7651977
1.3	.6200860
1.6	.4554022
1.9	.2818186
2.2	.1103623

جدول ۱.۳

با استفاده از $x_0 = 1.3$ و $x_1 = 1.6$ ، مقدار چندجمله‌ای درونیاب در ۱.۵ عبارت است از

$$P_1(1.5) = \frac{(1.5 - 1.6)}{(1.3 - 1.6)} (.6200860) + \frac{(1.5 - 1.3)}{(1.6 - 1.3)} (.4554022) = .5102968.$$

می‌توان از دو چندجمله‌ای درجه دو استفاده کرد، یکی با فرض $x_0 = 1.3$ ، $x_1 = 1.6$ ، و $x_2 = 1.9$ ، که نتیجه می‌دهد

$$\begin{aligned} P_2(1.5) &= \frac{(1.5 - 1.6)(1.5 - 1.9)}{(1.3 - 1.6)(1.3 - 1.9)} (.6200860) + \frac{(1.5 - 1.3)(1.5 - 1.9)}{(1.6 - 1.3)(1.6 - 1.9)} (.4554022) \\ &\quad + \frac{(1.5 - 1.3)(1.5 - 1.6)}{(1.9 - 1.3)(1.9 - 1.6)} (.2818186) \\ &= .5112857 \end{aligned}$$

و دیگری با فرض $x_0 = 1.0$ ، $x_1 = 1.3$ ، و $x_2 = 1.6$ ، که در این حالت

$$\hat{P}_2(1.5) = .5124715.$$

در حالت درجه سه نیز دو چندجمله‌ای انتخاب می‌شوند. یکی به ازای $x_0 = 1.3$ ،

$x_1 = 1.6$ ، و $x_2 = 1.9$ ، و $x_3 = 2.2$ ، که نتیجه می‌دهد

$$P_3(1.5) = .5118302.$$

دیگری با فرض $x_0 = 1.0$ ، $x_1 = 1.3$ ، $x_2 = 1.6$ ، و $x_3 = 1.9$ ، که نتیجه می‌دهد

$$\hat{P}_3(1.5) = .5118127.$$

چندجمله‌ای لاگرانژ درجه چهار از همه درایه‌های جدول استفاده می‌کند و انتظار

می‌رود که دقیقترین باشد. به ازای $x_0 = 1.0$ ، $x_1 = 1.3$ ، $x_2 = 1.6$ ، $x_3 = 1.9$ ، و

$x_4 = 2.2$ ، می‌توان نشان داد که

$$P_4(1.5) = .5118200.$$

چون $P_3(1.5)$ ، $\hat{P}_3(1.5)$ ، و $P_4(1.5)$ همه با تقریب 2×10^{-5} واحد یکسانند، انتظار

می‌رود که $P_4(1.5)$ دقیقترین تقریب بوده و احتمالاً تا 2×10^{-5} واحد دقیق باشد.

مقدار واقعی $f(1.5)$ مساوی 5118277 است؛ در نتیجه، دقت تقریبها به صورت

زیر است:

$$|P_1(1.5) - f(1.5)| \approx 1.53 \times 10^{-3},$$

$$|P_2(1.5) - f(1.5)| \approx 5.42 \times 10^{-4},$$

$$|\hat{P}_2(1.5) - f(1.5)| \approx 6.44 \times 10^{-4},$$

$$|P_3(1.5) - f(1.5)| \approx 2.5 \times 10^{-6},$$

$$|\hat{P}_3(1.5) - f(1.5)| \approx 1.50 \times 10^{-5},$$

$$|P_4(1.5) - f(1.5)| \approx 7.7 \times 10^{-6}.$$

توجه کنید که، در این حالت، P_3 دقیقترین تقریب است؛ بهر حال، با هیچ اطلاعی

از مقدار واقعی $f(1.5)$ ، P_4 به عنوان بهترین تقریب پذیرفته می‌شود. باید توجه داشت

که خطا یا جمله باقیمانده که در قضیه ۳.۳ بدست آمد را نمی‌توان اینجا اعمال کرد،

زیرا هیچ اطلاعی از مشتق چهارم f در دست نیست. متأسفانه، این حالتی است که عموماً

پیش می‌آید.

مجموعه تمرینات ۲.۳

۱. به دو طریق متفاوت نشان دهید که $f(x) = 3(x-1)(x-2)(x+1)$ و $g(x) = 3x^3 - 6x^2 - 3x + 6$ چندجمله‌ای یکسانی را نمایش می‌دهند.
۲. با استفاده از چندجمله‌ایهای درونیاب لاگرانژ درجه ۶، دو، سه، و چهار، $f(2.5)$ را تقریب کنید در صورتی که

$$\begin{aligned} f(2.0) &= .5103757 & f(2.6) &= .4813306 \\ f(2.2) &= .5207843 & f(2.8) &= .4359160 \\ f(2.4) &= .5104147 \end{aligned}$$

۳. چندجمله‌ای درجه ۳ یا کمتر که با $f(x) = x^2$ در $x_0 = 1$ ، $x_1 = 3$ ، $x_2 = 6$ ، و $x_3 = 7$ یکی است را محاسبه کنید.
۴. با استفاده از چندجمله‌ای درونیاب لاگرانژ از درجه ۳ یا کمتر و مقادیر زیر، $\cos .750$ را تقریب کنید. با استفاده از معادله (۶.۳)، کرانی برای خطا پیدا نمایید.

$$\begin{aligned} \cos .698 &= .7661 & \cos .768 &= .7193 \\ \cos .733 &= .7432 & \cos .803 &= .6946 \end{aligned}$$

- مقدار واقعی $\cos .750$ (تا چهار رقم اعشار) مساوی $.7317$ است. اگر بین خطای واقعی و کران خطای شما اختلاف بود، علت آن را توضیح دهید.
۵. فرض کنید $f(x) = 3xe^x - 2e^x$. با استفاده از چندجمله‌ای درونیاب از درجه ۳ یا بیشتر از دو و با استفاده از $x_0 = 1$ ، $x_1 = 1.05$ ، $x_2 = 1.07$ ، و $f(1.03)$ را تقریب کنید. خطای واقعی را با کران خطای حاصل از معادله (۶.۳) مقایسه نمایید.
۶. فرض کنید $f(x) = (4x - 7)/(x - 2)$ و $x_0 = 1.7$ ، $x_1 = 1.8$ ، $x_2 = 1.9$ ، و $x_3 = 2.1$.
 (آ) با استفاده از چندجمله‌ای درونیاب از درجه ۳ حداکثر دو برگره‌های x_0 ، x_1 ، و x_2 را تقریب نمایید.
 (ب) با استفاده از چندجمله‌ای درونیاب بر x_0 ، x_1 ، x_2 ، و x_3 ، $f(1.75)$ را تقریب نمایید.
- (پ) آیا کران خطا، از معادله (۶.۳)، را می‌توان در (آ) یا (ب) اعمال کرد؟ این کران، خطا را چقدر تخمین می‌زند؟
۷. فرض کنید $f(x) = e^x$ ، $0 \leq x \leq 2$. با استفاده از مقادیر جدول ذیل، این محاسبات

را انجام دهید:

(آ) $f(2.5)$ را با استفاده از درونیایی خطی به ازای $x_0 = 0$ و $x_1 = .5$ تقریب کنید؛

(ب) $f(7.5)$ را با استفاده از درونیایی خطی به ازای $x_0 = .5$ و $x_1 = 1$ تقریب

کنید؛

(پ) با استفاده از چندجمله‌ای درونیاب درجه دو به ازای $x_0 = 0$ ، $x_1 = 1$ ، و

$x_2 = 2$ ، $f(2.5)$ و $f(7.5)$ را تقریب کنید.

(ت) چه تقریبهایی بهترند؟ چرا؟

x	0	.5	1.0	2.0
$f(x)$	1.00000	1.64872	2.71828	7.38906

۸. فرض کنید بخواهیم برای تابع لگاریتمی معمولی یا در پایه ۱۰ از $x = 1$ تا $x = 10$

یک جدول شش رقمی بسازیم بطوری که درونیایی خطی تا شش رقم اعشار دقیق باشد.

بزرگترین طول گام ممکن برای این جدول را مشخص کنید.

۹. نشان دهید که

$$\sum_{k=0}^n L_k(x) = 1, \quad x \text{ هر بازای}$$

۱۰. فرض کنید $\omega(x) = \prod_{k=0}^n (x - x_k)$ نشان دهید که چندجمله‌ای درونیاب از درجه n

بر نقاط x_0, \dots, x_n را می‌توان به صورت زیر نوشت:

$$P(x) = \omega(x) \sum_{k=0}^n \frac{f(x_k)}{(x - x_k)\omega'(x_k)}$$

۱۱. قضیه ۱۳.۱ در صفحه ۹ را ثابت کنید. راهنمایی. فرض کنید

$$g(t) = f(t) - P(t) - [f(x) - P(x)] \cdot \frac{(t - x_0)^n}{(x - x_0)^n}$$

که در آن P چندجمله‌ای تیلور درجه n است، و قضیه ۱۱.۱ در صفحه ۸ را

بکار ببرید.

۱۲. نشان دهید

$$\max_{x_j \leq x < x_{j+1}} |g(x)| = \frac{h^2}{4}$$

کرد.

۱۳. در مقدمه^۶ این فصل، جدول زیر داده شده بود، که در آن جمعیت ایالات متحده از ۱۹۲۰ تا ۱۹۷۰ درج شده است.

سال	1920	1930	1940	1950	1960	1970
جمعیت (به هزار)	105,711	123,203	131,669	150,697	179,323	203,212

چندجمله‌ای لاگرانژ درجه 5 که مناسب این داده‌هاست را بیابید و، با استفاده از این چندجمله‌ای، جمعیت را در سالهای ۱۹۱۰، ۱۹۶۵، و ۲۰۰۰ تخمین بزنید. جمعیت در ۱۹۱۰ تقریباً 91,972,000 بود. فکر می‌کنید ارقام شما در ۱۹۶۵ و ۲۰۰۰ تا چه حد دقیق‌اند؟

۳.۳ درونیابی تکراری

یکی از مشکلاتی که استفاده از روش بخش ۲.۳ دارد این است که، چون کار با جمله خطا در قضیه ۳.۳ مشکل است، درجه چندجمله‌ای لازم برای دقت مطلوب عموماً "قبل از پایان کار معلوم نیست. معمولاً" نتایج از چندجمله‌ایهای مختلف محاسبه می‌شوند تا توافق مناسب بدست آید. این کار در مثال ۳ انجام شد. با بررسی این مثال، دیده می‌شود که کار لازم برای محاسبه تقریب به وسیله چندجمله‌ای درجه دو کار لازم برای محاسبه تقریب درجه سه را کم نمی‌کند، همچنین، تقریب درجه چهار با معلوم بودن تقریب درجه سه آسانتر بدست نمی‌آید. هدف این بخش یافتن این چندجمله‌ایهای تقریب ساز است بنحوی که از محاسبات قبلی حداکثر استفاده برده شود.

تعریف ۴.۳. فرض کنیم f تابعی باشد که در نقاط $x_0, x_1, x_2, \dots, x_n$ تعریف شده است، و m_1, m_2, \dots, m_k عدد صحیح متمایز باشد که، به ازای هر i ، $0 \leq m_i \leq n$. چندجمله‌ای لاگرانژ از درجه نایبتر از k که با f در $x_{m_1}, x_{m_2}, \dots, x_{m_k}$ یکی است را با P_{m_1, m_2, \dots, m_k} نشان می‌دهیم

$P_{1,2,4}$ چند جمله‌ای است که با f در $x_1 = 2$ ، $x_2 = 3$ ، و $x_4 = 6$ یکی است؛ یعنی،

$$P_{1,2,4}(x) = \frac{(x-3)(x-6)}{(2-3)(2-6)}(8) + \frac{(x-2)(x-6)}{(3-2)(3-6)}(27) + \frac{(x-2)(x-3)}{(6-2)(6-3)}(216).$$

قضیه ۵.۳. فرض کنیم f در نقاط x_0, x_1, \dots, x_k تعریف شده باشد، و x_j, x_i دو نقطه متمایز از این نقاط باشند. هرگاه

$$(9.3) \quad P(x) = \frac{(x-x_j)P_{0,1,\dots,j-1,j+1,\dots,k}(x) - (x-x_i)P_{0,1,\dots,i-1,i+1,\dots,k}(x)}{(x_i-x_j)}$$

آنگاه P چند جمله‌ای لاگرانژ از درجه نایبتر از k در x_0, x_1, \dots, x_k است.

برهان. برای ساده بودن نمادها، فرض می‌کنیم $Q \equiv P_{0,1,\dots,i-1,i+1,\dots,k}$ و $\hat{Q} \equiv P_{0,1,\dots,j-1,j+1,\dots,k}$. \hat{Q} چند جمله‌ایهایی از درجه $k-1$ یا کمترند؛ در نتیجه، باید از درجه نایبتر از k باشد. هرگاه $0 \leq r \leq k$ و $r \neq i, j$ ، آنگاه

$$P(x_r) = \frac{(x_r-x_j)\hat{Q}(x_r) - (x_r-x_i)Q(x_r)}{x_i-x_j} = \frac{(x_i-x_j)}{(x_i-x_j)} f(x_r) = f(x_r).$$

بعلاوه،

$$P(x_i) = \frac{(x_i-x_j)\hat{Q}(x_i) - (x_i-x_i)Q(x_i)}{x_i-x_j} = \frac{(x_i-x_j)}{(x_i-x_j)} f(x_i) = f(x_i),$$

و، بهمین نحو، $P(x_j) = f(x_j)$. اما، طبق تعریف، $P_{0,1,\dots,k}$ چند جمله‌ای منحصر بفردی است از درجه نایبتر از k که با f در x_0, x_1, \dots, x_k یکی است. بنابراین، $P = P_{0,1,\dots,k}$.

مثال ۲. در مثال ۲ از بخش ۲.۳، با استفاده از جدول ۲.۳، مقادیر چند جمله‌ایهای لاگرانژ مختلف در $x = 1.5$ بدست آمد.

هدف این مثال محاسبه $f(1.5)$ با استفاده از قضیه ۵.۳ است. اگر $x_0 = 1.0, x_1 = 1.3, x_2 = 1.6, x_3 = 1.9, x_4 = 2.2$ ایجاب می‌کند که $f(1.0) = P_0, f(1.3) = P_1, f(1.6) = P_2, f(1.9) = P_3, f(2.2) = P_4$ و

x	$f(x)$
1.0	.7651977
1.3	.6200860
1.6	.4554022
1.9	.2818186
2.2	.1103623

جدول ۲۰۳

است؛ در نتیجه، اینها پنج چندجمله‌ای از درجه صفر (ثابتها) هستند که $f(1.5)$ را تقریب می‌کنند.

با محاسبه $P_{0,1}(1.5)$ ، نتیجه می‌شود که

$$\begin{aligned} P_{0,1}(1.5) &= \frac{(1.5 - 1.0)P_1 - (1.5 - 1.3)P_0}{(1.3 - 1.0)} \\ &= \frac{.5(.6200860) - .2(.7651977)}{.3} \\ &= .5233449. \end{aligned}$$

به همین نحو،

$$\begin{aligned} P_{1,2}(1.5) &= \frac{(1.5 - 1.3)(.4554022) - (1.5 - 1.6)(.6200860)}{(1.6 - 1.3)} \\ &= .5102968, \end{aligned}$$

$$P_{2,3}(1.5) = .5132634.$$

و

$$P_{3,4}(1.5) = .5104270.$$

اینهاتقریبات را با استفاده از چندجمله‌ایهای درجه اول بدست می‌دهند. انتظار می‌رود که $P_{1,2}$ بهترین تقریب باشد، زیرا 1.5 بین $x_1 = 1.3$ و $x_2 = 1.6$ قرار دارد. تقریبات با استفاده از چندجمله‌ایهای درجه دو عبارتند از

$$P_{0,1,2}(1.5) = \frac{(1.5 - 1.0)(.5102968) - (1.5 - 1.6)(.5233449)}{(1.6 - 1.0)}$$

$$= .5124715,$$

$$P_{1,2,3}(1.5) = \frac{(1.5 - 1.3)(.5132634) - (1.5 - 1.9)(.5102968)}{(1.9 - 1.3)}$$

$$= .5112857,$$

$$P_{2,3,4}(1.5) = .5137361.$$

تقریبهای درجه سه عبارتند از

$$P_{0,1,2,3}(1.5) = \frac{(1.5 - 1.0)(.5112857) - (1.5 - 1.9)(.5124715)}{(1.9 - 1.0)}$$

$$= .5118127$$

$$P_{1,2,3,4}(1.5) = .5118302.$$

بالاخره، تقریب درجه چهار عبارت است از

$$P_{0,1,2,3,4}(1.5) = \frac{(1.5 - 1.0)(.5118302) - (1.5 - 2.2)(.5118127)}{(2.2 - 1.0)}$$

$$= .5118200.$$

مثل قبل، انتظار می رود که این تا 2×10^{-5} دقیق باشد.

با محاسبه نتایج فوق، می توان جدولی (جدول ۳.۳) به صورت زیر ساخت:

x_0	P_0				
x_1	P_1	$P_{0,1}$			
x_2	P_2	$P_{1,2}$	$P_{0,1,2}$		
x_3	P_3	$P_{2,3}$	$P_{1,2,3}$	$P_{0,1,2,3}$	
x_4	P_4	$P_{3,4}$	$P_{2,3,4}$	$P_{1,2,3,4}$	$P_{0,1,2,3,4}$

جدول ۳.۳

که، در این مثال، به صورت جدول ۴.۳ ظاهر می شود.

1.0	.7651977				
1.3	.6200860	.5233449			
1.6	.4554022	.5102968	.5124715		
1.9	.2818186	.5132634	.5112857	.5118127	
2.2	.1103623	.5104270	.5137361	.5118302	.5118200

جدول ۴.۳

فرض کنیم در این مرحله قطعی شده است که آخرین تقریب، یعنی $P_{0.1.2.3.4}$ ، دقت مطلوب را ندارد. می‌توان گره دیگر x_5 را اختیار، و سطر دیگری به جدول افزود:

$$x_5, P_5, P_{4.5}, P_{3.4.5}, P_{2.3.4.5}, P_{1.2.3.4.5}, P_{0.1.2.3.4.5},$$

و می‌توان $P_{0.1.2.3.4.5}$ ، $P_{1.2.3.4.5}$ ، $P_{0.1.2.3.4}$ را برای دقت بیشتر با هم مقایسه کرد. به آسانی از فرمول قضیه ۵.۳ دیده می‌شود که فقط سطر پنجم برای محاسبه سطر ششم لازم است. بطور کلی، فقط سطر قبلی برای افزودن سطر جدید به جدول لازم است؛ مثلاً،

$$P_{4.5}(x) = \frac{(x - x_4)P_5 - (x - x_5)P_4}{x_5 - x_4},$$

که فقط به x ، x_4 ، x_5 ، P_4 ، P_5 بستگی دارد.

در مثالی که در نظر گرفته‌ایم، مقدار تابع بسل نوع اول از مرتبه صفر در 2.5 مساوی 0.0483838 - است. با استفاده از این مقدار در ساختن سطر جدید، نتیجه می‌شود که

$$2.5, -0.0483838, .4807699, .5301984, .5119070, .5118430, .5118277.$$

درایه جدید تا شش رقم اعشار دقیق می‌باشد.

روند فوق روش نویل^۱ نامیده می‌شود. در نحوه ارائه این مطلب از نمادهای پر دردسراستفاده شده است، و اکنون شکل مفیدتر روش نویل را بررسی می‌کنیم. برای ایجاد یک آرایه، عملاً فقط دوزیرنویس لازم می‌شود. توجه کنید که از بالا به پایین جدول

آمدن متنظر است با استفاده از نقاط متوالی x_i با i بزرگتر، و رفتن به راست متنظر است با افزایش درجه چندجمله‌ای درونیاب.

فرض کنیم $i \geq j$ ، $Q_{i,j}$ چندجمله‌ای درونیاب از درجه j بر $(j+1)$ نقطه $x_{i-j}, x_{i-j+1}, \dots, x_{i-1}, x_i$ برای محاسبه

$$Q_{i,j} = P_{i-j, i-j+1, \dots, i-1, i}$$

به روش نویل،

$Q_{i-1, j-1} = P_{i-j, i-j+1, \dots, i-1}$ و $Q_{i, j-1} = P_{i-j+1, \dots, i-1, i}$
 را در (۹.۳) بکار می‌بریم تا به‌ازای هر $j = 1, 2, 3, \dots$ و $i = j, j+1, \dots$

$$Q_{i,j}(x) = \frac{(x - x_i)Q_{i-1, j-1}(x) - (x - x_{i-j})Q_{i, j-1}(x)}{x_{i-j} - x_i}$$

بعلاوه، فرض کنیم به‌ازای هر i ، $Q_{i,0} = f(x_i)$ ، در این صورت، آرایه جدول ۵.۳ ساخته می‌شود.

x_0	$Q_{0,0}$				
x_1	$Q_{1,0}$	$Q_{1,1}$			
x_2	$Q_{2,0}$	$Q_{2,1}$	$Q_{2,2}$		
x_3	$Q_{3,0}$	$Q_{3,1}$	$Q_{3,2}$	$Q_{3,3}$	
x_4	$Q_{4,0}$	$Q_{4,1}$	$Q_{4,2}$	$Q_{4,3}$	$Q_{4,4}$

جدول ۵.۳

این جدول و جدول ۳.۳ مربوط به P ها، صرف نظر از نمادها، یکی‌اند، اما جدول ۵.۳ با کامپیوتر خیلی آسانتر ساخته می‌شود.

الگوریتم زیر این جدول را سطر به سطر می‌سازد.

الگوریتم درونیابی تکراری نویل ۱.۳

برای محاسبه چندجمله‌ای درونیاب P بر $(n+1)$ نقطه متمایز x_0, \dots, x_n در نقطه x به‌ازای تابع f :

مرحله ۱. قرار می‌دهیم $Q_{0,0} = f(x_0)$ و $i = 1$.

مرحله ۲. قرار می‌دهیم $Q_{i,0} = f(x_i)$.

مرحله ۳. به ازای هر $i = 1, 2, \dots, j$ ، قرار می‌دهیم

$$Q_{i,j} = \frac{(x - x_i)Q_{i-1,j-1} - (x - x_{i-j})Q_{i,j-1}}{x_{i-j} - x_i}$$

مرحله ۴. به i یک می‌افزاییم.

مرحله ۵. اگر $i \leq n$ ، به مرحله ۲ می‌رویم.

مرحله ۶. $P(x) = Q_{n,n}$ ، و روند تمام خواهد بود.

این الگوریتم را می‌توان جهت اضافه کردن گره‌های درونیاب جدید اندکی پیراست.

مثلاً، مرحله ۴ را می‌توان با نامساوی

$$|Q_{i,i} - Q_{i-1,i-1}| < \epsilon,$$

عوض کرد، که در آن ϵ یک خطای قابل تحمل و از پیش معلوم است. اگر این نامساوی برقرار باشد، $Q_{i,i}$ یک تقریب معقول برای $f(x)$ است. اگر نامساوی برقرار نباشد، قبل از بازگشت به مرحله ۲، i یکی زیاد شده و یک نقطه درونیاب جدید مانند x_i افزوده می‌شود.

مجموعه تمرینات ۳-۳

۱. با استفاده از قضیه ۵.۳، تقریبهای مربوط به تمرین ۲ از بخش ۲.۳ را بدست آورید.

۲. با اعمال روش نویل برتابع $f(x) = 3x^3$ به ازای مقادیر $x_0 = -2$ ، $x_1 = -1$ ، $x_2 = 0$ ، $x_3 = 1$ و $x_4 = 2$ را تقریب کنید.

۳. با اعمال روش نویل برتابع $f(x) = x^2 e^x \cos x$ و استفاده از $x_0 = -1.0$ ، $x_1 = -0.9$ ، $x_2 = -0.8$ ، $x_3 = -0.7$ و $x_4 = -0.6$ ، $f(-0.78)$ را تقریب کنید.

۴. (A) با اعمال روش نویل برتابع $f(x) = 3xe^x - e^{2x}$ و استفاده از $x_0 = 1$ ، $x_1 = 1.05$ و $x_2 = 1.07$ ، $f(1.03)$ را با $P_{0,1,2}$ تقریب کنید.

(ب) فرض کنید تقریب (A) به قدر کافی دقیق نباشد بطوری که $P_{0,1,2,3}$ ، که $x_3 = 1.04$ ، باید محاسبه شود.

(پ) به افزودن گره‌های x_4, \dots, x_n, x_{n+1} ادامه دهید تا

$$|P_{0,1,\dots,n+1} - P_{0,1,\dots,n}| < 10^{-5}$$

۵. در روش ایتکن، که شبیه روش نویل است، جدول زیر از مقادیر درونیاب ساخته می‌شود:

x_0	P_0				
x_1	P_1	$P_{0,1}$			
x_2	P_2	$P_{0,2}$	$P_{0,1,2}$		
x_3	P_3	$P_{0,3}$	$P_{0,1,3}$	$P_{0,1,2,3}$	
x_4	P_4	$P_{0,4}$	$P_{0,1,4}$	$P_{0,1,2,4}$	$P_{0,1,2,3,4}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

برای محاسبه هر مقدار جدید، از مقدار بالای ستون قبل و مقدار واقع در همان سطر منتهای ستون قبل، استفاده می‌کنیم، مثلاً،

$$P_{0,1,3}(x) = \frac{(x - x_3)P_{0,1}(x) - (x - x_1)P_{0,3}(x)}{x_1 - x_3}$$

با استفاده از نماد $Q_{i,j}$ در روش نویل، الگوریتم را برای محاسبه $Q_{i,i}$ به روش ایتکن بسازید.

۶. با اعمال روش ایتکن بر تابع $f(x) = 3^x$ به ازای گره‌های

$x_0 = -2, x_1 = -1, x_2 = 0, x_3 = 1, x_4 = 2$ را تقریب کنید. نتیجه خود

را با تمرین ۲ مقایسه کنید. کدام روش را ترجیح می‌دهید؟

۷. تمرین ۴ را با استفاده از حساب چهار رقمی تکرار کنید. آیا به نظر شما روش نویل

نسبت به خطاهای گرد کردن حساس است؟

۸. برای $f(1 + \sqrt{10})$ ، که $f(x) = (1 + x^2)^{-1}$ به ازای $-5 \leq x \leq 5$ ، دنباله y_n از

مقادیر درونیاب را به صورت زیر بسازید: به ازای هر $n = 1, 2, \dots, 10$ فرض کنید

$h = 10/n$ و $y_n = P_n(1 + \sqrt{10})$ ، که در آن $P_n(x)$ چندجمله‌ای درونیاب برای $f(x)$ در

گره‌های $x_0^{(n)}, x_1^{(n)}, \dots, x_n^{(n)}$ است و به ازای هر $j = 0, 1, 2, \dots, n$ ، $x_j^{(n)} = -5 + jh$. آ پایه

نظر می‌رسد که دنباله $\{y_n\}$ به $f(1 + \sqrt{10})$ همگرا باشد؟

درونیابی معکوس. فرض کنید $f \in C^1[a, b]$ ، $f'(x) \neq 0$ بر $[a, b]$ ، و f در $[a, b]$

دارای یک صفر مانند p باشد. همچنین x_0, \dots, x_n $n + 1$ نقطه متمایز در $[a, b]$

باشد که به ازای هر $k = 0, 1, \dots, n$ ، $f(x_k) = y_k$. برای تقریب p ، چندجمله‌ای

درونیاب درجه n را بر گره‌های y_0, \dots, y_n به ازای f^{-1} می‌سازیم. چون $y_k = f(x_k)$ و

$0 = f(p)$ ، نتیجه می‌شود که $f^{-1}(y_k) = x_k$ و $f^{-1}(0) = p$. استفاده از درونیابی

تکراری برای تقریب $f^{-1}(0)$ درونیابی معکوس تکراری نامیده می‌شود.
 ۹. با استفاده از درونیابی معکوس تکراری و داده‌های زیر:

x	.3	.4	.5	.6
e^{-x}	.740818	.670320	.606531	.548812

یک جواب تقریبی $0 = e^{-x} - x$ را بدست آورید.
 ۱۰. الگوریتمی بسازید که بتوان آن را برای درونیابی معکوس بکار برد.

۴.۳ تفاضل‌های منقسم

روشهای درونیابی تکراری که در بخش قبل مطرح شدند برای تعیین مقادیر چندجمله‌ایهای درونیاب درجات بالاتر در یک نقطه خاص مفیدند. لیکن، هر درایه در جدول درونیابی به نقطه مورد محاسبه بستگی دارد، و در نتیجه، نمی‌توان از جدول برای ارائه نمایش صریح چندجمله‌ای درونیاب استفاده کرد.

روشهای تعیین نمایش صریح یک چندجمله‌ای درونیاب از داده‌های یک جدول به روشهای تفاضل منقسم معروفند. قبل از کامپیوترهای رقمی، از این روشها برای مقاصد محاسبه‌ای خیلی استفاده می‌شد، و از آن برای بدست آوردن روشهایی جهت تقریب مشتقها و انتگرالهای توابع و همچنین تقریب جوابهای معادلات دیفرانسیل نیز استفاده می‌شود. بحث ما از روشهای تفاضل منقسم کمی سطحی است، زیرا نتایج این بخش در بحثهای آتی مورد نیاز نیستند. چون در اغلب کتب درسی آنالیز عددی قبل از ۱۹۶۰، روشهای تفاضل منقسم به تفصیل مطرح شده‌اند، در صورت نیاز به بحث جامعتر، شما را به این کتابها ارجاع می‌دهیم. بالاخص، کتاب کلاسیک هیلدبراند^۱ [۳۶] مرجع مناسبی می‌باشد. فرض کنید P_n چندجمله‌ایی از درجه حداکثر n باشد که با تابع f در نقاط متمایز x_0, x_1, \dots, x_n یکی است. تفاضل‌های منقسم f ، نسبت به این نقاط، را می‌توان با نشان دادن اینکه P_n دارای نمایش

$$(10.3) \quad P_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n(x - x_0)(x - x_1)\dots(x - x_{n-1})$$

بمازای ثابتهای مناسب a_0, a_1, \dots, a_n است بدست آورد .

برای تعیین اولین ثابت a_0 ، توجه می‌کنیم که اگر بتوان $P_n(x)$ را به شکل معادله (۱۰.۳) نوشت، در این صورت، محاسبه P_n در x_0 فقط جمله ثابت a_0 را بجای می‌گذارد؛ یعنی ، $a_0 = P_n(x_0) = f(x_0)$.

بهین نحو، وقتی P_n در x_1 محاسبه شود، تنها جملات ناصفر در $P_n(x_1)$ جملات ثابت و خطی‌اند، یعنی

$$f(x_0) + a_1(x_1 - x_0) = P_n(x_1) = f(x_1),$$

بطوری که

$$(11.3) \quad a_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

در این مرحله نماد تفاضل منقسم را معرفی می‌کنیم . تفاضل منقسم صفرم تابع f ، نسبت به x_i ، با $f[x_i]$ نموده می‌شود و چیزی جز مقدار f در x_i نیست :

$$f[x_i] = f(x_i).$$

بقیه تفاضلهای منقسم به استقرا تعریف می‌شوند، اولین تفاضل منقسم f ، نسبت به x_i و x_j ، با $f[x_i, x_j]$ نموده و به صورت زیر تعریف می‌شود :

$$f[x_i, x_j] = \frac{f[x_j] - f[x_i]}{x_j - x_i}.$$

با معلوم بودن تفاضلهای منقسم $(k-1)$ م

$$f[x_i, x_{i+1}, x_{i+2}, \dots, x_{i+k-1}]$$

و

$$f[x_{i+1}, x_{i+2}, \dots, x_{i+k-1}, x_{i+k}]$$

تفاضل منقسم k ام f ، نسبت به $x_i, x_{i+1}, x_{i+2}, \dots, x_{i+k}$ ، از رابطه زیر بدست می‌آید :

$$f[x_i, x_{i+1}, \dots, x_{i+k-1}, x_{i+k}] = \frac{f[x_{i+1}, x_{i+2}, \dots, x_{i+k}] - f[x_i, x_{i+1}, \dots, x_{i+k-1}]}{x_{i+k} - x_i}.$$

با این نماد، می‌توان معادله (۱۱.۳) را به صورت $a_1 = f[x_0, x_1]$ و چند جمله‌ای درونیاب در معادله (۱۰.۳) را به صورت

$$P_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + a_2(x - x_0)(x - x_1) \\ + \dots + a_n(x - x_0)(x - x_1) \dots (x - x_{n-1})$$

بیان کرد.

ثابت‌های دیگر در P_n ، یعنی a_2, a_3, \dots, a_n ، را می‌توان متوالیا " و به روشی شبیه به روش محاسبه a_0 و a_1 بدست آورد، لیکن اعمال جبری آنها خسته کننده است. محاسبه a_2 در تمرین ۶ خواسته شده است. برای روند کلی محاسبه این ثابتها، خواننده می‌تواند به کتاب سابق الذکر هیلدبراند [۳۶] مراجعه نماید. همانطور که از محاسبه a_0 و a_1 انتظار می‌رود، ثابتهای مطلوب عبارتند از:

$$a_k = f[x_0, x_1, x_2, \dots, x_k],$$

بمازای $k = 0, 1, \dots, n$ ؛ در نتیجه، P_n را می‌توان به صورت زیر نوشت:

$$P_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \dots + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1) \dots (x - x_{n-1}).$$

تعیین تفاضلهای منقسم از نقاط داده شده در جدول ۶.۳ مختصراً " شرح داده شده است. دو تفاضل چهارم و یک تفاضل پنجم نیز از این داده‌ها بدست می‌آیند.

x	$f(x)$	تفاضلهای منقسم اول	تفاضلهای منقسم دوم	تفاضلهای منقسم سوم
x_0	$f[x_0]$			
x_1	$f[x_1]$	$f[x_0, x_1] = \frac{f[x_1] - f[x_0]}{x_1 - x_0}$	$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}$	$f[x_0, x_1, x_2, x_3] = \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0}$
x_2	$f[x_2]$	$f[x_1, x_2] = \frac{f[x_2] - f[x_1]}{x_2 - x_1}$	$f[x_1, x_2, x_3] = \frac{f[x_2, x_3] - f[x_1, x_2]}{x_3 - x_1}$	$f[x_1, x_2, x_3, x_4] = \frac{f[x_2, x_3, x_4] - f[x_1, x_2, x_3]}{x_4 - x_1}$
x_3	$f[x_3]$	$f[x_2, x_3] = \frac{f[x_3] - f[x_2]}{x_3 - x_2}$	$f[x_2, x_3, x_4] = \frac{f[x_3, x_4] - f[x_2, x_3]}{x_4 - x_2}$	$f[x_2, x_3, x_4, x_5] = \frac{f[x_3, x_4, x_5] - f[x_2, x_3, x_4]}{x_5 - x_2}$
x_4	$f[x_4]$	$f[x_3, x_4] = \frac{f[x_4] - f[x_3]}{x_4 - x_3}$	$f[x_3, x_4, x_5] = \frac{f[x_4, x_5] - f[x_3, x_4]}{x_5 - x_3}$	
x_5	$f[x_5]$	$f[x_4, x_5] = \frac{f[x_5] - f[x_4]}{x_5 - x_4}$		

وقتی نقاط x_0, x_1, \dots, x_n به فواصل مساوی قرار گرفته باشند، معادله^{۱۲.۳} را می‌توان طوری بیان کرد که برای مقاصد محاسبه‌ای کاملاً مفید باشد. با معرفی نماد $h = x_{i+1} - x_i$ به‌ازای هر $i = 0, 1, \dots, n-1$ ، تفاضلهای $x - x_i$ را می‌توان به‌صورت $x - x_i = (s - i)h$ نوشت؛ در نتیجه، معادله^{۱۲.۳} می‌شود:

$$P_n(x) = P_n(x_0 + sh) = f[x_0] + shf[x_0, x_1] + s(s-1)h^2f[x_0, x_1, x_2] + \dots + s(s-1)\dots(s-n+1)h^n f[x_0, x_1, \dots, x_n]$$

$$= \sum_{k=0}^n s(s-1)\dots(s-k+1)h^k f[x_0, x_1, \dots, x_k].$$

با استفاده از نماد ضریب دو جمله‌ای، یعنی

$$\binom{s}{k} = \frac{s!}{k!(s-k)!} = \frac{s(s-1)\dots(s-k+1)}{k!},$$

می‌توان $P_n(x)$ را به‌صورت فشرده^{۱۳} زیر بیان کرد:

$$(13.3) \quad P_n(x) = P_n(x_0 + sh) = \sum_{k=0}^n \binom{s}{k} h^k f[x_0, x_1, \dots, x_k].$$

این فرمول، وقتی $h > 0$ ، فرمول تفاضل منقسم پیشرو نیوتن و، وقتی $h < 0$ ، فرمول تفاضل منقسم پسرو نیوتن نامیده می‌شود. مثال زیر کاربرد این فرمولها در مقاصد محاسبه‌ای را نشان می‌دهد.

مثال ۱. جدول داده‌ها در مثالهای ۳ از بخش ۲.۳ و ۲ از بخش ۳.۳ را در نظر می‌گیریم. جدول ۷.۳ تفاضلهای منقسم نظیر به این داده‌ها را نشان می‌دهد.

اگر تقریبی به $f(1.1)$ لازم باشد، انتخاب معقول برای x_0, x_1, \dots, x_n عبارت خواهد بود از $x_0 = 1.0$ ، $x_1 = 1.3$ ، $x_2 = 1.6$ ، $x_3 = 1.9$ ، و $x_4 = 2.2$ ، زیرا، با این انتخاب از داده‌های مربوط به نزدیکترین نقاط به $x = 1.1$ بیشترین استفاده می‌شود و از تفاضل منقسم چهارم نیز استفاده می‌گردد. این امر ایجاب می‌کند که $h = 0.3$ و $s = \frac{1}{3}$ ؛ در نتیجه، فرمول تفاضل منقسم پیشرو نیوتن بکار می‌رود، با تفاضلهای منقسم که زیر آنها در جدول خط‌کشیده شده است:

$$\begin{aligned}
 f(1.1) &= P_4(1.0 + \frac{1}{3}(.3)) = .7651997 + (\frac{1}{3})(.3)(-.4837057) \\
 &+ (\frac{1}{3})(-\frac{2}{3})(.3)^2(-.1087339) \\
 &+ (\frac{1}{3})(-\frac{2}{3})(-\frac{5}{3})(.3)^3(.0658784) \\
 &+ (\frac{1}{3})(-\frac{2}{3})(-\frac{5}{3})(-\frac{5}{3})(.3)^4(.0018297) \\
 &= .7196480.
 \end{aligned}$$

	تفاضلهای منقسم اول	تفاضلهای منقسم دوم	تفاضلهای منقسم سوم	تفاضلهای منقسم چهارم
1.0	<u>.7651977</u>			
1.3	<u>.6200860</u>	<u>-.4837057</u>		
1.6	<u>.4554022</u>	<u>-.5489460</u>	<u>-.1087339</u>	
1.9	<u>.2818186</u>	<u>-.5786120</u>	<u>-.0494433</u>	<u>.0658784</u>
2.2	<u>.1103632</u>	<u>-.5715180</u>	<u>.0118233</u>	<u>.0680740</u>
				<u>.0018297</u>

جدول ۷.۳

برای تقریب یک مقدار، وقتی x به انتهای مقادیر در جدول نزدیک است، مثلاً " $x = 2.0$ ، مجدداً" مایلیم از داده‌های مربوط به نزدیکترین نقاط به x حداکثر استفاده را بکنیم. به این دلیل، $x_0 = 2.2$ و $h = -.3$ را انتخاب می‌کنیم، در نتیجه، $x_1 = 1.9$ ، $x_2 = 1.6$ ، $x_3 = 1.3$ ، $x_4 = 1.0$ ، و $s = \frac{2}{3}$. فرمول تفاضل منقسم پسر نیوتن با تفاضلهای منقسم در جدول که زیر آنها نقطه‌چین شده است بکار می‌رود.

$$\begin{aligned}
 f(2.0) &= P_4(2.2 - \frac{2}{3}(.3)) = .1103632 + (\frac{2}{3})(-.3)(-.5715180) \\
 &+ (\frac{2}{3})(-\frac{1}{3})(-.3)^2(.0118233) \\
 &+ (\frac{2}{3})(-\frac{1}{3})(-\frac{4}{3})(-.3)^3(.0680740) \\
 &+ (\frac{2}{3})(-\frac{1}{3})(-\frac{4}{3})(-\frac{7}{3})(-.3)^4(.0018297) \\
 &= .2238755.
 \end{aligned}$$

فرمولهای نیوتن برای تقریب مقدار x که در نزدیکی مرکز جدول قرار دارد مناسب نیستند، زیرا اعمال روش پسرو یا پیشرو با تفاضل از بالاترین مرتبه، x_0 را نزدیک x نمی سازند. با اینحال، چند فرمول تفاضل منقسم در این وضع وجود دارند، که از هر یک از آنها می توان در مواقعی حداکثر بهره را برد. بخاطر تعداد این روشها، فقط یکی، مثلاً "روش استرلینگ^۱"، را عرضه کرده، و برای توضیح بیشتر مجدداً "خواننده" علاقه مند را به هیلد براند [۳۶] ارجاع می دهیم.

روند معمول برای فرمولهای مرکزی این است که x_0 را نزدیک نقطه ای که تقریب می شود اختیار و نقاط بلافاصله زیر x_0 را به صورت x_1, x_2, \dots و نقاط بلافاصله بالای آن را به صورت x_{-1}, x_{-2}, \dots اندیسه گذاری می کنیم. با این قرارداد، فرمول استرلینگ به صورت زیر درمی آید: اگر $n = 2m + 1$ ، یعنی n فرد باشد،

$$P_n(x) = P_{2m+1}(x) = f[x_0] + \frac{sh}{2}(f[x_{-1}, x_0] + f[x_0, x_1]) + s^2 h^2 f[x_{-1}, x_0, x_1] + \frac{s(s^2 - 1)h^3}{2}(f[x_{-1}, x_0, x_1, x_2] + f[x_{-2}, x_{-1}, x_0, x_1]) + \dots + s^2(s^2 - 1)(s^2 - 4)\dots(s^2 - (m - 1)^2)h^{2m} f[x_{-m}, \dots, x_m] + \frac{s(s^2 - 1)(s^2 - 4)\dots(s^2 - m^2)h^{2m+1}}{2} \times (f[x_{-m}, \dots, x_{m+1}] + f[x_{-m-1}, \dots, x_m])$$

و اگر $n = 2m$ ، یعنی n زوج باشد، همین فرمول با حذف جمله آخر بکار می رود. در جدول ۸.۳،

	تفاضلهای منقسم اول	تفاضلهای منقسم دوم	تفاضلهای منقسم سوم	تفاضلهای منقسم چهارم
x	$f(x)$			
x_{-2}	$f[x_{-2}]$			
	$f[x_{-2}, x_{-1}]$			
x_{-1}	$f[x_{-1}]$	$f[x_{-2}, x_{-1}, x_0]$		
	$f[x_{-1}, x_0]$		$f[x_{-2}, x_{-1}, x_0, x_1]$	
x_0	$f[x_0]$	$f[x_{-1}, x_0, x_1]$	$f[x_{-1}, x_0, x_1, x_2]$	$f[x_{-2}, x_{-1}, x_0, x_1, x_2]$
	$f[x_0, x_1]$		$f[x_{-1}, x_0, x_1, x_2]$	
x_1	$f[x_1]$	$f[x_0, x_1, x_2]$		
	$f[x_1, x_2]$			
x_2	$f[x_2]$			

جدول ۸.۳

1. Stirling

درایه‌های بکار رفته برای این فرمول با منحنی بسته مشخص شده‌اند .

مثال ۲ . جدول داده‌های مثال پیش را در نظر می‌گیریم . برای استفاده از فرمول استرلینگ در تقریب $f(1.5)$ با مزای $x_0 = 1.6$ ، در جدول تفاضلی ۹.۳ از درایه‌هایی که زیرشان خط‌کشیده شده استفاده می‌کنیم .

x	f(x)	تفاضلهای			
		منقسم اول	منقسم دوم	منقسم سوم	منقسم چهارم
1.0	.7651977				
1.3	.6200860	-.4837057			
1.6	.4554022	-.5489460	-.1087339	.0658784	
1.9	.2818186	-.5786120	-.0494433	.0680740	.0018297
2.2	.1103632	-.5715180	-.0118233		

جدول ۹.۳

فرمول به‌مزای $h = .3$ ، $x_0 = 1.6$ ، و $s = -\frac{1}{3}$ خواهد شد

$$\begin{aligned}
 f(1.5) &\approx \hat{P}_4(1.6 + (-\frac{1}{3})(.3)) = .4554022 + (-\frac{1}{3})(\frac{.3}{3})(-.5489460 - .5786120) \\
 &+ (-\frac{1}{3})^2(.3)^2(-.0494433) \\
 &+ \frac{1}{2}(-\frac{1}{3})((-\frac{1}{3})^2 - 1)(.3)^3(.0658784 + .0680740) \\
 &+ (-\frac{1}{3})^2((-\frac{1}{3})^2 - 1)(.3)^4(.0018297) \\
 &= .5118200
 \end{aligned}$$

مجموعه تمرینات ۴.۳

۱ . شکل تفاضل منقسم نیوتن چندجمله‌ای درونیاب درجه چهار را برای نقاط غیرمتساوی الفاصله از جدول زیر بسازید :

x	f(x)
0.0	-7.00000
.1	-5.89483
.3	-5.65014
.6	-5.17788
1.0	-4.28172

۲. فرض کنید داده $f(1.1) = -3.99583$ به‌ترین ۱ اضافه‌شود، چندجمله‌ای درونیاب درجه پنج را بسازید.

۳. با استفاده از داده‌های زیر و فرمول تفاضل منقسم پیشرو نیوتن، $f(0.05)$ را تقریب کنید:

x	0.0	.2	.4	.6	.8
$f(x)$	1.00000	1.22140	1.49182	1.82212	2.22554

۴. با استفاده از داده‌های تمرین ۳ و فرمول تفاضل منقسم پسرو، $f(0.05)$ را تقریب کنید.

۵. با استفاده از داده‌های تمرین ۳ و فرمول استرلینگ، $f(0.43)$ را تقریب کنید.

۶. اگر

$$P_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + a_2(x - x_0)(x - x_1) + a_3(x - x_0)(x - x_1)(x - x_2) + \dots + a_n(x - x_0)(x - x_1) \dots (x - x_{n-1}).$$

با استفاده از $P_n(x_2)$ نشان دهید که $a_2 = f[x_0, x_1, x_2]$.

۷. نشان دهید که به‌ازای مقداری از ξ ،

$$f[x_0, x_1, \dots, x_n, x] = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}$$

[راهنمایی. از معادله (۶.۳)، یعنی

$$f(x) = P_n(x) + \frac{f^{(n+1)}(\xi(x))}{(n+1)!}(x - x_0) \dots (x - x_n)$$

با توجه به چندجمله‌ای درونیاب درجه $n+1$ بر x_0, x_1, \dots, x_n, x ، داریم

$$f(x) = P_{n+1}(x) = P_n(x) + f[x_0, x_1, \dots, x_n, x](x - x_0) \dots (x - x_n)$$

۸. جدول جمعیت زیر در مقدمه این فصل آمده بوده و در تمرین ۱۳ از بخش ۲.۳ مورد مطالعه قرار گرفت.

سال	1920	1930	1940	1950	1960	1970
جمعیت (به هزار)	105,711	123,203	131,669	150,697	179,323	203,212

با استفاده از یک روش تفاضل منقسم مناسب، مقادیر زیر را تقریب کنید:

(آ) جمعیت در سال ۱۹۶۵:

(ب) جمعیت در سال ۲۰۰۰.

۵.۳ درونیایی هرmit^۱

مجموعه چندجمله‌ایهای بوسان تعمیمی است از چندجمله‌ایهای تیلور و چندجمله‌ایهای لاگرانژ، که در قسمت اول این فصل مطالعه شدند. این چندجمله‌ایها این خاصیت را دارند که، به ازای $n + 1$ نقطه متمایز x_0, x_1, \dots, x_n و اعداد صحیح نامنفی m_0, m_1, \dots, m_n ، چند جمله‌ای بوسان تابع $f \in C^m[a, b]$ را تقریب می‌کند، که در آن $m = \max\{m_0, m_1, \dots, m_n\}$ و، به ازای هر $i = 0, \dots, n$ ، $x_i \in [a, b]$ ، چندجمله‌ای است از کمترین درجه با این خاصیت که با تابع f و همه مشتقات آن از مرتبه نابیشتر از m_i در x_i ، به ازای هر $i = 0, 1, \dots, n$ ، یکسان است. درجه این چندجمله‌ای بوسان حداکثر

$$M = \sum_{i=0}^n m_i + n$$

است، زیرا تعداد شرایطی که باید برقرار باشند مساوی $\sum_{i=0}^n m_i + (n + 1)$ است، و یک چندجمله‌ای از درجه M دارای $M + 1$ ضریب است که می‌توان با استفاده از آنها این شرایط را برقرار ساخت. برای روشن بودن وضع، تعریف صوری یک چندجمله‌ای بوسان در زیر داده شده است.

تعریف ۶.۳. فرض کنیم $n + 1$ نقطه متمایز در $[a, b]$ و m_i ، به ازای $i = 0, 1, \dots, n$ ، عدد صحیح نامنفی مربوط به x_i باشد. فرض کنیم

$$f \in C^m[a, b] \quad \text{و} \quad m = \max_{0 \leq i \leq n} m_i$$

چندجمله‌ای بوسان تقریب ساز f چندجمله‌ای P است، با کمترین درجه، بطوری که

$$\frac{d^k P(x_i)}{dx^k} = \frac{d^k f(x_i)}{dx^k}, \quad k = 0, 1, \dots, m_i \quad \text{و} \quad i = 0, 1, \dots, n$$

توجه کنید که، وقتی $n = 0$ ، چندجمله‌ای بوسان تقریب ساز f همان چندجمله‌ای تیلور درجه m_0 برای f در x_0 است، و وقتی به ازای n ، $m_i = 0$ ، $i = 0, 1, \dots, n$ ، چندجمله‌ای بوسان چندجمله‌ای درونیایی f بر x_0, x_1, \dots, x_n ، یعنی چندجمله‌ای لاگرانژ

در این نقاط، می باشد.

وقتی به ازای هر $m_i = 1, i = 0, 1, \dots, n$ ، ردهای از چند جمله‌ایها به نام چند جمله‌ایهای هرمیت بدست می‌آید. به ازای تابع معلوم f ، این چند جمله‌ایها نه تنها با مقادیر f در نقاط x_0, x_1, \dots, x_n یکی هستند، بلکه، چون مشتقات اول آنها با مشتقات اول f یکسانند، همان "شکل" تابع را در این نقاط دارند، به این معنی که خطوط مماس بر چند جمله‌ای و تابع یکی هستند. ما چند جمله‌ایهای بوسان را در این وضع بررسی می‌کنیم، و ابتدا به قضیه‌ای می‌پردازیم که روند بدست آوردن چند جمله‌ایهای از نوع هرمیت را دقیقاً "توصیف می‌کند".

قضیه ۷.۳. هرگاه $f \in C^1[a, b]$ و $x_0, \dots, x_n \in [a, b]$ متمایز باشند، چند جمله‌ای منحصر بفرد از کمترین درجه که با f و f' در x_0, \dots, x_n یکی باشد از رابطه زیر بدست می‌آید:

$$(14.3) \quad H_{2n+1}(x) = \sum_{j=0}^n f(x_j)H_{n,j}(x) + \sum_{j=0}^n f'(x_j)\hat{H}_{n,j}(x),$$

که در آن

$$(15.3) \quad H_{n,j}(x) = [1 - 2(x - x_j)L'_{n,j}(x_j)]L_{n,j}^2(x)$$

و

$$(16.3) \quad \hat{H}_{n,j}(x) = (x - x_j)L_{n,j}^2(x).$$

در این وضع، $L_{n,j}$ چند جمله‌ای ضریب لاگرانژ z م درجه n را نشان می‌دهد که با معادله (۴.۳) تعریف می‌شود. بعلاوه، اگر $f \in C^{(2n+2)}[a, b]$ ، به ازای نقطه‌ای چون ξ که $a < \xi < b$

$$(17.3) \quad f(x) - H_{2n+1}(x) = \frac{(x - x_0)^2 \cdots (x - x_n)^2}{(2n + 2)!} f^{(2n+2)}(\xi).$$

برهان. برای اثبات اینکه به ازای هر $j = 0, 1, \dots, n$ و $k = 0, 1, \dots, n$

$$(18.3) \quad \frac{d^k}{dx^k} H_{2n+1}(x_k) = \frac{d^k f(x_k)}{dx^k}$$

کافی است نشان دهیم $H_{n,j}$ و $\hat{H}_{n,j}$ ، که به ترتیب با معادلات (۱۵.۳) و (۱۶.۳) تعریف

شده‌اند، در شرایط (T) تا (ت) زیر صدق می‌کنند:

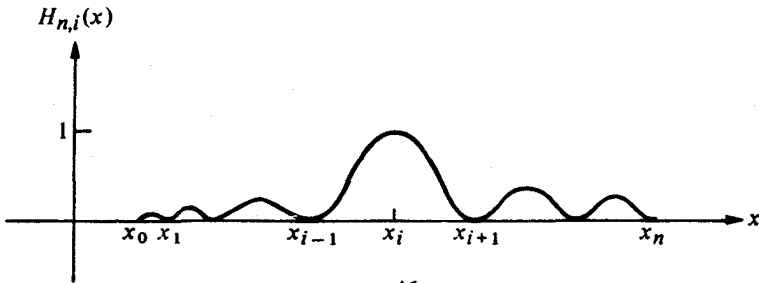
$$(T) \quad H_{n,j}(x_k) = \begin{cases} 0, & \text{اگر } j \neq k \\ 1, & \text{اگر } j = k \end{cases}$$

$$(ب) \quad \frac{d}{dx} H_{n,j}(x_k) = 0, \quad k \text{ به‌ازای هر } k$$

$$(پ) \quad H_{n,j}(x_k) = 0, \quad k \text{ به‌ازای هر } k$$

$$(ت) \quad \frac{d}{dx} \hat{H}_{n,j}(x_k) = \begin{cases} 0, & \text{اگر } j \neq k \\ 1, & \text{اگر } j = k \end{cases}$$

(ر.ک. شکل ۸.۳ و ۹.۳)



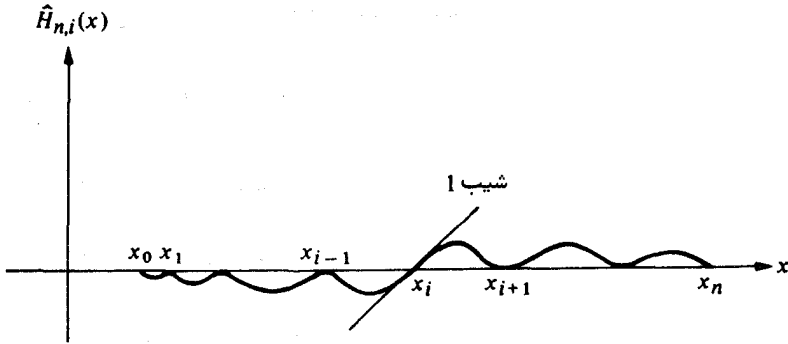
شکل ۸.۳

این شرایط متضمن آنند که وقتی H_{2n+1} یا H_{2n+1} در یکی از نقاط داده شده حساب شود، مقدار مناسب بدست آید، زیرا، به‌ازای هر $i = 0, 1, \dots, n$ ،

$$\begin{aligned} H_{2n+1}(x_i) &= \sum_{j=0}^n f(x_j) H_{n,j}(x_i) + \sum_{j=0}^n f'(x_j) \hat{H}_{n,j}(x_i) \\ &= f(x_i) \cdot 1 + \sum_{\substack{j=0, \\ j \neq i}}^n f(x_j) \cdot 0 + \sum_{j=0}^n f'(x_j) \cdot 0 \\ &= f(x_i), \end{aligned}$$

۹

$$\begin{aligned} H'_{2n+1}(x_i) &= \sum_{j=0}^n f(x_j) H'_{n,j}(x_i) + \sum_{j=0}^n f'(x_j) \hat{H}'_{n,j}(x_i) \\ &= \sum_{j=0}^n f(x_j) \cdot 0 + f'(x_i) \cdot 1 + \sum_{\substack{j=0, \\ j \neq i}}^n f'(x_j) \cdot 0 \\ &= f'(x_i), \end{aligned}$$



شکل ۹.۳

اگر ابتدا چند جمله‌ای $H_{n,j}$ را در نظر بگیریم، ملزومات (پ) و (ت) ایجاب می‌کنند که $H_{n,j}$ باید یک ریشه مضاعف در x_k به‌ازای $k \neq j$ و یک ریشه ساده در x_j داشته باشد. یک چند جمله‌ای از درجه حداکثر $(2n + 1)$ و حائز این شرایط عبارت است از

$$\begin{aligned} \hat{H}_{n,j}(x) &= \frac{(x - x_0)^2 \cdots (x - x_{j-1})^2 (x - x_j)(x - x_{j+1})^2 \cdots (x - x_n)^2}{(x_j - x_0)^2 \cdots (x_j - x_{j-1})^2 (1)(x_j - x_{j+1})^2 \cdots (x_j - x_n)^2} \\ &= L_{n,j}^2(x)(x - x_j). \end{aligned}$$

شرایط (ت) و (ب) ایجاب می‌کنند که x_k ، به‌ازای هر $k \neq j$ ، یک ریشه مضاعف $H_{n,j}(x)$ باشد و هر چند جمله‌ای از درجه حداکثر $(2n + 1)$ و صادق در (ت) و (ب) عبارت است از

$$H_{n,j}(x) = (x - x_0)^2 \cdots (x - x_j)^2 (x - x_{j+1})^2 \cdots (x - x_n)^2 (ax + b)$$

که در آن a و b ثابت هستند. با فرض

$$a = \hat{a} \prod_{\substack{i=0 \\ i \neq j}}^n (x_i - x_j)^2$$

$$b = \hat{b} \prod_{\substack{i=0 \\ i \neq j}}^n (x_i - x_j)^2$$

a و b را تعدیل می‌کنیم؛ خواهیم داشت

$$H_{n,j}(x) = L_{n,j}^2(x)(ax + b).$$

شرط (T) ایجاب می‌کند که

$$1 = H_{n,j}(x_j) = L_{n,j}^2(x_j)(ax_j + b) = ax_j + b$$

و، با استفاده از (ب) و معادله (۱۹.۳)، نتیجه می‌شود که

$$\begin{aligned} 0 &= \frac{dH_{n,j}(x_j)}{dx} = 2L_{n,j}(x_j)L'_{n,j}(x_j)(ax_j + b) + L_{n,j}^2(x_j)(a) \\ &= 2L'_{n,j}(x_j)(ax_j + b) + a \\ &= 2L'_{n,j}(x_j)(1) + a \end{aligned}$$

یا

$$a = -2L'_{n,j}(x_j).$$

از معادله (۱۹.۳) داریم

$$b = 1 - ax_j = 1 + 2L'_{n,j}(x_j) \cdot (x_j);$$

در نتیجه،

$$(ax + b) = -2L'_{n,j}(x_j)x + 1 + 2L'_{n,j}(x_j)(x_j) = 1 - 2(x - x_j)L'_{n,j}(x_j)$$

و

$$H_{n,j}(x) = (ax + b)L_{n,j}^2(x) = [1 - 2(x - x_j)L'_{n,j}(x)]L_{n,j}^2(x).$$

ما $H_{n,j}$ و $H_{n,j}$ ها را که به ترتیب با معادلات (۱۵.۳) و (۱۶.۳) داده شده‌اند ساختاریم. توجه کنید که، چون هر $H_{n,j}$ و هر $H_{n,j}$ از درجه حداکثر $(2n + 1)$ است، H_{2n+1} از درجه حداکثر $2n + 1$ می‌باشد. علاوه، شرایط (T)، (ب)، (پ)، و (ت) ایجاب می‌کنند که معادله (۱۸.۳) برقرار باشد.

حال یکتایی را نشان می‌دهیم؛ یعنی، نشان می‌دهیم که H_{2n+1} تنها چندجمله‌ای از درجه نابیشتر از $2n + 1$ است که در معادله (۱۸.۳) صدق می‌کند. فرض کنیم P یک چندجمله‌ای از درجه نابیشتر از $2n + 1$ باشد با این خاصیت که

$$P(x_k) = f'(x_k) \text{ و } P(x_k) = f(x_k), \quad k = 0, 1, \dots, n$$

چون $D = H_{2n+1} - P$ نیز یک چندجمله‌ای از درجه حداکثر $(2n + 1)$ است و

$$D(x_k) = 0 \text{ و } D'(x_k) = 0, \quad k = 0, 1, \dots, n$$

D باید، به ازای چندجمله‌ایی مانند $g(x)$ ، به شکل زیر باشد:

$$D(x) = (x - x_0)^2(x - x_1)^2 \cdots (x - x_n)^2 g(x),$$

اما D یک چندجمله‌ای از درجه حداکثر $(2n + 1)$ با $(2n + 2)$ ریشه است؛ در نتیجه، $D \equiv 0$. این ایجاب می‌کند که به ازای هر مقدار از x ، $P(x) = H_{2n+1}(x)$ ، و یکتایی H_{2n+1}

ثابت شده است .

محاسبه جمله خطا در معادله (۱۷.۳) در تمرین ۶ خواسته شده است .

مثال ۱. با استفاده از چندجمله‌ای با کمترین درجه که با داده‌های جدول ۱۵.۳ برای تابع بسط اول از مرتبه صفر سازگار است ، تقریبی برای $f(1.5)$ بیابید .

k	x_k	$f(x_k)$	$f'(x_k)$
0	1.3	.6200860	-.5220232
1	1.6	.4554022	-.5698959
2	1.9	.2818186	-.5811571

جدول ۱۵.۳

ابتدا چندجمله‌ایهای لاگرانژ و مشتقات آنها را حساب می‌کنیم :

$$L_{2,0}(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} = \frac{50}{9}x^2 - \frac{175}{9}x + \frac{152}{9},$$

$$L'_{2,0}(x) = \frac{100}{9}x - \frac{175}{9};$$

$$L_{2,1}(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} = \frac{-100}{9}x^2 + \frac{320}{9}x - \frac{247}{9},$$

$$L'_{2,1}(x) = \frac{-200}{9}x + \frac{320}{9};$$

$$L_{2,2}(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} = \frac{50}{9}x^2 - \frac{145}{9}x + \frac{104}{9},$$

و

$$L'_{2,2}(x) = \frac{100}{9}x - \frac{145}{9}.$$

دراین صورت ، چندجمله‌ایهای $H_{2,j}$ و $\hat{H}_{2,j}$ عبارتند از

$$H_{2,0}(x) = [1 - 2(x - 1.3)(-5)] \left(\frac{50}{9}x^2 - \frac{175}{9}x + \frac{152}{9} \right)^2$$

$$= (10x - 12) \left(\frac{50}{9} x^2 - \frac{175}{9} x + \frac{152}{9} \right)^2,$$

$$H_{2,1}(x) = 1 \cdot \left(\frac{-100}{9} x^2 + \frac{320}{9} x - \frac{247}{9} \right)^2,$$

$$H_{2,2}(x) = 10(2 - x) \left(\frac{50}{9} x^2 - \frac{145}{9} x + \frac{104}{9} \right)^2;$$

$$\hat{H}_{2,0}(x) = (x - 1.3) \left(\frac{50}{9} x^2 - \frac{175}{9} x + \frac{152}{9} \right)^2,$$

$$\hat{H}_{2,1}(x) = (x - 1.6) \left(\frac{-100}{9} x^2 + \frac{320}{9} x - \frac{247}{9} \right)^2,$$

$$\hat{H}_{2,2}(x) = (x - 1.9) \left(\frac{50}{9} x^2 - \frac{145}{9} x + \frac{104}{9} \right)^2.$$

بالاخره،

$$H_5(x) = .6200860H_{2,0}(x) + .4554022H_{2,1}(x) + .2818186H_{2,2}(x) \\ - .5220232\hat{H}_{2,0}(x) - .5698959\hat{H}_{2,1}(x) - .5811571\hat{H}_{2,2}(x)$$

$$H_5(1.5) = .6200860 \left(\frac{4}{27} \right) + .4554022 \left(\frac{64}{81} \right) + .2818186 \left(\frac{5}{81} \right) \\ - .5220232 \left(\frac{4}{405} \right) - .5698959 \left(\frac{-32}{405} \right) - .5811571 \left(\frac{-2}{405} \right) \\ = .5118277,$$

نتیجه‌ای که تا ارقام داده شده دقیق است.

مثال ۰۲. برای مقایسه تقریبهای حاصل از بکارگیری یک چندجمله‌ای هرمیت درجه بالاتر روی بازه‌های بزرگتر و یک چندجمله‌ای هرمیت درجه بالاتر بر بازه‌های کوچکتر، فرض کنیم $f(x) = e^{-1x^2}$. مقدار $f(1.25)$ را به وسیله $H_3(1.25)$ و $H_5(1.25)$ تقریب

می‌کنیم، که H_5 با استفاده از $x_0 = 1$ ، $x_1 = 2$ ، $x_2 = 3$ و H_3 با استفاده از $\bar{x}_0 = 1.5$ و $\bar{x}_1 = 1.5$ محاسبه می‌شوند. داده‌های مورد استفاده در جدول ۱۱.۳ آمده‌اند.

x	$f(x) = e^{-1x^2}$	$f'(x) = -2xe^{-1x^2}$
$x_0 = \bar{x}_0 = 1$	1.105170918	.2210341836
$\bar{x}_1 = 1.5$	1.252322716	.3756968148
$x_1 = 2$	1.491824698	.5967298792
$x_2 = 3$	2.459603111	1.475761867

جدول ۱۱.۳

محاسبات مربوط به $H_3(1.25)$ در زیر آمده‌اند:

$$L_{1,0}(x) = \frac{(x - \bar{x}_1)}{(\bar{x}_0 - \bar{x}_1)} = -2x + 3, \quad L'_{1,0}(x) = -2;$$

$$L_{1,1}(x) = \frac{(x - \bar{x}_0)}{(\bar{x}_1 - \bar{x}_0)} = 2x - 2, \quad L'_{1,1}(x) = 2;$$

$$H_{1,0}(x) = [1 - 2(x - \bar{x}_0)L'_{1,0}(\bar{x}_0)]L_{1,0}^2(x) = 16x^3 - 60x^2 + 72x - 27$$

$$H_{1,1}(x) = [1 - 2(x - \bar{x}_1)L'_{1,1}(\bar{x}_1)]L_{1,1}^2(x) = -16x^3 + 60x^2 - 72x + 28$$

$$\hat{H}_{1,0}(x) = (x - \bar{x}_0)L_{1,0}^2(x) = 4x^3 - 16x^2 + 21x - 9,$$

$$\hat{H}_{1,1}(x) = (x - \bar{x}_1)L_{1,1}^2(x) = 4x^3 - 14x^2 + 16x - 6;$$

$$\begin{aligned} H_3(x) &= f(\bar{x}_0)(16x^3 - 60x^2 + 72x - 27) + f(\bar{x}_1)(-16x^3 + 60x^2 - 72x + 28) \\ &\quad + f'(\bar{x}_0)(4x^3 - 16x^2 + 21x - 9) + f'(\bar{x}_1)(4x^3 - 14x^2 + 16x - 6) \\ &= 1.105170918[((16x - 60)x + 72)x - 27] \\ &\quad + 1.252322716[(-16x + 60)x - 72]x + 28] \\ &\quad + .2210341836[((4x - 16)x + 21)x - 9] \\ &\quad + .3756968148[((4x - 14)x + 16)x - 6]. \end{aligned}$$

در نتیجه،

$$\begin{aligned} H_3(1.25) &= .552585459 + .626161358 + .0138146365 - .0234810509 \\ &= 1.169080403. \end{aligned}$$

نتایج کلیدی در محاسبه $H_3(1.25)$ در زیر خلاصه شده‌اند:

$$\begin{aligned}
 H_{2,0}(x) &= .75x^5 - 8x^4 + 32.75x^3 - 63.5x^2 + 57x - 18 \\
 H_{2,0}(1.25) &= .7536621100; \\
 \hat{H}_{2,0}(x) &= .25x^5 - 2.75x^4 + 11.75x^3 - 24.25x^2 + 24x - 9 \\
 \hat{H}_{2,0}(1.25) &= .1076660130; \\
 H_{2,1}(x) &= x^4 - 8x^3 + 22x^2 - 24x + 9, \\
 H_{2,1}(1.25) &= .1914062500; \\
 \hat{H}_{2,1}(x) &= x^5 - 10x^4 + 38x^3 - 68x^2 + 57x - 18, \\
 \hat{H}_{2,1}(1.25) &= -.1435546900; \\
 H_{2,2}(x) &= -.75x^5 + 7x^4 - 24.75x^3 + 41.5x^2 - 33x + 10, \\
 H_{2,2}(1.25) &= .0549316370; \\
 \hat{H}_{2,2}(x) &= .25x^5 - 2.25x^4 + 7.75x^3 - 12.75x^2 + 10x - 3, \\
 \hat{H}_{2,2}(1.25) &= -.01538086000,
 \end{aligned}$$

و

$$H_5(1.25) = 1.169015871.$$

چون تا نه رقم اعشار $f(1.25) = 1.169118446$ ، خطاها عبارت خواهند بود از

$$|H_3(1.25) - f(1.25)| = |1.169080403 - 1.169118446| = 3.8043 \times 10^{-5},$$

$$|H_5(1.25) - f(1.25)| = |1.169015871 - 1.169118446| = 1.02575 \times 10^{-4}.$$

نتایج فوق نباید موجب تعجب شوند ، زیرا از کران خطا در قضیه ۷.۳ نتیجه

می‌شود که

$$\begin{aligned}
 |H_5(1.25) - f(1.25)| &\leq \frac{(1.25 - 1)^2(1.25 - 2)^2(1.25 - 3)^2}{6!} \max_{x \in [1, 3]} |f^{(6)}(x)| \\
 &\leq 1.4954 \times 10^{-4} \max_{x \in [1, 3]} [.000064x^6 + .0048x^4 + .072x^2 + .12]e^{-1x^2} \\
 &= (1.4954 \times 10^{-4})(2.9600) \\
 &\leq 4.43 \times 10^{-4},
 \end{aligned}$$

و

$$\begin{aligned}
 |H_3(1.25) - f(1.25)| &\leq \frac{(1.25 - 1)^2(1.25 - 1.5)^2}{4!} \max_{x \in [1, 1.5]} |f^{(4)}(x)| \\
 &\leq 1.6277 \times 10^{-4} \max_{x \in [1, 1.5]} [0.0016x^4 + 0.048^2 + .12]e^{-1x^2} \\
 &\leq (1.6277 \times 10^{-4})(.2957) \\
 &\leq 4.82 \times 10^{-5}.
 \end{aligned}$$

کران خطای بهتر برای H_3 نتیجه این امر است که مشتق ششم f خیلی از مشتق چهارم آن بزرگتر است. این مثال نشان می‌دهد که، در بعضی حالات، بهتر است برای تقریبهای بهتر بازه‌های کوچکتر و چند جمله‌ایهای از درجه پایین‌تر در نظر گرفت.

مجموعه تمرینات ۵.۳

۱. با داده‌های زیر و با استفاده از درونیابی هرمیت، تقریبی برای $f(2.5)$ بیابید.

x	$f(x)$	$f'(x)$
2.2	.5207843	-.0014878
2.4	.5104147	-.1004889
2.6	.4813306	-.1883635

۲. اگر $f(x) = 3xe^x - e^{2x}$ ، با استفاده از $x_0 = 1$ و $x_1 = 1.05$ ، $f(1.03)$ را به وسیله چند جمله‌ای درونیاب هرمیت از درجه حداکثر سه تقریب کنید. خطای واقعی را با کران خطای حاصل از معادله (۱۷.۳) مقایسه نمایید.

۳. با استفاده از $x_0 = 1$ ، $x_1 = 1.05$ و $x_2 = 1.07$ ، تمرین ۲ را با چند جمله‌ای درونیاب هرمیت از درجه حداکثر پنج تکرار کنید.

۴. فرض کنید $f(x) = 3xe^x - 2x$. آیا تقریب چند جمله‌ای درونیاب درجه دو، یعنی $P_2(1.03)$ ، که در تمرین ۵ از بخش ۲.۳ حساب شد، تقریب بهتری از تقریب چند جمله‌ای هرمیت درجه سه، یعنی $H_3(1.03)$ ، که در تمرین ۲ این بخش محاسبه شد، به $f(1.03)$ است؟ تخمینهای خطای نظری برای

$$|H_3(1.03) - f(1.03)| \quad \text{و} \quad |P_2(1.03) - f(1.03)|$$

را مقایسه نمایید.

۵. (T) فرض کنید $f(x) = e^x$ ، یعنی چند جمله‌ای هرمیتی تقریب ساز f را،

با استفاده از نقاط $x_0 = 0$ ، $x_1 = 1$ ، و $x_2 = 2$ پیدا کنید $H_3(2.5)$ را با $f(2.5)$ و با $P(2.5)$ بصورتی که در تمرین ۷ از بخش ۲.۳ داده شده مقایسه کنید .

(ب) با استفاده از قضیه ۷.۳ ، $|H_3(2.5) - f(2.5)|$ را تخمین بزنید .

۶. جمله خطا در قضیه ۷.۳ را بدست آورید . [راهنمایی . همان روش بدست آوردن خطای لاگرانژ را بکار برید ، طبق قضیه ۳.۳ از تعریف

$$g(t) = f(t) - H_{2n+1}(t) + \frac{(t - x_0)^2 \cdots (t - x_n)^2}{(x - x_0)^2 \cdots (x - x_n)^2} [f(x) - H_{2n+1}(x)]$$

و این امر که $g'(t)$ دارای $(2n + 2)$ ریشه متمایز در $[a, b]$ است استفاده کنید .

۶.۳ درونیابی اسپلاین مکعبی

در بخشهای پیشتر این فصل با تقریب توابع دلخواه بر بازه‌های بسته ، با استفاده از چندجمله‌ایها ، مواجه بودیم . این روش تقریب اگرچه در بسیاری مواقع مناسب است ، اما سرشت نوسانی چندجمله‌ایهای درجه بالا و این خاصیت که هر تغییر بر بخش کوچکی از بازه می‌تواند تغییرات زیادی را روی تمام برد موجب شود ، کاربرد آنها را در تقریب بسیاری از توابع که عملاً " در فیزیک پیش می‌آیند محدود می‌کند .

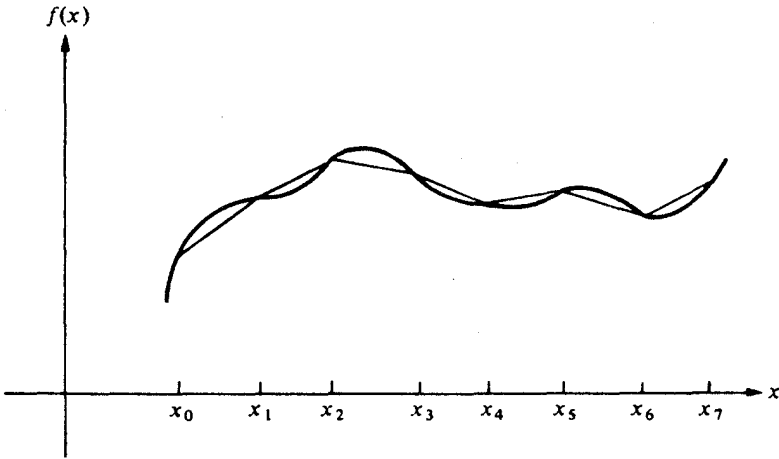
روش دیگری که می‌توان برای بدست آوردن توابع درونیاب بکار برد این است که بازه را به گردآیه‌ای از زیربازه‌ها تقسیم کرده و (عموماً " برهر زیربازه چندجمله‌ای تقریب ساز متفاوتی بسازیم . تقریب به‌وسیله توابع از این نوع تقریب قطعه قطعه چندجمله‌ای نامیده می‌شود .

ساده‌ترین نوع تقریب قطعه قطعه چندجمله‌ای ، درونیابی قطعه قطعه خطی نام دارد و عبارت است از اتصال مجموعه‌ای از نقاط داده‌ای مانند

$$\{(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))\}$$

با رشته‌ای از خطوط مستقیم بهم (ر.ک. شکل ۱۰.۳) . این روشی درونیابی است که در درسهای مقدماتی مربوط به توابع مثلثاتی یا لگاریتمی ، وقتی مقادیر میانی از یک گردآیه یا جدولی از مقادیر اختیار می‌شوند ، بکار می‌رود .

نقص حل یک مسئله تقریب با استفاده از توابع از این نوع ، یا مثلاً " چندجمله‌ای درجه دو بین سه نقطه متوالی ، این است که در هر نقطه انتهای زیربازه‌ها اطمینانی به مشتق‌پذیری نیست ، که ، به بیان هندسی ، یعنی تابع درونیاب در این نقاط "هموار"



شکل ۱۰.۳

نمی باشد .

چون اغلب یک تابع تقریب ساز به طور پیوسته مشتق پذیر مطلوب است ، روند دیگری که ارائه خواهد شد استفاده از یک قطعه قطعه چندجمله ای از نوع هرمیت می باشد . مثلاً ، اگر مقادیر تابعهای f و f' در هر نقطه $x_0 < x_1 < \dots < x_n$ معلوم باشند ، می توان با استفاده از یک چندجمله ای هرمیت درجه ۳ به هر زیر بازه $[x_{i-1}, x_i]$ ، $[x_0, x_1]$ ، $[x_1, x_2]$ ، \dots ، $[x_{n-1}, x_n]$ تابعی بدست آورد که بر بازه $[x_0, x_n]$ به طور پیوسته مشتق پذیر باشد . اشکال چنین روندی این است که داده های لازم برای تابع تقریب شو اغلب معلومند ، حال آنکه داده های مشتق آن چنین نیستند . به این دلیل ، درونیابی قطعه قطعه هرمیت عموماً " مناسب نیستند .

شاید ساده ترین نوع تابع قطعه-قطعه چندجمله ای که بر تمام بازه $[x_0, x_n]$ مشتق پذیر باشد تابع حاصل از تطبیق یک چندجمله ای درجه ۲ دو بین هر جفت گره متوالی است . این را می توان با ساختن یک چندجمله ای درجه ۲ دو بر $[x_0, x_1]$ که با تابع در x_0 و x_1 یکی است ، چندجمله ای درجه ۲ دو دیگر بر $[x_1, x_2]$ که با تابع در x_1 و x_2 یکی است ، الی آخر ، انجام داد . چون یک چندجمله ای درجه ۲ دو کلی سه ثابت دلخواه دارد ، جمله ثابت ، ضریب x ، و ضریب x^2 ، و فقط دو شرط برای تطبیق داده ها در نقاط انتهایی هر زیر بازه لازم است ، امکان اینکه چندجمله ای درجه ۲ دو طوری اختیار شود که ، علاوه بر این ، درونیاب بر $[x_0, x_n]$ مشتق پیوسته داشته باشد هست . بهر حال ، یکی از مشکلات

این روند زمانی است که تعیین یکسانی مشتق درونیاب با مشتق تابع در نقاط انتهایی x_0 و x_n لازم باشد. در این حالت، می‌توان نشان داد که ثابتهای کافی برای تضمین برقراری شرطها وجود ندارند. توضیح بیشتر در باب این مشکل و مشکلات دیگر رامی‌توان در تمرینهای ۷ و ۸ یافت.

تقریب قطعه قطعه چندجمله‌ای با استفاده از چندجمله‌ایهای مکعبی بین هر جفت گره متوالی درونیایی اسپلاین مکعبی نام دارد و در حال حاضر یکی از متداولترین روشهای مورد استفاده است. چون یک چندجمله‌ای مکعبی کلی چهار ثابت دارد، در روند اسپلاین مکعبی نه فقط برای آنکه درونیاب بر بازه به‌طور پیوسته مشتقپذیر باشد، بلکه مشتق دوم پیوسته نیز بر بازه داشته باشد انعطاف کافی وجود خواهد داشت. بهر حال، یادآوری می‌شویم که مشتقات درونیاب با مشتقات تابع، حتی در گرهها، یکسان فرض نشده است.

تعریف ۸.۳. فرض کنیم تابع f بر $[a, b]$ تعریف شده باشد، $a = x_0 < x_1 < \dots < x_n = b$ ، مجموعه‌ای از نقاط، یا گرهها، باشد. یک درونیاب اسپلاین مکعبی برای f تابعی است مانند S که در شرایط زیر صدق می‌کند:

(آ) S یک چندجمله‌ای مکعبی است، که به‌ازای هر $j = 0, 1, \dots, n-1$ بر زیربازه $[x_j, x_{j+1}]$ با S_j نموده می‌شود،

(ب) به‌ازای هر $j = 0, 1, \dots, n$ ، $S(x_j) = f(x_j)$ ؛

(پ) به‌ازای هر $j = 0, 1, \dots, n-2$ ، $S_{j+1}(x_{j+1}) = S_j(x_{j+1})$ ؛

(ت) به‌ازای هر $j = 0, 1, \dots, n-2$ ، $S'_{j+1}(x_{j+1}) = S'_j(x_{j+1})$ ؛

(ث) به‌ازای هر $j = 0, 1, \dots, n-2$ ، $S''_{j+1}(x_{j+1}) = S''_j(x_{j+1})$ ؛

(ج) یکی از شرایط مرزی زیر برقرار است،

(یک) $S''(x_0) = S''(x_n) = 0$ (مرزی آزاد)؛

(دو) $S'(x_0) = f'(x_0)$ و $S'(x_n) = f'(x_n)$ (مرزی مقید)

با آنکه می‌توان اسپلاینها مکعبی را با شرایط مرزی دیگر تعریف کرد، شرایط فوق برای مقاصد ما کافی خواهند بود. در حالت شرایط مرزی آزاد، اسپلاین را یک اسپلاین طبیعی می‌نامند و نمودارش تقریبی است برای شکل یک میله قابل انعطاف که باید از هریک از نقاط داده‌ای $\{(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))\}$ بگذرد. در

حالت کلی، شرایط مرزی مقید به تقریبهای دقیقتری منجر می‌شوند به این دلیل که شامل اطلاعات بیشتری از تابع هستند، با اینحال، برای برقراری این نوع شرط مرزی باید مقادیر مشتق در نقاط انتهایی یا یک تقریب دقیق به این مقادیر را داشته باشیم.

برای ساختن درونیاب اسپلاین مکعبی برای تابع معلوم f ، می‌توان شرایط تعریف را بر هر چند جمله‌ای مکعبی، یعنی $S_j(x)$ ، که به شکل

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3$$

به‌ازای هر $j = 0, 1, \dots, n-1$ نوشته می‌شود، اعمال کرد.

وقتی $S_j(x)$ به این صورت نوشته شود، واضح است که

$$S_j(x_j) = a_j = f(x_j)$$

و اگر شرط (پ) برقرار باشد، به‌ازای هر $j = 0, 1, \dots, n-2$ ،

$$\begin{aligned} a_{j+1} &= S_{j+1}(x_{j+1}) \\ &= S_j(x_{j+1}) \\ &= a_j + b_j(x_{j+1} - x_j) + c_j(x_{j+1} - x_j)^2 + d_j(x_{j+1} - x_j)^3 \end{aligned}$$

چون جمله $(x_{j+1} - x_j)$ در این بحث مکرر بکار می‌رود، شایسته است نماد ساده‌تر

$$h_j = x_{j+1} - x_j$$

را به‌ازای هر $j = 0, 1, \dots, n-1$ معرفی کنیم. علاوه بر این، اگر قرار دهیم $a_n = f(x_n)$ این ایجاب می‌کند که معادله

$$\begin{aligned} (22.3) \quad a_{j+1} &= a_j + b_j h_j + c_j h_j^2 + d_j h_j^3 \\ &\text{به‌ازای هر } j = 0, 1, \dots, n-1 \text{ برقرار باشد.} \end{aligned}$$

به همین نحو، تعریف می‌کنیم $b_n = S'(x_n)$ و می‌بینیم که

$$S'_j(x) = b_j + 2c_j(x - x_j) + 3d_j(x - x_j)^2$$

ایجاب می‌کند که به‌ازای هر $j = 0, 1, \dots, n-1$ ، $S'_j(x_j) = b_j$ ، اعمال شرط (ت)

نتیجه می‌دهد که به‌ازای هر $j = 0, 1, \dots, n-1$ ،

$$(23.3) \quad b_{j+1} = b_j + 2c_j h_j + 3d_j h_j^2$$

رابطه دیگر بین ضرایب S_j را می‌توان با تعریف $c_n = S''(x_n)/2$ و اعمال شرط

(ث) بدست آورد. در این حالت، به‌ازای هر $j = 0, 1, \dots, n-1$ ،

$$(24.3) \quad c_{j+1} = c_j + 3d_j h_j$$

از حل معادله (۱۴.۳) نسبت به d_j و گذاردن این مقدار در معادلات (۲۲.۳)

و (۲۳.۳)، معادلات جدید زیر بدست می‌آیند: به‌ازای هر $j = 0, 1, \dots, n-1$ ،

$$(25.2) \quad a_{j+1} = a_j + b_j h_j + \frac{h_j^2}{3} (2c_j + c_{j+1})$$

و

$$(26.2) \quad b_{j+1} = b_j + h_j (c_j + c_{j+1}).$$

رابطه نهایی مربوط به ضرایب را می‌توان با حل معادله مناسب به شکل معادله

(۲۵.۳)، ابتدا نسبت به b_j ، یعنی

$$(27.2) \quad b_j = \frac{1}{h_j} (a_{j+1} - a_j) - \frac{h_j}{3} (2c_j + c_{j+1}),$$

و سپس، با تقلیل اندیس، نسبت به b_{j-1} ، یعنی

$$b_{j-1} = \frac{1}{h_{j-1}} (a_j - a_{j-1}) - \frac{h_{j-1}}{3} (2c_{j-1} + c_j),$$

بدست آورد. با گذاردن این مقادیر در معادله حاصل از معادله (۲۵.۳)، وقتی اندیس

یکی کم شود، دستگاه معادلات خطی زیر بدست می‌آید: به‌ازای هر $j = 1, 2, \dots, n-1$ ،

$$h_{j-1} c_{j-1} + 2(h_{j-1} + h_j) c_j + h_j c_{j+1}$$

$$(28.2) \quad = \frac{3}{h_j} (a_{j+1} - a_j) - \frac{3}{h_{j-1}} (a_j - a_{j-1}).$$

این دستگاه فقط شامل $\{c_j\}_{j=0}^n$ به عنوان مجهولات است، زیرا مقادیر $\{h_j\}_{j=0}^{n-1}$ و

$\{a_j\}_{j=0}^n$ با قرار دادن نقاط $\{x_j\}_{j=0}^n$ و مقادیر f در این نقاط داده شده‌اند.

توجه کنید که، با معلوم بودن $\{c_j\}_{j=0}^n$ ، به سادگی می‌توان بقیه ثابتهای $\{b_j\}_{j=0}^{n-1}$

را از معادله (۲۷.۳)، و $\{d_j\}_{j=0}^{n-1}$ را از معادله (۲۴.۳) یافت، و چندجمله‌ایهای

مکعبی $\{S_j\}_{j=0}^{n-1}$ را ساخت.

سوال واضح در باب این ساخت آن است که آیا مقادیر $\{c_j\}_{j=0}^n$ را می‌توان با استفاده

از دستگاه معادلات (۲۸.۳) یافت، و اگر چنین است، آیا این مقادیر منحصر بفرداند؟

قضیه زیر نشان می‌دهد که، وقتی یکی از شرایط مرزی قسمت (ج) تعریف اعمال شود،

جواب هر دو سوال مثبت خواهد بود.

قضیه ۹.۳. هرگاه تابع f بر $[a, b]$ تعریف شده باشد، f دارای درونیاب اسپلاین

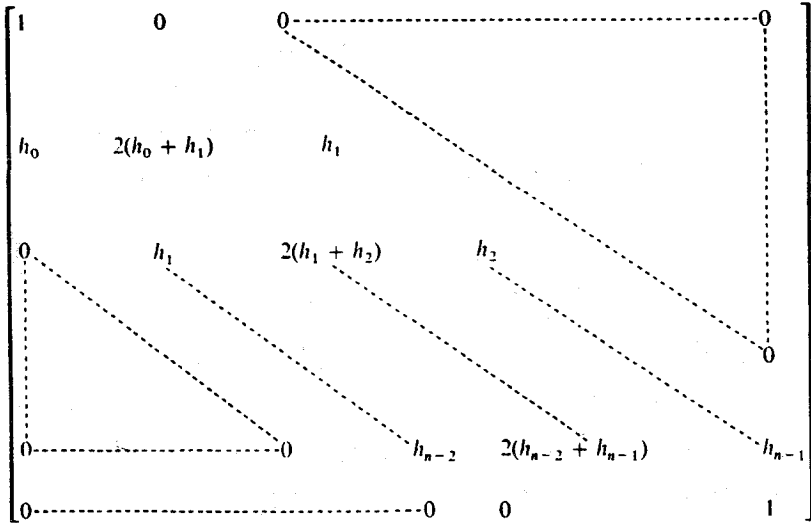
طبیعی منحصر بفرد است، یعنی، درونیاب اسپلاین منحصر بفردی که در شرایط مرزی آزاد $S''(a) = S''(b) = 0$ صدق می‌کند.

با نماد معمولی $a = x_0 < x_1 < \dots < x_n = b$ ، شرایط مرزی در این حالت ایجاب می‌کند که $c_n = S''(x_n)/2 = 0$ و

$$0 = S''(x_0) = 2c_0 + 6d_0(x_0 - x_0);$$

در نتیجه، $c_0 = 0$.

خلاصه برهان. قضیه را می‌توان با استفاده از نظریه ماتریسها، که در فصل ۷ عرضه خواهد شد، ساده‌تر اثبات کرد. معادلات (۲۸.۳)، همراه با $c_0 = c_n = 0$ ، ماتریسی به شکل زیر معین می‌کنند.



که بی‌درنگ (ر.ک. قضیه ۲۱.۷، ص ۴۷۱) قضیه را نتیجه خواهد داد.

جواب واقعی مسئله اسپلاین مکعبی با شرایط مرزی $S''(x_0) = S''(x_n) = 0$ را می‌توان با اعمال الگوریتم زیر بدست آورد.

الگوریتم اسپلاین مکعبی طبیعی ۲.۳

فرض کنیم تابع f در نقاط $x_0 < x_1 < \dots < x_n$ تعریف شده باشد. جهت ساختن درونیاب

اسپلاین مکعبی برای f که در $S''(x_0) = S''(x_n) = 0$ صدق کند .

مرحله ۱. به ازای هر $i = 1, 2, \dots, n-1$ ، قرار می‌دهیم

$$\alpha_i = \frac{3[f(x_{i+1})(x_i - x_{i-1}) - f(x_i)(x_{i+1} - x_{i-1}) + f(x_{i-1})(x_{i+1} - x_i)]}{(x_{i+1} - x_i)(x_i - x_{i-1})}$$

مرحله ۲. قرار می‌دهیم $l_0 = 1, \mu_0 = 0, z_0 = 0$

مرحله ۳. قرار می‌دهیم $i = 1$

مرحله ۴. قرار می‌دهیم

$$l_i = 2(x_{i+1} - x_{i-1}) - (x_i - x_{i-1})\mu_{i-1}$$

$$\mu_i = \frac{1}{l_i}(x_{i+1} - x_i),$$

$$z_i = \frac{1}{l_i}(\alpha_i - (x_i - x_{i-1})z_{i-1}).$$

مرحله ۵. ۱ را به i می‌افزاییم .

مرحله ۶. اگر $i < n$ ، به مرحله ۴ می‌رویم .

اگر $i = n$ ، به مرحله ۷ می‌رویم .

مرحله ۷. قرار می‌دهیم $l_n = 1, z_n = 0, c_n = z_n$

مرحله ۸. قرار می‌دهیم $j = n-1$

مرحله ۹. قرار می‌دهیم

$$c_j = z_j - \mu_j c_{j+1}$$

$$b_j = \frac{(f(x_{j+1}) - f(x_j))}{(x_{j+1} - x_j)} - \frac{(x_{j+1} - x_j)(c_{j+1} + 2c_j)}{3}$$

$$d_j = \frac{(c_{j+1} - c_j)}{3(x_{j+1} - x_j)}$$

مرحله ۱۰. ۱ را از j کم می‌کنیم .

مرحله ۱۱. اگر $j \geq 0$ ، به مرحله ۹ می‌رویم .

اگر $j < 0$ ، به مرحله ۱۲ می‌رویم .

مرحله ۱۲. روند تمام است. درونیای اسپلاین مکعبی طبیعی در هر بازه $[x_j, x_{j+1}]$ ،

به ازای $j = 0, 1, \dots, n-1$ ، عبارت است از

$$S_j(x) = f(x_j) + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3.$$

قضیه ۱۰.۳. هرگاه تابع f بر $[a, b]$ تعریف شده باشد، f درونیاب اسپلاین منحصر بفردی دارد که در شرایط مرزی مقید $S'(a) = f'(a)$ و $S'(b) = f'(b)$ صدق می‌کند.

دراین حالت، با استفاده از اینکه $S'(a) = S'(x_0) = b_0$ ، می‌توان دید که معادله (۲۷.۳)، به‌ازای $j = 0$ ، ایجاب می‌کند که

$$f'(a) = \frac{a_1 - a_0}{h_0} - \frac{h_0}{3}(2c_0 + c_1)$$

و

$$2h_0c_0 + h_0c_1 = \frac{3}{h_0}(a_1 - a_0) - 3f'(a).$$

به‌همین نحو،

$$\begin{aligned} f'(b) &= b_n \\ &= b_{n-1} + h_{n-1}(c_{n-1} + c_n); \end{aligned}$$

درنتیجه، معادله (۲۷.۳)، به‌ازای $j = n - 1$ ، ایجاب می‌کند که

$$\begin{aligned} f'(b) &= \frac{a_n - a_{n-1}}{h_{n-1}} - \frac{h_{n-1}}{3}(2c_{n-1} + c_n) + h_{n-1}(c_{n-1} + c_n) \\ &= \frac{a_n - a_{n-1}}{h_{n-1}} + \frac{h_{n-1}}{3}(c_{n-1} + 2c_n) \end{aligned}$$

و

$$h_{n-1}c_{n-1} + 2h_{n-1}c_n = 3f'(b) - \frac{3}{h_{n-1}}(a_n - a_{n-1}).$$

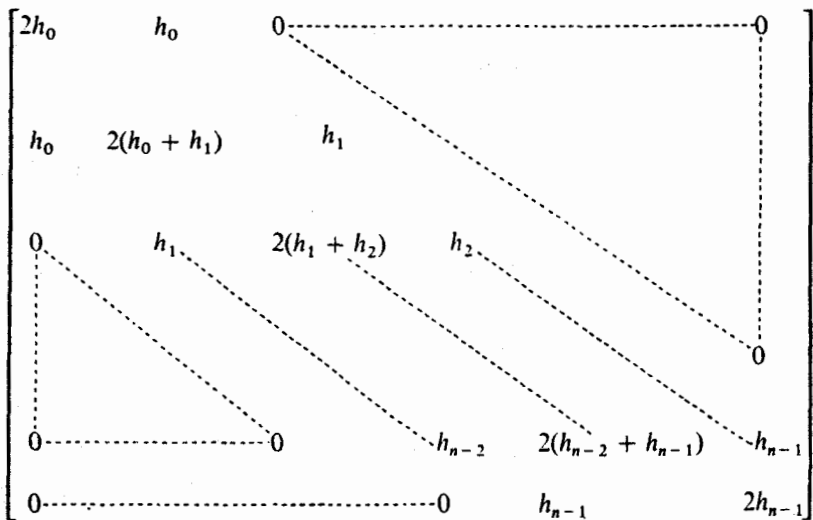
معادله (۲۸.۳) همراه با معادلات

$$2h_0c_0 + h_0c_1 = \frac{3}{h_0}(a_1 - a_0) - 3f'(a)$$

و

$$h_{n-1}c_{n-1} + 2h_{n-1}c_n = 3f'(b) - \frac{3}{h_{n-1}}(a_n - a_{n-1})$$

ماتریسی مشخص می‌کنند که، دراین حالت، به شکل زیر است:



نتایج نظریهٔ ماتریسها مجدداً قضیهٔ ۱۵.۳ را بدست خواهد دارد (ر.ک. قضیهٔ ۲۱.۷، ص ۳۴۴).

الگوریتم اسپلاین مکعبی مقید ۳.۳

تابع f در نقاط $x_0 < x_1 < \dots < x_n$ تعریف شده است. درونیاب اسپلاین مکعبی برای f می‌سازیم که در $S'(x_0) = f'(x_0)$ و $S'(x_n) = f'(x_n)$ صدق کند. مرحلهٔ ۱. قرار می‌دهیم

$$\alpha_0 = 3 \frac{(f(x_1) - f(x_0))}{(x_1 - x_0)} - 3f'(x_0),$$

$$\alpha_n = 3f'(x_n) - 3 \frac{(f(x_n) - f(x_{n-1}))}{(x_n - x_{n-1})},$$

و به‌ازای هر $i = 1, 2, \dots, n-1$

$$\alpha_i = \frac{3[f(x_{i+1})(x_i - x_{i-1}) - f(x_i)(x_{i+1} - x_{i-1}) + f(x_{i-1})(x_{i+1} - x_i)]}{(x_{i+1} - x_i)(x_i - x_{i-1})}$$

مرحلهٔ ۲. قرار می‌دهیم

$$l_0 = 2(x_1 - x_0),$$

$$\mu_0 = \frac{1}{2},$$

$$z_0 = \frac{x_0}{2(x_1 - x_0)},$$

$$b_0 = f'(x_0).$$

مرحله ۳. قرار می‌دهیم $i = 1$.

مرحله ۴. قرار می‌دهیم

$$l_i = 2(x_{i+1} - x_{i-1}) - (x_i - x_{i-1})\mu_{i-1},$$

$$\mu_i = \frac{1}{l_i}(x_{i+1} - x_i),$$

$$z_i = \frac{1}{l_i}[x_i - (x_i - x_{i-1})z_{i-1}].$$

مرحله ۵. به i یک می‌افزاییم.

مرحله ۶. اگر $i < n$ ، به مرحله ۴ می‌رویم.

اگر $i = n$ ، به مرحله ۷ می‌رویم.

مرحله ۷. قرار می‌دهیم

$$l_n = (x_n - x_{n-1})(2 - \mu_{n-1}),$$

$$z_n = \frac{1}{l_n}[x_n - (x_n - x_{n-1})z_{n-1}],$$

$$c_n = z_n.$$

مرحله ۸. قرار می‌دهیم $j = n - 1$.

مرحله ۹. قرار می‌دهیم

$$c_j = z_j - \mu_j c_{j+1}.$$

$$b_j = \frac{(f(x_{j+1}) - f(x_j)) - (x_{j+1} - x_j)(c_{j+1} + 2c_j)}{(x_{j+1} - x_j) \cdot 3},$$

$$d_j = \frac{(c_{j+1} - c_j)}{3(x_{j+1} - x_j)}.$$

مرحله ۱۰. از j یک کم می‌کنیم.

مرحله ۱۱. اگر $j \geq 0$ ، به مرحله ۹ می‌رویم.

اگر $j < 0$ ، به مرحله ۱۲ می‌رویم.

مرحله ۱۲. روند تمام است. درونیاب اسپلاین مکعبی صادق در شرایط مرزی مقید و

بر هر بازه $[x_j, x_{j+1}]$ ، به‌ازای $j = 0, 1, \dots, n - 1$ ، با

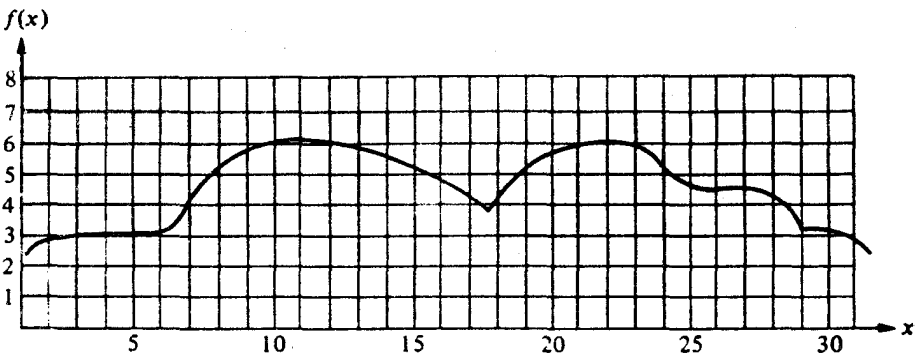
$$S_j(x) = f(x_j) + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3$$

داده می‌شود.

مثال ۰۱. درونیاب اسپلاین مکعبی بدست می‌آوریم که نمایش منحنی حاصل از قسمت بالایی شکل شناخته شده ۱۱.۳ باشد.

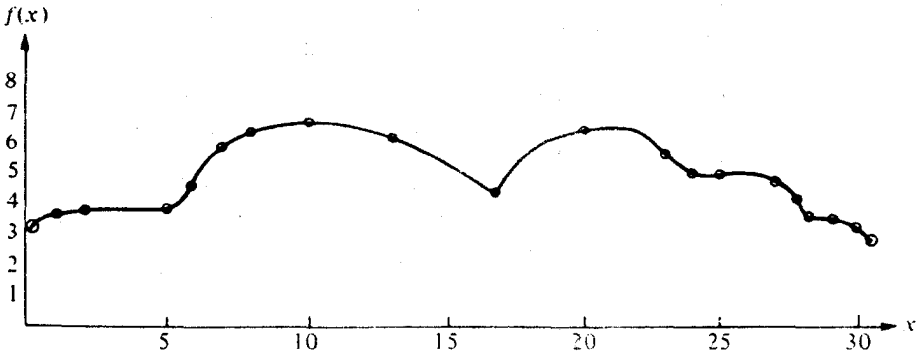


شکل ۱۱.۳



شکل ۱۲.۳

ابتدا یک طرح شطرنجی، مانند شکل ۱۲.۳ لازم است تا نقاط روی منحنی را مشخص کند. چون منحنی در نقاط تماس پنجه با سر و نقاط تماس سر با پشت مشتق پیوسته ندارد.



شکل ۱۳.۳

باید برای هر یک از این نواحی اسپلاینهای جداگانه‌ای بکار برد. با استفاده از شکل ۱۲.۲، داده‌های جدول ۱۲.۳ بدست می‌آیند، که در آنها مقادیر x در جاهایی که تغییرات ناگهانی صورت می‌گیرد فشرده‌تر اختیار شده‌اند.

#1 منحنی			#2 منحنی			#3 منحنی		
i	x_i	$f(x_i)$	i	x_i	$f(x_i)$	i	x_i	$f(x_i)$
0	1	3.0	0	17	4.5	0	27.7	4.1
1	2	3.7	1	20	7.0	1	28	4.3
2	5	3.9	2	23	6.1	2	29	4.1
3	6	4.2	3	24	5.6	3	30	3.0
4	7	5.7	4	25	5.8			
5	8	6.6	5	27	5.2			
6	10	7.1	6	27.7	4.1			
7	13	6.7						
8	17	4.5						

جدول ۱۲.۳

با استفاده از الگوریتم ۲.۳ در ساختن اسپلاینهای طبیعی نظیر این داده‌ها، جداول ۱۳.۳، ۱۴.۳ و ۱۵.۳ بدست می‌آیند.

#1 اسپلاین					
j	x_j	$a_j = f(x_j)$	b_j	c_j	d_j
0	1	3.0	.786	0	-.086
1	2	3.7	.529	-.257	.034
2	5	3.9	-.086	.052	.334
3	6	4.2	1.019	1.053	-.572
4	7	5.7	1.408	-.664	.156
5	8	6.6	.547	-.197	.024
6	10	7.1	.049	-.052	-.003
7	13	6.7	-.342	-.078	.007
8	17	4.5			

جدول ۱۳.۳

#2 اسپلاین					
j	x_j	$a_j = f(x_j)$	b_j	c_j	d_j
0	17	4.5	1.106	0	-.030
1	20	7.0	.289	-.272	.025
2	23	6.1	-.660	-.044	.204
3	24	5.6	-.137	.567	-.230
4	25	5.8	.306	-.124	-.089
5	27	5.2	-1.263	-.660	.314
6	27.7	4.1			

جدول ۱۴.۳

#3 اسپلاین					
j	x_j	$a_j = f(x_j)$	b_j	c_j	d_j
0	27.7	4.1	.749	0	-.910
1	28	4.3	.503	-.819	.116
2	29	4.1	-.787	-.470	.157
3	30	3.0			

جدول ۱۵.۳

نمودار این تابع قطعه‌قطعه اسپلاین در شکل ۱۴.۳ نموده شده است.

اسپلاین آزاد —

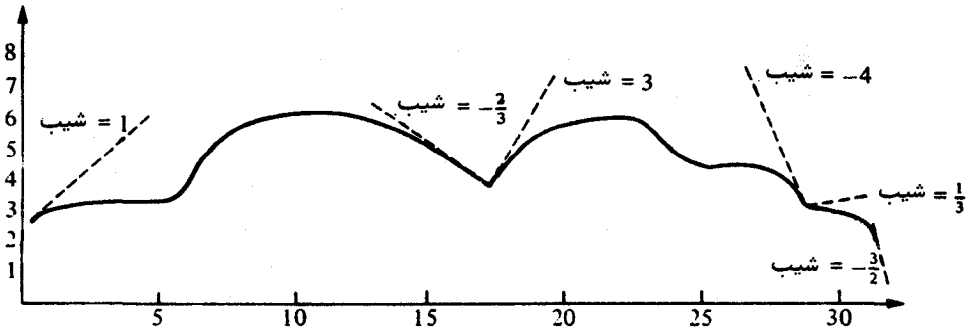
منحنی اصلی - - - -



شکل ۱۴.۳

اگر مشتق در هر نقطه انتهایی سه قطعه هموار از منحنی را تقریب کنیم، می‌توان

از سه شرط مرزی مقید برای تقریب منحنی استفاده کرد. ر.ک. شکل ۱۵.۳.



شکل ۱۵.۳

داده‌های ما در این وضع همان داده‌های مربوط به اسپلاین‌های طبیعی، مفروضات اضافی بر مشتقات در نقاط انتهایی، است و در جدول ۱۶.۳ داده شده‌اند.

منحنی #1				منحنی #2				منحنی #3			
i	x_i	$f(x_i)$	$f'(x_i)$	i	x_i	$f(x_i)$	$f'(x_i)$	i	x_i	$f(x_i)$	$f'(x_i)$
0	1	3.0	1	0	17	4.5	3	0	27.7	4.1	.33
1	2	3.7		1	20	7.0		1	28	4.3	
2	5	3.9		2	23	6.1		2	29	4.1	
3	6	4.2		3	24	5.6		3	30	3.0	-1.5
4	7	5.7		4	25	5.8					
5	8	6.6		5	27	5.2					
6	10	7.1		6	27.7	4.1	-4				
7	13	6.7									
8	17	4.5	-0.67								

جدول ۱۶.۳

#1 اسپلاین						
j	x_j	$a_j = f(x_j)$	b_j	c_j	d_j	$f'(x_j)$
0	1	3.0	1	-.347	.049	1
1	2	3.7	.447	-.206	.027	
2	5	3.9	-.074	.033	.342	
3	6	4.2	1.016	1.058	-.575	
4	7	5.7	1.409	-.665	.156	
5	8	6.6	.547	-.196	.024	
6	10	7.1	.048	-.053	-.003	
7	13	6.7	-.339	-.076	.006	
8	17	4.5				-0.67

جدول ۱۷.۳

با استفاده از الگوریتم ۳.۳ در ساختن اسپلاین با شرط مرزی مقید نظیر این داده‌ها، نتایج جدولهای ۱۷.۳، ۱۸.۳، و ۱۹.۳ بدست می‌آیند.

#۲ اسپلاین

j	x_j	$a_j = f(x_j)$	b_j	c_j	d_j	$f'(x_j)$
0	17	4.5	3	-1.101	.126	3
1	20	7.0	-.198	.035	-.023	
2	23	6.1	-.609	-.172	.280	
3	24	5.6	-.111	.669	-.357	
4	25	5.8	.154	-.403	.088	
5	27	5.2	-.401	.126	-2.568	
6	27.7	4.1				-4

جدول ۱۸.۳

#۳ اسپلاین

j	x_j	$a_j = f(x_j)$	b_j	c_j	d_j	$f'(x_j)$
0	27.7	4.1	.330	2.262	-3.800	.33
1	28	4.3	.661	-1.157	.296	
2	29	4.1	-.765	-.269	-.065	
3	30	3.0				-1.5

جدول ۱۹.۳

نمودار این تابع قطعه قطعه اسپلاین در شکل ۱۶.۳ نموده شده است.

— اسپلاین مقید
 ---- منحنی اصلی



شکل ۱۶.۳

دلیل اینکه اسپلاینهای طبیعی منحنی اصلی را بهتر از اسپلاینها با شرایط مرزی مقید تقریب می‌کند آن است که مشتقات در نقاط انتهایی بخشهای منحنی بدقت تخمین زده نشده‌اند. در این مثال، با این فرض که مشتق دوم در هر انتها صفر است، اسپلایین طبیعی تخمین بهتری برای این مشتقات در نظر گرفته است تا آنهایی که برای اسپلاینها با شرط مرزی مقید داده شده بودند.

عموماً در تقریب توابع به وسیله اسپلاینها مکعبی، استفاده از شرایط مرزی مقیدار ج است، در نتیجه، اغلب تخمین مشتق تابع در نقاط انتهایی بازه لازم می‌شود. در حالتی که گرهمها در نزدیکی هر دو نقطه انتهایی بازه متساوی الفاصله باشند، تقریبها را می‌توان با استفاده از معادلات (۲۲.۵) و (۲۶.۵)، یا هر فرمول مناسب دیگر از بخش ۱۰.۵، بدست آورد. در حالتی که گرهمها متساوی الفاصله نیستند، مسئله به مراتب مشکلتر خواهد بود.

برای اتمام این بخش، یک فرمول کران خطا برای حالت شرایط مرزی مقید ذکر می‌کنیم، که برهان آن را می‌توان در کتاب شولتز^۱ [۵۶]، صفحات ۵۷ تا ۵۸، یافت.

قضیه ۱۱.۳. فرض کنیم $f \in C^4[a, b]$ با این خاصیت که $\max_{a \leq x \leq b} |f^{(4)}(x)| \leq M$. هرگاه S درونیاب اسپلایین مکعبی منحصر بفردی به f نسبت به نقاط $a = x_0 < x_1 < \dots < x_n = b$ باشد که در $S'(a) = f'(a)$ و $S'(b) = f'(b)$ صدق می‌کند، آنگاه

$$(۲۹.۳) \quad \max_{a \leq x \leq b} |f(x) - S(x)| \leq \frac{5M}{384} \max_{0 \leq j \leq n-1} (x_{j+1} - x_j)^4.$$

مجموعه تمرینات ۶.۳

۱. با استفاده از اسپلایین مکعبی، به ازای تابع f حاصل از داده‌های تمرین ۱ در بخش ۵.۳ و فرض $f'(x_0) = f'(x_n) = 0$ ، $f(2.5)$ را تقریب کنید. مسئله را یکبار با شرایط مرزی مقید و یکبار با شرایط مرزی آزاد حل کنید و نتایج را با هم مقایسه نمایید.

۲. با سکار بردن اسپلاین مکعبی با شرایط مرزی مقید، تابع $f(x) = 3xe^x - e^{2x}$ را در $x = 1.03$ ، با استفاده از داده‌های زیر، تقریب کنید:

x	1.0	1.02	1.04	1.06
$f(x)$.76578939	.79536678	.82268817	.84752226

خطا را، با استفاده از معادله (۲۹.۳) ، تخمین بزنید، و آن را با خطای واقعی مقایسه کنید.

۳. (آ) به‌ازای افزایش $x_0 = 0, x_1 = .05, x_2 = .1$ از $[0, .1]$ ، اسپلاین مکعبی s را بیابید که درونیاب $f(x) = e^{2x}$ با شرایط مرزی مقید باشد.
 (-) با محاسبه $\int_0^1 s(x) dx$ ، تقریبی برای $\int_0^1 e^{2x} dx$ بیابید.
 (پ) با استفاده از قضیه ۱۱.۳ ، $\max_{0 \leq x \leq 1} |f(x) - s(x)|$ و

$$\left| \int_0^1 f(x) dx - \int_0^1 s(x) dx \right|$$

را تخمین بزنید.

(ت) اسپلاین مکعبی S را با شرایط مرزی آزاد معین کرده و $S(.02)$ ، $s(.02)$ ، و $e^{.04} = 1.04081077$ را مقایسه نمایید.

۴. به‌ازای افزایش $x_0 = 0, x_1 = .05, x_2 = .1$ از $[0, .1]$ ، تابع درونیاب قطعه قطعه خطی F را برای $f(x) = e^{2x}$ بیابید. $\int_0^1 e^{2x} dx$ را با $\int_0^1 F(x) dx$ تقریب کنید. نتایج را با نتایج تمرین ۳ مقایسه نمایید.

۵. فرض کنید داده‌های زیر برای تابع f مفروض باشند:

$$f(0) = 0, f(1) = -.1, f(2) = -.3, f(3) = -.5, f(4) = -.4,$$

$$f(5) = -.2, f(6) = 0, f(7) = .2, f(8) = .3, f(9) = .7, f(1) = .9.$$

اسپلاینی را بیابید که f را با شرایط مرزی آزاد نسبت به افزایش $x_j = .1$ به‌ازای $z = 0, 1, \dots, 10$ درونیایی کند.

۶. فرض کنید $f \in C^2[a, b]$ و گرهای $a = x_0 < x_1 < \dots < x_n = b$ داده شده باشند. برای تابع درونیاب قطعه قطعه خطی F خطای تخمینی شبیه به خطای تخمین قضیه ۱۱.۳ بیابید. با استفاده از این تخمین، کرانهای خطا در تمرین ۴ را نتیجه بگیرید.

۷. فرض کنید f بر $[a, b]$ تعریف شده باشد و گرهای $a = x_0 < x_1 < x_2 = b$ داده

شده باشند. یک تابع درونیاب اسپلاین درجه S دو مرکب است از چندجمله‌ای درجه دو

$$S_0(x) = a_0 + b_0(x - x_0) + c_0(x - x_0)^2 \quad \text{بر } [x_0, x_1]$$

و چندجمله‌ای درجه دو

$$S_1(x) = a_1 + b_1(x - x_1) + c_1(x - x_1)^2 \quad \text{بر } [x_1, x_2]$$

بطوری که

$$S(x_0) = f(x_0), \quad S(x_1) = f(x_1), \quad S(x_2) = f(x_2) \quad (\text{یک})$$

$$S \in C^1[x_0, x_2] \quad (\text{دو})$$

نشان دهید که شرایط (یک) و (دو) به پنج معادله شامل شش مجهول $a_0, b_0, c_0, a_1, b_1, c_1$ منجر می‌شوند. مسئله تعیین شرط اضافی است که جواب را منحصر بفرد کند، مثلاً،

$$f'(x_0) = S'(x_0) \quad \text{یا} \quad f'(x_2) = S'(x_2)$$

آیا شرط

$$S \in C^2[x_0, x_2]$$

به جوابی با معنی منجر می‌شود؟

۸. همانطور که در مقاله کامرر^۱، ردین^۲، و وارگا^۳ [۴۱] پیشنهاد شده است، می‌توان اسپلاینهای درونیاب درجه دو مفیدی را ساخت. اگر f تابعی بر $[a, b]$ بوده و $a = x_0 < x_1 < \dots < x_n = b$ اسپلاین درجه دو در شرایط زیر صدق کند:

$$s(x_0) = f(x_0) \quad (\text{یک})$$

$$s\left(\frac{x_i + x_{i+1}}{2}\right) = f\left(\frac{x_i + x_{i+1}}{2}\right), \quad i = 0, 1, \dots, n-1$$

$$s(x_n) = f(x_n):$$

$$s \in C^1[a, b] \quad (\text{دو})$$

$$i = 0, 1, \dots, n-1 \quad \text{به‌ازای هر } (سه)$$

$$s(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2, \quad x_i \leq x \leq x_{i+1}$$

(آ) نشان دهید که شرایط (یک)، (دو)، و (سه) به معادلات زیر منجر می‌شوند.

$$a_0 = f(x_0);$$

به ازای هر $i = 0, 1, \dots, n-1$ هر $a_i + \frac{1}{2}h_i b_i + \frac{1}{6}h_i^2 c_i = f(x_i + \frac{1}{2}h_i)$ ،

$$a_{n-1} + h_{n-1} b_{n-1} + h_{n-1}^2 c_{n-1} = f(x_n);$$

به ازای هر $i = 1, 2, \dots, n-1$ هر $a_i = a_{i-1} + b_{i-1} h_{i-1} + c_{i-1} h_{i-1}^2$ ،

به ازای هر $i = 1, 2, \dots, n-1$ هر $b_i = b_{i-1} + 2c_{i-1} h_{i-1}$ ،

(ب) نشان دهید که نتایج قسمت (آ) به معادلات زیر منجر می‌شوند.

$$\frac{3}{8}h_0 b_0 + \frac{1}{8}h_0 b_1 = f\left(x_0 + \frac{h_0}{2}\right) - f(x_0),$$

به ازای هر $i = 1, 2, \dots, n-2$ ،

$$\frac{1}{8}h_{i-1} b_{i-1} + \frac{3}{8}(h_i + h_{i-1})b_i + \frac{1}{8}h_i b_{i+1} = f\left(x_i + \frac{h_i}{2}\right) - f\left(x_{i-1} + \frac{h_{i-1}}{2}\right),$$

$$\frac{1}{8}h_{n-2} h_{n-2} + (\frac{3}{8}h_{n-2} + \frac{1}{8}h_{n-1})b_{n-1}$$

$$= (\frac{3}{8})f\left(x_{n-1} + \frac{h_{n-1}}{2}\right) - f\left(x_{n-2} + \frac{h_{n-2}}{2}\right) - \frac{1}{8}f(x_n).$$

که در آن، به ازای هر $i = 0, 1, \dots, n-1$ هر $h_i = x_{i+1} - x_i$ ،

(پ) برای یافتن b_0, \dots, b_{n-1} ، الگوریتمی شبیه الگوریتمهای ۲.۳ و ۳.۳ طرح کنید.

(ت) به فرض معلوم بودن b_0, \dots, b_{n-1} ، فرمولهایی برای $a_0, \dots, a_{n-1}, c_0, \dots, c_{n-1}$ طرح نمایید.

(ث) فرض کنید $f(x) = e^x$ ، $x_0 = 0$ ، $x_1 = .2$ ، $x_2 = .6$ ، $x_3 = .9$. اسپلین

درونیاب درجه دو s را برای f پیدا کرده و $s(5)$ را محاسبه نمایید. آیا $s(5)$

تقریب خوبی برای $f(5)$ است؟

(ج) تمرین ۲ را، با استفاده از اسپلین درونیاب درجه دو، تکرار کنید، و

نتایج را با نتایج تمرین ۲ مقایسه نمایید. آیا قابل قبول است که

$$|f(x) - s(x)| = O(h^k), \quad k = 2 \text{ یا } k = 4$$

جزئیات زیر را مربوط به تعیین فرمول دیگری برای اسپلینهای مکعبی کامل کنید.

به ازای افزایش

$$a = x_0 < x_1 < \dots < x_n = b, \quad [a, b]$$

اسپلین مکعبی s ، که f را درونیابی می‌کند، یک چندجمله‌ای مکعبی $s(x)$ بر

هر $[x_j, x_{j+1}]$ است. لذا، به ازای $j = 0, 1, \dots, n-1$ می توان نوشت

$$s''(x) = s_j''(x) = \frac{a_j(x_{j+1} - x)}{h} + \frac{a_{j+1}(x - x_j)}{h};$$

یعنی، $s_j''(x)$ بر هر زیربازه خطی است. نشان دهید که s'' بر $[a, b]$ پیوسته است. با استفاده از نمادهای $s_j = s(x_j)$ و $f_j = f(x_j)$ به ازای $j = 0, 1, \dots, n-1$ ، از $s_j''(x)$ دوبار انتگرال گرفته و، با استفاده از $s_j = f_j$ و $s_{j+1} = f_{j+1}$ ، با محاسبه ثابتهای انتگرالگیری

$$s_j(x) = a_j \frac{(x_{j+1} - x)^3}{6h_j} + \frac{a_{j+1}}{6h_j} (x - x_j)^3 + \left(f_{j+1} - \frac{a_{j+1}h_j^2}{6} \right) \frac{(x - x_j)}{h_j} + \left(f_j - \frac{a_j h_j^2}{6} \right) \frac{(x_{j+1} - x)}{h_j}$$

حال از این شرط که s' بر $[a, b]$ پیوسته است، یک دستگاه معادلات برای a_j ها بدست آورید.

۱۰. تعریف کنید

$$(x - \xi)_+^2 = \begin{cases} (x - \xi)^2, & x > \xi, \\ 0, & x \leq \xi. \end{cases}$$

فرض کنید $a = x_0 < x_1 < x_2 < x_3 = b$ و تعریف کنید

$$S(x) = c_1(x - x_1)_+^2 + c_2(x - x_2)_+^2, \quad x \in [a, b].$$

نشان دهید که $S \in C^2[a, b]$.

۱۱. فرض کنید $P(x) = a_0 + a_1x + a_2x^2 + a_3x^3$ یک چندجمله‌ای مکعبی بر $[a, b]$ بوده و

$a = x_0 < x_1 < \dots < x_n = b$ گرههای معلومی باشند. نشان دهید که

$$S(x) = P(x) + \sum_{j=1}^{n-1} c_j(x - x_j)_+^3$$

یک اسپلاین مکعبی است، S یک چندجمله‌ای مکعبی بر $[x_i, x_{i+1}]$ به ازای هر

$i = 0, 1, \dots, n-1$ و $S \in C^2[a, b]$.

۱۲. فرض کنید P یک چندجمله‌ای از درجه حداکثر سه بوده، $\{x_j\}_{j=0}^n$ افزایی از $[a, b]$

باشد، و

$$s(x) = P(x) + \sum_{j=1}^{n-1} c_j(x - x_j)_+^3.$$

نشان دهید که مشتق سوم s بر $[a, b]$ موجود و پیوسته است اگر و فقط اگر به‌ازای

$$c_j = 0, \quad j = 1, 2, \dots, n$$

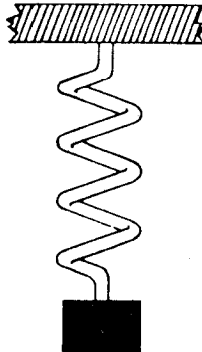
۱۳. جدول زیر، که جمعیت ایالات متحده را از سال ۱۹۲۰ تا ۱۹۷۰ نشان می‌دهد، در مقدمه^۴ این فصل و در تمرین ۱۳ از بخش ۲.۳ و تمرین ۸ از بخش ۴.۳ مورد بحث قرار گرفت.

سال	1920	1930	1940	1950	1960	1970
جمعیت	105,711	123,203	131,669	150,697	197,323	203,212

اسپلین مکعبی آزادی را بیابید که با این داده‌ها سازگار باشد و، با استفاده از آن، جمعیت در سالهای ۱۹۱۰، ۱۹۶۵، و ۲۰۰۰ را پیشگویی کنید. تقریبهای خود را با تقریبهای حاصل در تمرین ۱۳ از بخش ۲.۳ مقایسه نمایید. اگر اختیار با شما باشد، چه روند درونیایی را انتخاب می‌کنید؟

نظریه تقریب^۱

قانون هوک بیان می‌دارد که وقتی بر یک فنر ساخته شده از ماده‌ای یکنواخت نیرو وارد شود، طول فنر تابعی خطی از نیروی وارده است. این تابع خطی را می‌توان به صورت $F(l) = k \cdot l + E$ نوشت، که در آن $F(l)$ نیروی لازم برای کشیدن فنر به قدر l واحد است، ثابت E طول فنر است بی آنکه نیرویی بر آن وارد شود، و ثابت k ثابت فنر نام دارد.



فرض کنید بخواهیم ثابت فنر را برای فنری به طول اولیه^۲ 5.3 اینچ معین کنیم. نیروهای 2، 4، و 6 پوند را متوالیاً^۳ بر فنر وارد کرده و می‌بینیم طول آن بترتیب 7.0، 9.4، و 12.3 اینچ افزایش می‌یابد. یک امتحان سریع نشان می‌دهد که نقاط (0, 5.3)، (2, 7.0)، (4, 9.4)، و (6, 12.3) کاملاً^۴ در خط مستقیم

قرار ندارند. با آنکه فقط می‌شد یک جفت تصادفی از این نقاط را برای تقریب ثابت فنر بکار برد، ظاهراً "معقولتر است که، برای تعیین ثابت، خطی را بیابیم که همه نقاط داده شده را به بهترین وجه تقریب نماید. این نوع تقریب تقریب کمترین مربعات نام دارد و در این فصل مطرح خواهد شد.

بررسی نظریهٔ تقریب از دو نوع مسئلهٔ کلی تشکیل می‌شود. مسئلهٔ اول وقتی پیش می‌آید که تابع داده شده صریح است ولی می‌خواهیم نوع "ساده‌تری" از توابع، نظیر چند جمله‌ای، بیابیم که بتوان از آن برای تقریب مقادیر تابع داده شده استفاده کرد. مسئلهٔ دیگر نظریهٔ تقریب مربوط می‌شود به برازش توابع به داده‌های مفروض و یافتن "بهترین" تابع در رده‌ای معین که بتوان از آن برای نمایش داده‌ها استفاده کرد.

در فصل گذشته هر دو مسئله مطرح شدند. چند جمله‌ای تیلور درجه n حول نقطهٔ x_0 به عنوان یک تقریب عالی به تابع $(n+1)$ بار مشتق‌پذیر f در همسایگی کوچکی از x_0 مورد بحث قرار گرفت. چند جمله‌ایهای درونیاب لاگرانژ، یا کلیتر، چند جمله‌ایهای بوسان، هم به عنوان چند جمله‌ایهای تقریب ساز و هم چند جمله‌ایهایی که در داده‌های مشخصی صدق می‌کنند مورد بحث قرار گرفتند. اسپلاینهای مکعبی نیز در آن فصل مطرح شدند. در این فصل، به محدودیت‌های این روشها اشاره کرده و راههای دیگر دستیابی به آنها مورد بحث قرار می‌گیرند.

۱.۴ آشنایی با تقریب و نرمها

یکی از اولین نکاتی که در تقریب یک تابع معلوم با برازش داده‌های مفروض باید در نظر گرفت نوع تابع P است که باید بکار رود. P ، عموماً، ترکیبی خطی از یک مجموعه از توابع است که اعضایش دارای خواص معینی می‌باشند.

در این فصل، فقط مجموعه‌های متناهی $\{\phi_i\}_{i=1}^n$ از توابع را در نظر خواهیم گرفت،

و چنانچه P ترکیبی خطی از این مجموعه باشد، لازم است

$$P(x) = \sum_{i=1}^n c_i \phi_i(x),$$

که در آن c_i ها، $i = 1, 2, \dots, n$ ، ثابت‌هایی حقیقی هستند. مثلاً، ϕ_k می‌تواند x^{k-1} بوده و، در این صورت، تابع P یک چند جمله‌ای خواهد بود.

همینکه مجموعه $\{\phi_i\}_{i=1}^n$ اختیار شد، مسئله انتخاب اعداد حقیقی c_i ، $i = 1, 2, \dots, n$ ، است بطوری که P به قدر کافی نزدیک f بوده یا P به قدر کافی داده‌های $\{(x_i, y_i) | i = 1, 2, \dots, M\}$ را برازش کند. برای سنجش نزدیکی یا کفایت تقریب P ، از نرم تابع استفاده می‌شود. نرم‌های بکار رفته در این فصل در تعریف زیر داده شده‌اند. تعریف کلی نرم تابع در تعریف ۱ آورده شده است.

تعریف ۱.۴. به‌ازای تابع معلوم $f \in C[a, b]$ ، نرم‌های L_2 و L_∞ بترتیب با

$$(1.4) \quad \|f\|_2 = \left\{ \int_a^b [f(x)]^2 dx \right\}^{1/2}$$

$$(2.4) \quad \|f\|_\infty = \max_{x \in [a, b]} |f(x)|.$$

تعریف می‌شوند.

مثال ۱. فرض کنیم به‌ازای $f(x) = x^3 - 3x^2 - 9x + 7$ ، $x \in [-1, 5]$ ، برای نرم L_2 ، ابتدا محاسبه می‌کنیم

$$\begin{aligned} \|f\|_2^2 &= \int_{-1}^5 (x^3 - 3x^2 - 9x + 7)^2 dx \\ &= \int_{-1}^5 (x^6 - 6x^5 - 9x^4 + 68x^3 + 39x^2 - 126x + 49) dx \\ &= \left. \frac{x^7}{7} - x^6 - \frac{9}{5}x^5 + 17x^4 + 13x^3 - 63x^2 + 49x \right|_{-1}^5 \\ &= 830.7142815 - (-107.3428571) \\ &= 938.0571386, \end{aligned}$$

بنا بر این،

$$\|f\|_2 = 30.6277.$$

برای یافتن ماکزیمم مطلق $|f(x)|$ بر $[-1, 5]$ ، اولین کار یافتن نقاط بحرانی f

است. چون

$$f'(x) = 3x^2 - 6x - 9 = 3(x^2 - 2x - 3) = 3(x-3)(x+1),$$

نقاط بحرانی $x = 3$ و $x = -1$ بوده، و ماکزیمم $|x^3 - 3x^2 - 9x + 7|$ را در $x = -1$ ،

$x = 3$ ، یا $x = 5$ خواهیم داشت. چون

$$|f(-1)| = |12| = 12, \quad |f(3)| = |-20| = 20,$$

$$|f(5)| = |12| = 12,$$

نتیجه می‌شود که

$$\|f\|_\infty = 20.$$

مثال ۲. فرض کنیم f و g بر $[a, b]$ پیوسته باشند. طبق قضیه مقدار اکستریم، قضیه

۹.۱، صفحه ۶، هر دو تابع f و g بر $[a, b]$ ماکزیمم مطلق و مینیمم مطلق دارند.

در نتیجه، هردوی $|f(x)|$ و $|g(x)|$ ماکزیمم مطلق داشته و $\|f\|_\infty$ و $\|g\|_\infty$ تعریف شده‌اند.

نشان می‌دهیم که سه خاصیت زیر برقرارند:

$$(۳.۴) \quad \|f\|_\infty \geq 0 \quad \text{و} \quad \|f\|_\infty = 0 \quad \text{فقط اگر } f = 0 \text{ بر } [a, b];$$

$$(۴.۴) \quad \|\alpha f\|_\infty = |\alpha| \|f\|_\infty, \quad \alpha \text{ هر عدد حقیقی} ;$$

$$(۵.۴) \quad \|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty.$$

چون $\|f\|_\infty = \max |f(x)|$ ، و، به‌ازای هر x ، $|f(x)| \geq 0$ ، واضح است که

$\|f\|_\infty \geq 0$. هرگاه $\|f\|_\infty = 0$ ، آنگاه، به‌ازای هر x در $[a, b]$ ، $|f(x)| = 0$ ؛ در نتیجه،

به‌ازای هر $x \in [a, b]$ ، $f(x) = 0$. این (۳.۴) را ثابت می‌کند.

برای اثبات (۴.۴)، توجه می‌کنیم که

$$\begin{aligned} \|\alpha f\|_\infty &= \max_{x \in [a, b]} |\alpha f(x)| \\ &= \max_{x \in [a, b]} |\alpha| |f(x)| \\ &= |\alpha| \max_{x \in [a, b]} |f(x)| \\ &= |\alpha| \|f\|_\infty. \end{aligned}$$

بهمین نحو، می توان برقراری (۵.۴) را با ملاحظه اینکه

$$\begin{aligned} \|f + g\|_{\infty} &= \max_{x \in [a, b]} |f(x) + g(x)| \\ &\leq \max_{x \in [a, b]} [|f(x)| + |g(x)|] \\ &\leq \max_{x \in [a, b]} |f(x)| + \max_{x \in [a, b]} |g(x)| \\ &= \|f\|_{\infty} + \|g\|_{\infty}. \end{aligned}$$

ثابت کرد.

تمرین ۳ نشان می دهد که می توان خواص مشابهی را برای نرم (۱.۴) اثبات کرد.

با مراجعه به بحث کلی، مسئله تقریب تابع معلوم f به وسیله ترکیبی خطی از توابع $\{\phi_i\}_{i=1}^n$ را می توان به صورت یافتن ثابتهای c_i ، به ازای $i = 1, 2, \dots, n$ ، که

$$\left\| f - \sum_{i=1}^n c_i \phi_i \right\|$$

را مینیمم کنند تجسم کرد.

اگر این نرم L_2 باشد، تقریب $p(x) = \sum_{i=1}^n c_i \phi_i(x)$ را عموماً "تقریب کمترین مربعات می نامند. نام "بهترین" تقریب اغلب وقتی نرم L_{∞} بکار می رود بدان اطلاق می شود، ولی ما در کارهایمان از این نام استفاده نخواهیم کرد.

مجموعه تمرینات ۱۰۴

۱. به ازای مجموعه S از توابع، نرم $\|\cdot\|$ نگاشتی است که به هر عنصر S عدد حقیقی

نامنفی نسبت می دهد که از خواص زیر برخوردار است:

(یک) $\|f\| = 0$ اگر و فقط اگر $f \equiv 0$;

(دو) به ازای هر عدد حقیقی x و هر $f \in S$ ، $\|xf\| = |x| \|f\|$;

(سه) به ازای هر $f, g \in S$ ، $\|f + g\| \leq \|f\| + \|g\|$;

فرض کنید $S = C[a, b]$ و

به ازای هر $f \in C[a, b]$ ، تعریف کنید $\|f\|_1 = \int_a^b |f(x)| dx$

نشان دهید که $\| \cdot \|_1$ یک نرم بر $C[a, b]$ است. این نرم را نرم L_1 بر $C[a, b]$ می‌نامند.

۲. نرمهای L_1 ، L_2 ، و L_∞ هر یک از توابع زیر را بر بازه $[0, 3]$ بیابید:
 (ت) $f(x) = e^x$ ؛ (ب) $f(x) = x^2$ ؛ (پ) $f(x) = 3x + 1$ ؛ (ث) $f(x) = \sin x$.
۳. فرض کنید $f, g \in C[a, b]$. نامساوی کشی^۱ - بونیاکوفسکی^۲ - شوارتز^۳ برای انتگرالها (که اغلب فقط نامساوی شوارتز گفته می‌شود) را اثبات کنید:

$$\int_a^b |f(x)g(x)| dx \leq \left\{ \int_a^b |f(x)|^2 dx \right\}^{1/2} \left\{ \int_a^b |g(x)|^2 dx \right\}^{1/2}$$

[راهنمایی. به‌ازای هر λ حقیقی، $0 \leq \int_a^b \{|f(x)| + \lambda|g(x)|\}^2 dx$ ، آن را به عنوان یک معادله درجه دوم از λ به شکل $A\lambda^2 + B\lambda + C$ بسط داده و توجه کنید که مبین آن، یعنی $B^2 - 4AC$ ، به این دلیل که معادله تغییر علامت نمی‌دهد، باید نامثبت باشد.]

۴. نشان دهید که نرم L_2 یک نرم بر $C[a, b]$ است. [راهنمایی. برای اثبات $\|f + g\|_2 \leq \|f\|_2 + \|g\|_2$ ، را بسط داده و از نامساوی کشی - بونیاکوفسکی - شوارتز استفاده کنید.]

۵. $f \in C[0, 1]$ را در نظر بگیرید، و یک نامساوی بین $\|f\|_1$ ، $\|f\|_2$ ، و $\|f\|_\infty$ بدست آورید.

۶. (ت) تمرین ۵ را با $C[0, 2]$ به جای $C[0, 1]$ تکرار کنید.
 (ب) تمرین ۵ را با $C[0, .5]$ به جای $C[0, 1]$ تکرار کنید.
۷. نشان دهید که به‌ازای هر دو تابع f و g ی تعریف شده در x_0, \dots, x_n ،

$$\sum_{i=0}^n |f(x_i)g(x_i)| \leq \left(\sum_{i=0}^n |f(x_i)|^2 \right)^{1/2} \left(\sum_{i=0}^n |g(x_i)|^2 \right)^{1/2}$$

این نامساوی اغلب نامساوی کشی - بونیاکوفسکی - شوارتز برای مجموعه‌ها یا فقط نامساوی کشی نامیده می‌شود. به تشابه آن با نامساوی تمرین ۳ توجه کنید.

۸. اغلب مفید است توابعی را در نظر بگیریم که فقط در مجموعه‌ای متناهی از نقاط مانند $\{x_0, \dots, x_n\}$ تعریف شده‌اند و می‌توان مفهوم نرم را با توجه به این توابع

تغییر داد. فرض کنید S مجموعه‌ای از توابع باشد که بر نقاط x_0, \dots, x_n تعریف شده‌اند. نیم نرم گسسته $\|\cdot\|$ نگاهی است که به هر عنصر S عدد حقیقی نامنفی نسبت می‌دهد بطوری که از خواص زیر برخوردار است:

(یک) به ازای هر عدد حقیقی α و هر $f \in S$ ، $\|\alpha f\| = |\alpha| \|f\|$ ؛

(دو) به ازای هر $f, g \in S$ ، $\|f + g\| \leq \|f\| + \|g\|$.

(اصطلاح "نیم نرم" بدان جهت بکار می‌رود که این مفهوم همه خواص یک نرم را دارد جز آنکه ممکن است توابعی چون f که متحد صفر نیستند وجود داشته باشند که برای آنها $\|f\| = 0$.)

(ت) نشان دهید که $\|\cdot\|_{r_2}$ تعریف شده با

$$\|f\|_{r_2} = \left(\sum_{i=0}^n |f(x_i)|^2 \right)^{1/2}$$

یک نیم نرم گسسته است. [راهنمایی. از تمرین ۷ استفاده کنید.]

(ب) نشان دهید که $\|\cdot\|_{r_2}$ تعریف شده با

$$\|f\|_{r_1} = \sum_{i=0}^n |f(x_i)|$$

یک نیم نرم گسسته است.

(پ) نشان دهید که $\|\cdot\|_{r_\infty}$ تعریف شده با

$$\|f\|_{r_\infty} = \max_{0 \leq i \leq n} |f(x_i)|$$

یک نیم نرم گسسته است.

۹. $\|f\|_{r_1}$ ، $\|f\|_{r_2}$ ، $\|f\|_{r_\infty}$ را با استفاده از $x_0 = 0$ ، $x_1 = 1$ ، $x_2 = 2$ ، و $x_3 = 3$ ،

برای هر یک از توابع زیر حساب کنید:

(ت) $f(x) = e^x$ ؛ (ب) $f(x) = x^2$ ؛ (پ) $f(x) = 3x + 1$ ؛ (د) $f(x) = \sin x$ ؛

(ث) $f(x) = x^4 - 6x^3 + 11x^2 - 6x$.

۲.۴ تقریب کمترین مربعات گسسته

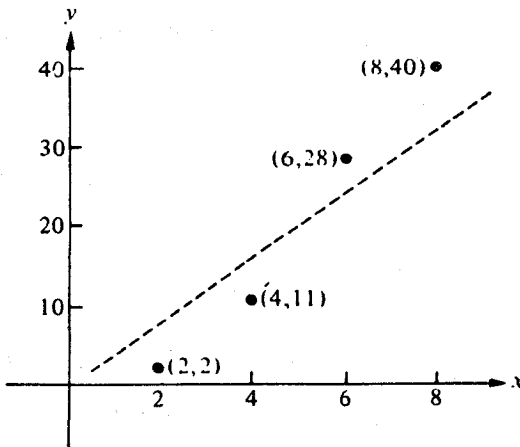
مسئله تخمین مقادیر یک تابع در نقاط غیر جدولی را برای داده‌های تجربی مفروض جدول

۱.۴ در نظر می‌گیریم .

i	x_i	y_i
1	2	2
2	4	11
3	6	28
4	8	40

جدول ۱.۴

روشهای مذکور در فصل ۳ نیاز به ساختن یک چندجمله‌ای درجه ۳ لاگرانژ یا یک اسپلاین مکعبی دارند بطوری که، در x_i ، به‌ازای هر $i = 1, 2, 3, 4$ ، مقدار y_i را بگیرد. لیکن، رسم مقادیر جدول ۱.۴ نشان می‌دهد که معقول است (ر.ک. شکل ۱.۴)



شکل ۱.۴

فرض شود که رابطه واقعی یک رابطه خطی است و هیچ خطی داده‌ها را بخاطر خطای روند جمع آوری داده‌ها، دقیقاً برازش نمی‌کند.

اگر واقعاً این حالت را داشته باشیم، جستجوی تابع تقریب سازی که بر داده‌های معلوم دقیقاً منطبق باشد نامعقول است، در واقع، چنین تابع تقریب ساز نوساناتی به-

بار می‌آورد که در اصل موجود نبوده‌اند. راه بهتر برای حل این نوع مسئله یافتن "بهترین" (به یک معنی) خطی است که بتواند به عنوان تابع تقریب ساز بکار رود ولو اینکه بر داده‌ها در هر نقطه دقیقاً "منطبق نباشد".

اولین مسئله‌ای که باید مورد توجه باشد یافتن بهترین خط تقریب ساز است که خطای مربوطه مساوی مجموع مربعات تفاضلهای بین مقادیر روی خط تقریب ساز و مقادیر معلوم است. اگر $ax_i + b$ مقدار i م روی خط تقریب ساز و y_i مقدار i م معلوم باشد، باید ثابتهای a, b را طوری بیابیم که کمیت

$$\sum_{i=1}^n (y_i - (ax_i + b))^2$$

را مینیمم کنند. این کار معادل مینیمم کردن نیم نرم گسسته داده شده در تمرین ۶ (T) بخش ۱.۴ است، جز آنکه در اینجا مربع خطا را مینیمم می‌کنیم. این مسئله خاص به یافتن ثابتهای a و b که عبارت

$$\begin{aligned} \sum_{i=1}^4 (y_i - (ax_i + b))^2 &= (2 - (2a + b))^2 + (11 - (4a + b))^2 \\ &+ (28 - (6a + b))^2 + (40 - (8a + b))^2. \end{aligned}$$

را مینیمم کنند تحویل می‌شود.

اگر $\sum_{i=1}^4 (y_i - (ax_i + b))^2$ را تابعی از دو متغیر a و b بگیریم، یک نتیجه مقدماتی از حساب چند متغیره ایجاب می‌کند که، برای داشتن مینیمم در (a, b) ، لازم است

$$0 = \frac{\partial}{\partial a} \left[\sum_{i=1}^4 (y_i - (ax_i + b))^2 \right]$$

و

$$0 = \frac{\partial}{\partial b} \left[\sum_{i=1}^4 (y_i - (ax_i + b))^2 \right].$$

در نتیجه،

$$\begin{aligned} 0 &= 2(2 - 2a - b)(-2) + 2(11 - 4a - b)(-4) \\ &+ 2(28 - 6a - b)(-6) + 2(40 - 8a - b)(-8), \end{aligned}$$

و

$$0 = 2(2 - 2a - b)(-1) + 2(11 - 4a - b)(-1) + 2(28 - 6a - b)(-1) + 2(40 - 8a - b)(-1),$$

که به روابط

$$30a + 5b = 134$$

و

$$20a + 4b = 81$$

ساده می‌شوند.

جواب این دستگاه معادلات عبارت است از $a = 6.55$ و $b = -12.5$ ؛ در نتیجه،

بهترین معادله خطی برحسب نرم l_2 خواهد بود

$$y = 6.55x - 12.5.$$

جدول ۲.۴ مقادیر تجربی را همراه با مقادیر حاصل از بکارگیری این تقریب نشان

می‌دهد.

i	x_i	y_i	$6.55x - 12.5$
1	2	2	.6
2	4	11	13.7
3	6	28	26.8
4	8	40	39.9

جدول ۲.۴

مسئله یافتن معادله بهترین تقریب خطی برحسب نیم نرم گسسته l_1 مذکور در

تمرین ۶ (پ) بخش ۱.۴ نیاز به مقادیر a و b ای دارد که

$$\text{Max}_{i=1,2,3,4} \{|y_i - (ax_i + b)|\}$$

را مینیمم کنند. این مسئله را معمولاً "مسئله مینیماکس می‌نامند و نمی‌توان آن را به روشهای مقدماتی حل کرد.

مسئله یافتن معادله بهترین تقریب خطی برحسب نیم نرم گسسته l_1 تعریف

شده در تمرین (۶ ب) بخش ۱.۴ مستلزم مقادیری از a و b است که کمیت

$$\sum_{i=1}^4 |y_i - (ax_i + b)| = |2 - (2a + b)| + |11 - (4a + b)| + |28 - (6a + b)| + |40 - (8a + b)|$$

را مینیمم کنند .

برای مینیمم کردن این کمیت ، لازم است که

$$0 = \frac{\partial}{\partial a} \sum_{i=1}^4 |y_i - (ax_i + b)|$$

$$0 = \frac{\partial}{\partial b} \sum_{i=1}^4 |y_i - (ax_i + b)|.$$

مشکل این روند این است که تابع قدر مطلق در صفر مشتقپذیر نیست و جوابهای این جفت معادله لزوماً قابل حصول نیستند .

از نکات فوق معلوم می شود که مناسبترین روند استفاده از نیم نرم l_2 برای یافتن بهترین تقریب خطی است ؛ خوشبختانه ، نکات نظری مهمی نیز وجود دارند که از این راه جانبداری می کنند . تقریب داده شده با نیم نرم l_∞ عموماً " به جزئی از داده ها وزن بسیار زیادی می دهد که خطای آن زیاد است ، ولی تقریب داده شده با نیم نرم l_1 خطای متوسطی در نقاط مختلف تولید کرده و به نقطه ای که به طور قابل ملاحظه خارج از تقریب است وزن کافی نمی دهد . روش کمترین مربعات اساساً " وزن بیشتر را روی نقطه ای که در مقایسه با بقیه داده ها خارج از خط است می گذارد ، لیکن اجازه تسلط کامل آن نقطه بر تقریب را نمی دهد .

دلیل دیگر برای در نظر گرفتن روش کمترین مربعات به بررسی توزیع آماری خطا مربوط می شود . اگر معلوم باشد یا فرض کنیم توزیع میانگین خطا به صورت خطی است ، مقادیر حاصل از یک روند کمترین مربعات خطی تخمینهایی برای معادله توصیف کننده میانگین می باشند . بعلاوه ، این مقادیر را می توان جهت محاسبه یک تخمین گر برای واریانس مربوط به توزیع بکار برد (یک معرفی از نظریه مربوطه را که به سادگی قابل خواندن است می توان در لارسن [۴۲] ، صفحات ۳۵۹ تا ۳۷۰ ، یافت .)

مسئله کلی برازش بهترین خط کمترین مربعات برای گرد آیه ای از داده های مربوطه مینیمم کردن $\sum_{i=1}^n (y_i - (ax_i + b))^2$ نسبت به پارامترهای a و b است . برای آنکه مینیمم روی دهد ، لازم است که

$$\begin{aligned} 0 &= \frac{\partial}{\partial a} \sum_{i=1}^n (y_i - (ax_i + b))^2 \\ &= 2 \sum_{i=1}^n (y_i - ax_i - b)(-x_i) \end{aligned}$$

$$\begin{aligned} 0 &= \frac{\partial}{\partial b} \sum_{i=1}^n (y_i - ax_i - b)^2 \\ &= 2 \sum_{i=1}^n (y_i - ax_i - b)(-1). \end{aligned}$$

این معادلات به معادلاتی به نام معادلات نرمال ساده می‌شوند.

$$a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i$$

$$a \sum_{i=1}^n x_i + b \cdot n = \sum_{i=1}^n y_i.$$

جواب این معادلات عبارت است از

$$(۶۰۴) \quad a = \frac{n(\sum_{i=1}^n x_i y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n(\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2}$$

$$(۷۰۴) \quad b = \frac{(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i) - (\sum_{i=1}^n x_i y_i)(\sum_{i=1}^n x_i)}{n(\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2}.$$

مثال ۱. داده‌های جدول ۳۰۴ را در نظر می‌گیریم.

برای بیافتن خط تقریب ساز کمترین مربعات به این داده‌ها، جدول را همانطور که جدول ۴۰۴ (صفحه ۱۹۳) نشان داده بسط و مجموع ستونها را تشکیل می‌دهیم.

x_i	y_i
1	1.3
2	3.5
3	4.2
4	5.0
5	7.0
6	8.8
7	10.1
8	12.5
9	13.0
10	15.6
11	16.1

جدول ۳.۴

x_i	y_i	x_i^2	$x_i y_i$
1	1.3	1	1.3
2	3.5	4	7.0
3	4.2	9	12.6
4	5.0	16	20.0
5	7.0	25	35.0
6	8.8	36	52.8
7	10.1	49	70.7
8	12.5	64	100.0
9	13.0	81	117.0
10	15.6	100	156.0
11	16.1	121	177.1
66	97.1	506	749.5

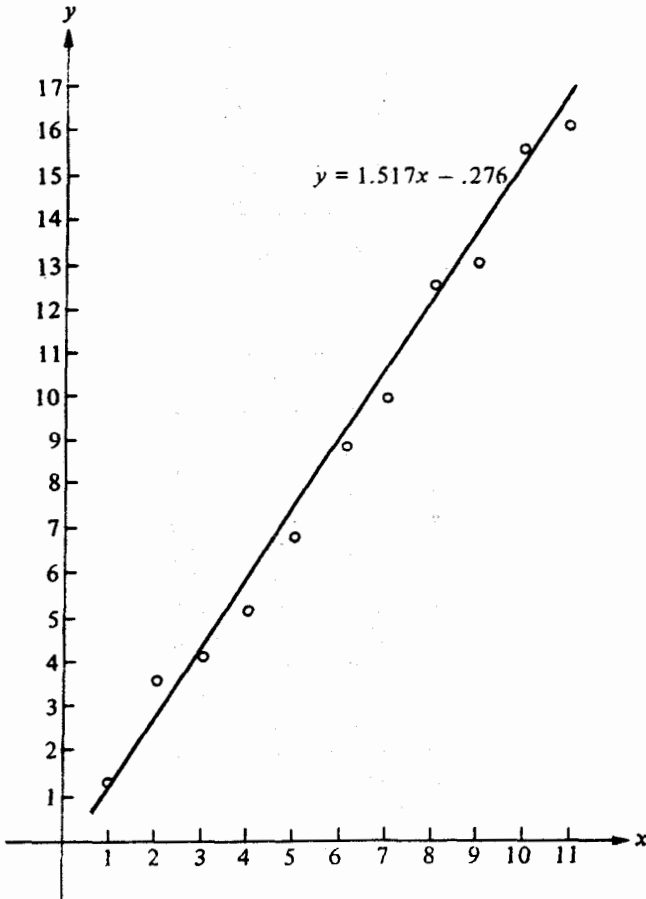
جدول ۴.۴

معادلات (۶.۴) و (۷.۴) ایجاب می‌کنند که

$$a = \frac{11(749.5) - 66(97.1)}{11(506) - (66)^2} = 1.517$$

$$b = \frac{506(97.1) - 66(749.5)}{11(506) - (66)^2} = -.276.$$

نمودار این خط همراه با نقاط داده شده در شکل ۲.۴ نموده شده است.



شکل ۲.۴

مقادیر تقریبی حاصل از روش کمترین مربعات و نقاط داده شده در جدول ۵.۴ ثبت

شده‌اند.

x_i	y_i	$1.517x_i - .276$
1	1.3	1.24
2	3.5	2.76
3	4.2	4.28
4	5.0	5.79
5	7.0	7.31
6	8.8	8.83
7	10.1	10.34
8	12.5	11.86
9	13.0	13.38
10	15.6	14.89
11	16.1	16.41

جدول ۵.۴

مسئله کلسی تقریب سازی مجموعه‌ای از داده‌ها، یعنی $\{(x_i, y_i) | i = 0, 1, \dots, M\}$ ، با یک چند جمله‌ای مانند $P_n(x) = \sum_{k=0}^n a_k x^k$ از درجه $n < M$ با استفاده از روند کمترین مربعات به همین نحو حل می‌شود و لازم است ثابتهای a_0, a_1, \dots, a_n طوری انتخاب شوند که

$$\begin{aligned}
 E &= \sum_{i=0}^M (y_i - P(x_i))^2 \\
 &= \sum_{i=0}^M y_i^2 - 2 \sum_{i=0}^M P(x_i) y_i + \sum_{i=0}^M (P(x_i))^2 \\
 &= \sum_{i=0}^M y_i^2 - 2 \sum_{i=0}^M \left(\sum_{j=0}^n a_j x_i^j \right) y_i + \sum_{i=0}^M \left(\sum_{j=0}^n a_j x_i^j \right)^2 \\
 &= \sum_{i=0}^M y_i^2 - 2 \sum_{j=0}^n a_j \left(\sum_{i=0}^M y_i x_i^j \right) + \sum_{j=0}^n \sum_{k=0}^n a_j a_k \left(\sum_{i=0}^M x_i^{j+k} \right)
 \end{aligned}$$

مینیمم گردد. مثل حالت خطی، برای آنکه E مینیمم شود، لازم است به‌ازای هر $j = 0, 1, \dots, n$ ، بنابراین $\partial E / \partial a_j = 0$.

$$0 = \frac{\partial E}{\partial a_j} = -2 \sum_{i=0}^M y_i x_i^j + 2 \sum_{k=0}^n a_k \sum_{i=0}^M x_i^{j+k}, \quad 0 \leq j \leq n.$$

از این رابطه $n + 1$ معادله با $n + 1$ مجهول a_j ، به نام معادلات نرمال، بدست می آیند:

$$\sum_{k=0}^n a_k \sum_{i=0}^M x_i^{j+k} = \sum_{i=0}^M y_i x_i^j, \quad j = 0, 1, \dots, n.$$

بسط این معادلات به صورت زیر مفید است:

$$a_0 \sum_{i=0}^M x_i^0 + a_1 \sum_{i=0}^M x_i^1 + a_2 \sum_{i=0}^M x_i^2 + \dots + a_n \sum_{i=0}^M x_i^n = \sum_{i=0}^M y_i x_i^0,$$

$$a_0 \sum_{i=0}^M x_i^1 + a_1 \sum_{i=0}^M x_i^2 + a_2 \sum_{i=0}^M x_i^3 + \dots + a_n \sum_{i=0}^M x_i^{n+1} = \sum_{i=0}^M y_i x_i^1$$

⋮

$$a_0 \sum_{i=0}^M x_i^n + a_1 \sum_{i=0}^M x_i^{n+1} + a_2 \sum_{i=0}^M x_i^{n+2} + \dots + a_n \sum_{i=0}^M x_i^{2n} = \sum_{i=0}^M y_i x_i^n.$$

در فصل ۷ نشان داده خواهد شد که معادلات نرمال در صورت متمایز بودن x_i ها، به ازای $i = 0, 1, \dots, M$ ، جواب منحصر بفرد دارند.

مثال ۲. داده های جدول ۴.۶ را با چند جمله ای کمترین مربعات گسسته درجه ۲ دوبرازش کنید.

i	0	1	2	3	4
x_i	0	.25	.5	.75	1.00
y_i	1.0000	1.2840	1.6487	2.1170	2.7183

جدول ۴.۶

در این مسئله $n = 2$ ، $M = 4$ ، و سه معادله نرمال عبارتند از

$$5a_0 + 2.5a_1 + 1.875a_2 = 8.7680$$

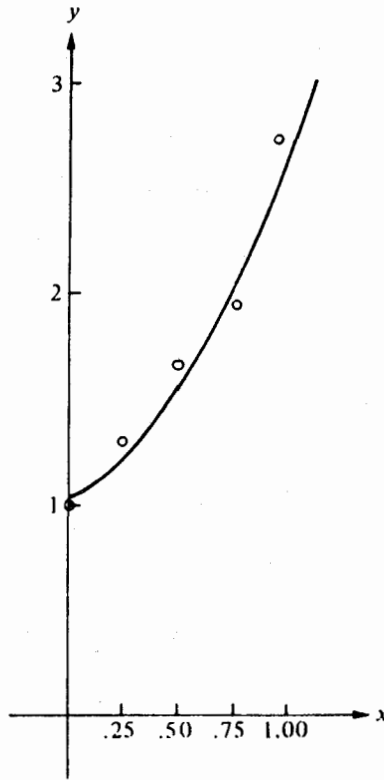
$$2.5a_0 + 1.875a_1 + 1.5625a_2 = 5.4514$$

$$1.875a_0 + 1.5625a_1 + 1.3828a_2 = 4.4015.$$

جواب این دستگاه عبارت است از

$$a_0 = 1.0052, \quad a_1 = .8641, \quad a_2 = .8437.$$

بنابراین، چند جمله‌ای کمترین مربعات درجه دو برازنده داده‌های فوق عبارت است از $P_2(x) = 1.0052 + .8641x + .8437x^2$ ، که نمودارش در شکل ۳.۴ نشان داده شده



شکل ۳.۴

است. در مقادیر داده شده x_i ، تقریبهای جدول ۷.۴ را خواهیم داشت.

i	0	1	2	3	4
x_i	0	.25	.50	.75	1.00
y_i	1.0000	1.2840	1.6487	2.1170	2.7183
$P(x_i)$	1.0052	1.2740	1.6482	2.1279	2.7130
$y_i - P(x_i)$	-.0052	.0100	.0005	-.0109	.0053

جدول ۷.۴

خطای

$$\left\{ \sum_{i=0}^4 |y_i - P(x_i)|^2 \right\}^{1/2} = 0.0166$$

بهترین خطایی است که می‌توان با استفاده از یک چندجمله‌ای درجه‌دو بدست آورد.

با آنکه کمترین مربعات با استفاده از چندجمله‌ایها متداولترین روندهای بکار رفته است، گاهی مقتضی است فرض کنیم که داده‌ها به‌طور نمایی بهم مربوطند. در این صورت، لازم است تابع تقریب ساز به شکل زیر باشد: a و b بازای ثابتهایی چون

$$y = be^{ax} \quad (۸.۴)$$

$$y = bx^a \quad (۹.۴)$$

مشکل اعمال روند کمترین مربعات در این نوع وضعیت از تلاش برای مینیم کردن مربع خطای مربوطه ناشی می‌شود، که

$$E = \sum_{i=0}^M (y_i - be^{ax_i})^2 \quad \text{در حالت معادله } (۸.۴) \text{ عبارت است از}$$

$$E = \sum_{i=0}^M (y_i - bx_i^a)^2 \quad \text{در حالت معادله } (۹.۴) \text{ عبارت است از}$$

معادلات نرمال مربوط به این روندها از

$$0 = \frac{\partial E}{\partial b} = 2 \sum_{i=0}^M (y_i - be^{ax_i})(-e^{ax_i})$$

$$0 = \frac{\partial E}{\partial a} = 2 \sum_{i=0}^M (y_i - be^{ax_i})(-bx_i e^{ax_i})$$

$$0 = \frac{\partial E}{\partial b} = 2 \sum_{i=0}^M (y_i - bx_i^a)(-x_i^a)$$

$$0 = \frac{\partial E}{\partial a} = 2 \sum_{i=0}^M (y_i - bx_i^a)(-b(\ln x_i)x_i^a)$$

بدست می‌آیند.

چون عموماً "برای هیچیک از این دستگانهانی توان جواب دقیق بدست آورد، روشی که معمولاً پس از این حدس که داده‌ها به‌طور نمایی بهم مربوطند حاصل می‌شود این است

که لگاریتم تابع تقریب ساز را در نظر بگیریم، که

(۱۰.۴) $\ln y = \ln b + ax$ درحالت معادله (۸.۴) عبارت است از و

(۱۱.۴) $\ln y = \ln b + a \ln x$ درحالت معادله (۹.۴) عبارت است از در هر حالت، یک مسئله خطی بدست می‌آید و جوابها برای $\ln b$ و a را می‌توان با تغییر مناسب معادلات (۶.۴) و (۷.۴) بدست آورد. با اینحال، باید بخاطر داشت که تقریب حاصل به این نحو تقریب کمترین مربعات برای مسئله اصلی نیست.

مثال ۳. گرد آیه داده‌های جدول ۸.۴ را در نظر می‌گیریم.

i	x_i	y_i
1	1.00	5.10
2	1.25	5.79
3	1.50	6.53
4	1.75	7.45
5	2.00	8.46

جدول ۸.۴

اگر x_i ها با $\ln y_i$ ها رسم شوند، معلوم می‌شود که داده‌ها رابطه‌ای خطی دارند، یا، معادلاً، اگر (x_i, y_i) ها روی کاغذ نیم لگاریتمی رسم شوند، یک شکل خطی بدست خواهد آمد؛ در نتیجه، معقول است تقریبی به شکل

$$y = \ln b + ax \quad \text{یا} \quad y = be^{ax}$$

فرض گردد.

با بسط جدول و جمع ستونها، داده‌های جدول ۹.۴ بدست می‌آیند.

با استفاده از معادلات (۶.۴) و (۷.۴)،

$$a = \frac{(5)(14.422) - (7.5)(9.404)}{(5)(11.875) - (7.5)^2} = .5056$$

$$\ln b = \frac{(11.875)(9.404) - (14.422)(7.5)}{(5)(11.875) - (7.5)^2} = 1.122.$$

i	x_i	y_i	$\ln y_i$	x_i^2	$x_i \ln y_i$
1	1.00	5.10	1.629	1.0000	1.629
2	1.25	5.79	1.756	1.5625	2.195
3	1.50	6.53	1.876	2.2500	2.814
4	1.75	7.45	2.008	3.0625	3.514
5	2.00	8.46	2.135	4.0000	4.270
7.50			9.404	11.875	14.422

جدول ۹.۴

بنابر این، تقریب شکل زیر را بخود می‌گیرد:

$$y = 3.071e^{.5056x}$$

که، در نقاط داده شده، مقادیر جدول ۱۰.۴ را بدست می‌دهد.

i	x_i	y_i	$3.071e^{.5056x_i}$
1	1.00	5.10	5.09
2	1.25	5.79	5.78
3	1.50	6.53	6.56
4	1.75	7.45	7.44
5	2.00	8.46	8.44

جدول ۱۰.۴

مجموعه تمرینات ۲.۴

- چند جمله‌ای کمترین مربعات گسسته^۲ درجه^۲ یک را برای داده‌های مثال ۲ حساب کنید.
- داده‌های مثال ۲ عملاً "از تابع $f(x) = e^x$ بدست می‌آیند. با در نظر گرفتن نمودار $f(x) = e^x$ بر $[0, 1]$ ، آیا قبول دارید که انتخاب چند جمله‌ای کمترین مربعات از درجه^۳ بالاتر نتایج مثال ۲ را اصلاح می‌کند؟ با استفاده از کامپیوتر، ادعای خود را با توجه به چند جمله‌ایهای کمترین مربعات از درجه^۳ سه و چهار پشتیبانی کنید.
- چند جمله‌ای درونیاب لاگرانژ از درجه^۲ دو را برای $f(x) = e^x$ بر گره‌های $x_0 = 0$ ،

۵. $x_1 = 1.0$ و $x_2 = 1.0$ حساب کنید. از مقایسه این چند جمله‌ای درجه دو با چند جمله‌ای مثال ۲، معین کنید کدامیک $f(x) = x^2$ را بهتر بر $[0, 1]$ تقریب می‌کند.

۴. برای داده‌های جدول زیر، چند جمله‌ایهای کمترین مربعات گسسته از درجات ۱، ۲، ۳ و ۴ را پیدا نمایید.

i	0	1	2	3	4	5
x_i	0	.15	.31	.5	.6	.75
y_i	1.0	1.004	1.031	1.117	1.223	1.422

چه درجه‌ای بهترین تقریب کمترین مربعات را بدست می‌دهد؟ یعنی، چه درجه‌ای کوچکترین خطا نسبت به نیم نرم l_2 را بدست می‌دهد؟ آیا می‌توانید جواب خود را توجیه کنید؟

۵. برای تعیین رابطه‌ای میان تعداد ماهیها و تعداد انواع ماهیها در نمونه‌هایی، پی. سیل^۱ و ا.ر. دیبدال^۲ [۵۲] یک چند جمله‌ای کمترین مربعات خطی به گرد آیه زیر از داده‌ها، که در زمانی بیش از دو سال در نمونه‌هایی جمع‌آوری شده‌اند، برازش کردند. فرض کنید x تعداد ماهیها در نمونه، و y تعداد انواع در نمونه باشد

x	y	x	y	x	y
13	11	29	12	60	14
15	10	30	14	62	21
16	11	31	16	64	21
21	12	36	17	70	24
22	12	40	13	72	17
23	13	42	14	100	23
25	13	55	22	130	34

۶. چند جمله‌ای کمترین مربعات خطی برای این داده‌ها را معین نمایید.

مجموعه داده‌های زیر، که در مارس ۱۹۲۵ به کمیته فرعی مخالف تراستهای سنا ارائه شده است، خصایص نسبی نجات از تصادف اتومبیلها را در انواع مختلف اتومبیل نشان می‌دهد. بهترین خط کمترین مربعات که این داده‌ها را تقریب می‌زند

پیدا کنید. (این جدول در صد تصادفات وسایط نقلیه را که در آنها شدیدترین جراحت کشنده یا خطرناک بوده است نشان می‌دهد.)

درصد وقوع	وزن متوسط به پوند	نوع
3.1	4800	۱. شخصی (لوکس) معمولی
4.0	3700	۲. شخصی (متوسط) معمولی
5.2	3400	۳. شخصی (کار) معمولی
6.4	2800	۴. شخصی کوچک
9.6	1900	۵. خارجی کوچک

۷. برای تعیین رابطه^۱ تابعی بین ضریب نازکی و ضخامت یک نمونه سنگ چخماق، وی. پی. سینگ^۱ [۵۹]، با استفاده از یک چندجمله‌ای کمترین مربعات خطی، گردآیه‌های ازداده‌ها را برازش کرده است. گردآیه^۲ زیر از داده‌ها از نموداری در آن مقاله اتخاذ شده است. بهترین چندجمله‌ای کمترین مربعات خطی که این داده‌ها را برازش

ضخامت (cm)	ضریب نازکی db/cm
26.5	.040
28.1	.041
25.2	.055
26.0	.056
24.0	.062
25.0	.071
26.4	.071
27.2	.078
25.6	.082
25.0	.090
26.8	.092
24.8	.100
27.0	.105
25.0	.120
27.3	.123
26.9	.130
26.2	.140

می‌کند بیاید .

۸ . در مقاله‌ای دربارهٔ قدرت سودمندی انرژی گرمایی نوروزی ، ال . شرودر [۵۵] از داده‌های زیر برای تعیین رابطهٔ لگاریتمی کمترین مربعات درجهٔ دوم بین W ، یعنی وزن زندهٔ کرمها به گرم ، و R ، یعنی مصرف اکسیژن کرمها به میلی‌گرم بر ساعت ، استفاده کرده است . این رابطه به شکل زیر بدست آمده است :

$$\log R = a + b \log W + c(\log W)^2,$$

که در آن a ، b ، و c ثابتهایی هستند که با این داده‌ها معین می‌شوند . با استفاده از این مقادیر ، ثابتهای a ، b ، و c را معین نمایید .

W	R	W	R	W	R	W	R	W	R
0.017	0.154	0.025	0.23	0.020	0.181	0.020	0.180	0.025	0.234
0.087	0.296	0.111	0.357	0.085	0.260	0.119	0.299	0.233	0.537
0.174	0.363	0.211	0.366	0.171	0.334	0.210	0.428	0.783	1.47
1.11	0.531	0.999	0.771	1.29	0.87	1.32	1.15	1.35	2.48
1.74	2.23	3.02	2.01	3.04	3.59	3.34	2.83	1.69	1.44
4.09	3.58	4.28	3.28	4.29	3.40	5.48	4.15	2.75	1.84
5.45	3.52	4.58	2.96	5.30	3.88			4.83	4.66
5.96	2.40	4.68	5.10					5.53	6.94

۳.۴ چند جمله‌ایهای متعامد و تقریب کمترین مربعات

در بخش پیش ، مسئلهٔ تقریب کمترین مربعات جهت برازش گردآیهای از داده‌ها مورد بحث قرار گرفت . مسئلهٔ دیگر تقریب که در مقدمه ذکر شد مربوط به تقریب توابع می‌باشد . فرض کنیم $f \in C[a, b]$ و یک چندجمله‌ای از درجهٔ حداکثر n ، مانند P_n مطلوب باشد که

$$(12.4) \quad \|f - P_n\|_2 = \left[\int_a^b (f(x) - P_n(x))^2 dx \right]^{1/2}$$

را مینیمم کند .

برای تعیین یک چندجمله‌ای تقریب ساز کمترین مربعات ، یعنی یک چندجمله‌ای که

(۱۲.۴) را مینیمم کند، فرض می‌کنیم

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = \sum_{k=0}^n a_k x^k,$$

و تعریف می‌کنیم

$$E(a_0, a_1, \dots, a_n) = \int_a^b \left(f(x) - \sum_{k=0}^n a_k x^k \right)^2 dx.$$

مسئله یافتن ضرایب حقیقی a_0, \dots, a_n است که E را مینیمم کند. از حساب توابع چند متغیره می‌دانیم که شرط لازم برای آنکه E به وسیله اعداد a_0, \dots, a_n مینیمم شود این است که

$$\frac{\partial E}{\partial a_j} = 0, \quad j = 0, 1, \dots, n$$

بهازای هر $n, 1, \dots, 0, j$

چون

$$E = \int_a^b [f(x)]^2 dx - 2 \sum_{k=0}^n a_k \int_a^b x^k f(x) dx + \int_a^b \left(\sum_{k=0}^n a_k x^k \right)^2 dx,$$

$$\frac{\partial E}{\partial a_j} = -2 \int_a^b x^j f(x) dx + 2 \sum_{k=0}^n a_k \int_a^b x^{j+k} dx.$$

لذا، برای یافتن P_n ، باید $(n+1)$ معادله خطی

$$(13.4) \quad \sum_{k=0}^n a_k \int_a^b x^{j+k} dx = \int_a^b x^j f(x) dx, \quad j = 0, 1, \dots, n,$$

نسبت به $(n+1)$ مجهول $a_j, j = 0, 1, \dots, n$ ، حل شوند. این معادلات را معادلات نرمال نامند. می‌توان نشان داد که معادلات نرمال همواره جواب منحصر بفرد دارند مشروط بر اینکه $f \in C[a, b]$ و $a \neq b$. (ر.ک. تمرین ۰۷)

مثال ۱. چند جمله‌ای تقریب ساز کمترین مربعات از درجه دو را برای تابع $f(x) = \sin \pi x$ بر بازه $[0, 1]$ پیدا کنید. معادلات نرمال برای $P_2(x) = a_2 x^2 + a_1 x + a_0$ عبارتند از

$$a_0 \int_0^1 1 dx + a_1 \int_0^1 x dx + a_2 \int_0^1 x^2 dx = \int_0^1 \sin \pi x dx,$$

$$a_0 \int_0^1 x dx + a_1 \int_0^1 x^2 dx + a_2 \int_0^1 x^3 dx = \int_0^1 x \sin \pi x dx,$$

$$a_0 \int_0^1 x^2 dx + a_1 \int_0^1 x^3 dx + a_2 \int_0^1 x^4 dx = \int_0^1 x^2 \sin \pi x dx.$$

انتگرالگیری نتیجه می دهد که

$$a_0 + \frac{1}{2}a_1 + \frac{1}{3}a_2 = \frac{2}{\pi},$$

$$\frac{1}{2}a_0 + \frac{1}{3}a_1 + \frac{1}{4}a_2 = \frac{1}{\pi},$$

$$\frac{1}{3}a_0 + \frac{1}{4}a_1 + \frac{1}{5}a_2 = \frac{\pi^2 - 4}{\pi^3}$$

سه معادله سه مجهولی را می توان حل کرد و بدست آورد که

$$a_0 = \frac{12\pi^2 - 120}{\pi^3} \approx -0.050465,$$

$$a_1 = \frac{720 - 60\pi^2}{\pi^3} \approx 4.12251,$$

$$a_2 = \frac{60\pi^2 - 720}{\pi^3} \approx -4.12251.$$

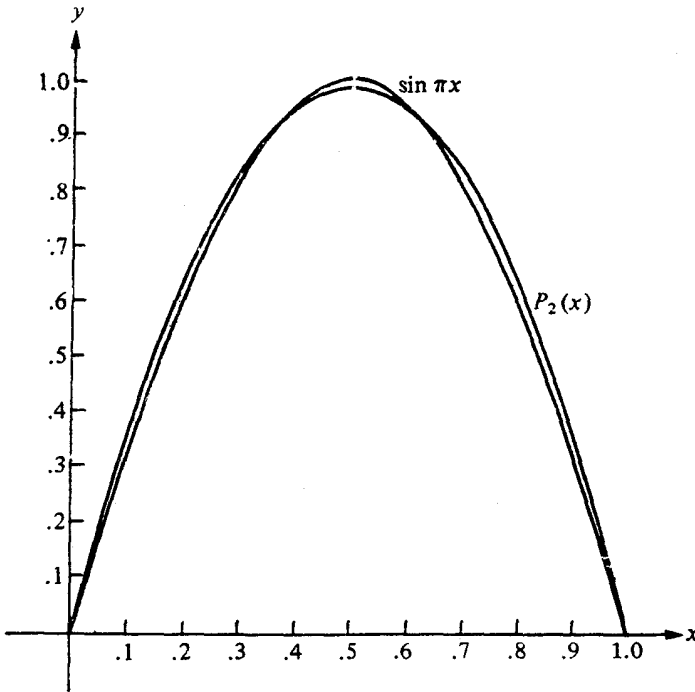
در نتیجه، چند جمله ای تقریب کمترین مربعات از درجه دو برای $f(x) = \sin \pi x$ بر $[0, 1]$ عبارت است از $P_2(x) = -4.12251x^2 + 4.12251x - 0.050465$ ؛ ر. ک. شکل

۴۰۴

مثال ۱ درجه سستی یافتن تقریب چند جمله ای کمترین مربعات را نشان می دهد. بایستی دستگاه $(n+1)$ در $(n+1)$ خطی نسبت به ضرایب a_0, \dots, a_n از P_n را حل کنیم. همچنین، در فصل ۷ نشان خواهیم داد که، چون ضرایب دستگاه خطی به شکل

$$\int_a^b x^{j+k} dx = \frac{b^{j+k+1} - a^{j+k+1}}{j+k+1}$$

می باشند، دستگاه خطی، به دلیل وجود خطاهای گرد کردن، جواب عددی مناسبی ندارد. نقص دیگر روش بالا شبیه نقیصی است که وقتی چند جمله ایهای لاگرانژ برای اولین



شکل ۴.۴

بار در بخش ۲.۳ معرفی شدند پدیدار شد؛ یعنی، محاسباتی که برای بدست آوردن بهترین چند جمله‌ای درجه n ، یعنی P_n ، انجام شد مقدار کار لازم برای بدست آوردن P_{n+1} ، یعنی چندجمله‌ای با یک درجه بیشتر، را تقلیل نمی‌دهد. حال برای بدست آوردن تقریبهای کمترین مربعات روش دیگر را بررسی می‌کنیم. خواهیم دید که این روش در محاسبات موثر همچنین از این نظر سودمند است که به محض معلوم شدن P_n ، به آسانی P_{n+1} معین می‌شود. برای تسهیل ادامه بحث، چند مفهوم جدید را معرفی می‌کنیم.

در مورد نمادها، فرض کنیم Π_n مجموعه تمام چندجمله‌ایها از درجه حداکثر n باشد. همه ϕ_0, \dots, ϕ_n نمایانگر چندجمله‌ای‌اند و ϕ_j از درجه j می‌باشد. در واقع، تحدید این بررسی به چند جمله‌ایها لازم نیست. آنچه مورد لزوم است این است که توابع ϕ_0, \dots, ϕ_n پیوسته بوده و در خاصیت زیر صدق کنند.

تعریف ۲.۴. گوئیم توابع ϕ_0, \dots, ϕ_n بر $[a, b]$ ، که $b > a$ ، مستقل خطی اند اگر هر وقت به ازای هر $x \in [a, b]$ ،

$$c_0 \phi_0(x) + c_1 \phi_1(x) + \dots + c_n \phi_n(x) = 0$$

داشته باشیم $c_0 = c_1 = \dots = c_n = 0$ ، در غیر این صورت، این توابع را وابسته خطی نامیم.

قضیه ۳.۴. اگر $\phi_j(x)$ ، به ازای هر $j = 0, 1, \dots, n$ ، یک چند جمله‌ای از درجه j باشد، آنگاه ϕ_0, \dots, ϕ_n بر هر بازه $[a, b]$ ، که $a < b$ ، مستقل خطی می‌باشند.

برهان. فرض کنیم c_0, \dots, c_n اعدادی حقیقی باشند بطوری که

$$c_0 \phi_0(x) + c_1 \phi_1(x) + \dots + c_n \phi_n(x) = 0, \quad x \in [a, b]$$

چون $P(x) = \sum_{k=0}^n c_k \phi_k(x)$ در هر نقطه در $[a, b]$ صفر می‌شود، P دارای بی‌نهایت ریشه است. P یک چند جمله‌ای از درجه حداکثر n است، پس نتیجه ۱۷.۲، صفحه ۸۴، ایجاب می‌کند که $P \equiv 0$ و ضریب هر توان x صفر باشد. هر ϕ_j از درجه دقیقاً j است؛ در نتیجه،

$$P(x) = c_n x^n + \dots$$

جملاتی از x با درجه پایین‌تر $c_n = 0$ را ایجاب می‌کند. با ادامه این بحث،

$$\begin{aligned} P(x) &= c_0 \phi_0(x) + \dots + c_{n-1} \phi_{n-1}(x) + 0 \cdot \phi_n(x) \\ &= c_{n-1} x^{n-1} + \dots \end{aligned}$$

که $c_{n-1} = 0$ را ایجاب خواهد کرد.

این عمل را می‌توان ادامه و نشان داد که

$$0 = c_n = c_{n-1} = \dots = c_1 = c_0$$

مثال ۲. فرض کنیم $\phi_0(x) = 2$ ، $\phi_1(x) = x - 3$ ، $\phi_2(x) = x^2 + 2x + 7$ بنا بر

قضیه ۳.۴، ϕ_0 ، ϕ_1 ، ϕ_2 بر بازه $[a, b]$ ، که $b > a$ ، مستقل خطی اند. فرض کنیم

$Q(x) = a_0 + a_1 x + a_2 x^2$ عضوی از Π_2 باشد. برای نشان دادن اینکه ثابتی چون

c_0, c_1, c_2 وجود دارند بطوری که $Q(x) = c_0 \phi_0(x) + c_1 \phi_1(x) + c_2 \phi_2(x)$ ، توجه

می‌کنیم که

$$\begin{aligned} \dot{1} &= \frac{1}{2}\phi_0(x), \\ x &= \phi_1(x) + 3 = \phi_1(x) + \frac{3}{2}\phi_0(x), \\ x^2 &= \phi_2(x) - 2x - 7 = \phi_2(x) - 2[\phi_1(x) + \frac{3}{2}\phi_0(x)] \\ &\quad - 7[\frac{1}{2}\phi_0(x)] \\ &= \phi_2(x) - 2\phi_1(x) - \frac{13}{2}\phi_0(x). \end{aligned}$$

بنابراین ،

$$\begin{aligned} Q(x) &= a_0[\frac{1}{2}\phi_0(x)] + a_1[\phi_1(x) + \frac{3}{2}\phi_0(x)] \\ &\quad + a_2[\phi_2(x) - 2\phi_1(x) - \frac{13}{2}\phi_0(x)] \\ &= [\frac{1}{2}a_0 + \frac{3}{2}a_1 - \frac{13}{2}a_2]\phi_0(x) \\ &\quad + [a_1 - 2a_2]\phi_1(x) + a_2\phi_2(x). \end{aligned}$$

مثال ۲ در محدوده بسیار کلیتری برقرار است. در واقع ، هر وقت ϕ_0, \dots, ϕ_n اعضای مستقل خطی از Π_n باشند ، هر عضو Π_n را می توان به صورت ترکیبی خطی از ϕ_0, \dots, ϕ_n و به طور منحصر بفرد نوشت . اثبات این مطلب در تمرین ۱۱ خواسته شده است . در این وضع ، مجموعه $\{\phi_0, \dots, \phi_n\}$ را اغلب یک پایه برای مجموعه Π_n می نامند . با مجموعه $\{\phi_0, \dots, \phi_n\}$ از چند جمله ایها به صورتی که قبلاً " توصیف شد ، اعداد حقیقی a_0, \dots, a_n لازم است تا

$$\begin{aligned} E(a_0, \dots, a_n) &= \int_a^b \left(f(x) - \sum_{k=0}^n a_k \phi_k(x) \right)^2 dx \\ &= \int_a^b [f(x)]^2 dx - 2 \sum_{k=0}^n a_k \int_a^b f(x) \phi_k(x) dx \\ &\quad + \sum_{i=0}^n \sum_{k=0}^n a_i a_k \int_a^b \phi_i(x) \phi_k(x) dx \end{aligned}$$

مینیمم شود . مثل قبل ، بایستی $\partial E / \partial a_j = 0$ برقرار شود ؛ در نتیجه ، معادلات نرمال به صورت زیر در می آیند :

$$\sum_{k=0}^n a_k \int_a^b \phi_f(x) \phi_k(x) dx = \int_a^b f(x) \phi_f(x) dx , \quad j = 0, 1, \dots, n$$

تابحال تعمیم از x^j به چند جمله‌ای ϕ_j از درجه j چیزی عایدمان نکرده است، ولی، اگر ϕ_0, \dots, ϕ_n را می‌شد با این خاصیت که

$$\int_a^b \phi_j(x) \phi_k(x) dx = 0, \quad j \neq k$$

هر وقت

اختیار کرد، دستگاه معادلات نرمال به صورت زیر تحویل می‌شد:

بمازای هر $j = 0, 1, \dots, n$

$$a_j \int_a^b [\phi_j(x)]^2 dx = \int_a^b f(x) \phi_j(x) dx$$

و

$$a_j = \frac{\int_a^b f(x) \phi_j(x) dx}{\int_a^b [\phi_j(x)]^2 dx}, \quad j = 0, 1, \dots, n$$

بمازای هر $j = 0, 1, \dots, n$

مشروط بر اینکه بمازای هر j

$$\int_a^b [\phi_j(x)]^2 dx \neq 0.$$

تعریف ۴.۴. مجموعه توابع (نه لزوماً چند جمله‌ای) ϕ_0, \dots, ϕ_n بر $[a, b]$ متعامد نامیده می‌شود اگر

$$\int_a^b \phi_j(x) \phi_k(x) dx = \begin{cases} 0, & j \neq k \\ \alpha_k > 0, & j = k \end{cases}$$

هر وقت $j \neq k$ ، هر وقت $j = k$

این تعریف، همراه با نکات پیش از آن، قضیه زیر را بدست می‌دهد.

قضیه ۵.۴. هرگاه $\{\phi_0, \dots, \phi_n\}$ یک مجموعه متعامد از توابع بر بازه $[a, b]$ ، که $a < b$ ، باشد، آنگاه تقریب کمترین مربعات به f بر $[a, b]$ ، با استفاده از ϕ_0, \dots, ϕ_n عبارت است از

$$P(x) = \sum_{k=0}^n a_k \phi_k(x),$$

که در آن

$$a_k = \frac{\int_a^b \phi_k(x) f(x) dx}{\int_a^b [\phi_k(x)]^2 dx}$$

$$= \frac{1}{\alpha_k} \int_a^b \phi_k(x) f(x) dx.$$

مثال ۳. مجموعه^۴ توابع $\{\phi_0, \phi_1, \dots, \phi_{2n}\}$ ، که در آن

$$\phi_0(x) \equiv \frac{1}{\sqrt{2\pi}},$$

$$\phi_{2k}(x) = \frac{1}{\sqrt{\pi}} \cos(kx) \quad , k = 1, 2, \dots, n$$

به‌ازای هر $k = 1, 2, \dots, n$

$$\phi_{2k-1}(x) = \frac{1}{\sqrt{\pi}} \sin(kx) \quad , k = 1, 2, \dots, n$$

به‌ازای هر $k = 1, 2, \dots, n$

یک مجموعه^۴ متعامد از توابع بر $[-\pi, \pi]$ تشکیل می‌دهند. (ر.ک. تمرین ۰۸) به‌علاوه، این توابع دارای این خاصیت‌اند که

$$\int_{-\pi}^{\pi} [\phi_k(x)]^2 dx = 1, \quad k = 0, 1, \dots, 2n.$$

به‌ازای $f \in C[-\pi, \pi]$ ، تقریب کمترین مربعات (به نام چند جمله‌ای مثلثاتی) که با

$$S_n(x) = \sum_{k=0}^{2n} a_k \phi_k(x)$$

تعریف می‌شود، که در آن

$$a_k = \int_{-\pi}^{\pi} f(x) \phi_k(x) dx \quad , k = 0, 1, \dots, 2n$$

وقتی $n \rightarrow \infty$ ، یک سری فوریه نام دارد. سریهای فوریه برای توصیف جواب معادلات دیفرانسیل معمولی و جزئی مختلف که در شرایط فیزیکی به‌وجود می‌آیند ابزار فوق‌العاده مفیدی می‌باشند.

برای تعیین چند جمله‌ای مثلثاتی از درجه^۴ $2n$ که تابع

$$f(x) = |x|, \quad -\pi < x < \pi,$$

را تقریب می‌کند، لازم است ضرایب زیر حساب شوند:

$$\begin{aligned} a_0 &= \int_{-\pi}^{\pi} |x| \frac{1}{\sqrt{2\pi}} dx \\ &= -\frac{1}{\sqrt{2\pi}} \int_{-\pi}^0 x dx + \frac{1}{\sqrt{2\pi}} \int_0^{\pi} x dx \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\pi} x dx \\ &= \frac{\sqrt{2\pi^2}}{2\sqrt{\pi}}, \end{aligned}$$

به‌ازای هر $k = 1, 2, \dots, n$

$$\begin{aligned} a_{2k} &= \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} |x| \cos kx dx \\ &= \frac{2}{\sqrt{\pi}} \int_0^{\pi} x \cos kx dx \\ &= \frac{2}{\sqrt{\pi k^2}} [(-1)^k - 1], \end{aligned}$$

و، به‌ازای هر $k = 1, 2, \dots, n$

$$\begin{aligned} a_{2k-1} &= \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} |x| \sin kx dx \\ &= 0, \end{aligned}$$

بنابراین، چند جمله‌ای مثلثاتی که f را تقریب می‌کند عبارت است از

$$S_n(x) = \frac{\pi}{2} - \frac{2}{\pi} \sum_{k=1}^n \frac{(-1)^k - 1}{k^2} \cos(kx).$$

مثال ۴. به ساختن گردآیه‌ای از چند جمله‌ایهای متعامد از Π_2 برای بازه $[0, 1]$ می‌پردازیم.

بدلخواه اختیار می‌کنیم $\phi_0(x) \equiv 1$. برای بدست آوردن ϕ_1 ، یک چند جمله‌ای از

درجه یک بر $[0, 1]$ ، به‌ازای ثابتی چون B_1 ، قرار می‌دهیم:

$$\begin{aligned}\phi_1(x) &= (x - B_1)\phi_0(x) \\ &= x - B_1\end{aligned}$$

ثابت B_1 را می‌توان از شرط

$$\begin{aligned}0 &= \int_0^1 \phi_1(x)\phi_0(x) dx \\ &= \int_0^1 (x - B_1) dx \\ &= \frac{1}{2} - B_1,\end{aligned}$$

بدست آورد. که ایجاب می‌کند $B_1 = \frac{1}{2}$ ؛ در نتیجه،

$$\phi_1(x) = x - \frac{1}{2}.$$

برای یافتن $\phi_2(x)$ ، فرض می‌کنیم

$$\phi_2(x) = (x - B_2)\phi_1(x) - C_2\phi_0(x),$$

که در آن B_2 و C_2 طوری اختیار شده‌اند که

$$\begin{aligned}0 &= \int_0^1 \phi_2(x)\phi_1(x) dx \\ &= \int_0^1 (x - B_2)[\phi_1(x)]^2 dx - \int_0^1 C_2\phi_0(x)\phi_1(x) dx \\ &= \int_0^1 (x - B_2)(x - \frac{1}{2})^2 dx \\ &= \int_0^1 x(x - \frac{1}{2})^2 dx - B_2 \int_0^1 (x - \frac{1}{2})^2 dx\end{aligned}$$

$$\begin{aligned}0 &= \int_0^1 \phi_2(x)\phi_0(x) dx \\ &= \int_0^1 (x - B_2)\phi_1(x)\phi_0(x) dx - \int_0^1 C_2[\phi_0(x)]^2 dx \\ &= \int_0^1 x\phi_1(x)\phi_0(x) dx - C_2 \int_0^1 [\phi_0(x)]^2 dx \\ &= \int_0^1 x(x - \frac{1}{2}) dx - C_2 \int_0^1 dx.\end{aligned}$$

بنابراین ،

$$B_2 = \frac{\int_0^1 x(x - \frac{1}{2})^2 dx}{\int_0^1 (x - \frac{1}{2})^2 dx} = \frac{1}{2}$$

و

$$C_2 = \int_0^1 x(x - \frac{1}{2}) dx = \frac{1}{12}.$$

این نتیجه می دهد که

$$\begin{aligned} \phi_2(x) &= (x - \frac{1}{2})\phi_1(x) - \frac{1}{12}\phi_0(x) \\ &= x^2 - x + \frac{1}{6}, \end{aligned}$$

و گردآیه متعامد عبارت است از

$$\{1, x - \frac{1}{2}, x^2 - x + \frac{1}{6}\}.$$

برای آنکه از این مجموعه از چند جمله ایهای متعامد جهت تعیین چند جمله ایهای

کمترین مربعات از درجات 0 ، 1 ، و 2 برای تابع $f(x) = e^x$ بر $[0, 1]$ استفاده کنیم ،

ابتدا قرار می دهیم

$$P_0(x) = a_0 \phi_0(x),$$

که در آن

$$a_0 = \frac{\int_0^1 e^x \phi_0(x) dx}{\int_0^1 [\phi_0(x)]^2 dx} = \int_0^1 e^x dx = e - 1.$$

بعد فرض می کنیم

$$P_1(x) = a_0 + a_1 \phi_1(x)$$

که در آن

$$a_1 = \frac{\int_0^1 e^x \phi_1(x) dx}{\int_0^1 [\phi_1(x)]^2 dx} = \frac{\int_0^1 e^x (x - \frac{1}{2}) dx}{\int_0^1 (x - \frac{1}{2})^2 dx} = 18 - 6e,$$

در نتیجه ،

$$\begin{aligned} P_1(x) &= e - 1 + (18 - 6e)\phi_1(x) \\ &= e - 1 + (18 - 6e)(x - \frac{1}{2}) \\ &= (18 - 6e)x + 4e - 10. \end{aligned}$$

بالاخره ،

$$P_2(x) = a_0 \phi_0(x) + a_1 \phi_1(x) + a_2 \phi_2(x)$$

9

$$\begin{aligned} a_2 &= \frac{\int_0^1 e^x \phi_2(x) dx}{\int_0^1 [\phi_2(x)]^2 dx} \\ &= \frac{\int_0^1 e^x (x^2 - x + \frac{1}{6}) dx}{\int_0^1 [x^2 - x + \frac{1}{6}]^2 dx} = 210e - 570, \end{aligned}$$

در نتیجه،

$$\begin{aligned} P_2(x) &= P_1(x) + a_2 \phi_2(x) \\ &= (18 - 6e)x + 4e - 10 + (210e - 570)(x^2 - x + \frac{1}{6}) \\ &= (210e - 570)x^2 + (588 - 216e)x + 39e - 105. \end{aligned}$$

مجموعه مسائل ۳.۴

۱. تقریب چند جمله‌ای کمترین مربعات از درجه یک به $f(x) = x^2 - 2x + 3$ را بر $[0, 1]$ بیابید.
۲. تقریب چند جمله‌ای کمترین مربعات از درجه دو به $f(x) = x^2 - 2x + 3$ را بر $[0, 1]$ بیابید.
۳. به روش مثال ۴، چند جمله‌ایهای $\psi_0(t)$ ، $\psi_1(t)$ ، و $\psi_2(t)$ را برای بازه $[-1, 1]$ بسازید.
۴. با استفاده از تمرین ۳، چند جمله‌ای کمترین مربعات از درجه دو را برای $f(t) = e^{(t+1)/2}$ بر $[-1, 1]$ حساب کنید.
۵. چند جمله‌ایهای ψ_0, ψ_1, ψ_2 در تمرین ۳ را از توابع ϕ_0, ϕ_1, ϕ_2 داده شده در مثال ۴ با تغییر متغیر $x = \frac{1}{2}(t+1)$ ، که $[0, 1]$ را بتوی $[-1, 1]$ می‌نگارد، بدست آورید. چون می‌خواهیم ضرایب پیشرو ψ_1 و ψ_2 مساوی 1 باشند، باید در یک عامل مقیاس ضرب کنیم. توجه کنید که این کار بر خاصیت تعامد اثری نمی‌گذارد.
۶. چند جمله‌ای کمترین مربعات از درجه دو برای $f(t) = e^{(t+1)/2}$ را بر $[-1, 1]$ از مثال ۴ با تغییر متغیر $x = \frac{1}{2}(t+1)$ بدست آورید. جواب خود را با جواب تمرین ۴ مقایسه کنید.
۷. نشان دهید که معادلات نرمال (۱۳.۴) جواب منحصر بفرد دارند. [راهنمایی:

نشان دهید که تنها جواب تابع $f(x) \equiv 0$ عبارت است از $a_j = 0$ ، $j = 0, 1, \dots, n$ ،
 (۱۳۰۴) را در a_j ضرب کرده و روی همه z ها جمع ببندید . علامات انتگرال و
 جمع بندی را با هم عوض کرده بدست آورید $\int_0^b [P(x)]^2 dx = 0$. بنابراین ، به ازای
 $P(x) = 0$ ، $j = 0, \dots, n$ یا $a_j = 0$. در نتیجه ، ماتریس ضرایب نامفرد است ، و
 جواب منحصر بفردی برای (۱۳۰۴) وجود دارد .

۸ . نشان دهید که توابع $\phi_0(x) = 1/\sqrt{2\pi}$ ، $\phi_1(x) = (1/\sqrt{\pi}) \sin x$ ، \dots ، $\phi_{2n-1}(x) = (1/\sqrt{\pi}) \sin nx$ ، $\phi_{2n}(x) = (1/\sqrt{\pi}) \cos nx$
 بر $[-\pi, \pi]$ متعامدند . [راهنمایی . با استفاده از
 اتحادهای مثلثاتی برای $\cos(mx \pm nx)$ و $\sin(mx \pm nx)$ ، انتگرالهای مربوطه را
 ساده کنید .]

۹ . چند جمله ای تقریب ساز کمترین مربعات مثلثاتی برای $f(x) = x$ را بر $[-\pi, \pi]$ ، با
 استفاده از مثال ۳ به ازای $n = 2$ ، بدست آورید .

۱۰ . چند جمله ای مثلثاتی کمترین مربعات کلی $S_n(x)$ را برای

$$f(x) = \begin{cases} -1 & \text{اگر } -\pi < x < 0, \\ 1 & \text{اگر } 0 < x < \pi. \end{cases}$$

بیابید .

۱۱ . فرض کنید $\{\phi_0, \phi_1, \dots, \phi_n\}$ یک مجموعه^۴ مستقل خطی از Π_n باشد . نشان دهید که ،
 به ازای هر عنصر $Q \in \Pi_n$ ، ثابتهای منحصر بفردی مانند c_0, c_1, \dots, c_n وجود دارند
 بطوری که

$$Q(x) = \sum_{k=0}^n c_k \phi_k(x).$$

۱۲ . فرض کنید ϕ_0, \dots, ϕ_n بر $[a, b]$ متعامد باشند . نشان دهید که $\{\phi_0, \phi_1, \dots, \phi_n\}$ یک
 مجموعه^۴ مستقل خطی است .

۴۰۴ تقریب کمترین مربعات وزندار

تقریب کمترین مربعات برای تابع f تعریف شده بر بازه^۵ $[a, b]$ ، که در بخش ۳۰۴ ارائه
 شد ، با توجه به انحراف تابع تقریب ساز از f ، به نقاط بازه وزنه ای مساوی می دهد . این
 امر در بسیاری از موارد مناسب است ، ولی ، در حالات زیادی ، معقولتر آن است که ، در
 بخشی از بازه ، به خطا اهمیت بیشتری داده شود . تابع w ، که بر بازه^۵ $[a, b]$ انتگرال پذیر

بوده و بر (a, b) در $w(x) \geq 0$ صدق کند ولی بر هر زیربازه از (a, b) ، $w(x) \neq 0$ ، یک تابع وزن برای این بازه نامیده می‌شود. تعریف تعامد را می‌توان به صورت زیر به حالتی که شامل توابع وزن باشد تعمیم داد:

تعریف ۶.۴. گوئیم $\{\phi_0, \phi_1, \dots, \phi_n\}$ یک مجموعه متعامد از توابع برای بازه $[a, b]$ نسبت به تابع وزن $w(x)$ است اگر

$$\int_a^b w(x)\phi_j(x)\phi_k(x) dx = \begin{cases} 0, & j \neq k \\ \alpha_k > 0, & j = k \end{cases}$$

هروقت $j \neq k$
هروقت $j = k$

توجه کنید که در حالت $w(x) \equiv 1$ ، تعریف ۶.۴ به تعریف ۴.۴ تحویل می‌شود. فرض کنیم توابع ϕ_0, \dots, ϕ_n (که لزوماً چند جمله‌ای نیستند) بر $[a, b]$ نسبت به تابع وزن $w(x)$ متعامد بوده، و $f \in C[a, b]$. ترکیب خطی

$$P(x) = \sum_{k=0}^n a_k \phi_k(x)$$

را جستجو می‌کنیم که

$$E(a_0, \dots, a_n) = \int_a^b w(x) \left[f(x) - \sum_{k=0}^n a_k \phi_k(x) \right]^2 dx$$

را مینیمم کند. جواب $P(x) = \sum_{k=0}^n a_k \phi_k(x)$ یک تقریب کمترین مربعات وزندار به f نامیده می‌شود. در اینجا می‌توان از روشی که قبلاً برای بدست آوردن a_0, \dots, a_n بکار رفت استفاده کرد، و معادلات نرمال زیر را بدست آورد:

بمازای هر $j = 0, \dots, n$

$$(14.4) \quad \sum_{k=0}^n a_k \int_a^b w(x)\phi_k(x)\phi_j(x) dx = \int_a^b w(x)\phi_j(x)f(x) dx,$$

و، بخاطر تعامد،

$$(15.4) \quad a_j = \frac{\int_a^b w(x)\phi_j(x)f(x) dx}{\int_a^b w(x)[\phi_j(x)]^2 dx}, \quad j = 0, \dots, n$$

بمازای هر $j = 0, \dots, n$

(ر.ک. تعریف ۰.۵)

در بقیه این بخش، فقط به مجموعه‌های متعامد از چند جمله‌ایها توجه داریم.

قضیه بعد، که بر روشی به نام فرایند گرام^۱ - اشمیت^۲ استوار است، طرز ساختن چند جمله‌ایهای متعامد بر $[a, b]$ نسبت به تابع وزن $w(x)$ را نشان می‌دهد.

قضیه ۷.۴. مجموعه چند جمله‌ایهای $\{\phi_0, \phi_1, \dots, \phi_n\}$ تعریف شده به طریق زیر، بر $[a, b]$ نسبت به تابع وزن $w(x)$ متعامد است:

$$\phi_0(x) \equiv 1,$$

$$\phi_1(x) = x - B_1, \quad a \leq x \leq b \quad \text{بمازای هر}$$

که در آن

$$B_1 = \frac{\int_a^b xw(x)[\phi_0(x)]^2 dx}{\int_a^b w(x)[\phi_0(x)]^2 dx},$$

وقتی $k \geq 2$

$$\phi_k(x) = (x - B_k)\phi_{k-1}(x) - C_k\phi_{k-2}(x), \quad a \leq x \leq b \quad \text{بمازای هر}$$

که در آن

$$B_k = \frac{\int_a^b xw(x)[\phi_{k-1}(x)]^2 dx}{\int_a^b w(x)[\phi_{k-1}(x)]^2 dx},$$

و

$$C_k = \frac{\int_a^b xw(x)\phi_{k-1}(x)\phi_{k-2}(x) dx}{\int_a^b w(x)[\phi_{k-2}(x)]^2 dx}$$

برهان. هر ϕ_k به شکل جملاتی با مراتب پایین‌تر + x^k است، در نتیجه، همه مخرجهای B_k و C_k ناصفرند.

به استقرا بر k ، نشان می‌دهیم که،

$$\int_a^b w(x)\phi_k(x)\phi_i(x) dx = 0, \quad i < k$$

بمازای $k = 1$

$$\begin{aligned}
 \int_a^b w(x)\phi_1(x)\phi_0(x) dx &= \int_a^b w(x)(x - B_1) dx \\
 &= \int_a^b xw(x) dx - B_1 \int_a^b w(x) dx \\
 &= \int_a^b xw(x)[\phi_0(x)]^2 dx \\
 &\quad - \left[\frac{\int_a^b xw(x)[\phi_0(x)]^2 dx}{\int_a^b w(x)[\phi_0(x)]^2 dx} \right] \int_a^b w(x)[\phi_0(x)]^2 dx \\
 &= 0.
 \end{aligned}$$

فرض کنیم نتیجه به ازای $k = n - 1$ درست باشد. در این صورت،

$$\begin{aligned}
 \int_a^b w(x)\phi_n(x)\phi_{n-1}(x) dx &= \int_a^b w(x)[(x - B_n)\phi_{n-1}(x) \\
 &\quad - C_n\phi_{n-2}(x)] \cdot \phi_{n-1}(x) dx \\
 &= \int_a^b w(x)(x - B_n)[\phi_{n-1}(x)]^2 dx \\
 &= \int_a^b xw(x)[\phi_{n-1}(x)]^2 dx \\
 &\quad - B_n \int_a^b w(x)[\phi_{n-1}(x)]^2 dx \\
 &= 0.
 \end{aligned}$$

به همین نحو (ر.ک. تمرین ۸)، می توان نشان داد که

$$(۱۶۰۴) \quad \int_a^b w(x)\phi_n(x)\phi_{n-2}(x) dx = 0.$$

به ازای $i < n - 2$

$$\begin{aligned}
 \int_a^b w(x)\phi_n(x)\phi_i(x) dx &= \int_a^b w(x)[(x - B_n)\phi_{n-1}(x) - C_n\phi_{n-2}(x)]\phi_i(x) dx \\
 &= \int_a^b w(x)x\phi_{n-1}(x)\phi_i(x) dx \\
 &= \int_a^b w(x)\phi_{n-1}(x)[\phi_{i+1}(x) + B_{i+1} \cdot \phi_i(x) \\
 &\quad + C_{i+1}\phi_{i-1}(x)] dx \\
 &= 0.
 \end{aligned}$$

نتیجه ۸.۴. بهازای هر $n > 0$ ، چند جمله‌ایهای ϕ_0, \dots, ϕ_n در قضیه ۷.۴ بر $[a, b]$ مستقل خطی اند و یک پایه برای Π_n می‌سازند، و، بهازای هر چند جمله‌ای Q_k از درجه $k < n$ ،

$$\int_a^b w(x) \phi_n(x) Q_k(x) dx = 0.$$

برهان. برای اثبات استقلال خطی ϕ_0, \dots, ϕ_n ، فرض کنیم بهازای هر $x \in [a, b]$ $0 = c_0 \phi_0(x) + \dots + c_n \phi_n(x)$ ، با ضرب در $w(x) \phi_k(x)$ ، بهازای هر $k = 0, 1, \dots, n$ ، بدست می‌آوریم

$$0 = \sum_{j=0}^n c_j w(x) \phi_j(x) \phi_k(x).$$

پس

$$0 = \sum_{j=0}^n c_j \int_a^b w(x) \phi_j(x) \phi_k(x) dx,$$

$$0 = c_k \int_a^b w(x) [\phi_k(x)]^2 dx,$$

و نتیجه می‌شود که $c_k = 0$. چون این بهازای هر $k = 0, 1, \dots, n$ درست است، $\{\phi_0, \dots, \phi_n\}$ یک مجموعه مستقل خطی است و، طبق تبصره‌های بعد از مثال ۲ از بخش ۳.۴، یک پایه برای Π_n تشکیل می‌دهد.

فرض کنیم $Q_k(x)$ یک چند جمله‌ای از درجه k باشد. در تمرین ۱۱ از بخش ۳.۴ گفته شد که اعدادی مانند c_0, \dots, c_k وجود دارند بطوری که

$$Q_k(x) = \sum_{j=0}^k c_j \phi_j(x).$$

لذا،

$$\int_a^b w(x) Q_k(x) \phi_n(x) dx = \sum_{j=0}^k c_j \int_a^b w(x) \phi_j(x) \phi_n(x) dx = 0,$$

زیرا ϕ_n ، بهازای هر $j = 0, 1, \dots, k$ ، متعامد به ϕ_j است.

مثال ۰۱. یکی از متداولترین مجموعه‌ها از چند جمله‌ایهای متعامد مجموعهٔ چند جمله‌ایهای لژاندر^۱ $\{P_n\}$ است، که بر $[-1, 1]$ نسبت به تابع وزن $w(x) \equiv 1$ متعامدند. در تعریف کلاسیک چند جمله‌ایهای لژاندر باید به‌ازای هر n ، $P_n(1) = 1$ ، و می‌توان، با استفاده از یک رابطهٔ بازگشتی، چند جمله‌ایها را وقتی $n \geq 2$ تولید کرد. این نرمالسازی در بحث ما لازم نمی‌شود، و چند جمله‌ایهای تقریب ساز کمترین مربعات تولید شده در هر حالت یکی هستند. با استفاده از روند قضیهٔ ۷.۴،

$$P_0(x) \equiv 1$$

$$B_1 = \frac{\int_{-1}^1 x \, dx}{\int_{-1}^1 dx} = 0$$

در نتیجه،

$$P_1(x) = (x - B_1)P_0(x) = x.$$

همچنین،

$$B_2 = \frac{\int_{-1}^1 x^3 \, dx}{\int_{-1}^1 x^2 \, dx} = 0$$

و

$$C_2 = \frac{\int_{-1}^1 x^2 \, dx}{\int_{-1}^1 1 \, dx} = \frac{1}{3}.$$

در نتیجه،

$$P_2(x) = (x - B_2)P_1(x) - C_2P_0(x) = (x - 0)x - \frac{1}{3} \cdot 1 = x^2 - \frac{1}{3}.$$

بالاخره،

$$B_3 = \frac{\int_{-1}^1 x \cdot (x^2 - \frac{1}{3})^2 \, dx}{\int_{-1}^1 (x^2 - \frac{1}{3})^2 \, dx} = 0,$$

و

$$C_3 = \frac{\int_{-1}^1 x \cdot x \cdot (x^2 - \frac{1}{3}) \, dx}{\int_{-1}^1 x^2 \, dx} = \frac{\frac{8}{45}}{\frac{2}{3}} = \frac{4}{15},$$

و در نتیجه،

$$P_3(x) = (x - B_3)P_2(x) - C_3P_1(x) \\ = x \cdot (x^2 - \frac{1}{3}) - \frac{4}{15}x = x^3 - \frac{4}{15}x.$$

مثال ۲. مجموعه دیگری از چند جمله‌ای‌های متعامد که بعداً " در این فصل بکار می‌رود چند جمله‌ای‌های چبیشف^۱ $\{T_n\}$ نام دارند. آنها را می‌توان، با استفاده از تابع وزن $w(x) = (1 - x^2)^{-1/2}$ ، به وسیله قضیه ۷.۴ ، بر بازه $(-1, 1)$ بدست آورد. ما در اینجا چند جمله‌ای‌های چبیشف را طور دیگر بدست آورده، و بعد نشان می‌دهیم که در خواص لازم تعامد صدق می‌کنند.

بمازای $x \in [-1, 1]$ ، تعریف می‌کنیم :

$$T_n(x) = \cos[n \arccos x] , n \geq 0$$

با جایگذاری $\theta = \arccos x$ ، آن را به

$$T_n(x) = \cos(n\theta)$$

تغییر می‌دهیم ، که در آن $\theta \in [0, \pi]$.

با توجه به

$$T_{n+1}(x) = \cos((n+1)\theta) \\ = \cos(n\theta)\cos\theta - \sin(n\theta)\sin\theta$$

$$T_{n-1}(x) = \cos((n-1)\theta) \\ = \cos(n\theta)\cos\theta + \sin(n\theta)\sin\theta,$$

می‌توان یک رابطه بازگشتی بدست آورد. در نتیجه ،

$$T_{n+1}(x) = 2 \cos(n\theta)\cos\theta - T_{n-1}(x).$$

اگر به متغیر x بازگردیم ، خواهیم داشت :

$$(۱۷.۴) \quad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) , 1 \leq n$$

چون

$$T_0(x) = \cos(0 \cdot \arccos x) = 1$$

$$T_1(x) = \cos(1 \arccos x) = x$$

با استفاده از معادله (۱۷.۴) ، چند جمله‌ای‌های چبیشف به روشی تسلسلی به آسانی بدست

می‌آیند :

$$\begin{aligned}T_2(x) &= 2xT_1(x) - T_0(x) = 2x^2 - 1, \\T_3(x) &= 2xT_2(x) - T_1(x) = 4x^3 - 3x, \\T_4(x) &= 2xT_3(x) - T_2(x) = 8x^4 - 8x^2 + 1,\end{aligned}$$

و غیره،

برای اثبات تعامد چند جمله‌ایهای چبیشف،

$$\int_{-1}^1 \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}} dx = \int_{-1}^1 \frac{\cos(n \arccos x)\cos(m \arccos x)}{\sqrt{1-x^2}} dx$$

را در نظر می‌گیریم. با جایگذاری مجدد $\theta = \arccos x$ ، انتگرال زیر بدست می‌آید:

$$\begin{aligned}\int_{-1}^1 \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}} dx &= \int_{\pi}^0 \frac{\cos(n\theta)\cos(m\theta)}{\sin \theta} (-\sin \theta d\theta) \\&= - \int_{\pi}^0 \cos(n\theta)\cos(m\theta) d\theta \\&= \int_0^{\pi} \cos(n\theta)\cos(m\theta) d\theta.\end{aligned}$$

فرض کنیم $n \neq m$ ، چون

$$\cos n\theta \cos m\theta = \frac{1}{2}[\cos(n+m)\theta + \cos(n-m)\theta],$$

$$\begin{aligned}\int_{-1}^1 \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}} dx &= \frac{1}{2} \int_0^{\pi} \cos((n+m)\theta) d\theta \\&\quad + \frac{1}{2} \int_0^{\pi} \cos((n-m)\theta) d\theta \\&= \left[\frac{1}{2(n+m)} \sin((n+m)\theta) \right. \\&\quad \left. + \frac{1}{2(n-m)} \sin((n-m)\theta) \right] \Big|_0^{\pi} \\&= 0.\end{aligned}$$

همچنین، به روش مشابه می‌توان نشان داد که،

$$(18.4) \quad \int_{-1}^1 \frac{T_n^2(x)}{\sqrt{1-x^2}} dx = \frac{\pi}{2}, \quad n \geq 1$$

بمازای هر

انگیزه اصلی برای توجه به کمترین مربعات وزندار این است که تابع وزن بربخشیهای متمایز یک بازه تاکید متفاوت می‌گذارد. مثلاً، تابع وزن

$$w(x) = \frac{1}{\sqrt{1-x^2}}$$

نزدیک مرکز بازه $(-1, 1)$ تاکید کمتر، و وقتی $|x|$ نزدیک یک است تاکید بیشتری دارد. ممکن است محاسبه انتگرالها در روشهای تقریب کمترین مربعات پرزحمت باشد؛ با اینحال، روشهای عددی انتگرالگیری در دسترس اند و در فصل ۵ ارائه خواهند شد. فرایند بحث شده در قضیه ۷.۴، به ازای هر بازه $[a, b]$ و تابع وزن $w(x)$ ، روشی برای رسیدن به یک فرمول بازگشتی از نوع زیر به ما می‌دهد:

$$(19.4) \quad \phi_k(x) = (x - B_k)\phi_{k-1}(x) - C_k\phi_{k-2}(x), \quad k \geq 2$$

اگر این فرمول معلوم باشد، و ضرایب

$$a_k = \int_a^b w(x)\phi_k(x)f(x) dx \Big/ \int_a^b w(x)[\phi_k(x)]^2 dx$$

بدست آمده باشند، با استفاده از الگوریتم زیر، می‌توان تقریب کمترین مربعات

$$P_n(x) = \sum_{k=0}^n a_k \phi_k(x)$$

برای تابع f را در هر نقطه به آسانی حساب کرد.

الگوریتم محاسبه کمترین مربعات ۱۰.۴

برای محاسبه تقریب کمترین مربعات $P_n(x) = \sum_{k=0}^n a_k \phi_k(x)$ به تابع f ، در نقطه x ، که به ازای هر k ، $a_k = \int_a^b w(x)\phi_k(x)f(x) dx / \int_a^b w(x)[\phi_k(x)]^2 dx$ ، و مجموعه‌ای از توابع متعامد نسبت به تابع وزن $w(x)$ است:

مرحله ۱. قرار می‌دهیم

$$b_{n+2} = 0,$$

$$b_{n+1} = 0,$$

$$B_{n+1} = 0,$$

$$C_{n+2} = 0,$$

$$C_{n+1} = 0.$$

مرحله ۲. به ازای $k = n, n-1, \dots, 1$ ، قرار می‌دهیم $B_k = \frac{\int_a^b xw(x)[\phi_{k-1}(x)]^2 dx}{\int_a^b w(x)[\phi_{k-1}(x)]^2 dx}$

مرحله ۳. به ازای هر $k = n, n-1, \dots, 2$ ، قرار می‌دهیم

$$C_k = \frac{\int_a^b xw(x)\phi_{k-1}(x)\phi_{k-2}(x) dx}{\int_a^b w(x)[\phi_{k-2}(x)]^2 dx}$$

مرحله ۴. به ازای هر $k = n, n-1, \dots, 0$ ، قرار می‌دهیم

$$b_k = (x - B_{k+1})b_{k+1} - C_{k+2}b_{k+2} + a_k$$

مرحله ۵. روند تمام خواهد بود؛ $b_0 = P_n(x)$.

برای تحقیق صحت الگوریتم ۱.۴، توجه می‌کنیم که

$$a_k = b_k - (x - B_{k+1})b_{k+1} + C_{k+2}b_{k+2},$$

در نتیجه،

$$\begin{aligned} P_n(x) &= \sum_{k=0}^n a_k \phi_k(x) \\ &= \sum_{k=0}^n [b_k - (x - B_{k+1})b_{k+1} + C_{k+2}b_{k+2}] \phi_k(x) \\ &= b_0 \phi_0(x) + b_1 [\phi_1(x) - (x - B_1)\phi_0(x)] \\ &\quad + \sum_{k=2}^n b_k [\phi_k(x) - (x - B_k)\phi_{k-1}(x) + C_k \phi_{k-2}(x)] \\ &= b_0 \phi_0(x) \\ &= b_0. \end{aligned}$$

مثال ۳. با بازگشت به چند جمله‌ایهای چبیشف تعریف شده در مثال ۲ بر $[-1, 1]$ با

$$T_0(x) = 1, \quad T_1(x) = x$$

۹

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad n \geq 1$$

ملاحظه می‌کنیم که $T_n(x)$ ، به ازای هر $n \geq 1$ ، دارای ضریب پیشرو 2^{n-1} است. این به

استقرا نتیجه می شود ، زیرا

$$T_2(x) = 2x^2 - 1 = 2^{2-1}x^2 - 1,$$

$$T_3(x) = 2xT_2(x) - T_1(x) = 2^{3-1}x^3 + \dots$$

$$T_4(x) = 2xT_3(x) - T_2(x) = 2^{4-1}x^4 + \dots$$

و غیره. استفاده از چند جمله ایها با ضریب پیشرو 1 اغلب مطلوب است ، که نیاز به تعریف زیر دارد :

$$\tilde{T}_n(x) = 2^{1-n}T_n(x) , \quad n \geq 1$$

فرمول بازگشتی برای \tilde{T}_n عبارت خواهد بود از

$$\tilde{T}_0(x) \equiv 1, \quad \tilde{T}_1(x) \equiv x, \quad \tilde{T}_2(x) = x^2 - \frac{1}{2},$$

و

$$(۲۰۰۴) \quad \tilde{T}_{n+1}(x) = x\tilde{T}_n(x) - \frac{1}{2}\tilde{T}_{n-1}(x) , \quad n \geq 2$$

برای استفاده از الگوریتم ۱۰۴ جهت محاسبه

$$P_4(x) = \sum_{k=0}^4 \frac{(-1)^k}{k^2 + 1} \tilde{T}_k(x)$$

در $x = \frac{1}{2}$ ، ابتدا ملاحظه می کنیم که چون به ازای هر $k = 0, 1, 2, 3, 4$ ، $a_k = (-1)^k / (k^2 + 1)$ معادله بازگشتی (۲۰۰۴) ایجاب می کند که

$$B_1 = B_2 = B_3 = B_4 = 0, \quad C_2 = \frac{1}{2}, \quad C_3 = C_4 = \frac{1}{4}.$$

طبق الگوریتم ۱۰۴ ، $n = 4$ ، در نتیجه ، $B_5 = 0$ ، $C_6 = C_5 = 0$ ، و $b_6 = b_5 = 0$ ، بنابراین ،

$$b_4 = (x - B_5)b_5 - C_6b_6 + a_4 = .05882353,$$

$$b_3 = (x - B_4)b_4 - C_5b_5 + a_3 = -.07058824,$$

$$b_2 = (x - B_3)b_3 - C_4b_4 + a_2 = .14999999,$$

$$b_1 = (x - B_2)b_2 - C_3b_3 + a_1 = -.40735294,$$

و

$$P_4(.5) = b_0 = (x - B_1)b_1 - C_2b_2 + a_0 = .7213235.$$

بمازای هر چند جمله‌ای چیشف $T_n(x)$ ،

$$\int_{-1}^1 \frac{T_n^2(x)}{\sqrt{1-x^2}} dx = \frac{\pi}{2} .$$

۲ . مقدار چند جمله‌ای لژاندر $P_4(x)$ در $x = .75$ را

(آ) با جایگذاری مستقیم در $P_4(x)$ ؛

(ب) با استفاده از الگوریتم ۱۰۴

حساب کنید .

۳ . با استفاده از قضیه ۷۰۴ و الگوریتم ۱۰۴ ، برای هریک از توابع زیر ، چند جمله‌ای

تقریب‌ساز کمترین مربعات از درجه سه را تعیین ، و L_2 - نرم خطا را بدست آورید :

$$(آ) f(x) = \cos x, 0 \leq x \leq 1$$

$$(ب) f(x) = \ln x, 1 \leq x \leq 2$$

$$(پ) f(x) = x^4, 0 \leq x \leq 1$$
 ، با تابع وزن $w(x) = e^x$ ؛

$$(ت) f(x) = e^{-1x}, 0 \leq x \leq 1$$

۴ . نشان دهید که اگر $\{\phi_0, \dots, \phi_n\}$ یک مجموعه متعامد از توابع بر $[a, b]$ نسبت به

تابع وزن $w(x)$ باشد ، $\{\phi_0, \dots, \phi_n\}$ یک مجموعه مستقل خطی است .

۵ . معادلات (۱۴۰۴) و (۱۵۰۴) را بدست آورید .

۶ . با استفاده از روش قضیه ۷۰۴ ، L_2 ، L_3 ، و L_1 را طوری حساب کنید که

$\{L_0, L_1, L_2, L_3\}$ یک مجموعه متعامد از چند جمله‌ایهای تعریف شده بر $(0, \infty)$ نسبت

به تابع وزن $w(x) = e^{-x}$ بوده و $L_0(x) \equiv 1$. چند جمله‌ایهای حاصل از این روند چند

جمله‌ایهای لاگور^۱ نامیده می‌شوند .

۷ . P_3 و P_4 ، یعنی چند جمله‌ایهای لژاندر درجه چهار و پنج که در مثال ۱ معرفی

شدند ، را حساب کنید .

۸ . نشان دهید که مرحله استقرای

$$\int_a^b w(x) \phi_n(x) \phi_{n-2}(x) dx = 0$$

در قضیه ۷۰۴ برقرار است .

۵.۴ چند جمله‌ایهای چبیشف و با صرفه‌سازی سریهای توانی

در این بخش، بررسی چند جمله‌ایهای چبیشف را که در مثالهای ۲ و ۳ از بخش ۴.۴ آغاز شد ادامه می‌دهیم. این بررسی به دو نتیجه زیر منجر می‌شود: "اولا"، جای بهینه نقاط درونیاب برای مینیمم سازی خطا در درونیابی لاگرانژ؛ "ثانیا"، ابزارهای تقلیل درجه یک چند جمله‌ای تقریب ساز با کمترین زبان در دقت.

به یاد می‌آوریم که تعریف چند جمله‌ای چبیشف به صورت زیر است:

$$(21.4) \quad T_n(x) = \cos[n \arccos x], \quad n = 0, 1, 2, \dots \text{ و } x \in [-1, 1]$$

جایگذاری $\theta = \arccos x$ ایجاب می‌کند که،

$$(22.4) \quad T_n(x) = \cos(n\theta), \quad n = 0, 1, 2, \dots \text{ و } \theta \in [0, \pi]$$

به کمک معادله (۲۲.۴) و اتحادهای مثلثاتی مختلف، در مثال ۲ از بخش ۴.۴ نشان دادیم که،

$$(23.4) \quad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad x \in [-1, 1] \text{ و } n = 1, 2, \dots$$

با استفاده از $T_0(x) = 1$ و $T_1(x) = x$ می‌توان از رابطه بازگشتی (۲۳.۴) نتیجه

گرفت که، به ازای هر $n \geq 1$ ، $T_n(x)$ یک چند جمله‌ای از درجه n با ضریب پیشرو 2^{n-1} است.

اولین نتیجه در این بخش مربوط است به صفرهای چند جمله‌ای T_n

قضیه ۹.۴. چند جمله‌ای چبیشف T_n از درجه $n \geq 1$ دارای n صفر ساده در $[-1, 1]$ در نقاط زیر است.

$$(24.4) \quad \bar{x}_k = \cos\left(\frac{2k-1}{2n}\pi\right), \quad k = 1, 2, \dots, n \text{ به ازای هر}$$

بعلاوه، T_n اکسترممهای مطلق خود را در نقاط زیر می‌گیرد:

$$(25.4) \quad \bar{x}'_k = \cos\left(\frac{k\pi}{n}\right), \quad k = 0, \dots, n \text{ به ازای هر}$$

بطوری که،

$$(26.4) \quad T_n(\bar{x}'_k) = (-1)^k, \quad k = 0, 1, \dots, n \text{ به ازای هر}$$

$$\bar{x}_k = \cos\left(\frac{2k-1}{2n}\pi\right), \quad k = 1, 2, \dots, n$$

آنگاه

$$\begin{aligned} T_n(\bar{x}_k) &= \cos(n \arccos \bar{x}_k) \\ &= \cos\left(n \arccos\left(\cos\left(\frac{2k-1}{2n}\pi\right)\right)\right) \\ &= \cos\left(\frac{2k-1}{2}\pi\right) = 0; \end{aligned}$$

در نتیجه، \bar{x}_k ، به ازای هر $k = 1, 2, \dots, n$ ، یک صفر T_n است. چون T_n یک چندجمله‌ای از درجه n است، همه صفرهای T_n باید به این شکل باشند. برای اثبات قسمت دوم این قضیه، ابتدا ملاحظه می‌کنیم که

$$\begin{aligned} T'_n(x) &= \frac{d}{dx} [\cos(n \arccos x)] \\ &= \frac{n \sin(n \arccos x)}{\sqrt{1-x^2}}, \end{aligned}$$

و، وقتی $1 \leq k \leq n-1$ ،

$$\begin{aligned} T'_n(\bar{x}_k) &= \frac{n \sin\left(n \arccos\left(\cos\left(\frac{k\pi}{n}\right)\right)\right)}{\sqrt{1 - \cos^2\left(\frac{k\pi}{n}\right)}} \\ &= \frac{n \sin(k\pi)}{\sin\left(\frac{k\pi}{n}\right)} = 0. \end{aligned}$$

بعلاوه، چون T_n یک چندجمله‌ای از درجه n است، T'_n یک چندجمله‌ای از درجه $(n-1)$ بوده و همه صفرهای T'_n در این نقاط رخ می‌دهند. تنها اکسترممهای ممکن دیگر تابع T_n در نقاط انتهایی بازه $[-1, 1]$ رخ خواهند داد؛ یعنی، در $\bar{x}_0 = 1$ و $\bar{x}_n = -1$.

چون

$$\begin{aligned} T_n(\bar{x}_k) &= \cos\left(n \arccos\left(\cos\left(\frac{k\pi}{n}\right)\right)\right) \\ &= \cos(k\pi) = (-1)^k, \end{aligned}$$

در هر مقدار زوج k ماکزیمم و در هر مقدار فرد مینیمم خواهیم داشت .

در مثال ۳ از بخش ۴.۴ ، چند جمله‌ایهای چبیشف تکین (چند جمله‌ایهای با ضریب پیشرو ۱) \tilde{T}_n به صورت زیر تعریف شدند .

$$(۲۷.۴) \quad \tilde{T}_0(x) = 1, \quad \tilde{T}_n(x) = 2^{1-n}T_n(x), \quad n \geq 1$$

در آن مثال همچنین نشان دادیم که این چند جمله‌ایها در رابطه بازگشتی زیر صدق می‌کنند :

$$(۲۸.۴) \quad \tilde{T}_2(x) = x\tilde{T}_1(x) - \frac{1}{2}\tilde{T}_0(x); \quad \tilde{T}_{n+1}(x) = x\tilde{T}_n(x) - \frac{1}{4}\tilde{T}_{n-1}(x), \quad n \geq 2$$

بخاطر رابطه خطی بین T_n و \tilde{T}_n ، قضیه ۹.۴ ایجاب می‌کند که صفرهای \tilde{T}_n نیز نقاط زیر خواهند بود :

$$\bar{x}_k = \cos\left(\frac{k\pi}{n}\right), \quad k = 1, 2, \dots, n \text{ هر بازای}$$

و مقادیر اکستریم \tilde{T}_n در نقاط زیر رخ خواهند داد :

$$\bar{x}_k = \cos\left(\frac{2k-1}{2n}\pi\right), \quad k = 0, 1, 2, \dots, n \text{ هر بازای}$$

در این مقادیر اکستریم ، بازای $n \geq 1$ ،

$$(۲۹.۴) \quad \tilde{T}_n(\bar{x}_k) = \frac{(-1)^k}{2^{n-1}}, \quad k = 0, 1, 2, \dots, n \text{ هر بازای}$$

اگر $\tilde{\Pi}_n$ مجموعه تمام چند جمله‌ایهای تکین از درجه n باشد ، رابطه (۲۹.۴) به خاصیت مینیمم سازی مهمی منجر می‌شود که چند جمله‌ایهای \tilde{T}_n را از سایر اعضای $\tilde{\Pi}_n$ متمایز می‌سازد .

قضیه ۱۰.۴ . از همه چند جمله‌ایهای موجود در رده $\tilde{\Pi}_n$ ، چند جمله‌ایهای به شکل \tilde{T}_n ، وقتی $n \geq 1$ ، دارای کمترین L_∞ نرم بر $[-1, 1]$ هستند ؛ یعنی ،

$$(۳۰.۴) \quad \frac{1}{2^{n-1}} = \|\tilde{T}_n\|_\infty \leq \|P_n\|_\infty, \quad P_n \in \tilde{\Pi}_n \text{ هر بازای}$$

تساوی در (۳۰.۴) فقط وقتی می‌تواند رخ دهد که $P_n = \tilde{T}_n$.

برهان . فرض کنیم $P_n \in \tilde{\Pi}_n$ و

$$\max_{x \in [-1, 1]} |P_n(x)| = \|P_n\|_\infty \leq \frac{1}{2^{n-1}}.$$

همچنین، $Q = \tilde{T}_n - P_n$ ، چون \tilde{T}_n و P_n هر دو چند جمله‌ایهایی تکین از درجه n هستند، Q یک چند جمله‌ای از درجه حداکثر $(n-1)$ است. بعلاوه، در نقاط اکستریم \tilde{T}_n ،

$$Q(\bar{x}_k) = \tilde{T}_n(\bar{x}_k) - P_n(\bar{x}_k) = \frac{(-1)^k}{2^{n-1}} - P_n(\bar{x}_k).$$

این امر که

$$|P_n(\bar{x}_k)| \leq \frac{1}{2^{n-1}}$$

ایجاب می‌کند که، به‌ازای $k = 1, 2, \dots, n$ ،

$$Q(\bar{x}_k) \leq 0 \text{ وقتی } k \text{ فرد باشد،}$$

و

$$Q(\bar{x}_k) \geq 0 \text{ وقتی } k \text{ زوج باشد.}$$

چون Q پیوسته است، می‌توان با استفاده از قضیه مقدار میانی نشان داد که چند جمله‌ای درجه $(n-1)$ Q باید دست کم n صفر در بازه $[-1, 1]$ داشته باشد، که بوضوح ناممکن است مگر آنکه $Q \equiv 0$. این ایجاب می‌کند که $P_n = \tilde{T}_n$ ، که قضیه را ثابت خواهد کرد.

با استفاده از قضیه ۱۰.۴، می‌توان به این سوال که گرهای درونیاب را باید کجا قرارداد تا در درونیابی لاگرانژ خطا مینیمم شود پاسخ داد. قضیه ۳.۳ (ص ۱۲۵)، وقتی بر $[-1, 1]$ اعمال شود، می‌گوید که، هرگاه x_0, \dots, x_n نقاط متمایز در بازه $[-1, 1]$ بوده و $f \in C^{n+1}[-1, 1]$ ، آنگاه، به‌ازای هر $x \in [-1, 1]$ ، نقطه‌ای مانند $\xi(x)$ در $(-1, 1)$ هست بطوری که

$$f(x) = P(x) + \frac{f^{(n+1)}(\xi(x))}{(n+1)!} (x-x_0)(x-x_1)\cdots(x-x_n),$$

که در آن P چند جمله‌ای درونیاب لاگرانژ است. مینیمم سازی خطا با انتخاب دقیق نقاط x_0, \dots, x_n عموماً "معادل است با یافتن نقاط x_0, \dots, x_n که کمیت

$$|(x - x_0)(x - x_1) \cdots (x - x_n)|$$

را در سراسر بازه $[-1, 1]$ مینیمم کند. این را می توان با انتخاب نقاط x_0, x_1, \dots, x_n طوری انجام داد که مقدار ماکزیمم $|(x - x_0)(x - x_1) \cdots (x - x_n)|$ را مینیمم کند. چون $(x - x_0)(x - x_1) \cdots (x - x_n)$ یک چند جمله ای تکین از درجه $(n + 1)$ است، بنابراین قضیه ۱۰۴، این مینیمم بدست می آید اگر فقط اگر

$$(x - x_0)(x - x_1) \cdots (x - x_n) = \tilde{T}_{n+1}(x).$$

وقتی نقطه x_k ، $(k + 1)$ مین صفر \tilde{T}_{n+1} ، به ازای هر $k = 0, 1, \dots, n$ ، اختیار شود، یعنی وقتی x_k به صورت زیر اختیار شود:

$$(۳۱۰۴) \quad \bar{x}_{k+1} = \cos \frac{2k + 1}{2(n + 1)} \pi,$$

مقدار ماکزیمم $|(x - x_0)(x - x_1) \cdots (x - x_n)|$ مینیمم می شود. چون $\|\tilde{T}_{n+1}\|_\infty = 1/2^n$ این ایجاب می کند که، به ازای هر انتخاب نقاط x_0, x_1, \dots, x_n در بازه $[-1, 1]$ ،

$$\begin{aligned} \frac{1}{2^n} &= \max_{x \in [-1, 1]} |(x - \bar{x}_1)(x - \bar{x}_2) \cdots (x - \bar{x}_{n+1})| \\ &\leq \max_{x \in [-1, 1]} |(x - x_0)(x - x_1) \cdots (x - x_n)|, \end{aligned}$$

با تغییر متغیر $\bar{x} = \frac{1}{2}[(b - a)x + a + b]$ و تبدیل نقاط \bar{x}_k در بازه $[-1, 1]$ به نقاط نظیرشان \bar{x}_k در بازه $[a, b]$ ، می توان روش فوق را برای انتخاب نقاطی که خطای درونیابی را مینیمم می کند به آسانی به بازه بسته کلی $[a, b]$ تعمیم داد.

مثال ۱. فرض کنیم بر $[0, 1.5]$ ، $f(x) = xe^x$. دو چند جمله ای درونیاب از درجه حداکثر سه را می سازیم. ابتدا، با استفاده از گره های متساوی الفاصله درونیابی $x_0 = 0, x_1 = .5, x_2 = 1, x_3 = 1.5$ بدست می آوریم

$$L_0(x) = -1.3333x^3 + 4.0000x^2 - 3.6667x + 1,$$

$$L_1(x) = 4.0000x^3 - 10.000x^2 + 6.0000x,$$

$$L_2(x) = -4.0000x^3 + 8.0000x^2 - 3.0000x,$$

$$L_3(x) = 1.3333x^3 - 2.0000x^2 + .66667x.$$

با مقادیر مذکور در جدول ۱۱.۴ ، چندجمله‌ای درونیاب عبارت خواهد بود از

$$P_3(x) = 1.3875x^3 + .057570x^2 + 1.2730x.$$

x	$f(x) = xe^x$
0	.0000
.5	.824361
1.0	2.71828
1.5	6.72253

جدول ۱۱.۴

برای ارائه چندجمله‌ای درونیاب دوم ، صفرهای $\bar{x}_k = \cos((2k + 1)/8)\pi$ ، $k = 0, 1, 2, 3$ از \bar{T}_4 را ، با استفاده از تبدیل

$$\bar{x}_k = .75 + .75\bar{x}_k$$

از $[-1, 1]$ به $[0, 1.5]$ منتقل می‌کنیم ، خواهیم داشت

$$\bar{x}_0 = 1.44291, \quad \bar{x}_1 = 1.03701, \quad \bar{x}_2 = .46299,$$

و

$$\bar{x}_3 = .05709.$$

در این صورت ، چندجمله‌ایهای ضرایب لاگرانژ به صورت زیر محاسبه می‌شوند :

$$\bar{L}_0(x) = 1.8142x^3 - 2.8249x^2 + 1.0264x - .049728,$$

$$\bar{L}_1(x) = -4.3799x^3 + 8.5977x^2 - 3.4026x + .16705,$$

$$\bar{L}_2(x) = 4.3799x^3 - 11.112x^2 + 7.1738x - .37415,$$

$$\bar{L}_3(x) = -1.8142x^3 + 5.3390x^2 - 4.7976x + 1.2568.$$

مقادیر تابعی لازم برای این چندجمله‌ایها در جدول ۱۲.۴ آمده‌اند ، و چند -

جمله‌ای درونیاب از درجه حداکثر سه عبارت است از

$$\bar{P}_3(x) = 1.3811x^3 + .044445x^2 + 1.3030x - .014357.$$

برای مقایسه ، در جدول ۱۳.۴ مقادیر مختلف x ، همراه با مقادیر

x	$f(x) = xe^x$
$\tilde{x}_0 = 1.44291$	6.10783
$\tilde{x}_1 = 1.03701$	2.92517
$\tilde{x}_2 = .46299$.73560
$\tilde{x}_3 = .05709$.060444

جدول ۱۲.۴

x	$f(x) = xe^x$	$P_3(x)$	$ xe^x - P_3(x) $	$\bar{P}_3(x)$	$ xe^x - \bar{P}_3(x) $
.15	.1743	.1969	.0226	.1868	.0125
.25	.3210	.3435	.0225	.3358	.0148
.35	.4967	.5121	.0154	.5064	.0097
.65	1.245	1.233	.0120	1.231	.0140
.75	1.588	1.572	.0160	1.571	.0170
.85	1.989	1.976	.0130	1.973	.0160
1.15	3.632	3.650	.0180	3.643	.0110
1.25	4.363	4.391	.0280	4.381	.0180
1.35	5.208	5.237	.0290	5.224	.0160

جدول ۱۳.۴

نتیجه ۱۱.۴ . هرگاه P چند جمله‌ای درونیاب از درجه n حداکثر n با گره‌هایی در صفرهای $T_{n+1}(x)$ باشد، آنگاه، به‌زای هر $f \in C^{n+1}[-1, 1]$ ،

نتیجه ۱۲.۴ . از این جدول می‌توان دید که، اگرچه خطای استفاده از P_3 نزدیک مرکز جدول از خطای استفاده از \bar{P}_3 کمتر است، خطای ماکزیمم استفاده از \bar{P}_3 به‌طور قابل ملاحظه‌ای کمتر می‌باشد.

نتیجه ۱۳.۴ . زیرا می‌توان با استفاده از قضیه ۱۰.۴ به‌آسانی بدست آورد. برهان در تمرین ۱ خواسته شده است.

نتیجه ۱۱.۴ . هرگاه P چند جمله‌ای درونیاب از درجه n حداکثر n با گره‌هایی در صفرهای $T_{n+1}(x)$ باشد، آنگاه، به‌زای هر $f \in C^{n+1}[-1, 1]$ ،

$$\max_{x \in [-1, 1]} |f(x) - P(x)| \leq \frac{1}{2^n(n+1)!} \max_{x \in [-1, 1]} |f^{(n+1)}(x)|$$

همچنین، از چند جمله‌ایهای چبیشف می‌توان برای تقلیل درجه یک چند جمله‌ای تقریب‌ساز با کمترین زیان در دقت استفاده کرد. این روش بخصوص وقتی مفید است که چند جمله‌ای تقریب‌ساز مورد استفاده یک چند جمله‌ای تیلور باشد. با آنکه چند جمله‌ایهای تیلور در مجاورت نقطه‌ای که بسط داده شده‌اند خیلی دقیق‌اند، اثرشان وقتی در نقاط دورتر از این نقطه بکار می‌روند سرعت از بین می‌رود. به این دلیل، ممکن است برای تحمل خطایی معین به یک چند جمله‌ای تیلور از درجه بالا نیاز باشد. چون چند جمله‌ایهای چبیشف کمترین مقدار قدر مطلق ماکزیمم را دارند که بر یک بازه به‌طور یکنواخت پخش شده است، می‌توان از آنها برای تقلیل درجه چند جمله‌ای تیلور بدون تجاوز از تحمل خطا استفاده کرد.

مثال زیر روش مربوط را توضیح می‌دهد.

مثال ۲. تابع $f(x) = e^x$ را می‌توان بر بازه $[-1, 1]$ با چند جمله‌ای تیلور درجه چهار که حول صفر بسط داده شده است:

$$P_4(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}$$

و با خطای

$$|R_4(x)| = \frac{|f^{(5)}(\xi(x))||x^5|}{120} \leq \frac{e}{120} \approx 0.023, \quad -1 \leq x \leq 1$$

تقریب کرد.

فرض کنیم خطای واقعی 0.05 قابل تحمل باشد، و حالتی را در نظر می‌گیریم که در آن جمله شامل x^4 در چند جمله‌ای تیلور با چند جمله‌ایهای چبیشف معادل از درجه نایب‌تر از چهار عوض شده است.

پیش از ادامه این مثال، نمایش صریح x^4 را بر حسب T_0, T_1, \dots, T_4 ، به‌ازای چند عدد صحیح مثبت کوچکتر k ، ثبت می‌کنیم. صحت جدول ۱۴.۴ در تمرین ۲ خواسته شده است.

بنابراین،

$$\begin{aligned} P_4(x) &= 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}\left[\frac{3}{8}T_0(x) + \frac{1}{2}T_2(x) + \frac{1}{8}T_4(x)\right] \\ &= 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{64}T_0(x) + \frac{1}{48}T_2(x) + \frac{1}{192}T_4(x) \end{aligned}$$

$$\begin{aligned}
 &= 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{64} + \frac{1}{48}(2x^2 - 1) + \frac{1}{192}T_4(x) \\
 &= (1 + \frac{1}{64} - \frac{1}{48}) + x + (\frac{1}{2} + \frac{1}{24})x^2 + \frac{1}{6}x^3 + \frac{1}{192}T_4(x) \\
 &= \frac{191}{192} + x + \frac{13}{24}x^2 + \frac{1}{6}x^3 + \frac{1}{192}T_4(x).
 \end{aligned}$$

k	T_k	x^k
0	1	T_0
1	x	T_1
2	$2x^2 - 1$	$\frac{1}{2}T_0 + \frac{1}{2}T_2$
3	$4x^3 - 3x$	$\frac{3}{4}T_1 + \frac{1}{4}T_3$
4	$8x^4 - 8x^2 + 1$	$\frac{3}{8}T_0 + \frac{1}{2}T_2 + \frac{1}{8}T_4$
5	$16x^5 - 20x^3 + 5x$	$\frac{5}{8}T_1 + \frac{5}{16}T_3 + \frac{1}{16}T_5$
6	$32x^6 - 48x^4 + 18x^2 - 1$	$\frac{5}{16}T_0 + \frac{15}{16}T_2 + \frac{3}{16}T_4 + \frac{3}{32}T_6$
7	$64x^7 - 112x^5 + 56x^3 - 7x$	$\frac{35}{64}T_1 + \frac{7}{64}T_3 + \frac{7}{64}T_5 + \frac{1}{64}T_7$

جدول ۱۴.۴

اما، چون $\|T_4\|_\infty = 1$ ، داریم $|\frac{1}{192}T_4(x)| \leq \frac{1}{192} = .0053$ ، در نتیجه،

$$|R_4(x)| + |\frac{1}{192}T_4(x)| \leq .023 + .0053 = .0283$$

که هنوز از خطای قابل تحمل 05. کمتر است. در نتیجه، جمله درجه چهارم $(1/192)T_4(x)$ را می توان از چند جمله ای تقریب ساز حذف کرد و تقریب مطلوب را باز هم بدست آورد. چند جمله ای درجه سه

$$P_3(x) = \frac{191}{192} + x + \frac{13}{24}x^2 + \frac{1}{6}x^3$$

تقریب مطلوب بر بازه $[-1, 1]$ خواهد بود.

در کوشش برای حذف جمله درجه سه، باید x^3 را با $\frac{3}{4}T_1 + \frac{1}{4}T_3$ عوض کرد که

نتیجه می شود

$$\begin{aligned}
 P_3(x) &= \frac{191}{192} + x + \frac{13}{24}x^2 + \frac{1}{6}[\frac{3}{4}T_1(x) + \frac{1}{4}T_3(x)] \\
 &= \frac{191}{192} + \frac{9}{8}x + \frac{13}{24}x^2 + \frac{1}{24}T_3(x).
 \end{aligned}$$

بهرحال، $\max_{x \in [-1, 1]} |\frac{1}{24}T_3(x)| = .0417$ ، که، وقتی با خطای ممکن 0283. که قبلاً بدست آمد تلفیق شود، کران خطای 07. بدست می آید، که از خطای قابل تحمل 05. بیشتر است. لذا، $P_3(x)$ چند جمله ای با کمترین درجه و مناسب این تقریب است. در

جدول ۱۵.۴ ، تابع و چندجمله‌ایهای تقریب ساز در چند نقطه انتخابی ثبت شده‌اند ، که در آن $P_2(x)$ چندجمله‌ای

$$P_2(x) = \frac{191}{192} + \frac{9}{8}x + \frac{13}{24}x^2.$$

می‌باشد .

با آنکه کران خطا برای چندجمله‌ای

$$P_2(x) = \frac{191}{192} + \frac{9}{8}x + \frac{13}{24}x^2$$

از خطای قابل تحمل 05. بیشتر بود ، ولی درایه‌های جدول کمتر از آن کران می‌باشند .

x	e^x	$P_4(x)$	$P_3(x)$	$P_2(x)$
-0.75	.47237	.47412	.47917	.45573
-0.25	.77880	.77881	.77604	.74740
0	1.00000	1.00000	.99479	.99479
.25	1.28403	1.28402	1.28125	1.30990
.75	2.11700	2.11475	2.11979	2.14323

جدول ۱۵.۴

مجموعه تمرینات ۵.۴

- ۱ . نتیجه ۱۱.۴ را ثابت کنید .
- ۲ . درایه‌های جدول ۱۴.۴ را بدست آورید .
- ۳ . چندجمله‌ای تیلور درجه ۳ برای xe^x را یافته و ، با استفاده از با صرفه‌سازی چیبیشف ، ضمن اینکه خطا بر $[-1, 1]$ را کمتر از 01. نگه می‌دارید ، یک تقریب چند جمله‌ای از درجه کمتر بدست آورید .
- ۴ . با استفاده از صفرهای $T_3(x)$ به عنوان گره‌های درونیاب ، چندجمله‌ای درونیاب درجه ۳ به $P_3(x)$ را

$$f(x) = \sqrt{1+x}, -1 \leq x \leq 1.$$

- ۵ . با بیابید $P_3(1)$ را حساب کرده ، و با مثال ۲ از بخش ۱۰.۳ مقایسه نمایید .
- با استفاده از صفرهای $T_3(x)$ و تبدیل بازه $[0, 1]$ بروی $[-1, 1]$ ، چندجمله‌ای درونیاب درجه ۳ به $P_3(x)$ به $f(x) = \sqrt{1+x}, 0 \leq x \leq 1$ را پیدا کنید . $P_3(1)$ را حساب کرده ، و با تمرین ۴ و مثال ۲ از بخش ۱۰.۳ مقایسه نمایید .

مشتق‌گیری و انتگرال‌گیری عددی ^۵

می‌خواهیم با یک ماشین که یک ورقه صاف آلومینیم را طوری پرس می‌کند که مقطع عرضی‌اش به شکل موج سینوسی است، یک ورقه شیروانی موجدار بسازیم.



فرض کنیم یک ورقه موجدار به طول ۱۲۰ سانتیمتر لازم باشد، که ارتفاع هر موج $2/5$ سانتیمتر از مرکز خط بوده، و هر موج با دوره تناوب تقریباً 5π سانتیمتر باشد. مسئله یافتن طول ورقه صاف اولیه عبارت است از یافتن طول قوس منحنی داده شده با $f(x) = \sin x$ از 0 سانتیمتر x تا 120 سانتیمتر x . از حساب دیفرانسیل و انتگرال می‌دانیم که این طول را می‌توان به صورت زیر بیان کرد:

$$L = \int_0^{48} \sqrt{1 + \left(\frac{df(x)}{dx}\right)^2} dx = \int_0^{48} \sqrt{1 + (\cos x)^2} dx,$$

در نتیجه، مسئله به محاسبه این انتگرال تحویل می‌شود. اما، با آنکه تابع سینوس یکی از معمولی‌ترین توابع ریاضی است، محاسبه طول قوس آن به یک انتگرال بیضوی نوع دوم منجر می‌شود، که نمی‌توان آن را با روشهای معمول حساب

کرد. در این فصل، روشهایی تقریبی ارائه می‌شوند که مسائلی از این نوع را به تمریناتی مقدماتی تحویل می‌سازند.

چون در بسیاری از کاربردها از مشتق و انتگرال توابع استفاده می‌شود، انتظار می‌رود که تقریب سازی به کمک این مفاهیم مفید باشد.

در مقدمه فصل ۳ گفتیم که یک دلیل برای استفاده از رده چند جمله‌ایهای جبری در تقریب یک مجموعه دلخواه از داده‌ها این خاصیت جبری چند جمله‌ایهاست که، به ازای هر تابع پیوسته تعریف شده بر یک بازه بسته، یک چند جمله‌ای وجود دارد که در هر نقطه در بازه بدلخواه به تابع نزدیک است. خاصیت دیگری که این رده دارد این است که مشتق و انتگرال اعضای آن نسبتاً ساده بدست می‌آیند و محاسبه می‌شوند. لذا، نباید تعجب کرد که اغلب روندهای تقریب انتگرالها و مشتقها با چند جمله‌ایهای جبری که تابع را تقریب می‌کنند شروع شود.

۱۰۵ مشتقگیری عددی

برای معرفی مشتقگیری عددی، ابتدا به دو مثال می‌پردازیم.

مثال ۱. به ازای تابع $f \in C^2[a, b]$ و نقطه دلخواه x_0 در $[a, b]$ ، می‌خواهیم روشی برای تقریب $f'(x_0)$ بدست آوریم. با $x_1 = x_0 + h$ به ازای $h \neq 0$ ، آنقدر کوچک که $x_1 \in [a, b]$ ، $P_{0,1}(x)$ را حساب می‌کنیم. با استفاده از نمادگذاری بخش ۳.۳ و نتایج قضیه ۳.۳ (ص ۱۲۵)، به ازای ξ ای، وابسته به x ، در $[a, b]$ داریم

$$f(x) = P_{0,1}(x) + \frac{(x-x_0)(x-x_1)}{2!} f''(\xi)$$

$$= \frac{f(x_0)(x - x_0 - h)}{-h} + \frac{f(x_0 + h)(x - x_0)}{h} + \frac{(x - x_0)(x - x_0 - h)}{2} f''(\xi),$$

مشتگیری نتیجه می دهد که

$$\begin{aligned} f'(x) &= \frac{f(x_0 + h) - f(x_0)}{h} + \frac{d}{dx} \left[\frac{(x - x_0)(x - x_0 - h)}{2} f''(\xi) \right] \\ &= \frac{f(x_0 + h) - f(x_0)}{h} + \frac{2(x - x_0) - h}{2} f''(\xi) \\ &\quad + \frac{(x - x_0)(x - x_0 - h)}{2} \frac{d}{dx} (f''(\xi)). \end{aligned}$$

با آنکه در این مشتگیری باید $(d/dx)[f''(\xi)]$ موجود باشد، فرمول تقریب را می توان به صورت دیگری دون این فرض، با استفاده از قضیه تیلور، بدست آورد (ر.ک. تمرین ۹).

به ازای $x = x_0$ ، این معادله به

$$(1.5) \quad f'(x_0) = \frac{f(x_0 + h) - f(x_0)}{h} - \frac{h}{2} f''(\xi)$$

ساده می شود. به ازای مقادیر کوچک h ، می توان از خارج قسمت تفاضلی $(f(x_0 + h) - f(x_0))/h$ برای تقریب $f'(x_0)$ با خطایی کراندار به $(h/2)M$ ، که در آن M یک کران $f''(x)$ به ازای $x \in [a, b]$ است، استفاده کرد. این فرمول به فرمول تفاضل پیشرو اگر $h > 0$ ، و فرمول تفاضل پسرو اگر $h < 0$ معروف است.

مثال ۲. فرض کنیم $f(x) = \ln x$ و $x_0 = 1.8$. با استفاده از خارج قسمت

$$\frac{f(1.8 + h) - f(1.8)}{h}, \quad h > 0,$$

$f'(1.8)$ را با خطای

$$1.8 < \xi < 1.8 + h, \quad \frac{|hf''(\xi)|}{2} \leq \frac{|h|}{2(1.8)^2}$$

تقریب می کنیم. نتایج برای حالاتی که h مساوی ۱، ۰.۰۱، و ۰.۰۰۱ است در جدول

h	$f(1.8 + h)$	$\frac{f(1.8 + h) - f(1.8)}{h}$	$\frac{ h }{2(1.8)^2}$
.1	.64185389	.5406722	.0154321
.01	.59332685	.5540180	.0015432
.001	.58834207	.5554000	.0001543

جدول ۱۰۵

چون $f'(x) = 1/x$ ، مقدار دقیق $f'(1.8)$ مساوی ۰.۵۵۵ است و کرانه‌های خطای بالا مناسب می‌باشند .

برای در نظر گرفتن این روند مشتقگیری در حالت کلی ، فرض کنیم $\{x_0, \dots, x_n\}$ ، $(n + 1)$ نقطه متمایز در بازه‌های مانند I بوده و $f \in C^{n+1}(I)$. طبق قضیه ۳.۳ ، بازای ξ ای . وابسته به x ، در I داریم

$$f(x) = \sum_{k=0}^n f(x_k)L_k(x) + \frac{(x - x_0) \cdots (x - x_n)}{(n + 1)!} f^{(n+1)}(\xi)$$

مشتقگیری نتیجه می‌دهد که

$$\begin{aligned} f'(x) &= \sum_{k=0}^n f(x_k)L'_k(x) + \frac{d}{dx} \left[\frac{(x - x_0) \cdots (x - x_n)}{(n + 1)!} \right] f^{(n+1)}(\xi) \\ (۲.۰۵) \quad &+ \frac{(x - x_0) \cdots (x - x_n)}{(n + 1)!} \frac{d}{dx} [f^{(n+1)}(\xi)]. \end{aligned}$$

از این فرمول می‌شود برای تقریب $f'(x)$ بازای هر $x \in I$ استفاده کرد ، ولی تحلیل جمله خطای برشی ، یعنی

$$\begin{aligned} t_n(x) &= \frac{d}{dx} \left[\frac{(x - x_0) \cdots (x - x_n)}{(n + 1)!} \right] f^{(n+1)}(\xi) \\ (۳.۰۵) \quad &+ \frac{(x - x_0) \cdots (x - x_n)}{(n + 1)!} \frac{d}{dx} [f^{(n+1)}(\xi)], \end{aligned}$$

مشکل است، زیرا از $(d/dx)f^{(n+1)}(\xi)$ اطلاع کمی در دست است. با اینحال، از معادله (۲۰۵) می‌توان وقتی x یک گره درونیاب است استفاده کرد. در این حالت، با فرض وجود و کراندار بودن $[f^{(n+1)}(\xi)]$ ، واضح است که جمله خطابه

$$t_n(x_k) = \left\{ \frac{d}{dx} \left[\frac{(x-x_0)\cdots(x-x_n)}{(n+1)!} \right] f^{(n+1)}(\xi) \right\} \Bigg|_{x=x_k}$$

$$= \frac{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}{(n+1)!} \times f^{(n+1)}(\xi).$$

ساده می‌شود. از اینرو، معادله (۲۰۵) خواهد شد

$$(۴۰۵) \quad f'(x_k) = \sum_{j=0}^n f(x_j)L'_j(x_k) + t_n(x_k)$$

معادله (۴۰۵) یک فرمول $(n+1)$ نقطه‌ای برای تقریب $f'(x_k)$ نامیده می‌شود، زیرا یک ترکیب خطی از $(n+1)$ مقدار $f(x_j)$ ، به‌ازای $0 \leq j \leq n$ ، بکار رفته است. استفاده از نقاط بیشتر x_j ظاهراً "به دقت بیشتری منجر می‌شود، زیرا t_n نزول می‌کند، اما از دیدگاه عملی به محاسبه تابع f در نقاط دیگر منجر می‌شود و، حتی اگر محاسبه تابع آسان باشد، خطاهای گرد کردن ما را از استفاده از تعداد زیادی گره درونیاب برحذر می‌کند. برای توضیح این نکته، چند فرمول سه‌نقطه‌ای مفید بدست آورده، و خطاهای گرد کردن آنها را در نظر می‌گیریم. چون

$$L_0(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)},$$

این ایجاب می‌کند که

$$L'_0(x) = \frac{2x-x_1-x_2}{(x_0-x_1)(x_0-x_2)}.$$

به‌همین نحو،

$$L'_1(x) = \frac{2x-x_0-x_2}{(x_1-x_0)(x_1-x_2)}$$

$$L'_2(x) = \frac{2x-x_0-x_1}{(x_2-x_0)(x_2-x_1)}.$$

از اینرو، بهازای هر $j = 0, 1, 2$ ،

$$(5.5) \quad \begin{aligned} f'(x_j) = & f(x_0) \left[\frac{2x_j - x_1 - x_2}{(x_0 - x_1)(x_0 - x_2)} \right] \\ & + f(x_1) \left[\frac{2x_j - x_0 - x_2}{(x_1 - x_0)(x_1 - x_2)} \right] \\ & + f(x_2) \left[\frac{2x_j - x_0 - x_1}{(x_2 - x_0)(x_2 - x_1)} \right] + \frac{1}{6} f^{(3)}(\xi_j) \prod_{\substack{i=0 \\ i \neq j}}^2 (x_j - x_i), \end{aligned}$$

که نماد ξ_j نشان می‌دهد که این نقطه به x_j بستگی دارد.

فرمول (5.5) بخصوص وقتی مفید است که گرهها متساوی الفاصله باشند؛ یعنی، وقتی

$$x_2 = x_0 + 2h \text{ و } x_1 = x_0 + h, \quad h \neq 0$$

با استفاده از معادله (5.5) بهازای $x_j = x_0 + h$ ، $x_1 = x_0 + h$ ، و $x_2 = x_0 + 2h$ ، نتیجه می‌شود که

$$\begin{aligned} f'(x_0) = & f(x_0) \left[\frac{2x_0 - x_0 - h - x_0 - 2h}{(x_0 - x_0 - h)(x_0 - x_0 - 2h)} \right] \\ & + f(x_1) \left[\frac{x_0 - x_0 - 2h}{(x_0 + h - x_0)(x_0 + h - x_0 - 2h)} \right] \\ & + f(x_2) \left[\frac{x_0 - x_0 - h}{(x_0 + 2h - x_0)(x_0 + 2h - x_0 - h)} \right] \\ & + \frac{1}{6} f^{(3)}(\xi_0)(x_0 - x_0 - h)(x_0 - x_0 - 2h), \end{aligned}$$

یا

$$f'(x_0) = \frac{1}{h} \left[-\frac{3}{2}f(x_0) + 2f(x_1) - \frac{1}{2}f(x_2) \right] + \frac{h^2}{3} f^{(3)}(\xi_0).$$

اگر همین کار بهازای $x_j = x_1$ و $x_j = x_2$ انجام شود، نتیجه خواهد شد که

$$f'(x_1) = \frac{1}{h} \left[-\frac{1}{2}f(x_0) + \frac{1}{2}f(x_2) \right] - \frac{h^2}{6} f^{(3)}(\xi_1),$$

$$f'(x_2) = \frac{1}{h} \left[\frac{1}{2}f(x_0) - 2f(x_1) + \frac{3}{2}f(x_2) \right] + \frac{h^2}{3} f^{(3)}(\xi_2).$$

با نوشتن معادلات برحسب x_0 ، خواهیم داشت

$$(۶۰۵) \quad f'(x_0) = \frac{1}{h} \left[-\frac{3}{2}f(x_0) + 2f(x_0 + h) - \frac{1}{2}f(x_0 + 2h) \right] + \frac{h^2}{3} f^{(3)}(\xi_0),$$

$$(۷۰۵) \quad f'(x_0 + h) = \frac{1}{h} \left[-\frac{1}{2}f(x_0) + \frac{1}{2}f(x_0 + 2h) \right] - \frac{h^2}{6} f^{(3)}(\xi_1),$$

$$(۸۰۵) \quad f'(x_0 + 2h) = \frac{1}{h} \left[\frac{1}{2}f(x_0) - 2f(x_0 + h) + \frac{3}{2}f(x_0 + 2h) \right] + \frac{h^2}{3} f^{(3)}(\xi_2).$$

از تعویض x_0 با $(x_0 - h)$ در (۷۰۵) و x_0 با $(x_0 - 2h)$ در (۸۰۵)، سه فرمول برای تقریب $f'(x_0)$ بدست می‌آیند:

$$(۹۰۵) \quad f'(x_0) = \frac{1}{2h} \left[-3f(x_0) + 4f(x_0 + h) - f(x_0 + 2h) \right] + \frac{h^2}{3} f^{(3)}(\xi_0),$$

$$(۱۰۰۵) \quad f'(x_0) = \frac{1}{2h} \left[-f(x_0 - h) + f(x_0 + h) \right] - \frac{h^2}{6} f^{(3)}(\xi_1),$$

$$(۱۱۰۵) \quad f'(x_0) = \frac{1}{2h} \left[f(x_0 - 2h) - 4f(x_0 - h) + 3f(x_0) \right] + \frac{h^2}{3} f^{(3)}(\xi_2).$$

بالاخره، ملاحظه می‌کنیم که چون معادله (۱۱۰۵) را می‌توان از (۹۰۵) به‌سادگی به‌وسیله تعویض h با $-h$ بدست آورد، عملاً "دو فرمول بیشتر وجود ندارند".

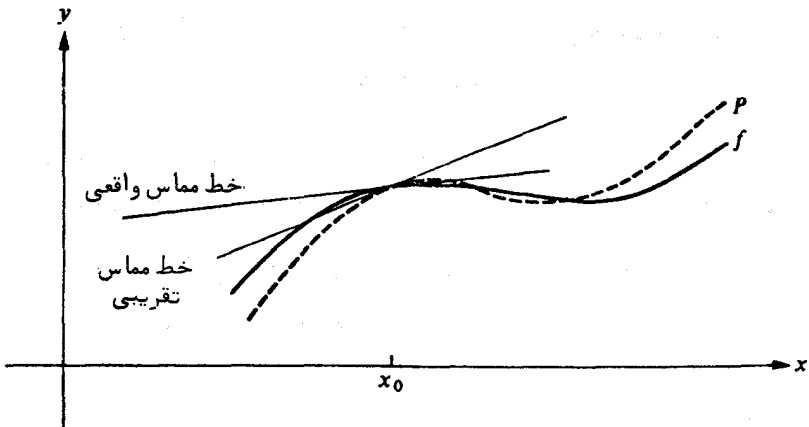
$$(12.5) \quad f'(x_0) = \frac{1}{2h} [-3f(x_0) + 4f(x_0 + h) - f(x_0 + 2h)] \\ + \frac{h^2}{3} f^{(3)}(\xi_0), \quad h \neq 0,$$

و

$$(13.5) \quad f'(x_0) = \frac{1}{2h} [-f(x_0 - h) + f(x_0 + h)] - \frac{h^2}{6} f^{(3)}(\xi_1), \quad h \neq 0,$$

که در آن ξ_0 و ξ_1 هر دو بین $(x_0 - h)$ و $(x_0 + h)$ قرار دارند.

در معادله (۱۳.۵) انتظار معقول است، زیرا، در معادله (۱۳.۵)، داده‌ها در طرفین x_0 ولی در (۱۲.۵) فقط در یک طرف امتحان می‌شوند. تقریب در معادله (۱۲.۵) نزدیک نقاط انتهایی بازه I مفید واقع می‌شود، زیرا اطلاع از f در خارج بازه لزوماً در دست نیست. همچنین، توجه کنید که در (۱۳.۵)، f فقط باید در دو نقطه حساب شود، در حالی که، در (۱۲.۵)، سه نقطه لازم خواهد بود.



شکل ۱-۵

روشهای ارائه شده در معادلات (۱۲.۵) و (۱۳.۵) به فرمولهای سه نقطه‌ای معروفند. بهمین نحو، روشهایی هستند، به نام فرمولهای پنج نقطه‌ای، که تابع در نقاط بیشتری

حساب می‌شود ولی جمله خطایش به شکل $O(h^4)$ می‌باشد. برای بدست آوردن فرمولی از این نوع، تابع f را به صورت یک چندجمله‌ای تیلور مرتبه چهار بسط داده و آن را در $x_0 - 2h$ ، $x_0 - h$ ، $x_0 + h$ ، و $x_0 + 2h$ حساب می‌کنیم تا بدست آید که

$$(14.5) \quad f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \frac{1}{6}f'''(x_0)h^3 + \frac{1}{24}f^{(4)}(x_0)h^4 + \frac{1}{120}f^{(5)}(\xi_1)h^5$$

$$(15.5) \quad f(x_0 - h) = f(x_0) - f'(x_0)h + \frac{1}{2}f''(x_0)h^2 - \frac{1}{6}f'''(x_0)h^3 + \frac{1}{24}f^{(4)}(x_0)h^4 - \frac{1}{120}f^{(5)}(\xi_{-1})h^5$$

که در آن $x_0 - h < \xi_{-1} < x_0 < \xi_1 < x_0 + h$ بهمین ترتیب،

$$(16.5) \quad f(x_0 + 2h) = f(x_0) + 2f'(x_0)h + 2f''(x_0)h^2 + \frac{4}{3}f'''(x_0)h^3 + \frac{2}{3}f^{(4)}(x_0)h^4 + \frac{4}{15}f^{(5)}(\xi_2)h^5,$$

$$(17.5) \quad f(x_0 - 2h) = f(x_0) - 2f'(x_0)h + 2f''(x_0)h^2 - \frac{4}{3}f'''(x_0)h^3 + \frac{2}{3}f^{(4)}(x_0)h^4 - \frac{4}{15}f^{(5)}(\xi_{-2})h^5$$

که در آن $x_0 - 2h < \xi_{-2} < x_0 < \xi_2 < x_0 + 2h$

باکم کردن (15.5) از (14.5) و معادله (17.5) از (16.5)، نتیجه می‌شود که

$$(18.5) \quad f(x_0 + h) - f(x_0 - h) = 2f'(x_0)h + \frac{1}{3}f'''(x_0)h^3 + \frac{1}{120}[f^{(5)}(\xi_1) + f^{(5)}(\xi_{-1})]h^5$$

$$(19.5) \quad f(x_0 + 2h) - f(x_0 - 2h) = 4f'(x_0)h + \frac{8}{3}f'''(x_0)h^3 + \frac{4}{15}[f^{(5)}(\xi_2) + f^{(5)}(\xi_{-2})]h^5.$$

اگر معلوم باشد که $f^{(5)}$ بر بازه $[x_0 - 2h, x_0 + 2h]$ پیوسته است، می‌توان با استفاده از قضیه مقدار میانی (قضیه ۱۲.۱، ص ۹) مطمئن شد که نقاطی مانند

$\xi_1 \leq \xi_{-1} < \xi_1, \xi_{-1} < \xi_2 < \xi_2 < \xi_2$ و $\xi_2 < \xi_2$ با این خاصیت وجود دارند که

$$\frac{f^{(5)}(\xi_1) + f^{(5)}(\xi_{-1})}{2} = f^{(5)}(\xi_1)$$

و

$$\frac{f^{(5)}(\xi_2) + f^{(5)}(\xi_{-2})}{2} = f^{(5)}(\xi_2)$$

حال می توان معادلات (۱۸.۵) و (۱۹.۵) را به صورت زیر بازنویسی کرد .

$$(۲۰.۵) f(x_0 + h) - f(x_0 - h) = 2f'(x_0)h + \frac{1}{3}f'''(x_0)h^3 + \frac{1}{60}f^{(5)}(\xi_1)h^5$$

و

$$(۲۱.۵) f(x_0 + 2h) - f(x_0 - 2h) = 4f'(x_0)h + \frac{8}{3}f'''(x_0)h^3 + \frac{8}{15}f^{(5)}(\xi_2)h^5$$

با حذف جملات شامل $f'''(x_0)$ در این دو معادله، نتیجه می شود که

$$8f(x_0 + h) - 8f(x_0 - h) - f(x_0 + 2h) + f(x_0 - 2h) = 12hf'(x_0) + \frac{2}{15}[f^{(5)}(\xi_1) - 4f^{(5)}(\xi_2)]h^5$$

یا

$$f'(x_0) = \frac{1}{12h} [f(x_0 - 2h) - 8f(x_0 - h) + 8f(x_0 + h) - f(x_0 + 2h)] + \frac{h^4}{90} [4f^{(5)}(\xi_2) - f^{(5)}(\xi_1)].$$

جمله خطایی که به این طریق بدست می آید کاملاً " پیچیده است، ولی می توان نشان

داد که به ازای نقطه ای مانند ξ ، که $(x_0 - 2h) < \xi < (x_0 + 2h)$ ، مساوی $(h^4/30)f^{(5)}(\xi)$

است. روش معمول برای بدست آوردن جمله خطا در این شکل مشتقگیری از چند جمله ای

لاگرانژ و جمله خطا در قضیه ۳.۳ است، که روند بسیار پرهزمت تری است. برای مراجعه

در آینده، همه فرمولهای پنج نقطه ای را ذکر می کنیم، که هر یک از آنها را می توان از

روندی مشابه فوق بدست آورد.

$$(۲۲.۵) f'(x_0) = \frac{1}{12h} [-25f(x_0) + 48f(x_0 + h) - 36f(x_0 + 2h) + 16f(x_0 + 3h) - 3f(x_0 + 4h)] + \frac{h^4}{5} f^{(5)}(\xi),$$

$$(۲۳.۵) \quad f'(x_0) = \frac{1}{12h} [-3f(x_0 - h) - 10f(x_0) + 18f(x_0 + h) \\ - 6f(x_0 + 2h) + f(x_0 + 3h)] - \frac{h^4}{20} f^{(5)}(\xi),$$

$$(۲۴.۵) \quad f'(x_0) = \frac{1}{12h} [f(x_0 - 2h) - 8f(x_0 - h) + 8f(x_0 + h) \\ - f(x_0 + 2h)] + \frac{h^4}{30} f^{(5)}(\xi),$$

$$(۲۵.۵) \quad f'(x_0) = \frac{1}{12h} [-f(x_0 - 3h) + 6f(x_0 - 2h) - 18f(x_0 - h) \\ + 10f(x_0) + 3f(x_0 + h)] - \frac{h^4}{20} f^{(5)}(\xi),$$

و

$$(۲۶.۵) \quad f'(x_0) = \frac{1}{12h} [3f(x_0 - 4h) - 16f(x_0 - 3h) + 36f(x_0 - 2h) \\ - 48f(x_0 - h) + 25f(x_0)] + \frac{h^4}{5} f^{(5)}(\xi),$$

که در آنها ξ بین مقادیر اکستریم گره‌های آمده در هر فرمول قرار دارد.

مثال ۳. به‌ازای $h = \pm 1$ ، برای تقریب سازی $f'(3.15)$ از فرمولهای (۱۲.۵) و (۱۳.۵) به‌کمک داده‌های جدول ۲.۵ استفاده کنید.

x	$f(x)$
2.95	13.364875
3.05	14.665125
3.15	16.088375
3.25	17.640625
3.35	19.327875

جدول ۲.۵

با استفاده از معادله^۱ (۱۲.۵) بهازای $h = 1$ ،

$$f'(3.15) \approx \frac{1}{2} [-3f(3.15) + 4f(3.25) - f(3.35)] = 14.8475.$$

با استفاده از معادله^۱ (۱۲.۵) بهازای $h = -1$ ،

$$f'(3.15) \approx -\frac{1}{2} [-3f(3.15) + 4f(3.05) - f(2.95)] = 14.8475.$$

با استفاده از معادله^۱ (۱۳.۵) بهازای $h = 1$ ،

$$f'(3.15) \approx \frac{1}{2} [-f(3.05) + f(3.25)] = 14.8775.$$

دراین مثال ، $f(x) = x^3 - 3x^2 + 4x + 2$ ، در نتیجه ، $f'(3.15) = 14.8675$.
 خطای استفاده از (۱۲.۵) بهازای $h = \pm 1$ مساوی 0.02 است ، و خطای استفاده از معادله^۱ (۱۳.۵) مساوی 0.01 است . لذا ، دراین مثال ، استفاده از (۱۳.۵) دوبار دقیقتر از استفاده از (۱۲.۵) است . باید توجه داشت که همه^۱ خطاهای این تقریب را می توان به آسانی از فرمول خطا تعیین کرد ، زیرا $f^{(3)}(x) = 6$ ، و در نتیجه ، مقادیر خاص ξ_0 و ξ_1 مهم نیستند .

تنها فرمول پنج نقطه ای که داده های کافی برایش ارائه شده معادله^۱ (۲۴.۵) است . استفاده از این روش منجر به

$$\begin{aligned} f'(3.15) &\approx \frac{1}{1.2} [f(2.95) - 8f(3.05) + 8f(3.25) - f(3.35)] \\ &= 14.8675, \end{aligned}$$

یعنی مقدار دقیق ، می شود ، که انتظار آن می رفت ، زیرا جمله^۱ خطا دراین حالت شامل $f^{(5)}(\xi)$ است و $f^{(5)}(\xi) \equiv 0$.

موضوعی که در بررسی مشتقگیری عددی اهمیت خاص دارد اثر خطای گرد کردن است . به بررسی دقیقتر فرمول

$$f'(x_0) = \frac{f(x_0 + h) - f(x_0 - h)}{2h} - \frac{h^2}{6} f^{(3)}(\xi_1)$$

می پردازیم . فرض کنید در محاسبه^۱ $f(x_0 + h)$ و $f(x_0 - h)$ با گرد کردن خطاهای

و $e(x_0 + h)$ و $e(x_0 - h)$ مواجه شویم؛ یعنی، مقادیر محاسبه شده ما $\tilde{f}(x_0 + h)$ و $\tilde{f}(x_0 - h)$ با مقادیر واقعی $f(x_0 + h)$ و $f(x_0 - h)$ به وسیله فرمولهای

$$f(x_0 + h) = \tilde{f}(x_0 + h) + e(x_0 + h)$$

و

$$f(x_0 - h) = \tilde{f}(x_0 - h) + e(x_0 - h)$$

به هم مربوط شوند. در این حالت، خطای تقریب، یعنی

$$f'(x_0) - \frac{\tilde{f}(x_0 + h) - \tilde{f}(x_0 - h)}{2h} = \frac{e(x_0 + h) - e(x_0 - h)}{2h} - \frac{h^2}{6} f^{(3)}(\xi_1)$$

قسمتی ناشی از گرد کردن و قسمتی ناشی از برش است. اگر فرض کنیم خطاهای گرد کردن $e(x_0 \pm h)$ به وسیله $\varepsilon > 0$ ی کراندار بوده، و مشتق سوم f به وسیله عددی مانند $M > 0$ کراندار باشد،

$$\left| f'(x_0) - \frac{\tilde{f}(x_0 + h) - \tilde{f}(x_0 - h)}{2h} \right| \leq \frac{\varepsilon}{h} + \frac{h^2}{6} M.$$

اگر h کوچک باشد، نتیجه می شود که خطای ناشی از گرد کردن، یعنی ε/h ، ممکن است بزرگ باشد. لذا، در عمل h را خیلی کوچک گرفتن بندرت مفید واقع می شود، زیرا خطاهای گرد کردن بزرگ خواهند شد.

مثال ۴. با استفاده از مقادیر جدول ۳.۵، به تقریب $f'(0.900)$ برای $f(x) = \sin x$ می پردازیم.

x	sin x	x	sin x
.800	.71736	.901	.78395
.850	.75128	.902	.78457
.880	.77074	.905	.78643
.890	.77707	.910	.78950
.895	.78021	.920	.79560
.898	.78208	.950	.81342
.899	.78270	1.000	.84147

با استفاده از فرمول

$$f'(900) \approx \frac{f(900 + h) - f(900 - h)}{2h}$$

بمازای مقادیر مختلف h ، مقادیر جدول ۴.۵ بدست می‌آیند.

h	تقریب به $f'(900)$
.001	.62500
.002	.62250
.005	.62200
.010	.62150
.020	.62150
.050	.62140
.100	.62055

جدول ۴.۵

چون $f'(900) = \cos(900) = .62161$ ، ظاهراً "یک انتخاب بهینه برای h بین .005 و .05 قرار دارد. فرض کنید جمله خطا، یعنی

$$\frac{\varepsilon}{h} + \frac{h^2}{6} M,$$

را برای $f(x) = \sin x$ بر $[.800, 1.00]$ بنوعی تحلیل کرده باشیم. با فرض

$$e(h) = \frac{\varepsilon}{h} + \frac{h^2}{6} M,$$

می‌توان تحقیق کرد (ر.ک. تمرین ۱۰) که یک مینیمم برای e در $h = \sqrt[3]{3\varepsilon/M}$ رخ می‌دهد، که در آن

$$M = \max_{x \in [.800, 1.00]} |f'''(x)| = \max_{x \in [.800, 1.00]} |\cos x| = .69671.$$

چون مقادیر f تا پنج رقم اعشار داده شده‌اند، فرض $\varepsilon = .000005$ معقول می‌باشد.

لذا، انتخاب بهینه h تقریباً "عبارت است از

$$h = \sqrt[3]{\frac{3(0.000005)}{.69671}} \approx .028$$

که با نتایج ما سازگار است .

لیکن ، در عمل محاسبه یک مقدار بهینه h برای استفاده در تقریب مشتق ممکن نیست ، زیرا اطلاعاتی از مشتق سوم تابع نداریم .
 در این بخش ، متداولترین فرمولها را برای تقریب $f'(x)$ به تفصیل بررسی کردیم . باید به یاد داشت که مشتگیری عددی ، به عنوان یک روش تقریب ، پایدار نیست ، زیرا مقادیر کوچک h می توانند به خطاهای بزرگی منجر شوند . این اولین روش با این خاصیت است که بررسی کرده ایم ، و در عمل مشتگیری عددی حتی الامکان اجتناب می شود . بهرحال ، فرمولهای بدست آمده مفیدند ، و در تقریب جوابهای معادلات دیفرانسیل معمولی و جزئی کاربرد دارند .

مجموعه تمرینات ۱-۵

- ۱ . به ازای $f(x) = e^x$ ، با استفاده از معادلات (۱۲.۵) و (۱۳.۵) به ازای $h = .01$ ، تقریبهایی برای $f'(1.8)$ بدست آورید . نتایج خود را با مقدار دقیق $e^{1.8}$ مقایسه کنید .
- ۲ . فرض کنید $f(x) = x^3 e^{x^2} - \sin x$. با استفاده از معادلات (۱۲.۵) و (۱۳.۵) به ازای $h = .1$ و $h = .01$ ، $f'(2.19)$ را تقریب کنید .
- ۳ . فرض کنید $f(x) = 2^x \sin x$. با استفاده از $h = .05$ و $h = .01$ در فرمول (۱۳.۵) با

داده های زیر

x	1.0	1.04	1.06	1.10
$f(x)$	1.6829420	1.7732994	1.8188014	1.9103448

$f'(1.05)$ را تقریب کنید .

- ۴ . تمرین ۳ را با استفاده از حساب چهاررقمی و مقایسه نتایج با نتایج قبل تکرار کنید .
- ۵ . فرض کنید داده های زیر به طور تجربی جمع آوری شده باشند :

x	1.00	1.01	1.02
$f(x)$	1.27	1.32	1.38

(ت) با استفاده از (1.005) ، $f'(1.005)$ و $f'(1.015)$ را تقریب کنید .

(ب) با استفاده از (1.01) ، $f''(1.01)$ را تقریب کنید .

(پ) فرض کنید داده‌ها تا ± 0.005 دقیق باشند . خطای ماکزیم ناشی از داده‌های غیردقیق در قسمتهای (ت) و (ب) را بیابید .

۶. در یک مدار با ولتاژ $\mathcal{E}(t)$ و ضریب القای L ، قانون اول کیرشف^۱ رابطه

$$\mathcal{E} = L \frac{di}{dt} + Ri$$

را بدست می‌دهد، که در آن R مقاومت مدار و i شدت جریان است . فرض کنید شدت جریان به‌ازای چند مقدار از t اندازه‌گیری شده، و بدست آمده باشد :

t	1.00	1.01	1.02	1.03	1.04
i	3.10	3.12	3.14	3.18	3.24

که در آن t به ثانیه، i به آمپر، ضریب القای L ثابت و مساوی 98. هانری، و مقاومت مساوی 142 اهم است . با استفاده از فرمولهای سه‌نقطه‌ای مناسب، ولتاژ \mathcal{E} را در مقادیر $t = 1, 1.01, 1.02, 1.03, 1.04$ تقریب کنید .

۷. به‌ازای $x_1 = x_0 + h$ و $x_2 = x_0 + 3h$ ، و با استفاده از معادله^۴ (4.5) تقریبی برای $f'(x_0)$ بدست آورید . با استفاده از این فرمول، $f'(1.8)$ را برای $f(x) = e^x$ تقریب کرده، و نتایج خود را با نتایج تمرین ۱ مقایسه کنید .

۸. برای فرمول

$$f'(x_0) = \frac{f(x_0 + h) - f(x_0)}{h} - \frac{h}{2} f''(\xi_0),$$

خطاهای گردکردن را مثل مثال ۴ تحلیل کنید . برای تابع مثال ۲ یک $h > 0$ بهینه پیدا نمایید .

۹. با استفاده از چندجمله‌ای تیلور از درجه^۵ یک که حول $x = x_0$ بسط داده شده، معادله^۵ (1.5) را بدست آورید .

۱۰. تابع

$$e(h) = \frac{\varepsilon}{h} + \frac{h^2}{6} M$$

را در نظر بگیرید، که در آن M یک کران برای

$$\left| \frac{d^3(\sin x)}{dx^3} \right| \text{ بر } [0.800, 1.000]$$

است. نشان دهید که $e(h)$ یک مینیمم در $\sqrt[3]{3\epsilon/M}$ دارد.

۱۱. برای ساختن یک اسپلین مکعبی با شرایط مرزی آزاد، به مقادیر f' در نقاط انتهایی بازه نیاز داریم.

(آ) با استفاده از شرایط مرزی مقید، تمرین ۲ از بخش ۶.۳ را تکرار کنید. با استفاده از معادله (۹.۵)، $f'(x_0)$ ، و با استفاده از (۱۱.۵)، $f'(x_3)$ را تقریب کنید. نتایج خود را با تمرین ۲ از بخش ۶.۳ مقایسه نمایید.

(ب) با استفاده از شرایط مرزی مقید، تمرین ۵ از بخش ۶.۳ را تکرار کنید. با استفاده از معادله (۲۲.۵)، $f'(x_0)$ ، و با استفاده از معادله (۲۶.۵)، $f'(x_{10})$ را تقریب کنید.

۱۲. (آ) با استفاده از مشتق اسپلین مکعبی ساخته شده در تمرین ۲ از بخش ۶.۳، $f''(1.03)$ را تقریب کنید.

(ب) با استفاده از مشتق مکعبی ساخته شده در تمرین ۱۱ (آ)، $f''(1.03)$ را تقریب کرده و جواب را با نتیجه قسمت (آ) مقایسه نمایید.

۲.۵ مشتقات بالاتر و برونمایی

روشهای بخش ۱.۵ را می توان، فقط با استفاده از مقادیر جدولی تابع در نقاط مختلف، برای یافتن تقریبهایی به مشتقات بالاتر یک تابع تعمیم داد. روشهای مربوط از لحاظ جبری ملال آورند؛ در نتیجه، فقط یک روند در اینجا ارائه می شود.

با بسط تابع f به صورت یک چند جمله ای درجه سه حول نقطه x_0 و محاسبه آن در $x_0 + h$ و $x_0 - h$ ، بدست می آید

$$(27-5) \quad f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \frac{1}{6}f'''(x_0)h^3 + \frac{1}{24}f^{(4)}(\xi_1)h^4$$

$$(28-5) \quad f(x_0 - h) = f(x_0) - f'(x_0)h + \frac{1}{2}f''(x_0)h^2 - \frac{1}{6}f'''(x_0)h^3 + \frac{1}{24}f^{(4)}(\xi_{-1})h^4,$$

که در آنها $x_0 - h < \xi_{-1} < x_0 < \xi_1 < x_0 + h$. به جای تفریق برای حذف جملات با بالاترین مشتق، که در بخش ۱.۵ انجام شد، معادلات (۲۷.۵) و (۲۸.۵) را جمع می‌کنیم تا بدست آید

$$(29.5) \quad f(x_0 + h) + f(x_0 - h) = 2f(x_0) + f''(x_0)h^2 + \frac{1}{24}[f^{(4)}(\xi_1) + f^{(4)}(\xi_{-1})]h^4$$

یا

$$(30.5) \quad f''(x_0) = \frac{1}{h^2} [f(x_0 - h) - 2f(x_0) + f(x_0 + h)] - \frac{h^2}{24} [f^{(4)}(\xi_1) + f^{(4)}(\xi_{-1})].$$

اگر $f^{(4)}$ بر $[x_0 - h, x_0 + h]$ پیوسته باشد، طبق قضیه مقدار میانی (قضیه ۱۲.۱، ص ۹)، فرمول فوق را می‌توان به صورت زیر بازنویسی کرد: به‌ازای یک $x_0 - h < \xi < x_0 + h$

$$(31.5) \quad f''(x_0) = \frac{1}{h^2} [f(x_0 - h) - 2f(x_0) + f(x_0 + h)] - \frac{h^2}{12} f^{(4)}(\xi),$$

چون این فرمول جمله خطایی شامل h^2 دارد، در عمل احتمالاً لازم است روندی را بیابیم که شامل جمله‌ای با مرتبه بالاتر، جهت بهتر کردن دقت تقریب، باشد. در اینجا یک روند متفاوت برای بدست آوردن فرمول مرتبه بالاتر ارائه می‌شود. این روش به روند برونابی ریچاردسون^۱ معروف است، و در بسیاری از حالات دیگر نیز مفید واقع می‌شود.

فرض کنیم f به صورت یک چند جمله‌ای تیلور درجه پنج، به‌جای درجه سه، بسط داده شده بود. به آسانی می‌توان نشان داد (ر. ک. تمرین ۸) که یک معادله شبیه (۲۹.۵) بدست می‌آید با

$$(32.5) \quad \frac{1}{h^2} [f(x_0 + h) - 2f(x_0) + f(x_0 - h)] = f''(x_0) + \frac{h^2}{12} f^{(4)}(x_0) + \frac{h^4}{360} f^{(6)}(\xi),$$

که در آن، مثل قبل، $x_0 - h < \xi < x_0 + h$ ، فرض کنیم در معادله (۳۲.۵) با qh تعویض شده باشد، که در آن q مقداری غیر از ± 1 است. با استفاده از دو معادله (۳۲.۵) و

$$\begin{aligned} & \frac{1}{q^2 h^2} [J(x_0 + qh) - 2f(x_0) + f(x_0 - qh)] \\ &= f''(x_0) + \frac{q^2 h^2}{12} f^{(4)}(x_0) + \frac{q^4 h^4}{360} f^{(6)}(\xi) \end{aligned}$$

می‌توان جملهء شامل h^2 را حذف کرد،

$$\begin{aligned} & \frac{q^2}{h^2(q^2 - 1)} [f(x_0 + h) - 2f(x_0) + f(x_0 - h)] \\ & - \frac{1}{(q^2 - 1)q^2 h^2} [f(x_0 + qh) - 2f(x_0) + f(x_0 - qh)] \\ &= f''(x_0) - \frac{q^2 h^4}{(1 - q^2)360} [q^2 f^{(6)}(\xi) - f^{(6)}(\xi)] \end{aligned}$$

یا

$$\begin{aligned} (۳۴.۵) \quad f''(x_0) &= \frac{1}{h^2 q^2 (q^2 - 1)} [-f(x_0 + qh) + q^4 f(x_0 + h) \\ & - (2q^4 - 2)f(x_0) + q^4 f(x_0 - h) - f(x_0 - qh)] \\ & + O(h^4). \end{aligned}$$

بدین ترتیب، تقریبی از $f''(x_0)$ با خطای برشی $O(h^4)$ بدست می‌آید.

مثال ۱. برای یافتن تقریبی از مشتق دوم در 3.15 از تابعی که مقادیرش در جدول مثال ۳ از بخش ۱.۵ داده شده، لازم است $q = \pm 2$ اختیار شود. اینها تنها مقادیر q هستند که به‌ازای آنها داده‌ها کافی‌اند. با $q = 2$ ، معادله (۳۴.۵) شکل زیر را می‌گیرد.

$$\begin{aligned} f''(3.15) &= \frac{1}{(1)^2 2^2 (2^2 - 1)} [-f(3.35) + 2^4 f(3.25) - (2 \cdot 2^4 - 2)f(3.15) \\ & + 2^4 f(3.05) - f(2.95)] + \frac{2^2 (1)^4}{360} f^{(6)}(\xi) \end{aligned}$$

این ایجاب می‌کند که

$$f''(3.15) \approx 12.9.$$

درواقع، این مقدار دقیق است زیرا، به‌ازای هر مقدار از ξ ، اگر $2.95 < \xi < 3.35$ ،
 $f^{(6)}(\xi) = 0$.

مثالی از روند دیگر که از روش برونمایی ریچاردسون استفاده می‌کند، و دو تخمین برای مشتق اول یک تابع را تلفیق می‌نماید، در تمرین ۹ داده شده است.

مجموعه تمرینات ۲۰۵

۱. به‌ازای $f(x) = e^x$ ، با استفاده از معادله (31.05) به‌ازای $h = 0.01$ ، تقریبی برای $f''(1.8)$ بدست آورید. این نتیجه را با مقدار $e^{1.8}$ مقایسه کنید.
۲. فرض کنید $f(x) = 3xe^x - \cos x$. با استفاده از داده‌های زیر و معادله (31.05) ، $f''(1.3)$ را به‌ازای $h = 0.1$ و $h = 0.01$ تقریب کنید.

x	1.20	1.29	1.30	1.31	1.40
$f(x)$	11.59006	13.78176	14.04276	14.30741	16.86187

نتایج خود را با $f''(1.3)$ مقایسه نمایید.

۳. با استفاده از فرمول برونمایی ریچاردسون (34.05) به‌ازای $q = 2$ ، تمرین ۱ را تکرار کنید.
۴. با بسط تابع f به‌صورت یک چند جمله‌ای تیلور درجه چهار حول x_0 و محاسبه آن در $x_0 \pm h$ و $x_0 \pm 2h$ ، روشی برای تقریب $f'''(x_0)$ بدست آورید که جمله خطا از مرتبه h^2 باشد.
۵. به‌ازای $f(x) = e^x$ ، با استفاده از تمرین ۴ و $h = 0.01$ ، تقریبی از $f'''(1.8)$ بدست آورید. این نتیجه را با مقدار $e^{1.8}$ مقایسه کنید.
۶. با استفاده از روند برونمایی ریچاردسون در فرمول تمرین ۴، فرمولی از مرتبه h^4 برای تقریب $f'''(x_0)$ ارائه دهید.
۷. با استفاده از فرمول برونمایی ریچاردسون بدست آمده در تمرین ۶، تمرین ۵ را تکرار کنید.

- ۸ . با استفاده از چندجمله‌ای تیلور درجه پنج، فرمول (۳۲.۵) را بدست آورید.
- ۹ . با استفاده از فرمول (۱۳.۵)، چندجمله‌ای تیلور درجه پنج، و برونیاپی ریچاردسون، فرمول (۲۴.۵) را بدست آورید.

۳.۵ مبانی انتگرالگیری عددی

اغلب لازم است انتگرال معین تابعی حساب شود که هیچ پادمشتق صریحی نداشته یا مقادیر پادمشتق آن به آسانی بدست نمی‌آیند. روش تقریب $I(f) = \int_a^b f(x) dx$ روش انتگرالگیری عددی نام دارد و عموماً شامل فرمولی از نوع

$$I_{n+1}(f) = \sum_{i=0}^n a_i f(x_i)$$

برای تقریب $I(f)$ است. مثل گذشته، به بررسی و مینیم سازی خطای این تقریب توجه داریم، این خطا را با

$$E_{n+1}(f) = I(f) - I_{n+1}(f)$$

نشان خواهیم داد.

در روشهای انتگرالگیری عددی اغلب مفید است بدانیم که تقریب به ازای چه رده از چندجمله‌ایها نتیجه دقیق می‌دهد. تعریف زیر برای این منظور داده شده است.

تعریف ۱.۵ . درجه دقت، یا صحت، یک فرمول انتگرالگیری عدد صحیح و مثبت n است بطوریکه، به ازای هر چندجمله‌ای P_k از درجه نایبتر از n ، $E(P_k) = 0$ ، ولی به ازای چندجمله‌ای از درجه $(n+1)$ ، $E(P_{n+1}) \neq 0$.

چون انتگرالگیری و جمع بندی اعمالی خطی اند، یعنی به ازای هر جفت تابع انتگرال پذیر f و g و هر جفت ثابت حقیقی α و β ،

$$\int_a^b (\alpha f(x) + \beta g(x)) dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx$$

و

$$\sum_{i=0}^n (\alpha f(x_i) + \beta g(x_i)) = \alpha \sum_{i=0}^n f(x_i) + \beta \sum_{i=0}^n g(x_i)$$

می‌توان دید (ر.ک. تمرین ۵) که درجه دقت یک فرمول انتگرالگیری n است اگر و فقط

اگر، بهازای هر $k = 0, 1, \dots, n$ ، $E(x^k) = 0$ ولی $E(x^{n+1}) \neq 0$.

روش انتگرالگیری را بر چند جمله‌ایهای درونیاب داده شده در فصل ۳ بنا می‌کنیم. برای این‌کار، ابتدا مجموعه $\{x_0, \dots, x_n\}$ از گره‌های متمایز از بازه $[c, d]$ را که شامل $[a, b]$ است اختیار می‌کنیم. فرض کنیم P_n چند جمله‌ای درونیاب لاگرانژ

$$P_n(x) = \sum_{i=0}^n f(x_i)L_i(x)$$

باشد، و با انتگرالگیری از P_n روی $[a, b]$ ، فرمول انتگرالگیری

$$I_{n+1}(f) = \int_a^b \sum_{i=0}^n f(x_i)L_i(x) dx = \sum_{i=0}^n a_i f(x_i),$$

را بدست می‌آوریم، که در آن

$$a_i = \int_a^b L_i(x) dx \quad , \quad i = 0, 1, \dots, n$$

بهازای هر $i = 0, 1, \dots, n$

مثال ۱. فرض کنیم $x_0 = a, h = (b - a)/2$ ؛ در نتیجه $x_1 = x_0 + h$ و $x_2 = x_0 + 2h = b$ برای تقریب

$$\int_a^b f(x) dx = \int_{x_0}^{x_2} f(x) dx,$$

از فرمول زیر استفاده می‌کنیم.

$$\begin{aligned} \int_{x_0}^{x_2} P_2(x) dx &= \int_{x_0}^{x_2} \left[\frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} f(x_0) + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} f(x_1) \right. \\ &\quad \left. + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} f(x_2) \right] dx \\ &= \frac{f(x_0)}{2h^2} \int_{x_0}^{x_2} (x - x_1)(x - x_2) dx - \frac{f(x_1)}{h^2} \int_{x_0}^{x_2} (x - x_0)(x - x_2) dx \\ &\quad + \frac{f(x_2)}{2h^2} \int_{x_0}^{x_2} (x - x_0)(x - x_1) dx. \end{aligned}$$

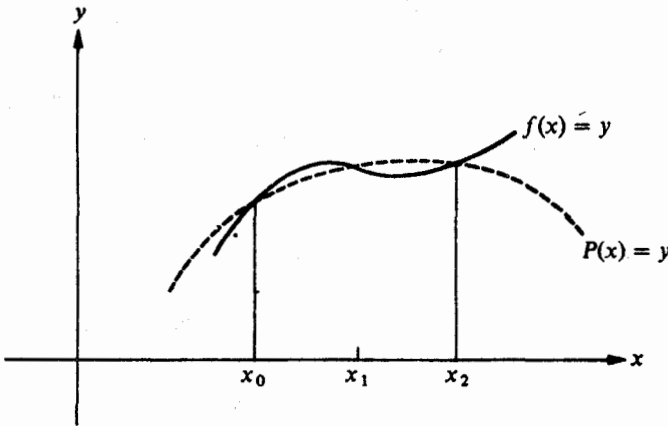
با استفاده از تغییر متغیر $x = x_0 + th$ ،

$$\int_{x_0}^{x_2} P_2(x) dx = \frac{f(x_0)h}{2h^2} \int_0^2 (ht - h)(ht - 2h) dt - \frac{f(x_1)h}{h^2} \int_0^2 (ht)(ht - 2h) dt + \frac{f(x_2)h}{2h^2} \int_0^2 (ht)(ht - h) dt = \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)].$$

فرمول انتگرالگیری

$$(۳۵.۵) \quad \int_{x_0}^{x_2} f(x) dx \approx \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)]$$

قاعده سیمپسون^۱ نامیده می شود (ر.ک. شکل ۲.۵).



شکل ۲.۵

برای بررسی خطای فرمولهای انتگرالگیری، فرض می کنیم $[c, d]$ شامل مجموعه $\{x_0, \dots, x_n\}$ و بازه انتگرالگیری $[a, b]$ باشد. هرگاه $f \in C^{n+1}[c, d]$ ، آنگاه، طبق قضیه ۳.۳، نقطه‌ای مانند $\xi \in [c, d]$ ، وابسته به x ، هست بطوری که

$$f(x) = \sum_{i=0}^n f(x_i)L_i(x) + \prod_{i=0}^n (x - x_i) \frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

انتگرالگیری روی $[a, b]$ نتیجه می دهد که

$$(۳۶.۵) \int_a^b f(x) dx = \sum_{i=0}^n a_i f(x_i) + \frac{1}{(n+1)!} \int_a^b f^{(n+1)}(\xi) \prod_{i=0}^n (x - x_i) dx,$$

که در آن، مثل قبل،

$$a_i = \int_a^b L_i(x) dx, \quad i = 0, 1, \dots, n$$

بنابراین،

$$E_{n+1}(f) = \frac{1}{(n+1)!} \int_a^b f^{(n+1)}(\xi) \prod_{i=0}^n (x - x_i) dx,$$

و کرانی به شکل

$$|E_{n+1}(f)| \leq \frac{1}{(n+1)!} \max_{x \in [a,b]} |f^{(n+1)}(x)| \int_a^b \prod_{i=0}^n |x - x_i| dx.$$

بدست می‌آید.

این نتیجه، همراه با نکات بعد از تعریف ۱.۵، ایجاب می‌کند که یک فرمول انتگرالگیری بر $(n+1)$ نقطه دارای درجه دقت دست کم n است، زیرا مشتق $(n+1)$ م x^n متحد صفر می‌باشد. کران $|E_{n+1}(f)|$ را می‌توان در اغلب حالات دقیقتر کرد، و با آنکه قضیه زیر برای حالت بسیار خاصی داده شده، نتیجه در بسیاری از فرمولهای انتگرالگیری یکی است.

قضیه ۲.۵. فرض کنیم بازه $[c, d]$ شامل گرههای $\{x_0, \dots, x_n\}$ و بازه $[a, b]$ باشد. هرگاه $f \in C^{n+1}[c, d]$ و حاصل ضرب $(x - x_0) \cdots (x - x_n)$ بر (a, b) تغییر علامت ندهد، آنگاه $\mu \in [a, b]$ ای هست که به ازای آن

$$(۳۷.۵) \quad I(f) - I_{n+1}(f) \equiv E_{n+1}(f) = \frac{f^{(n+1)}(\mu)}{(n+1)!} \int_a^b \prod_{j=0}^n (x - x_j) dx.$$

برهان. از معادله (۳۶.۵) داریم

$$E_{n+1}(f) = \frac{1}{(n+1)!} \int_a^b f^{(n+1)}(\xi) \prod_{j=0}^n (x - x_j) dx.$$

چون $f^{(n+1)}(\xi)$ تابع پیوسته‌ای از x است و $\prod_{j=0}^n (x - x_j)$ تغییر علامت نمی‌دهد، با

استفاده از قضیه مقدار میانگین برای انتگرالها (قضیه ۱۰.۱، ص ۷)، می توان $\mu \in [a, b]$ ای یافت که به ازای آن

$$\int_a^b f^{(n-1)}(\xi) \prod_{j=0}^n (x - x_j) dx = f^{(n+1)}(\mu) \int_a^b \prod_{j=0}^n (x - x_j) dx.$$

مثال ۲. فرض کنیم $x_1 = b$ ، $x_0 = a$ و $h = b - a$. با استفاده از فرمول انتگرالگیری

$$I_2(f) = \int_{x_0}^{x_1} \left[\frac{(x - x_1)}{(x_0 - x_1)} f(x_0) + \frac{(x - x_0)}{(x_1 - x_0)} f(x_1) \right] dx,$$

تقریبی برای $\int_a^b f(x) dx$ بدست می آوریم و، با انتگرال گرفتن، خواهیم داشت

$$I_2(f) = \frac{x_1 - x_0}{2} [f(x_0) + f(x_1)].$$

چون $(x - x_0)(x - x_1)$ همیشه بر (x_0, x_1) منفی است، قضیه ۲.۵ ایجاب می کند که، به ازای $f \in C^2[x_0, x_1]$ ، $\mu \in [x_0, x_1]$ ای هست که به ازای آن

$$E_2(f) = \frac{f''(\mu)}{2} \int_{x_0}^{x_1} (x - x_0)(x - x_1) dx.$$

با استفاده از تغییر متغیر $x = x_0 + th$ و انتگرالگیری، بدست می آید

$$(۳۸.۵) \quad E_2(f) = \frac{-f''(\mu)h^3}{12}$$

این فرمول انتگرالگیری

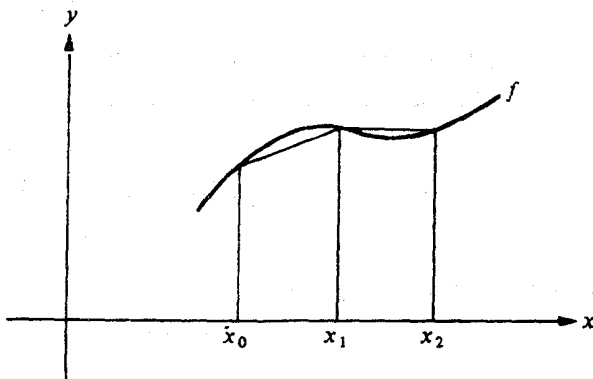
$$(۳۹.۵) \quad \int_{x_0}^{x_1} f(x) dx = \frac{h}{2} [f(x_0) + f(x_1)] - \frac{f''(\mu)h^3}{12}$$

قاعده دوزنقه ای با جمله خطا نامیده می شود (ر. ک. شکل ۳.۵).

مثال ۳. برای مقایسه قاعده سیمپسون، فرمول (۳۵.۵)، و قاعده دوزنقه ای، یعنی فرمول (۳۹.۵)، روی $[0, 2]$ ، ملاحظه می کنیم که در این حالت قاعده سیمپسون عبارت

است از

$$\int_0^2 f(x) dx \approx \frac{1}{3} [f(0) + 4f(1) + f(2)],$$



شکل ۹.۵

وقاعده دوزنقهای عبارت است از

$$\int_0^2 f(x) dx \approx f(0) + f(2).$$

این نتایج برای چند نمونه از توابع در جدول ۵.۵ خلاصه شده‌اند.

	$f(x) = 1$	$f(x) = x$	$f(x) = x^2$	$f(x) = x^3$	$f(x) = x^4$	$f(x) = e^x$
مقدار دقیق	2	2	2.67	4	6.40	6.389
دوزنقهای	2	2	4	8	16	8.389
سیمپسون	2	2	2.67	4	6.67	6.421

جدول ۵.۵

می‌توان دید که درجه دقت قاعده دوزنقهای یک و درجه دقت قاعده سیمپسون سه است. لذا، همانطور که در حالت $f(x) = e^x$ می‌بینیم، از قاعده سیمپسون انتظار نتایج فوق‌العاده بهتری نسبت به قاعده دوزنقهای می‌رود.

در بخش بعد، جملات خطا برای نقاط هم فاصله با جزئیات بیشتری بررسی می‌شود.

مجموعه تمرینات ۳.۵

۱. با استفاده از قواعد سیمپسون و دوزنقهای، $\int_0^{\pi/2} \sin^2 x \, dx$ را تقریب کنید. در هر حالت یک کران ماکزیمم برای خطا پیدا کنید. تقریبهای حاصل را با نتیجه واقعی مقایسه نمایید.
۲. با استفاده از قواعد دوزنقهای و سیمپسون، $\int_1^2 \ln x \, dx$ را تقریب کنید. این تقریبها را با مقدار واقعی مقایسه نمایید.
۳. با استفاده از قواعد دوزنقهای و سیمپسون، $\int_0^1 x^{1/3} \, dx$ را تقریب کنید. این تقریبها را با مقدار واقعی مقایسه نمایید.
۴. با استفاده از جدول زیر و بکار بردن (آ) قاعده دوزنقهای با $x_0 = 1.1$ و $x_1 = 1.5$ ؛
(ب) قاعده سیمپسون با $x_0 = 1.1$ ، $x_1 = 1.3$ ، و $x_2 = 1.5$ تقریبی برای $\int_{1.1}^{1.5} e^x \, dx$ بدست آورید.

x	e^x
1.1	3.0042
1.3	3.6693
1.5	4.4817

۵. عبارت بعد از تعریف ۱.۵ را اثبات کنید؛ یعنی، نشان دهید که درجه دقت یک فرمول انتگرالگیری وقتی بهازای تمام $k = 0, 1, \dots, n$ ، $E(x^k) = 0$ و $E(x^{n+1}) \neq 0$ دقیقاً برابر n است.
۶. با انجام انتگرالگیری در معادله (۳۸.۵)، نشان دهید که $E_2(f) = -f''(\mu)h^3/12$.

۴.۵ فرمولهای نیوتن - کاتس^۱

در فرمولهای انتگرالگیری این بخش نقاط هم فاصله یا گرههای x_i بکار می‌روند که در آنها بهازای هر $i = 0, 1, \dots, n$ ، $x_i = x_0 + ih$ ، که در آن $x_0 = a$ ، $x_n = b$ ، $h = (b - a)/n$ ، فرمول بسته نیوتن - کاتس نامیده می‌شود. لذا،

$$a_i = \int_a^b L_i(x) dx$$

$$= \int_{x_0}^{x_n} \prod_{\substack{j=0 \\ j \neq i}}^n \frac{(x_i - x_j)}{(x_i - x_j)} dx$$

با یکار بردن تغییر متغیر $x = x_0 + th$ نتیجه می شود که

$$a_i = h \int_0^n \prod_{\substack{j=0 \\ j \neq i}}^n \frac{(t-j)}{(i-j)} dt$$

$$= h \int_0^n \frac{(t)(t-1) \cdots (t-i-1)(t-i+1) \cdots (t-n)}{i(i-1) \cdots 1 \cdot (-1) \cdots (i-n)} dt$$

$$= \frac{h}{i!} \int_0^n \frac{t(t-1) \cdots (t-i-1)(t-i+1) \cdots (t-n)}{-1(-2) \cdots (-(n-i))} dt$$

$$= \frac{(-1)^{n-i} h}{i!(n-i)!} \int_0^n t(t-1) \cdots (t-i-1)(t-i+1) \cdots (t-n) dt$$

فرمول انتگرالگیری، نظیر قبل، عبارت است از

$$\int_a^b f(x) dx = \sum_{i=0}^n a_i f(x_i) + E_{n+1}(f)$$

انتظار می رود که جمله خطا به شکل

$$E_{n+1}(f) = \frac{1}{(n+1)!} \int_a^b f^{(n+1)}(\xi) \prod_{j=0}^n (x-x_j) dx$$

باشد، ولی به علت انتخاب خاص گرهبها، تحلیل خطا را می توان به صورت زیر بهتر کرد:

قضیه ۳۰۵. فرض کنیم $x_0 = a$ ، $x_n = b$ ، و $h = (b-a)/n$. نقطه های مانند $\xi \in [a, b]$ وجود دارد که برای آن

$$(40.5) \int_a^b f(x) dx = \sum_{i=0}^n a_i f(x_i) + \frac{h^{n+3} f^{(n+2)}(\xi)}{(n+2)!} \int_0^n t^2(t-1) \cdots (t-n) dt$$

اگر n زوج و $f \in C^{n+2}[a, b]$ و

$$(40.5) \int_a^b f(x) dx = \sum_{i=0}^n a_i f(x_i) + \frac{h^{n+2} f^{(n+1)}(\xi)}{(n+1)!} \int_0^n t(t-1) \cdots (t-n) dt$$

اگر n فرد و $f \in C^{n+1}[a, b]$

هرچند عامل

$$(x - x_0)(x - x_1) \cdots (x - x_n) = h^{n+1}t(t-1) \cdots (t-n)$$

$(n-1)$ بار در $[a, b]$ تغییر علامت می دهد، می توان تخمینهایی نظیر (۳۷.۵) بدست آورد. در قضیه ۳.۵ وقتی n عدد صحیح زوجی باشد، درجه دقت $(n+1)$ است، اگرچه درجه چندجمله‌ای درونیاب حداکثر n می باشد. درحالتی که n فرد است، قسمت دوم قضیه نشان می دهد که درجه دقت n می باشد. در نتیجه، اگر n زوج بوده و برای بهتر کردن دقت باید گرهای بیشتری اضافه شوند، فقط با افزودن یک گره هیچ افزایشی در دقت حاصل نمی شود و گرهای اضافه شده باید مضاربی از دو باشند.

تعدادی از فرمولهای بسته نیوتن - کاتس برای n کوچک در زیر آمده اند:

$$n = 1 \quad (\text{قاعده دوزنقه‌ای}).$$

$$(41.5) \quad \int_{x_0}^{x_1} f(x) dx = \frac{h}{2} (f(x_0) + f(x_1)) - \frac{h^3}{12} f^{(2)}(\xi)$$

که در آن $x_0 < \xi < x_1$

$$n = 2 \quad (\text{قاعده سیمپسون}):$$

$$(42.5) \quad \int_{x_0}^{x_2} f(x) dx = \frac{h}{3} (f(x_0) + 4f(x_1) + f(x_2)) - \frac{h^5}{90} f^{(4)}(\xi)$$

که در آن $x_0 < \xi < x_2$

$$n = 3 \quad (\text{قاعده سه هشت سیمپسون}):$$

$$(43.5) \quad \int_{x_0}^{x_3} f(x) dx = \frac{3h}{8} (f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)) - \frac{3h^5}{80} f^{(4)}(\xi)$$

که در آن $x_0 < \xi < x_3$

$$n = 4$$

$$(44.5) \quad \int_{x_0}^{x_4} f(x) dx = \frac{2h}{45} (7f(x_0) + 32f(x_1) + 12f(x_2) + 32f(x_3) + 7f(x_4)) - \frac{8h^7}{945} f^{(6)}(\xi)$$

که در آن $x_0 < \xi < x_4$

فرمولهای بالا را می‌توان مستقیماً با استفاده از معادلات $(\bar{A}40.5)$ و $(\bar{A}40.5)$ بدست آورد، ولی در عوض از طریقی دیگر یکی از این فرمولها را بدست می‌آوریم. فرض کنید بخواهیم قاعدهٔ سیمپسون با جملهٔ خطا، یعنی معادلهٔ $(\bar{A}20.5)$ ، را بدست آوریم. چون n زوج است، از نقاط درونیابی $x_0, x_1 = x_0 + h, x_2 = x_0 + 2h$ استفاده می‌شود. از معادلهٔ $(\bar{A}40.5)$ نتیجه می‌شود که ثابتی چون K وجود دارد بطوری که

$$\int_{x_0}^{x_2} f(x) dx = a_0 f(x_0) + a_1 f(x_1) + a_2 f(x_2) + Kf^{(4)}(\xi).$$

چون خطا مربوط به مشتق چهارم f است، درجهٔ دقت حداقل سه بوده و این فرمول برای $1, x, x^2$ باید دقیق باشد. معادلات خطی بدست آمده از

$$\int_{x_0}^{x_2} dx = a_0 + a_1 + a_2,$$

$$\int_{x_0}^{x_2} x dx = a_0 x_0 + a_1 x_1 + a_2 x_2$$

و

$$\int_{x_0}^{x_2} x^2 dx = a_0 x_0^2 + a_1 x_1^2 + a_2 x_2^2$$

را در نظر می‌گیریم. برای تعیین ضرایب a_0, a_1, a_2

$$a_0 + a_1 + a_2 = x_2 - x_0,$$

$$a_0 x_0 + a_1 x_1 + a_2 x_2 = \frac{1}{2}(x_2^2 - x_0^2)$$

و

$$a_0 x_0^2 + a_1 x_1^2 + a_2 x_2^2 = \frac{1}{3}(x_2^3 - x_0^3)$$

را بکار می‌بریم. این دستگاه به سادگی حل و

$$a_0 = \frac{h}{3}, \quad a_1 = \frac{4h}{3}, \quad a_2 = \frac{h}{3}$$

بدست می‌آیند. برای تعیین خطای $Kf^{(4)}(\xi)$ ، از این مطلب که درجهٔ دقت سه است

استفاده کرده بدست می آوریم

$$\int_{x_0}^{x_2} x^4 dx = \frac{h}{3} [x_0^4 + 4x_1^4 + x_2^4] + K \frac{d^4}{dx^4} (x^4)$$

در نتیجه،

$$\frac{1}{5}(x_2^5 - x_0^5) = \frac{h}{3} [x_0^4 + 4x_1^4 + x_2^4] + 4!K$$

چون $x_2 = x_0 + 2h$ و $x_1 = x_0 + h$ ، این عبارت به صورت

$$\begin{aligned} 2hx_0^4 + 8x_0^3h^2 + 16x_0^2h^3 + 16x_0h^4 + \frac{32}{5}h^5 \\ = 2hx_0^4 + 8x_0^3h^2 + 16x_0^2h^3 + 16x_0h^4 + \frac{20}{3}h^5 + 4!K \end{aligned}$$

یا

$$K = \frac{(32/5 - 20/3)}{4!} h^5 = \frac{-h^5}{90}$$

خلاصه می شود . در نتیجه ،

$$\int_{x_0}^{x_2} f(x) dx = \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)] - \frac{h^5}{90} f^{(4)}(\xi)$$

در فرمولهای باز نیوتن-گاتس ، نقاط $x_j = x_0 + jh$ به ازای هر $j = 0, 1, \dots, n$ که در آن $h = (b - a)/(n + 2)$ و $x_0 = a + h$ ، بکار می روند . این ایجاب می کند که $x_n = b - h$ ؛ در نتیجه ، اگر نقاط انتهایی را با $x_{-1} = a$ و $x_{n+1} = b$ نشان دهیم ، فرمولها به صورت

$$\int_{x_{-1}}^{x_{n+1}} f(x) dx = \sum_{i=0}^n a_i f(x_i) + E_{n+1}(f)$$

درمی آیند ، که مجدداً " در آنها

$$a_i = \int_a^b L_i(x) dx$$

قضیه ای شبیه قضیه ۳.۵ به صورت زیر است :

قضیه ۴.۵ . فرض کنیم $x_{n+1} = b$ ، $x_{-1} = a$ و $h = (b - a)/(n + 2)$. نقطه ای

مانند $\xi \in [a, b]$ وجود دارد که برای آن

$$(۳۴۵.۵) \int_a^b f(x) dx = \sum_{i=0}^n a_i f(x_i) + \frac{h^{n+3} f^{(n+2)}(\xi)}{(n+2)!} \int_{-1}^{n+1} t^2(t-1)\cdots(t-n) dt$$

اگر n زوج باشد و $f \in C^{n+2}[a, b]$ ، و

$$(۳۵۵.۵) \int_a^b f(x) dx = \sum_{i=0}^n a_i f(x_i) + \frac{h^{n+2} f^{(n+1)}(\xi)}{(n+1)!} \int_{-1}^{n+1} t(t-1)\cdots(t-n) dt$$

اگر n فرد باشد و $f \in C^{n+1}[a, b]$ ،

ضرایب و جملات خطا را می توان با انتگرالگیری مستقیم یا با روشی مبتنی بر قضیه ۳۰۵ تعیین کرد .

تعدادی از فرمولهای باز نیوتن - کاتس با جملات خطای آنها در زیر آمده اند :
 $n = 0$ (قاعده نقطه میانی) :

$$(۴۶.۵) \int_{x_{-1}}^{x_1} f(x) dx = 2hf(x_0) + \frac{h^3}{3} f''(\xi)$$

که در آن $x_{-1} < \xi < x_1$ ؛
 $n = 1$

$$(۴۷.۵) \int_{x_{-1}}^{x_2} f(x) dx = \frac{3h}{2} [f(x_0) + f(x_1)] + \frac{3h^3}{4} f''(\xi)$$

که در آن $x_{-1} < \xi < x_2$ ؛
 $n = 2$

$$(۴۸.۵) \int_{x_{-1}}^{x_3} f(x) dx = \frac{4h}{3} (2f(x_0) - f(x_1) + 2f(x_2)) + \frac{14h^5}{45} f^{(4)}(\xi)$$

که در آن $x_{-1} < \xi < x_3$.

انتخاب n و اینکه فرمولهای باز یا بسته بکار رود تحت تاثیر عوامل متعددی است . توجه کنید که ، گرچه فرمولهای باز دقت کمتری می دهند محاسبات کمتری نیز لازم دارند . انتخاب تعداد نقاط به دقت مورد نیاز و سهولت محاسبه تابع f بستگی دارد . همچنین ،

وقتی انتگرالده در a یا b تعریف نشده، فرمول باز بکار می‌رود. برای تحلیل خطاهای گرد کردن مربوط به روشهای انتگرالگیری عددی، فرمول انتگرالگیری

$$(۴۹.۵) \quad \int_a^b f(x) dx \approx \sum_{i=0}^n a_i f(x_i)$$

که یکی از فرمولهای نیوتن-کاتس از نوع باز یا بسته است، را در نظر می‌گیریم. با $a = x_0 + qh$ و $b = x_0 + ph$ ، که در آنها $q = 0$ و $p = n$ برای فرمولهای بسته و $q = -1$ و $p = n + 1$ برای فرمولهای باز، داریم

$$a_i = \frac{(-1)^{n-i} h}{i!(n-i)!} \int_a^p t(t-1) \cdots (t-i+1)(t-i-1) \cdots (t-n) dt$$

در نتیجه،

$$a_i = hm_i$$

که در آن

$$m_i = \frac{(-1)^{n-i}}{i!(n-i)!} \int_a^p t(t-1) \cdots (t-i+1)(t-i-1) \cdots (t-n) dt$$

بنابراین، فرمول انتگرالگیری (۴۹.۵) به صورت زیر درمی‌آید:

$$\int_a^b f(x) dx \approx h \left(\sum_{i=0}^n m_i f(x_i) \right)$$

فرض کنیم به‌ازای هر $i = 0, 1, \dots, n$ ، $f(x_i) = \tilde{f}(x_i) + e_i$ ، که در آن e_i خطای گرد کردن مربوط به استفاده از $\tilde{f}(x_i)$ برای تقریب $f(x_i)$ است. در این صورت،

$$\int_a^b f(x) dx \approx h \left(\sum_{i=0}^n m_i \tilde{f}(x_i) \right) + h \left(\sum_{i=0}^n m_i e_i \right)$$

که در آن جمله

$$r_{n+1} = h \left(\sum_{i=0}^n m_i e_i \right)$$

خطای گرد کردن مجتمع است. اگر به‌ازای هر $i = 0, 1, \dots, n$ ، $|e_i| < \varepsilon$ ، داریم

$$(۵۰.۵) \quad |r_{n+1}| \leq h\varepsilon \left(\sum_{i=0}^n |m_i| \right)$$

در حالت خاصی که همه $m_i \geq 0$ ها، که در تمام فرمولهای بسته مرتبه پایین و بعضی از فرمولهای باز رخ می دهد،

$$|r_{n+1}| \leq h\epsilon \left(\sum_{i=0}^n m_i \right)$$

چون همه فرمولها برای $f(x) = 1$ دقیق می باشند، نتیجه می شود که

$$\int_a^b dx = h \left(\sum_{i=0}^n m_i \right)$$

و

$$h \left(\sum_{i=0}^n m_i \right) = b - a$$

در نتیجه،

$$|r_{n+1}| \leq \epsilon(b - a)$$

توجه کنید که این کران به h بستگی ندارد و، در نتیجه، وقتی $h \rightarrow 0$ ، این روند

عدد پایدار است، چیزی که در روشهای مشتقگیری عددی مورد بحث در بخش ۱۰۵ درست نبود.

وقتی بعضی از m_i ها منفی اند، کران

$$|r_{n+1}| \leq h\epsilon \sum_{i=0}^n |m_i|$$

به آن خوبی نیست؛ اما مجدداً می توان نشان داد که این روند پایدار است.

مجموعه تمرینات ۴.۵

۱. با استفاده از فرمول بسته نیوتن - کاتس برای $n = 3$ و فرمول باز برای $n = 2$ ،

$$\int_0^1 e^{-x^2} dx$$

را تقریب کنید.

۲. با استفاده از فرمول بسته برای $n = 4$ ، تمرین ۱ را تکرار کنید.

۳. با استفاده از فرمولهای (۴۲.۵)، (۴۳.۵)، (۴۴.۵)، انتگرالهای زیر را تقریب

کنید. آیا دقت تقریبها با فرمولهای خطا سازگار است؟ از قسمتهای (ت) و (ث)

کدام تقریب بهتری را ارائه می دهد؟

$$\int_0^1 \sqrt{1+x} dx \quad (T)$$

(ب) $\int_0^{\pi/2} \sin^2 x \, dx$:

(پ) $\int_{1.1}^{1.5} e^x \, dx$:

(ت) $\int_1^{10} \frac{1}{x} \, dx$:

(ث) $\int_1^{5.5} \frac{1}{x} \, dx + \int_{5.5}^{10} \frac{1}{x} \, dx$:

(ج) $\int_0^1 x^{1/3} \, dx$.

۴ . تابع f با مقادیر جدول بندی شده در زیر مفروض است :

x	1.8	2.0	2.2	2.4	2.6
$f(x)$	3.12014	4.42569	6.04241	8.03014	10.46675

با بکار بردن تمام فرمولهای انتگرالگیری قابل استفاده در این بخش، $\int_{1.8}^{2.6} f(x) \, dx$ را تقریب کنید .

۵ . فرض کنید داده‌های تمرین ۴ دارای خطاهای گرد کردن مندرج در جدول زیر باشد :

x	1.8	2.0	2.2	2.4	2.6
$f(x)$ در خطا	2×10^{-6}	-2×10^{-6}	-9×10^{-6}	-9×10^{-6}	2×10^{-6}

خطاهای مربوط به گرد کردن در تمرین ۴ را محاسبه کنید .

۶ . با هریک از روشهای ارائه شده به وسیله معادلات (۴۱.۵) تا (۴۴.۵) و (۴۶.۵) تا (۴۸.۵)، تقریبی برای $\int_0^{0.5} (1+x)^{-1} \, dx$ تعیین و نتایج حاصله را با مقدار دقیق 9.162907 مقایسه کنید .

۷ . قاعده سه هشتم سیمپسون، معادله (۴۳.۵)، با جمله خطا را با استفاده از قضیه ۳.۵ نتیجه بگیرید .

۸ . معادله (۴۷.۵) با جمله خطا را با استفاده از قضیه ۴.۵ نتیجه بگیرید .

۹ . معادله (۴۸.۵) با جمله خطا را به روشی مشابه روش بدست آوردن قاعده سیمپسون

به گونه‌ای دیگر نتیجه بگیرید .

۵.۵ انتگرالگیری عددی مرکب

برای محاسبه انتگرال $\int_a^b f(x) dx$ ، وقتی $b - a$ نسبتاً بزرگ است ، می‌توان مجموعه‌ای از گره‌های $\{x_0, \dots, x_n\}$ برای یک مقدار بزرگ n را انتخاب کرد ، گرچه این روش موانع متعدد در پیش دارد . اگر ضرایب a_i در فرمول انتگرالگیری $\sum_{i=0}^n a_i f(x_i)$ بخواهند به سادگی تعیین شوند باید نقاط هم فاصله $x_i = x_0 + ih$ به‌ازای $h > 0$ ی انتخاب گردند . متأسفانه ، چندجمله‌ای درونیاب $P_n(x)$ با درجه زیاد روی نقاط هم فاصله نمی‌توانند تقریب مناسبی به $f(x)$ باشند و در تمام حالات ممکن فرمول انتگرالگیری دقیق نیست . در عوض ، اگر گره‌های x_0, \dots, x_n طوری انتخاب شوند که P_n به‌ازای n بزرگی تقریب مناسبی به $f(x)$ باشد ، ممکن است ضرایب a_i به سختی تعیین شوند . به این دلایل ، معمولاً "استفاده تنها از یک چندجمله‌ای با درجه زیاد برای تقریب یک‌انتگرال روی یک بازه بزرگ مناسب نیست .

فرض کنیم تعیین تقریبی $\int_0^4 e^x dx$ مورد نظر است . بازه $[0, 4]$ بازه نسبتاً بزرگی بنظر می‌رسد ؛ طبق دلایل بالا ، یک چندجمله‌ای از درجه زیاد احتمالاً "تقریب نادقیقی برای این انتگرال ارائه می‌دهد یا محاسبه آن مشکل است . لیکن ، بکار بردن یک چندجمله‌ای از درجه کم مسئله را حل نمی‌کند . فرض کنیم از قاعده سیمپسون با $h = 2$ استفاده شده باشد . در این صورت ،

$$\int_0^4 e^x dx \approx \frac{2}{3}(e^0 + 4e^2 + e^4) = 56.76958$$

چون جواب دقیق در این حالت $e^4 - e^0 = 53.59815$ است . خطای -3.17143 خیلی بیشتر از خطایی است که معمولاً به عنوان خطای قابل قبول در نظر گرفته می‌شود .

در بخش ۶.۳ ، وقتی تقریب چندجمله‌ای مناسب نبود ، از قطعه قطعه چندجمله‌ایها برای تقریب استفاده شد . در اینجا ، برای استفاده از چنین روشی ، $[0, 4]$ را به $[2, 4]$ و $[0, 2]$ تقسیم و با دوبار استفاده از قاعده سیمپسون ، یعنی معادله (۴۲.۵) ، به‌ازای $h = 1$ بدست می‌آید

$$\int_0^4 e^x dx =: \int_0^2 e^x dx + \int_2^4 e^x dx$$

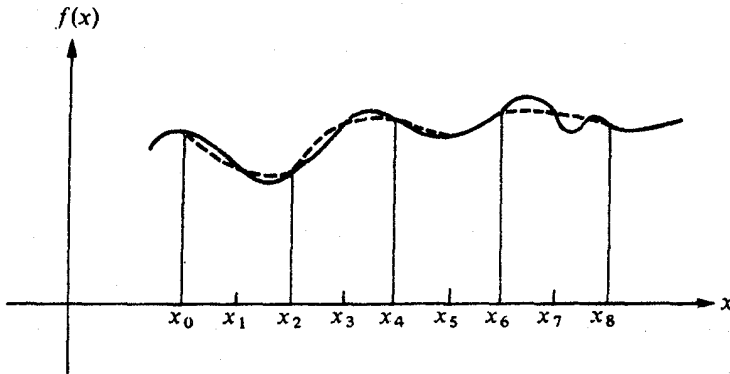
$$\begin{aligned} &\approx \frac{1}{3}[e^0 + 4e + e^2] + \frac{1}{3}[e^2 + 4e^3 + e^4] \\ &= \frac{1}{3}[e^0 + 4e + 2e^2 + 4e^3 + e^4] \\ &= 53.86385. \end{aligned}$$

خطای واقعی به 26570- . با رضایت از این نتایج، هریک از بازه‌های بالا را تقسیم و قاعده سیمپسون به‌ازای $h = \frac{1}{2}$ را بکار می‌بریم، نتیجه می‌شود

$$\begin{aligned} \int_0^4 e^x dx &= \int_0^1 e^x dx + \int_1^2 e^x dx + \int_2^3 e^x dx + \int_3^4 e^x dx \\ &\approx \frac{1}{6}[e^0 + 4e^{1/2} + e] + \frac{1}{6}[e + 4e^{3/2} + e^2] \\ &\quad + \frac{1}{6}[e^2 + 4e^{5/2} + e^3] + \frac{1}{6}[e^3 + 4e^{7/2} + e^4] \\ &= \frac{1}{6}[e^0 + 4e^{1/2} + 2e + 4e^{3/2} + 2e^2 + 4e^{5/2} + 2e^3 + 4e^{7/2} + e^4] \\ &= 53.61622. \end{aligned}$$

خطای این تقریب 01807- است.

تعمیمی از این روند به این ترتیب است: بازه $[a, b]$ را به $2m$ زیربازه تقسیم و قاعده سیمپسون را روی هر جفت متوالی از آنها بکار می‌بریم. (ر.ک. شکل ۴.۵)



شکل ۴.۵

فرض کنیم $h = (b - a)/2m$ و $a = x_0 < x_1 < \dots < x_{2m} = b$ که در آن به‌ازای هر $j = 0, 1, \dots, 2m$ ، $x_j = x_0 + jh$ و به‌ازای j ای با $x_{2j-2} < \xi_j < x_{2j}$

$$\int_a^b f(x) dx = \sum_{j=1}^m \int_{x_{2j-2}}^{x_{2j}} f(x) dx$$

$$= \sum_{j=1}^m \left\{ \frac{h}{3} [f(x_{2j-2}) + 4f(x_{2j-1}) + f(x_{2j})] - \frac{h^5}{90} f^{(4)}(\xi_j) \right\}$$

با استفاده از این امر که به ازای هر $j = 1, 2, \dots, m-1$ ، $f(x_{2j})$ در جمله متناظر با بازه $[x_{2j-2}, x_{2j}]$ ، و نیز در بازه $[x_{2j}, x_{2j+2}]$ ظاهر می‌شود، این فرمول به صورت زیر درمی‌آید:

$$\int_a^b f(x) dx = \frac{h}{3} \left[f(x_0) + 2 \sum_{j=1}^{m-1} f(x_{2j}) + 4 \sum_{j=1}^m f(x_{2j-1}) + f(x_{2m}) \right]$$

$$(51.5) \quad - \frac{h^5}{90} \sum_{j=1}^m f^{(4)}(\xi_j).$$

قضیه ۵.۵. هرگاه $\mu \in [a, b]$ ، $f \in C^4[a, b]$ ای هست که به ازای آن قاعده مرکب سیمپسون روی $2m$ زیر بازه از $[a, b]$ را می‌توان با جمله خطا به صورت زیر بیان کرد:

$$\int_a^b f(x) dx = \frac{h}{3} \left[f(a) + 2 \sum_{j=1}^{m-1} f(x_{2j}) + 4 \sum_{j=1}^m f(x_{2j-1}) + f(b) \right]$$

$$(52.5) \quad - \frac{(b-a)h^4}{180} f^{(4)}(\mu),$$

که در آن به ازای هر $j = 0, 1, \dots, 2m$ و $x_j = x_0 + jh$ و $u = x_0 < x_1 < \dots < x_{2m} = b$ ، $h = (b-a)/2m$. برهان. طبق معادله (۵۱.۵)، خطای مربوط به این تقریب برابر

$$E(f) = -\frac{h^5}{90} \sum_{j=1}^m f^{(4)}(\xi_j)$$

است، که در آن به ازای هر $j = 1, 2, \dots, m$ ، $x_{2j-2} < \xi_j < x_{2j}$ ، چون $f \in C^4[a, b]$ ،

$$\min_{x \in [a, b]} f^{(4)}(x) \leq f^{(4)}(\xi_j) \leq \max_{x \in [a, b]} f^{(4)}(x);$$

در نتیجه،

$$m \min_{x \in [a, b]} f^{(4)}(x) \leq \sum_{j=1}^m f^{(4)}(\xi_j) \leq m \max_{x \in [a, b]} f^{(4)}(x)$$

$$\min_{x \in [a, b]} f^{(4)}(x) \leq \frac{1}{m} \sum_{j=1}^m f^{(4)}(\xi_j) \leq \max_{x \in [a, b]} f^{(4)}(x)$$

طبق قضیه مقدار میانی (قضیه ۱۲.۱ ، صفحه ۹) ، $\mu \in [a, b]$ ای وجود دارد بطوری که

$$f^{(4)}(\mu) = \frac{1}{m} \sum_{j=1}^m f^{(4)}(\xi_j)$$

بنابراین ،

$$E(f) = \frac{-h^5}{90} m f^{(4)}(\mu)$$

چون $h = (b - a)/2m$ ،

$$E(f) = \frac{-h^4(b - a)}{180} f^{(4)}(\mu)$$

الگوریتم زیر قاعده مرکب سیمپسون روی $2m$ زیربازه را بکار می برد و یک الگوریتم انتگرالگیری است که اغلب در عمل بکار می رود .

الگوریتم مرکب سیمپسون ۱.۵

برای تقریب $I = \int_a^b f(x) dx$ ، عدد صحیح $m > 0$ را انتخاب می کنیم .

مرحله ۱ . قرار می دهیم $h = (b - a)/2m$.

مرحله ۲ . به ازای هر $i = 0, 1, \dots, 2m$ ، قرار می دهیم $x_i = a + ih$.

مرحله ۳ . قرار می دهیم $i = 0$ و $I_2 = 0$.

مرحله ۴ . اگر $i = 0$ ، $f(x_i)$ را به I_2 افزوده و به مرحله ۵ می رویم .

اگر $i = 2m$ ، $f(x_i)$ را به I_2 افزوده و به مرحله ۵ می رویم .

اگر i فرد باشد ، $4f(x_i)$ را به I_2 افزوده و به مرحله ۵ می رویم .

$2f(x_i)$ را به I_2 می افزاییم .

مرحله ۵ . i را به $i + 1$ می افزاییم .

مرحله ۶ . اگر $i \leq 2m$ ، به مرحله ۴ می رویم .

مرحله ۷ . I_2 را در $\frac{1}{3}h$ ضرب می کنیم .

مرحله ۸ . روند تمام خواهد بود . I_1 ، I_2 را با خطای $O(h^4)$ تقریب می کند .

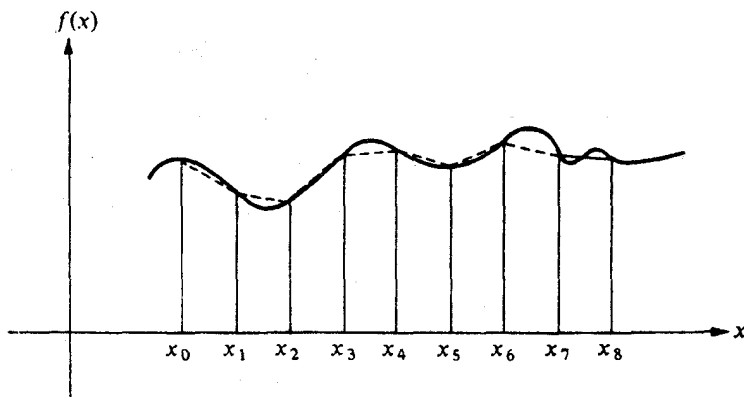
این روش را می توان به طور مشابه برای هر فرمول مرتبه^۵ پایین بکار برد. تعمیمهایی از قواعد ذوزنقه‌ای و نقطه^۵ میانی (ر.ک. شکل ۵.۵)، بدون اثبات، در دو قضیه^۵ زیر ارائه می شوند.

قضیه^۵ ۶.۰۵. فرض کنیم $f \in C^2[a, b]$. قاعده^۵ ذوزنقه‌ای برای m زیربازه با $x_j = a + jh$ و $h = (b - a)/m$ به ازای هر $j = 0, 1, \dots, m$ ، برای $\mu \in [a, b]$ ای عبارت است از

$$(۵۳.۰۵) \quad \int_a^b f(x) dx = \frac{h}{2} \left[f(a) + f(b) + 2 \sum_{j=1}^{m-1} f(x_j) \right] - \frac{(b-a)h^2}{12} f''(\mu)$$

قضیه^۵ ۷.۰۵. فرض کنیم $f \in C^2[a, b]$. قاعده^۵ نقطه^۵ میانی برای $2m$ زیربازه، با $x_j = a + (j+1)h$ و $h = (b-a)/(2m+1)$ به ازای هر $j = -1, 0, \dots, 2m+1$ ، برای $\mu \in [a, b]$ ای عبارت است از

$$(۵۴.۰۵) \quad \int_a^b f(x) dx = 2h \sum_{j=0}^m f(x_{2j}) + \frac{b-a}{6} h^2 f''(\mu)$$



شکل ۵.۵

مثال ۱. با استفاده از قاعده^۵ ترکیبی سیمپسون، تقریب $\int_0^{\pi} \sin x dx$ با حداکثر خطای ۰.۰۰۰۰۰۲ مورد نظر است. با اعمال معادله^۵ (۵۲.۰۵) بر انتگرال $\int_0^{\pi} \sin x dx$ نتیجه

می‌شود که

$$\int_0^\pi \sin x \, dx = \frac{h}{3} \left[2 \sum_{j=1}^{m-1} \sin x_{2j} + 4 \sum_{j=1}^m \sin x_{2j-1} \right] - \frac{\pi h^4}{180} \sin \mu.$$

چون لازم است خطای برشی کمتر از 0.00002 باشد، نامعادله

$$\left| \frac{\pi h^4}{180} \sin \mu \right| \leq \frac{\pi h^4}{180} = \frac{\pi^5}{2880m^4} \leq 0.00002$$

برای تعیین m و h بکار می‌رود. با اتمام محاسبات، نتیجه می‌شود که $m \geq 9$. به‌ازای $m = 10, h = \pi/20$ ، فرمول بدست آمده از (۵۲.۵) به‌صورت زیر درمی‌آید:

$$\int_0^\pi \sin x \, dx \approx \frac{\pi}{60} \left[2 \sum_{j=1}^9 \sin\left(\frac{j\pi}{10}\right) + 4 \sum_{j=1}^{10} \sin\left(\frac{(2j-1)\pi}{20}\right) \right]$$

برای اطمینان از این درجه دقت با استفاده از قاعده دوزنقه‌ای، یعنی معادله

(۵۳.۵)، لازم می‌آید که

$$\left| \frac{\pi h^2}{12} \sin \mu \right| \leq \frac{\pi h^2}{12} = \frac{\pi^3}{12m^2} < 0.00002$$

یا که $m \geq 360$. چون این محاسبات خیلی از محاسبات در قاعده سیمپسون بیشتر است، استفاده از قاعده دوزنقه‌ای در این مسئله نامطلوب می‌باشد. به‌منظور مقایسه، از قاعده دوزنقه‌ای با $m = 20$ و $h = \pi/20$ نتیجه می‌شود که

$$\begin{aligned} \int_0^\pi \sin x \, dx &\approx \frac{\pi}{40} \left[2 \sum_{j=1}^{19} \sin\left(\frac{j\pi}{20}\right) + \sin 0 + \sin \pi \right] \\ &= \frac{\pi}{40} \left[2 \sum_{j=1}^{19} \sin\left(\frac{j\pi}{20}\right) \right]. \end{aligned}$$

از قاعده سیمپسون جواب 2.00000679 و از قاعده دوزنقه‌ای جواب 1.9958860 بدست می‌آید. پاسخ دقیق 2 است؛ در نتیجه، از قاعده سیمپسون با کران خطای مورد نیاز جوابی بهتر عاید می‌شود.

مجموعه تمرینات ۵.۵

۱. با استفاده از قاعده^۳ ترکیبی سیمپسون، مقادیر m و h لازم برای تقریب $\int_0^1 e^x \sin x \, dx$ ، با خطای کمتر از 10^{-6} ، را بدست آورید. تقریب را مشخص کنید.
۲. با استفاده از تعمیم قاعده^۳ دوزنقهای، تمرین ۱ را تکرار کنید.
۳. با استفاده از تعمیم قاعده^۳ نقطه^۳ میانی، تمرین ۱ را تکرار کنید.
۴. در تمرین ۶ از بخش ۴.۵ خواسته شد که همه روشهای آن بخش برای تقریب $\int_0^5 (1+x)^{-1} \, dx$ بکاربرده شود. با استفاده از الگوریتم ۱.۵، با $m = 5$ ، تقریب دیگری برای این انتگرال بدست آورید. این نتیجه را با مقادیر بدست آمده در آن تمرین مقایسه کنید.
۵. با استفاده از قضیه^۳ ۶.۵، به جای الگوریتم ۱.۵، با $m = 2$ ، تمرین ۴ را تکرار نمایید.
۶. با استفاده از قاعده^۳ مرکب سیمپسون، $\int_1^{10} \ln x \, dx$ را تا 10^{-4} تقریب کنید.
۷. فرض کنید f به صورت زیر تعریف شده باشد:

$$f(x) = \begin{cases} x^3 + 1, & 0 \leq x \leq .1, \\ 1.001 + .03(x - .1) + .3(x - .1)^2 + 2(x - .1)^3, & .1 \leq x \leq .2, \\ 1.009 + .15(x - .2) + .9(x - .2)^2 + 2(x - .2)^3, & .2 \leq x \leq .3. \end{cases}$$

- (آ) پیوستگی مشتقهای f را بررسی کنید.
- (ب) با استفاده از تعمیم قاعده^۳ دوزنقهای، با $m = 6$ ، $\int_0^3 f(x) \, dx$ را تقریب کنید و با بکار بردن معادله^۳ (۵۳.۵) خطا را تخمین بزنید.
- (پ) با استفاده از قاعده^۳ مرکب سیمپسون، با $m = 6$ ، $\int_0^3 f(x) \, dx$ را تقریب کنید. آیا نتایج دقیقتر از قسمت (ب) هستند؟
۸. در یک مدار شامل ولتاژ موثر \mathcal{E} و یک خازن به ظرفیت C ، رابطه^۳

$$\mathcal{E} = Ri + \frac{1}{C} \int_0^t i \, dt$$

- برقرار است، که در آن R مقاومت و i شدت جریان مدار می باشد. فرض کنید در مداری $R = 1$ اهم و $C = 1$ فاراد و $i = \ln(2t + 1)$ آمپر باشد. با استفاده از قاعده^۳ مرکب سیمپسون، با $h = .1$ ، ولتاژ را پس از ۲ ثانیه بدست آورید.
۹. ذره ای به جرم m که در سیالی متحرک است تحت تاثیر مقاومت کششی R می باشد

که تابعی از سرعت v است. رابطه بین مقاومت R ، سرعت v ، و زمان t با معادله

$$t = \int_{v(0)}^{v(t)} \frac{m}{R(u)} du$$

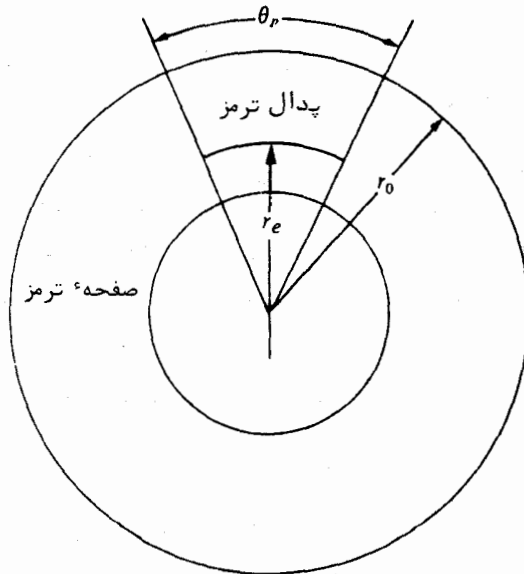
داده می‌شود. فرض کنید برای سیال خاصی $R(v) = -v\sqrt{v}$ ، که در آن R برحسب نیوتن و v برحسب متر بر ثانیه است. اگر $m = 10$ کیلوگرم و $v(0) = 10$ متر بر ثانیه باشد، زمان لازم برای اینکه سرعت ذره به $v = 5$ متر بر ثانیه برسد را تقریب کنید: با استفاده از

(آ) قاعده مرکب سیمپسون؛ $h = 2$.

(ب) تعمیم قاعده ذوزنق‌های با $h = 2$.

(پ) این تقریبها را با مقدار واقعی مقایسه نمایید.

۱۰. برای شبیه‌سازی مشخصات حرارتی صفحات ترمز، دی.ا.سکریست^۱ و ا.دبلیو. هورن‌بک^۲ [۵۸]، لازم دیدند که T ، یعنی "دمای سطح متوسط آستر" پدال



1. D. A. Secrist

2. R. W. Hornbeck

ترمز، را از فرمول زیر به طور عددی تقریب کنند :

$$T = \frac{\int_{r_e}^{r_0} Tr\theta_p dr}{\int_{r_e}^{r_0} r\theta_p dr}$$

که در آن r_e شعاع تماس اولیه، صفحه پدال، r_0 شعاع خارجی تماس صفحه پدال، θ_p زاویه مقابل به قطاع پدالهای ترمز، و T دمای هر نقطه از پدال است، که به طور عددی از تحلیل معادله حرارت (ر. ک. بخش ۱۱. ۳) بدست آمد. اگر $r_e = 308$ فوت و $r_0 = 478$ فوت و $\theta_p = 7051$ رادیان و دمای حساب شده در نقاط مختلف صفحه به شرح زیر باشد، با استفاده از قاعده مرکب سیمپسون، تقریبی برای T پیدا کنید.

r (فوت)	$T(r)$ (فارنهایت)
308	640
325	794
342	885
359	943
376	1034
393	1064
410	1114
427	1152
444	1204
461	1222
478	1239

۱۱. ناپیوستگی با جهش متناهی. فرض کنید f به صورت زیر تعریف شده باشد :

$$f(x) = \begin{cases} x^2 - 1, & 0 \leq x \leq .5, \\ e^x \sin x, & .5 < x \leq 1. \end{cases}$$

(آ) آیا f بر $[0, 1]$ پیوسته است؟

(ب) $\int_0^1 f(x) dx$ را حساب کنید.

(پ) با استفاده از قاعده مرکب سیمپسون، با $h = .05$ ، انتگرال موجود در

(ب) را تقریب کنید.

(ت) با استفاده از $(.5, .5\Delta)$ ، با $h = .04$ ، این انتگرال را تقریب کنید.

(ث) با استفاده از $(\Delta 2.5)$ بر $[0, .5]$ (با $f(.5) = -.75$) و بر $[.5, 1]$ ، با $h = .05$

$$\int_0^{.5} f(x) dx + \int_{.5}^1 f(x) dx$$

را تقریب کنید و نتایج بدست آمده را با نتایج قسمت (پ) مقایسه نمایید .

(ج) در کدام قسمت‌های (پ) ، (ت) ، و (ث) بهترین تقریب نتیجه می شود؟ آیا می توان در هریک از این سه قسمت فرمول خطایی بکار برد؟

۱۲ . ناپیوستگیهای نامتناهی . انتگرال

$$\int_a^b \frac{dx}{(x-a)^p}$$

همگراست (یا وجود دارد) اگر و فقط اگر $p < 1$ (ر.ک. فولکس [۲۹] ، صفحه ۴۷۳) . اگر بتوان تابع f را به شکل

$$f(x) = \frac{g(x)}{(x-a)^p}$$

نوشت ، که در آن $g \in C[a, b]$ ، انتگرال

$$\int_a^b f(x) dx$$

نیز وجود دارد . اگر بازای $n \geq 0$ ، $g \in C^{n+1}[a, b]$ ، چند جمله‌ای تیلور P از درجه n به صورت

$$P(x) = \sum_{k=0}^n \frac{g^{(k)}(a)}{k!} (x-a)^k$$

را می توان تشکیل داد . سپس می توان انتگرال f را به صورت زیر تقریب کرد :

$$\int_a^b f(x) dx = \int_a^b \frac{g(x) - P(x)}{(x-a)^p} dx + \int_a^b \frac{P(x)}{(x-a)^p} dx$$

که در آن تابع $(g(x) - P(x))/(x-a)^p$ در $x = a$ دارای n مشتق است . انتگرال

$$\int_a^b \frac{g(x) - P(x)}{(x-a)^p} dx$$

را می توان با هریک از فرمولهای انتگرالگیری در بخشهای ۴.۵ و ۵.۵ تقریب کرد ،

درحالی که انتگرال

$$\int_a^b \frac{P(x)}{(x-a)^p} dx$$

را می توان به طور صریح حساب کرد .

(\bar{T}) نشان دهید که

$$\int_a^b \frac{P(x)}{(x-a)^p} dx = \sum_{k=0}^n \frac{g^{(k)}(a)}{k!(k+1-p)} (b-a)^{k+1-p}$$

(ب) انتگرال

$$\int_0^1 \frac{e^x}{\sqrt{x}} dx$$

را با استفاده از قاعده مرکب سیمپسون ، با $h = .05$ ، تقریب کنید . آیا می توان

فرمول خطای (۵۲.۵) را بکار برد ؟

(پ) انتگرال

$$\int_0^1 \frac{e^{-x}}{\sqrt{1-x}} dx$$

را با استفاده از تعمیم قاعده دوزنقهای ، با $h = .05$ ، تقریب کنید .

۱۳ . انتگرالگیری با حدود نامتناهی . انتگرال $\int_a^\infty f(x) dx$ ، $a > 0$ ، در صورت وجود ، را

اغلب می توان با استفاده از فرمول انتگرالگیری پس از تغییر متغیر $t = x^{-1}$ تقریب

کرد .

(\bar{T}) نشان دهید که

$$\int_a^\infty f(x) dx = \int_0^{1/a} t^{-2} f\left(\frac{1}{t}\right) dt$$

(ب) با استفاده از قسمت (\bar{T}) و تعمیم قاعده نقطه میانی (۵۴.۵) ، با $h = .05$ ،

انتگرال

$$\int_1^\infty x^{-2} \sin x dx$$

را تقریب کنید .

(پ) با استفاده از قسمت (\bar{T}) و تعمیم قاعده نقطه میانی (۵۴.۵) ، با $h = .05$ ،

انتگرال

$$\int_0^{\infty} \sqrt{x} e^{-x} dx$$

را تقریب کنید.

(ن) با استفاده از قسمت (آ) و تعمیم قاعده نقطه میانی (۵۴.۵)، با $h = .05$ ،

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^2}$$

را تقریب کنید.

۱۴. فرض کنید جسمی به جرم m از سطح $x=R$ از زمین به طور قائم به بالا پرتاب شده باشد. اگر از همه مقاومتها جز جاذبه زمین صرف نظر شود، سرعت گریز v به وسیله

$$v^2 = 2gR \int_1^{\infty} f(z) dz$$

داده می شود، که در آن $f(z) = z^{-2}$ ، $z = \frac{x}{R}$ ، و g نیروی جاذبه در سطح زمین است. اگر $g = .00609$ متر بر مجذور ثانیه و $R = 3960$ کیلومتر باشد، با استفاده از قاعده نقطه میانی، با $h = .1$ ، سرعت گریز v را تقریب کنید.

۱۵. تحقیق کنید که قضیه مقدار میانی (قضیه ۱۲.۰۱) را می توان برای بدست آوردن معادله

$$f^{(4)}(\mu) = \frac{1}{m} \sum_{j=1}^m f^{(4)}(\xi_j)$$

که در برهان قضیه ۵.۵ لازم شد، بکار برد.

۶.۵ انتگرالگیری رامبرگ^۱

با آنکه قاعده دوزنقهای عملاً "ساده ترین فرمول نیوتن - کاتس است، کرارا" در بخشهای قبل نشان داده ایم که درجه دقتی که معمولاً لازم می شود ندارد. انتگرالگیری رامبرگ روشی است که کاربرد زیادی دارد، چون از قاعده دوزنقهای جهت تعیین تقریبهای مقدماتی استفاده کرده و با بکار بردن روند بیرونیابی ریچاردسون (که در بخش ۲۰.۵ مورد بحث قرار گرفت)

تقریبهای بهتری بدست می دهد.

برای شروع روش انتگرالگیری رامبرگ، بخاطر می آوریم که (قضیه ۶.۵) تعمیم قاعده دوزنقهای برای تقریب انتگرال تابع f بر بازه $[a, b]$ با بکار بردن m زیربازه عبارت است از

$$(۵۵.۵) \int_a^b f(x) dx = \frac{h}{2} \left[f(a) + f(b) + 2 \sum_{j=1}^{m-1} f(x_j) \right] - \frac{(b-a)}{12} h^2 f''(\mu)$$

که در آن $x_j = a + jh$ ، $j = 0, 1, \dots, m$ و بهمازای هر $h = (b-a)/m$ و $a < \mu < b$ اولین گام در فرایند رامبرگ بدست آوردن تقریبهایی به قاعده دوزنقهای است با $m_1 = 1, m_2 = 2, m_3 = 4, \dots, m_n = 2^{n-1}$ که در آن n عدد صحیح مثبتی است. مقادیر طول گام h_k ، متناظر با m_k ، برابر $h_k = (b-a)/m_k = (b-a)/2^{k-1}$ است، و با این نمادگذاری قاعده دوزنقهای عبارت است از

$$\int_a^b f(x) dx = \frac{h_k}{2} \left[f(a) + f(b) + 2 \left(\sum_{i=1}^{2^{k-1}-1} f(a + ih_k) \right) \right] - \frac{(b-a)}{12} h_k^2 f''(\mu_k).$$

که در آن μ_k نقطه‌ای در $[a, b]$ است.

هرگاه نماد $R_{k,1}$ نمایش بخشی از (۵۵.۵) باشد که برای تقریب دوزنقهای بکار

می رود، آنگاه

$$R_{1,1} = \frac{h_1}{2} [f(a) + f(b)],$$

$$= \frac{(b-a)}{2} [f(a) + f(b)];$$

$$R_{2,1} = \frac{h_2}{2} [f(a) + f(b) + 2f(a + h_2)]$$

$$= \frac{(b-a)}{4} \left[f(a) + f(b) + 2f\left(a + \frac{(b-a)}{2}\right) \right]$$

$$= \frac{1}{2} \left[R_{1,1} + h_1 f\left(a + \frac{1}{2} h_1\right) \right];$$

$$\begin{aligned}
 R_{3,1} &= \frac{h_3}{2} \left\{ f(a) + f(b) + 2 \left[f \left(a + \frac{(b-a)}{4} \right) \right. \right. \\
 &\quad \left. \left. + f \left(a + \frac{(b-a)}{2} \right) + f \left(a + \frac{3(b-a)}{4} \right) \right] \right\} \\
 &= \frac{(b-a)}{8} \left\{ f(a) + f(b) + 2 \left[f \left(a + \frac{(b-a)}{4} \right) \right. \right. \\
 &\quad \left. \left. + f \left(a + \frac{(b-a)}{2} \right) + f \left(a + \frac{3(b-a)}{4} \right) \right] \right\} \\
 &= \frac{1}{2} \left\{ R_{2,1} + h_2 \left[f \left(a + \frac{h_2}{2} \right) + f \left(a + \frac{3h_2}{2} \right) \right] \right\}
 \end{aligned}$$

و، در حالت کلی، به‌ازای هر $k = 2, 3, \dots, n$

$$(56.5) \quad R_{k,1} = \frac{1}{2} \left[R_{k-1,1} + h_{k-1} \sum_{i=1}^{2^{k-2}} f \left(a + \left(i - \frac{1}{2} \right) h_{k-1} \right) \right]$$

برای بدست آوردن معادله (56.5) خواننده را به تمرینهای ۶ و ۷ ارجاع می‌دهیم.

مثال ۱. با استفاده از (56.5) جهت تشکیل اولین مرحله روش انتگرالگیری رامبرگ برای تقریب $\int_0^{\pi} \sin x \, dx$ با $n = 6$ نتیجه می‌شود:

$$R_{1,1} = \frac{\pi}{2} [\sin 0 + \sin \pi] = 0;$$

$$\begin{aligned}
 R_{2,1} &= \frac{1}{2} \left[R_{1,1} + \pi \sin \frac{\pi}{2} \right] \\
 &= 1.57079633;
 \end{aligned}$$

$$\begin{aligned}
 R_{3,1} &= \frac{1}{2} \left[R_{2,1} + \frac{\pi}{2} \left(\sin \frac{\pi}{4} + \sin \frac{3\pi}{4} \right) \right] \\
 &= 1.89611890;
 \end{aligned}$$

$$R_{4,1} = \frac{1}{2} \left[R_{3,1} + \frac{\pi}{4} \left(\sin \frac{\pi}{8} + \sin \frac{3\pi}{8} + \sin \frac{5\pi}{8} + \sin \frac{7\pi}{8} \right) \right]$$

$$= 1.97423160$$

$$R_{5,1} = 1.99357034$$

9

$$R_{6,1} = 1.99839336$$

چون مقدار دقیق این انتگرال 2 است، واضح است که، اگرچه محاسبات مشکل نیستند، همگرایی خیلی کند است.

حال، برای سرعت دادن به همگرایی، روش برونیاپی ریچاردسون ارائه می شود. می توان، البته نه به سادگی، نشان داد که (برای توضیح کامل مطلب، ر.ک. رالستون [۵۰] صفحات ۱۳۵ - ۱۳۱) تعمیم قاعده دوزنقه‌ای داده شده در معادله (۵۵.۵) با جمله خطای دیگری به‌مازای هر $k = 1, 2, \dots, n$ و h_k^2 ای که $a < \mu_k < b$ به صورت زیر است:

$$\int_a^b f(x) dx = \frac{h_k}{2} \left[f(a) + f(b) + 2 \sum_{i=1}^{2^{k-1}-1} f(a + ih_k) \right] - \frac{h_k^2}{12} [f'(b) - f'(a)] + \frac{(b-a)h_k^4}{720} f^{(4)}(\mu_k).$$

با تعمیم قاعده دوزنقه‌ای به این شکل، می توان جمله شامل h_k^2 را از ترکیب

معادلات

$$\int_a^b f(x) dx = R_{k-1,1} - \frac{h_{k-1}^2}{12} [f'(b) - f'(a)] + \frac{(b-a)h_{k-1}^4}{720} f^{(4)}(\mu_{k-1})$$

9

$$\begin{aligned} \int_a^b f(x) dx &= R_{k,1} - \frac{h_k^2}{12} [f'(b) - f'(a)] + \frac{(b-a)h_k^4}{720} f^{(4)}(\mu_k) \\ &= R_{k,1} - \frac{h_{k-1}^2}{48} [f'(b) - f'(a)] + \frac{(b-a)h_k^4}{720} f^{(4)}(\mu_k) \end{aligned}$$

$$\begin{aligned} \int_a^b f(x) dx &= \frac{4R_{k,1} - R_{k-1,1}}{3} \\ &+ \frac{(b-a)}{2160} [4h_k^4 f^{(4)}(\mu_k) - h_{k-1}^4 f^{(4)}(\mu_{k-1})] \\ &= \frac{4R_{k,1} - R_{k-1,1}}{3} + O(h_k^4) \end{aligned}$$

را بدست آورد.

به آسانی می توان نشان داد که (ر.ک. تمرین ۵) تقریب حاصل از این روش در واقع تقریبی است که قاعده مرکب سیمپسون، یعنی قضیه ۵.۵ در بخش ۵.۵ با $h = h_k$ را می دهد. لذا، خطایی از مرتبه h_k^4 انتظار می رود.

برای ادامه روش رامبرگ، به ازای هر $k = 2, 3, \dots, n$ تعریف می کنیم

$$R_{k,2} = \frac{4R_{k,1} - R_{k-1,1}}{3}$$

و روش برونابی ریچاردسون را بر این مقادیر اعمال می کنیم. می توان نشان داد (ر.ک. رالستون [۴۹]) که این فرایند به ازای هر $i = 2, 3, 4, \dots, n$ و $j = 1, 2, \dots, i$

$$R_{i,j} = \frac{4^{j-1} R_{i,j-1} - R_{i-1,j-1}}{4^{j-1} - 1}$$

را نتیجه می دهد، که در آن مقادیر با اندیس j بزرگتر به فرمولهای نیوتن - کاتس از مرتبه متوالی بالاتر نظیر می شود. تقریبها اغلب در جدولی به شکل جدول ۵.۶ ارائه می شوند.

$R_{1,1}$				
$R_{2,1}$	$R_{2,2}$			
$R_{3,1}$	$R_{3,2}$	$R_{3,3}$		
$R_{4,1}$	$R_{4,2}$	$R_{4,3}$	$R_{4,4}$	
\vdots	\vdots	\vdots	\ddots	
$R_{n,1}$	$R_{n,2}$	$R_{n,3}$	\dots	$R_{n,n}$

برای نشان دادن اینکه جملات قطری به این انتگرال همگرا هستند اگر مقادیر $R_{n,1}$ به این عدد همگرا باشند، می توان قضیه جمعپذیری زیبایی از سیلورمن^۱ و تپلیتز^۲ را بکار برد. برهانی از این همراه با شرایط لازم برای همگرایی را می توان در صفحات ۱۲۵ - ۱۲۱ رالستون [۵۵] یافت. معمولاً انتظار داریم که دنباله قطری $\{R_{m,m}\}_{m=1}^{\infty}$ خیلی سریعتر از دنباله $\{R_{n,1}\}_{n=1}^{\infty}$ همگرا باشد. روش رامبرگ دارای این خاصیت مطلوب اضافی است که اجازه می دهد تمام درایه های سطری جدید در جدول به سادگی با اعمال قاعده ذوزنقه ای و سپس استفاده از مقادیر حساب شده قبلی محاسبه شوند تا درایه های متوالی در این سطر بدست آیند. روشی که معمولاً در ساختن جدولی از این نوع بکار می رود این خاصیت را با محاسبه سطر به سطر درایه ها ثبت می کند؛ یعنی، به ترتیب $R_{1,1}$ ، $R_{2,1}$ ، $R_{2,2}$ ، $R_{3,1}$ ، $R_{3,2}$ ، $R_{3,3}$ ، و غیره. الگوریتم زیر جزئیات این روش را شرح می دهد.

الگوریتم رامبرگ ۲.۵

برای تقریب $\int_a^b f(x) dx$ ، ابتدا عدد صحیح مثبت n را انتخاب می کنیم. فرض می کنیم $h_1 = b - a$ و مراحل زیر را انجام می دهیم:

مرحله ۱. قرار می دهیم $i = 1$ ، $j = 1$ ، و

$$R_{1,1} = \frac{h_1}{2} [f(a) + f(b)]$$

مرحله ۲. ۱ را به i می افزاییم.

مرحله ۳. اگر $i > n$ ، به مرحله ۹ می رویم.

مرحله ۴. قرار می دهیم $h_i = \frac{1}{2} h_{i-1}$ و

$$R_{i,1} = \frac{1}{2} \left[R_{i-1,1} + h_{i-1} \sum_{k=1}^{2^{i-2}} f \left(a + \left(k - \frac{1}{2} \right) h_{i-1} \right) \right]$$

مرحله ۵. ۱ را به j می افزاییم.

مرحله ۶. اگر $j > i$ ، قرار می دهیم $j = 1$ و به مرحله ۲ می رویم.

مرحله ۷. قرار می دهیم

$$R_{i,j} = \frac{2^{2(j-1)}R_{i,j-1} - R_{i-1,j-1}}{2^{2(j-1)} - 1}$$

مرحله ۸. به مرحله ۵ می‌رویم .

مرحله ۹. روند تمام است و $R_{n,n}$ تقریبی از $\int_a^b f(x) dx$ می‌باشد .

اغلب مفید است به‌جای داشتن مقدار ویژه^۶ از قبل تعیین شده‌ای برای n و پیراستن جزئی الگوریتم اجازه دهیم این روند تا وقتی ادامه یابد که مقداری از n که برای ϵ قابل تحمل مفروضی که در $\epsilon < |R_{n,n} - R_{n-1,n-1}|$ صدق می‌کند بدست آید .

مثال ۲. در مثال ۱ مقادیری برای $R_{1,1}$ تا $R_{n,1}$ برای تقریب $\int_0^\pi \sin x dx$ با $n = 6$ بدست آمد . با الگوریتم ۲۰.۵ ، جدول رامبرگ در جدول ۲۰.۵ ارائه می‌شود .

0					
1.57079633	2.09439511				
1.89611890	2.00455976	1.99857073			
1.97423160	2.00026917	1.99998313	2.00000555		
1.99357034	2.00001659	1.99999975	2.00000001	1.99999999	
1.99839336	2.00000103	2.00000000	2.00000000	2.00000000	2.00000000

جدول ۲۰.۵

مجموعه تمرینات ۶.۵

۱. با استفاده از روش انتگرالگیری رامبرگ ، تقریبی با دقت 10^{-6} برای $\int_1^3 e^x \sin x dx$ بدست آورید . جوابهای خود را با مقدار دقیق و مقدار بدست آمده در تمرین ۱ از بخش ۵.۵ مقایسه کنید .

۲. با استفاده از روش انتگرالگیری رامبرگ ، تقریبهایی برای $\int_0^{1.5} (1+x)^{-1} dx$ ، که جدول را برای $n = 6$ تکمیل می‌کند ، بدست آورید . این نتیجه را با مقادیر حاصل از تمرین ۴ از بخش ۵.۵ و تمرین ۶ از بخش ۴.۵ مقایسه کنید .

۳. $\int_1^{10} \ln x dx$ را با استفاده از روند انتگرالگیری رامبرگ ، با $n = 10$ ، تقریب کنید .

۴. فرض کنید $N = 10$. روند انتگرالگیری رامبرگ را بر انتگرال $\int_a^b f(x) dx$ ، برای توابع و مقادیر a و b در زیر ، تا $|R_{k,k} - R_{k-1,k-1}| \leq 10^{-5}$ ، یا تا $k > N$ ، اعمال کنید :

$$f(x) = x^{1/3}, \quad a = 0, b = 1 \quad (1)$$

$$f(x) = \begin{cases} x^3 + 1, & 0 \leq x \leq .1, \\ 1.001 + .03(x - .1) + .3(x - .1)^2 + 2(x - .1)^3, & .1 < x \leq .2, \quad (+) \\ 1.009 + .15(x - .2) + .9(x - .2)^2 + 2(x - .2)^3, & .2 < x \leq .3, \end{cases}$$

$$a = 0, b = .3 \quad \text{با}$$

۵. نشان دهید که تقریب حاصل از $R_{k,2}$ همان تقریبی است که قاعده مرکب سیمپسون

در قضیه ۵.۵ با $h = h_k$ بدست می دهد.

۶. نشان دهید که، به ازای هر k ,

$$\sum_{i=1}^{2^{k-1}-1} f\left(a + \frac{i}{2}h_{k-1}\right) = \sum_{i=1}^{2^{k-2}} f\left(a + \left(i - \frac{1}{2}\right)h_{k-1}\right) + \sum_{i=1}^{2^{k-2}-1} f\left(a + ih_{k-1}\right)$$

۷. با استفاده از تمرین ۶، درستی معادله (۵۶.۵) را تحقیق کنید؛ یعنی، نشان

دهید که، به ازای هر k ,

$$R_{k,1} = \frac{1}{2} \left[R_{k-1,1} + h_{k-1} \sum_{i=1}^{2^{k-2}} f\left(a + \left(i - \frac{1}{2}\right)h_{k-1}\right) \right]$$

۷.۵ انتگرالگیری گاوسی

فرمولهای بسته نیوتن - کاتس در بخش ۴.۵ با انتگرالگیری چندجمله‌ایهای درونیاب

لاگرانژ بدست آمدند. چون جمله خطا در چندجمله‌ای درونیاب لاگرانژ از درجه n

مستلزم مشتق $(n+1)$ م تابع مورد تقریب است، قبلاً "تذکر داده‌ایم که این فرمول در

تقریب هر چندجمله‌ای از درجه نابیشتر از n دقیق است. در نتیجه، فرمولهای بسته

نیوتن - کاتس دارای درجه دقت حداقل n می باشند. در واقع، درجه دقت فرمولهای

فرد دقیقاً n است، در حالی که فرمولهای زوج دارای درجه دقت $(n+1)$ می باشند.

وضعیت مشابهی برای فرمولهای باز نیوتن - کاتس، که بعداً "در همان بخش ارائه شدند،

موجود است.

همه فرمولهای نیوتن - کاتس نیاز دارند که مقادیری از تابع، که انتگرال آن

تقریب می شود، در نقاط مساوی الفاصله معلوم باشند، که این وضع وقتی انتظار می رود

که برای تابع داده‌های جدولی بکار برده شود، اما اگر تابع صریحاً "داده شده باشد،

می توان، برای محاسبه تابع، این نقاط را به روشی دیگر انتخاب کرد، که منجر به افزایش دقت تقریب می شود. **انتگرالگیری گاوسی** مربوط به انتخاب این نقاط برای محاسبه در یک روش بهینه است. این انتگرالگیری روندی برای انتخاب مقادیر x_1, x_2, \dots, x_n در بازه $[a, b]$ و انتخاب مقادیر ثابت c_1, c_2, \dots, c_n ارائه می دهد که انتظار می رود خطای حاصل از انجام عمل تقریب

$$(۶۱.۵) \quad \int_a^b f(x) dx \approx \sum_{i=1}^n c_i f(x_i)$$

به ازای هر تابع دلخواه f را به حداقل برساند. برای اندازه گیری این دقت، معمولاً "فرض می شود بهترین انتخاب برای این مقدار انتخابی باشد که درجه دقت فرمول را ماکزیمم کند.

چون مقادیر c_1, c_2, \dots, c_n کاملاً "دلخواه و مقادیر x_1, x_2, \dots, x_n تنها مقید به قابل تعریف بودن تابعی که انتگرال آن تقریب می شود در این نقاطند، حداکثر $2n$ پارامتر، n پارامتر به وسیله مقادیر ثابت c_1, c_2, \dots, c_n و n پارامتر به وسیله x_1, x_2, \dots, x_n وجود دارند.

اگر ضرایب چند جمله ای را نیز پارامتر بگیریم، رده چند جمله ایهای از درجه n یا بیشتر از $(2n - 1)$ شامل $2n$ پارامتر می باشد و بزرگترین رده از چند جمله ایهاست که برای آن انتظار دقیق بودن معادله (۶۱.۵) معقول است. در واقع، برای انتخاب مناسبی از این مقادیر و ثابتها، درستی انتخاب روی این جهت می تواند حاصل شود که نتیجه دهد معادله (۶۱.۵) دارای درجه دقت $(2n - 1)$ است.

پیش از ادامه بحث انتگرالگیری گاوسی، لازم است مطالبی چند از مجموعه توابع متعامد، که در بخش ۴.۴ بررسی شد، یادآور شویم که $\{\phi_0, \phi_1, \dots, \phi_n\}$ مجموعه ای از توابع متعامد بر $[a, b]$ نسبت به تابع وزن $w(x)$ است، مشروط بر اینکه

$$\int_a^b \phi_k(x) \phi_j(x) w(x) dx$$

اگر $j \neq k$ صفر و اگر $j = k$ مثبت است.

از این در بخشهای ۳.۴ و ۴.۴ استفاده شد، و در تمرین ۱۱، بخش ۳.۴، دیدیم که اگر $\{\phi_0, \phi_1, \dots, \phi_n\}$ یک مجموعه متعامد از چند جمله ایهای تعریف شده بر $[a, b]$ بوده و ϕ_i ، به ازای هر $i = 0, 1, \dots, n$ ، از درجه i باشد، آنگاه، به ازای هر چند

جمله‌ای در Π_n مانند Q ، مقادیر ثابت منحصر بفرد $\alpha_0, \alpha_1, \dots, \alpha_n$ وجود دارند بطوری که $Q(x) = \sum_{i=0}^n \alpha_i \phi_i(x)$. این نتیجه در اثبات قضیه بعدی بکار می‌رود ، نتیجه‌ای که در تعیین انتخاب بهینه مقادیر x_1, x_2, \dots, x_n اهمیت بسیار دارد .

قضیه ۸.۵ . اگر $\{\phi_0, \phi_1, \dots, \phi_n\}$ مجموعه‌ای از چند جمله‌ای‌های متعامد ، تعریف شده بر $[a, b]$ ، نسبت به تابع وزن w ، که بر (a, b) پیوسته است ، بوده و ϕ_k به‌ازای هر $k = 0, 1, \dots, n$ از درجه k باشد ، آنگاه ϕ_k دارای k ریشه متمایز در بازه (a, b) است .

برهان . چون ϕ_0 یک چندجمله‌ای از درجه صفر است ، ثابت C با $\phi_0(x) = C$ وجود دارد . از اینجا نتیجه می‌شود که به‌ازای $k \geq 1$ ،

$$0 = \int_a^b \phi_k(x) \phi_0(x) w(x) dx$$

$$= C \int_a^b \phi_k(x) w(x) dx.$$

و چون $w(x) \geq 0$ ولی $w(x) \neq 0$ ، ϕ_k باید دست کم یک‌بار در (a, b) تغییر علامت دهد . فرض کنیم ϕ_k دقیقاً z بار در (a, b) در نقاط $\{r_i\}_{i=1}^z$ ، که $a < r_1 < r_2 < \dots < r_j < b$ و $z < k$ ، تغییر علامت دهد . بی‌آنکه کلیت از کف برود ، می‌توان فرض کرد که $\phi_k(x) > 0$ بر (a, r_1) (ر. ک. تمرین ۷) . در نتیجه ، $\phi_k(x) < 0$ بر (r_1, r_2) ، $\phi_k(x) > 0$ بر (r_2, r_3) و ، در حالت کلی ، ϕ_k بر هر یک از بازه‌های مجاور $(a, r_1), (r_1, r_2), \dots, (r_j, b)$ مختلف‌العلامه است .

چند جمله‌ای P از درجه z را با

$$P(x) = \prod_{i=1}^z (x - r_i)$$

تعریف می‌کنیم . توجه کنید که P بر هر یک از زیر بازه‌های $(a, r_1), (r_1, r_2), \dots, (r_j, b)$ هم‌علامت ϕ_k است ؛ و در نتیجه ، بر هر یک از این بازه‌ها ، $P(x) \phi_k(x) > 0$. چون بر (a, b) ، $w(x) \geq 0$ ولی $w(x) \neq 0$ ، نتیجه می‌شود که

$$\int_a^b P(x) \phi_k(x) w(x) dx > 0 \quad (۶۲.۵)$$

لیکن، P یک چندجمله‌ای از درجه $k < z$ است؛ در نتیجه، به‌ازای گردآیمای از ثابتهای $\alpha_0, \alpha_1, \dots, \alpha_j$ ،

$$P(x) = \sum_{i=0}^j \alpha_i \phi_i(x)$$

که نتیجه می‌دهد

$$\int_a^b P(x) \phi_k(x) w(x) dx = \sum_{i=0}^j \alpha_i \int_a^b \phi_i(x) \phi_k(x) w(x) dx = 0$$

که با معادله (۶۲.۵) متناقض است.

تنها فرض ما در این روند این بود که ϕ_k در (a, b) دقیقاً z بار تغییر علامت می‌دهد، که در آن $k < z$ ؛ در نتیجه، این عبارت باید نادرست باشد. از اینرو، ϕ_k در (a, b) دست کم k بار تغییر علامت می‌دهد. چون قضیه مقدار میانی (قضیه ۱۲.۱) ایجاب می‌کند که در هر تغییر علامت ریشه‌ای وجود دارد، ϕ_k باید در (a, b) ، k صفر متمایز داشته باشد.

برای بکار بردن قضیه ۸.۵، فرض کنیم f تابعی باشد که برای آن تقریبی به $\int_{-1}^1 f(x) dx$ لازم است. مجموعه چند جمله‌ایهای لژاندر $\{P_0, P_1, \dots, P_n\}$ ، که در مثال ۱ از بخش ۴.۴ تعریف شده و بر $[-1, 1]$ نسبت به $w(x) \equiv 1$ متعامدند، را در نظر می‌گیریم. x_1, x_2, \dots, x_n را n ریشه متمایز P_n ، که طبق قضیه ۸.۵ در $(-1, 1)$ قرار دارند، اختیار می‌کنیم.

فرض کنیم P یک چندجمله‌ای دلخواه از درجه k باشد، که $k \leq 2n - 1$. با تقسیم P بر P_n بدست می‌آوریم

$$P(x) = Q(x)P_n(x) + R(x)$$

که در آن Q و R هر دو چندجمله‌ایهایی از درجه کمتر از n می‌باشند. اگر یک فرمول انتگرالگیری با درجه دقت حداقل $(n - 1)$ به شکل زیر باشد:

$$\int_{-1}^1 f(x) dx = \sum_{i=1}^n c_i f(x_i)$$

که در آن به‌ازای هر $i = 1, 2, \dots, n$

$$c_i = \int_{-1}^1 \prod_{j=1}^n \frac{(x - x_j)}{(x_i - x_j)} dx$$

و مقادیر x_1, \dots, x_n به شرح فوق باشند، آنگاه، چون Q و R از درجه کمتر از n هستند

$$\begin{aligned} \int_{-1}^1 P(x) dx &= \int_{-1}^1 Q(x)P_n(x) dx + \int_{-1}^1 R(x) dx \\ &= 0 + \int_{-1}^1 R(x) dx \\ &= \sum_{i=1}^n c_i R(x_i). \end{aligned}$$

اما، چون x_1, x_2, \dots, x_n همگی صفرهای P_n اند،

$$P(x_i) = Q(x_i)P_n(x_i) + R(x_i) = R(x_i)$$

در نتیجه،

$$\int_{-1}^1 P(x) dx = \sum_{i=1}^n c_i P(x_i)$$

و این فرمول برای تمام چندجمله‌ایهای موجود در Π_{2n-1} دقیق است.

ثابت‌های c_1, c_2, \dots, c_n را می‌توان با حل n معادله خطی n مجهولی زیر بدست

آورد:

$$\beta_0 = \int_{-1}^1 P_0^2(x) dx = \sum_{i=1}^n c_i P_0^2(x_i),$$

$$\beta_1 = \int_{-1}^1 P_1^2(x) dx = \sum_{i=1}^n c_i P_1^2(x_i),$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

$$\beta_{n-1} = \int_{-1}^1 P_{n-1}^2(x) dx = \sum_{i=1}^n c_i P_{n-1}^2(x_i)$$

اما، در واقع، این ثابتها و صفرهای P_n در هم‌جا با جدول داده می‌شوند؛ مثلاً،

ر.ک. استرادل^۱ و سکرست^۲ [۶۳].

چون تبدیل خطی ساده^۳ $t = [1/(b-a)](2x - a - b)$ ، به شرط $a \neq b$ ، هر

بازه $[a, b]$ را به $[-1, 1]$ تبدیل می‌کند، چند جمله‌ایهای لژاندر را می‌توان برای تقریب

$$\int_a^b f(x) dx = \int_{-1}^1 f\left(\frac{(b-a)t + b + a}{2}\right) \frac{(b-a)}{2} dt$$

به‌ازای هر تابع قابل محاسبه در نقاط مورد نیاز بکار برد.

مثال ۱. تعیین تقریبهایی به $\int_1^{1.5} e^{-x^2} dx$ مورد نظر است. فرمولهای نیوتن - کاتس مذکور در بخش ۴.۵ را می‌توان بکار برد؛ تقریبهای حاصل عبارتند از:
فرمولهای بسته نیوتن - کاتس:

$$\underline{n = 1, h = \frac{1}{2}}:$$

$$\int_1^{1.5} e^{-x^2} dx \approx \frac{1}{4}[e^{-1} + e^{-(3/2)^2}] = .1183197;$$

$$\underline{n = 2, h = \frac{1}{4}}:$$

$$\int_1^{1.5} e^{-x^2} dx \approx \frac{1}{12}[e^{-1} + 4e^{-(5/4)^2} + e^{-(3/2)^2}] = .1093104;$$

$$\underline{n = 3, h = \frac{1}{6}}:$$

$$\int_1^{1.5} e^{-x^2} dx \approx \frac{1}{16}[e^{-1} + 3e^{-(7/6)^2} + 3e^{-(4/3)^2} + e^{-(3/2)^2}]$$

$$= .1093404$$

$$\underline{n = 4, h = \frac{1}{8}}:$$

$$\int_1^{1.5} e^{-x^2} dx \approx \frac{1}{160}[7e^{-1} + 32e^{-(9/8)^2} + 32e^{-(5/4)^2} + 32e^{-(11/8)^2}$$

$$+ 7e^{-(3/2)^2}] = .1093643$$

فرمولهای باز نیوتن - کاتس:

$$\underline{n = 0, h = \frac{1}{4}}:$$

$$\int_1^{1.5} e^{-x^2} dx \approx \frac{1}{2}e^{-(5/4)^2} = .1048057;$$

$n = 1, h = \frac{1}{6}$:

$$\int_1^{1.5} e^{-x^2} dx \approx \frac{1}{4}[e^{-(7/6)^2} + e^{-(4/3)^2}] = .1063473;$$

$n = 2, h = \frac{1}{8}$:

$$\int_1^{1.5} e^{-x^2} dx \approx \frac{1}{8}[2e^{-(9/8)^2} - e^{-(5/4)^2} + 2e^{-(11/8)^2}] = .1094116.$$

برای بکاربردن روند انتگرالگیری گاوسی در این مسئله، ابتدا لازم است این انتگرال به مسئله‌ای که بازه انتگرالگیری در آن $[-1, 1]$ است تغییر کند. با استفاده از معادله (۶۴.۵)، بدست می‌آید

$$\int_1^{1.5} e^{-x^2} dx = \frac{1}{4} \int_{-1}^1 e^{-(t+5)^2/16} dt$$

در جدول ۸.۵ صفرهای چند جمله‌ایهای لژاندر برای مقادیر مختلف n همراه با ضرایبی که به مقادیری از تابع وابسته می‌شوند تا در این صفرها محاسبه شوند فهرست شده است.

n	صفرها	ضرایب
2	.5773502692	1.0000000000
	-.5773502692	1.0000000000
3	.7745966692	.5555555556
	.0000000000	.8888888889
	-.7745966692	.5555555556
4	.8611363116	.3478548451
	.3399810436	.6521451549
	-.3399810436	.6521451549
	-.8611363116	.3478548451
5	.9061798459	.2369268850
	.5384693101	.4786286705
	.0000000000	.5688888889
	-.5384693101	.4786286705
	-.9061798459	.2369268850

با استفاده از این مقادیر، تقریبهای انتگرالگیری گاوسی برای این مسئله عبارتند از:

$n = 2:$

$$\int_1^{1.5} e^{-x^2} dx \approx \frac{1}{4} [e^{-(5+.5773502692)^2/16} + e^{-(5-.5773502692)^2/16}] = .1094003,$$

$n = 3:$

$$\int_1^{1.5} e^{-x^2} dx \approx \frac{1}{4} [(.5555555556)e^{-(5+.7745966692)^2/16} + (.8888888889)e^{-(5)^2/16} + (.5555555556)e^{-(5-.7745966692)^2/16}] = .1093642.$$

مقادیر حاصل از بکاربردن روند رامبرگ با $n = 4$ ، برای مقایسه، در جدول ۹.۵ فهرست شده‌اند.

.1183197			
.1115627	.1093104		
.1099114	.1093610	.1093643	
.1095009	.1093641	.1093643	.1093643.

جدول ۹.۵

جواب درست، تا 7 رقم اعشار، برابر است با **.1093643**.

مجموعه‌های دیگری از چند جمله‌ایهای متعامد را می‌توان در روش انتگرالگیری گاوسی بکار برد؛ مثالهایی از آنها در تمرینات داده شده‌اند.

مجموعه تمرینات ۷.۵

۱. با استفاده از روند انتگرالگیری گاوسی با $n = 2, 3, 4$ ، تقریبهایی برای $\int_1^3 e^x \sin x dx$ بدست آورید. جوابهای خود را با نتایج تمرین ۱ از بخش ۶.۵ و تمرین ۱ از بخش

۵.۵ و نیز با مقدار دقیق مقایسه کنید .

- ۲ . با استفاده از روند انتگرالگیری گاوسی با $n = 2, 3, 4$ ، تقریبهایی برای $\int_0^1 (1+x)^{-1} dx$ بدست آورید . جوابهای خود را با نتایج تمرین ۲ از بخش ۶.۵ و تمرین ۴ از بخش ۵.۵ و تمرین ۶ از بخش ۴.۵ و نیز با مقدار دقیق مقایسه کنید .
- ۳ . درستی درایه‌های جدول ۸.۵ برای مقادیر $n = 2, 3$ را ، با یافتن ریشه‌های چند جمله‌ایهای لژاندر ، تحقیق کنید . ضرایب وابسته به این مقادیر را با استفاده از معادلات (۶۳.۵) بدست آورید .

- ۴ . مجموعه $\{L_0, L_1, L_2, L_3\}$ از چندجمله‌ایهای لاگور ، که بر بازه $(0, \infty)$ نسبت به تابع وزن e^{-x} متعامدند ، در تمرین ۶ ، بخش ۴.۴ ، معرفی شد . این چندجمله‌ایها را می‌توان برای بدست آوردن تقریبهایی از $\int_0^{\infty} e^{-x} f(x) dx$ ، به شرط موجود بودن این انتگرال مجازی ، بکار برد . نتیجه‌گیری نظیر آنچه در مورد چندجمله‌ایهای لژاندر بعد از اثبات قضیه ۸.۵ شد می‌باشد . نشان دهید که این مجموعه از چندجمله‌ایها فرمولی با درجه ۶ دقت سه برای تقریب

$$\int_0^{\infty} e^{-x} f(x) dx$$

بدست می‌دهد .

- ۵ . ریشه‌های چندجمله‌ایهای لاگور L_1 ، L_2 ، و L_3 مورد بحث در تمرین ۴ را بدست آورده ، و با استفاده از ضرایب نظیر حاصل از معادله (۶۳.۵) ، تقریبهایی برای

$$\int_0^{\infty} e^{-x} \sin x dx$$

وقتی 3 و $n = 2$ ، بدست آورید .

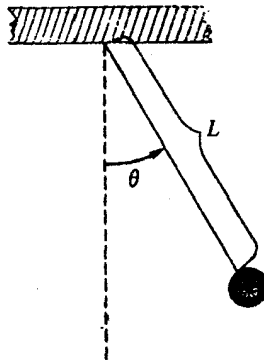
- ۶ . $\int_0^{\infty} \sqrt{x} e^{-x} dx$ را با استفاده از چندجمله‌ایهای لاگور ، مانند تمرین ۵ ، تقریب کنید ، و این تقریب را با آنچه در تمرین ۱۳ (پ) ، بخش ۵.۵ ، بدست آمد مقایسه کنید .
- ۷ . در برهان قضیه ۸.۵ این عبارت آمده بود که ، با فرض مثبت بودن $\phi_1(x)$ در (a, r_1) کلیتی از دست نمی‌رود . نشان دهید که این حکم درست است .

مسائل مقدار اولیه در معادلات دیفرانسیل معمولی^۶

حرکت یک آونگ در حال نوسان تحت مفروضات ساده^۷ معینی را می‌توان با معادله^۸ دیفرانسیل مرتبه^۹ دوم

$$\frac{d^2\theta}{dt^2} + \frac{g}{L} \sin \theta = 0$$

توصیف کرد، که در آن L طول آونگ، g ثابت ثقل زمین، و θ زاویه^{۱۰} آونگ با وضعیت قائم یا تعادل است. اگر، علاوه بر این، وضعیت آونگ در شروع حرکت را با $\theta(t_0) = \theta_0$ و سرعت در آن نقطه را با $\theta'(t_0) = \theta'_0$ مشخص کنیم، با مسئله^{۱۱} مقدار اولیه مواجه هستیم.



بهازای مقادیر کوچک θ ، تقریب $\sin \theta \approx \theta$ را می‌توان برای ساده کردن

این مسئله به معادله خطی

$$\frac{d^2\theta}{dt^2} + \frac{g}{L}\theta = 0$$

بکار برد، که با یک روش متعارف در معادلات دیفرانسیل حل می‌شود. لیکن، به ازای مقادیر بزرگتر θ باید از تقریبهایی از نوع بدست آمده در این فصل استفاده کرد.

گرچه هر کتاب درسی در معادلات دیفرانسیل معمولی روشهایی برای تعیین جوابهای مسائل مقدار اولیه^۶ مرتبه^۷ اول را شرح می‌دهد، در عمل معدودی از مسائل ناشی از پدیده‌های فیزیکی دقیقاً قابل حلند.

در قسمت اول این فصل به تقریب جواب $y(t)$ از مسئله‌ای به شکل زیر علاقه‌مندیم:

$$(۱۰۶) \quad \frac{dy}{dt} = f(t, y), \quad a \leq t \leq b$$

با شرط اولیه^۸

$$(۲۰۶) \quad y(a) = \alpha$$

در بخش آخر این فصل به تعمیم این روشها به دستگاهی از معادلات دیفرانسیل مرتبه^۹ اول به شکل

$$(۳۰۶) \quad \begin{aligned} \frac{dy_1}{dt} &= f_1(t, y_1, y_2, \dots, y_n), \\ \frac{dy_2}{dt} &= f_2(t, y_1, y_2, \dots, y_n), \\ &\vdots \\ \frac{dy_n}{dt} &= f_n(t, y_1, y_2, \dots, y_n), \end{aligned}$$

به‌ازای هر $a \leq t \leq b$ ، با شرایط اولیه^{۱۰}

$$(۴۰۶) \quad y_1(a) = \alpha_1, \quad y_2(a) = \alpha_2, \quad \dots, \quad y_n(a) = \alpha_n,$$

و رابطه^{۱۱} یک دستگاه از این نوع با مسئله^{۱۲} مقدار اولیه^{۱۳} مرتبه^{۱۴} n م‌کلی به شکل

$$(۵.۶) \quad y^{(n)} = f(t, y, y', y'', \dots, y^{(n-1)})$$

به‌ازای $a \leq t \leq b$ ، با شرایط اولیه

$$(۶.۶) \quad y(a) = \alpha_0, \quad y'(a) = \alpha_1, \quad \dots, \quad y^{(n-1)}(a) = \alpha_{n-1}$$

می‌پردازیم.

۱.۶ نظریهٔ مقدماتی مسائل مقدار اولیه

پیش از مبادرت به حل یک مسئلهٔ مقدار اولیه، می‌خواهیم بدانیم آیا جوابی وجود دارد و اگر چنین است، جواب منحصر بفرد است. بعلاوه، چون مسائل موجود در بررسی پدیده‌های فیزیکی معمولاً "حالت واقعی را تقریب می‌کنند، جالب است بدانیم آیا تغییرات کوچکی در صورت مسئله موجب تغییرات کوچکی در جواب می‌شوند. این مطلب نیز مهم است که به دلیل استفاده از روشهای عددی احتمال خطاهای گرد کردن وجود دارد. برای بحث در این مسائل به چند تعریف و نتایجی از نظریهٔ معادلات دیفرانسیل معمولی نیاز داریم. اولین تعریف تعمیم تعریف داده شده در تمرین ۱.۱ از بخش ۱.۱ به تابع دو متغیره است.

تعریف ۱.۶. گوئیم تابع $f(t, y)$ با متغیر y بر مجموعه $R^2 \subset D$ در شرط لیپ‌شیتس صدق می‌کند در صورتی که یک ثابت مانند $L > 0$ با این خاصیت وجود داشته باشد که هر وقت $(t, y_1), (t, y_2) \in D$

$$|f(t, y_1) - f(t, y_2)| \leq L|y_1 - y_2|$$

ثابت L را یک ثابت لیپ‌شیتس برای f گوئیم.

مثال ۱. هرگاه $\{3 \leq v \leq 4, 1 \leq t \leq 2\} = D$ و $f(t, y) = t|y|$ ، آنگاه، به‌ازای هر $(t, y_1), (t, y_2) \in D$

$$\begin{aligned} |f(t, y_1) - f(t, y_2)| &= |t|y_1| - t|y_2|| \\ &= |t| ||y_1| - |y_2|| \\ &\leq 2|y_1 - y_2|. \end{aligned}$$

لذا، f با متغیر y بر D در شرط لیپ‌شیتس با ثابت لیپ‌شیتس ۲ صدق می‌کند. در واقع، در این مسئله، کوچکترین مقدار ممکن برای ثابت لیپ‌شیتس عبارت است از $L = 2$.

تعریف ۲۰۶. گوییم مجموعه $D \subset R^2$ محدب است اگر، هر وقت (t_1, y_1) و (t_2, y_2) متعلق به D باشند، نقطه

$$((1 - \lambda)t_1 + \lambda t_2, (1 - \lambda)y_1 + \lambda y_2)$$

نیز بازای هر λ ، که $0 \leq \lambda \leq 1$ ، متعلق به D باشد.

به عبارت هندسی، تعریف ۲۰۶ بیان می‌کند که یک مجموعه محدب است در صورتی که هر وقت دو نقطه متعلق به این مجموعه باشد، تمام پاره خط مستقیم بین آن دو نقطه نیز متعلق به این مجموعه باشد.

قضیه ۳۰۶. فرض کنیم $f(t, y)$ بر یک مجموعه محدب $D \subset R^2$ تعریف شده باشد. اگر ثابتی چون $L > 0$ وجود داشته باشد که به ازای هر $(t, y) \in D$ ،

$$(۷.۶) \quad \left| \frac{\partial f}{\partial y}(t, y) \right| \leq L$$

نگاه f نسبت به متغیر y بر D در شرط لیپ‌شیتس با ثابت لیپ‌شیتس L صدق می‌کند.

اثبات قضیه ۳۰۶ در تمرین ۴ مورد بحث قرار می‌گیرد، و مانند اثبات نتیجه

متناظر برای توابع یک متغیره است که در تمرین ۱۱، بخش ۱۰۱، بررسی شد.

همانطور که قضیه زیر نشان می‌دهد، اغلب سودمند است تعیین شود که آیا تابعی با یک مسئله مقدار اولیه نسبت به متغیر دوم آن در شرط لیپ‌شیتس صدق می‌کند، و آیا بکار بردن شرط (۷.۶) عموماً از تعریف خیلی ساده‌تر است یا نه. بهر حال، باید گفت که قضیه ۳۰۶ فقط شرایط کافی برای برقراری شرط لیپ‌شیتس را ارائه می‌دهد؛ بررسی مجدد مثال ۱ نشان می‌دهد که این شرایط حتماً الزامی نیستند.

قضیه زیر شرحی است از قضیه اساسی وجود و یکتایی برای معادلات دیفرانسیل معمولی مرتبه اول. گرچه این قضیه، با تقلیل مفروضات به نحوی، قابل اثبات است، ولی برای اهداف ما این شکل قضیه کفایت می‌کند. [اثبات این قضیه، تقریباً به این شکل، را می‌توان در بیرکف^۱ و روتا^۲ [۱۱]، صفحات ۱۰۳ و ۱۱۵-۱۱۲ یافت.]

قضیه ۴.۶. فرض کنیم $D = \{(t, y) | a \leq t \leq b, -\infty < y < \infty\}$ و $f(t, y)$ بر D پیوسته باشد. هرگاه f نسبت به متغیر y بر D در شرط لیب شیتس صدق کند، آنگاه مسئله مقدار اولیه

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

دارای جواب منحصر بفرد $y(t)$ ، به ازای $a \leq t \leq b$ ، است.

مثال ۲. مسئله مقدار اولیه

$$y' = 1 + t \sin(ty), \quad 0 \leq t \leq 2, \quad y(0) = 0$$

را در نظر می‌گیریم. با ثابت گرفتن t و بکار بردن قضیه مقدار میانگین (قضیه ۸.۱، ص ۵) در تابع

$$f(t, y) = 1 + t \sin(ty)$$

در می‌یابیم که هر وقت $y_1 < y_2$ ، نقطه‌ای مانند ξ ، که $y_1 < \xi < y_2$ ، وجود دارد که

$$t^2 \cos(\xi t) = \frac{\partial}{\partial y} f(t, \xi) = \frac{f(t, y_2) - f(t, y_1)}{y_2 - y_1}$$

لذا، به ازای تمام $y_1 < y_2$ ،

$$|f(t, y_2) - f(t, y_1)| = |y_2 - y_1| |t^2 \cos(\xi t)| \leq 4|y_2 - y_1|,$$

و f نسبت به متغیر y در شرط لیب شیتس با ثابت لیب شیتس چهار صدق می‌کند.

چون، علاوه، $f(t, y)$ وقتی $0 \leq t \leq 2$ و $-\infty < y < \infty$ پیوسته است، قضیه ۴.۶ ایجاب می‌کند که یک جواب منحصر بفرد برای این مسئله مقدار اولیه وجود دارد. ممکن است تعیین جواب دقیق این مسئله برای خواننده‌ای که یک دوره درس معادلات دیفرانسیل را گذرانده است جالب باشد.

حال که تاحدی از این سوال که چه زمانی مسائل مقدار اولیه جواب منحصر بفرد دارند فارغ شده‌ایم، به سوال دیگری که قبلاً "در این بخش مطرح شده بود می‌پردازیم. آيا روشی برای تعیین اینکه مسئله، بخصوصی این خاصیت را دارا باشد که تغییرات کوچکی در صورت مسئله موجب تغییرات کوچکی در جواب شود وجود دارد؟ طبق معمول، ابتدا تعریف قابل استفاده‌ای برای بیان این مفهوم مورد نیاز است.

تعریف ۵.۶. گوئیم مسئله مقدار اولیه

$$(۸.۰۶) \quad \frac{dy}{dt} = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha,$$

یک مسئله خوش وضع است اگر:

(یک) یک جواب منحصر بفرد، مثلاً " $y(t)$ "، برای این مسئله وجود داشته باشد؛
(دو) عددی مانند $\varepsilon > 0$ با این خاصیت باشد که جواب منحصر بفرد $z(t)$ برای مسئله

$$(۹.۰۶) \quad \frac{dz}{dt} = f(t, z) + \delta(t), \quad a \leq t \leq b, \quad z(a) = \alpha + \varepsilon_0$$

هرگاه $\varepsilon < |\varepsilon_0|$ و به ازای هر $a \leq t \leq b$ ، $|\delta(t)| < \varepsilon$ ، وجود داشته باشد؛

(سه) ثابتی مانند $k > 0$ با این خاصیت باشد که

$$|z(t) - y(t)| < k\varepsilon, \quad a \leq t \leq b$$

مسئله‌ای که با معادله (۹.۰۶) مشخص شد را اغلب یک مسئله منحرف شده وابسته

به مسئله اصلی (۸.۰۶) می‌نامند. قضیه زیر شرایطی را مشخص می‌کند که خوش وضع بودن یک مسئله مقدار اولیه را بدست می‌دهد. [برهان این قضیه را در مرجعی که قبلاً ذکر شد می‌توان یافت، بیرکف و روتا [۱۱]، صفحات ۱۰۷ - ۱۰۳.]

قضیه ۶.۶. فرض کنیم $D = \{(t, y) | a \leq t \leq b, c \leq y \leq d\}$. مسئله مقدار اولیه

$$(۱۰.۰۶) \quad \frac{dy}{dt} = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

خوش وضع است اگر f پیوسته و نسبت به متغیر y بر مجموعه D در شرط لیب شیتس صدق کند.

مثال ۳. فرض کنیم $D = \{(t, y) | 0 \leq t \leq 1, -\infty < y < \infty\}$ و مسئله مقدار اولیه

$$(۱۱.۰۶) \quad \frac{dy}{dt} = -y + t + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

را در نظر می‌گیریم. چون

$$\left| \frac{\partial(-y + t + 1)}{\partial y} \right| = 1$$

قضیه ۳.۶ نتیجه می دهد که $f(t, y) = -y + t + 1$ بر D در شرط لیب شیتس با ثابت لیب شیتس 1 صدق می کند. چون، بعلاوه، f بر D پیوسته است، طبق قضیه ۶.۶، معادله (۱۱.۶) یک مسئله خوش وضع است. مسئله منحرف شده

$$(12.6) \quad \frac{dz}{dt} = -z + t + 1 + \delta, \quad 0 \leq t \leq 1, \quad z(0) = 1 + \varepsilon_0$$

را در نظر می گیریم، که در آن δ و ε_0 ثابت اند. می توان نشان داد که جوابهای معادلات (۱۱.۶) و (۱۲.۶)، بترتیب، عبارتند از

$$y(t) = e^{-t} + t$$

$$z(t) = (1 + \varepsilon_0 - \delta)e^{-t} + t + \delta$$

به آسانی معلوم می شود که اگر $|\delta| < \varepsilon$ و $|\varepsilon_0| < \varepsilon$ ، آنگاه، برای هر t ،

$$\begin{aligned} |y(t) - z(t)| &= |(\delta - \varepsilon_0)e^{-t} - \delta| \\ &\leq |\varepsilon_0| + |\delta||1 - e^{-t}| \\ &\leq 2\varepsilon, \end{aligned}$$

که نتیجه حاصل از بکار بردن قضیه ۶.۶ را تأیید می کند.

مجموعه تمرینات ۱.۶

۱. برای هر یک از توابع داده شده $f(t, y)$ در (T) تا $(ت)$:
(یک) آیا f بر

$$D = \{(t, y) | 0 \leq t \leq 1, -\infty < y < \infty\}$$

در شرط لیب شیتس صدق می کند؟

(دو) تعیین کنید آیا مسئله مقدار اولیه

$$y' = f(t, y), \quad 0 \leq t \leq 1,$$

$$y(0) = 1,$$

خوش وضع است:

$$(T) \quad f(t, y) = t^2y + 1 \quad ; \quad (ب) \quad f(t, y) = ty \quad ; \quad (پ) \quad f(t, y) = \frac{1}{y^2 + 1}$$

$$\cdot f(t, y) = \sqrt{y+1} \quad (ت)$$

۲. با استفاده از قضیه ۴.۶، نشان دهید که

$$y' = y \cos t, \quad 0 \leq t \leq 1,$$

$$y(0) = 1$$

دارای جواب منحصر بفرد است. این جواب را بدست آورید.

۳. با استفاده از قضیه ۴.۶، نشان دهید که

$$y' - \frac{2}{t}y = t^2 e^t, \quad 1 \leq t \leq 2,$$

$$y(1) = 0$$

دارای جواب منحصر بفرد است. این جواب را بدست آورید.

۴. با استفاده از قضیه مقدار میانگین (قضیه ۸.۱، ص ۵) برای $f(t, y)$ و ثابت گرفتن

قضیه ۳.۶ را ثابت کنید.

۲.۶ روش اویلر^۱

گرچه روش اویلر در عمل بندرت بکار می‌رود، سادگی بدست آوردن آن را می‌توان برای تشریح تکنیکهای ساختن بعضی از روشهای پیشرفته‌تر، بدون وارد شدن در جبر ملال‌آور این ساختنها، بکار گرفت.

هدف این روش تعیین تقریبی برای مسئله مقدار اولیه خوش وضع

$$(13.6) \quad \frac{dy}{dt} = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

است. در عمل، یک تقریب پیوسته به جواب $y(t)$ بدست نمی‌آید؛ در عوض، تقریبهایی به y در نقاط متعدد بازه $[a, b]$ ، به نام نقاط شبکه‌ای، پدید می‌آیند. وقتی جواب تقریبی در این نقاط بدست آمد، جواب تقریبی در سایر نقاط بازه با استفاده از یکی از روندهای درونیایی، بررسی شده در فصل ۳، حاصل می‌شود.

در این بخش مفیدیم که توزیع نقاط شبکه‌ای در طول بازه $[a, b]$ به‌طور یکسان باشد.

این شرط با انتخاب عدد صحیح مثبت N و نقاط $\{t_0, t_1, t_2, \dots, t_N\}$ که در آن

بمازای هر $t_i = a + ih$ ، $i = 0, 1, 2, \dots, N$

حاصل می‌شود. $h = (b - a)/N$ ، یعنی فاصله مشترک بین این نقاط، اندازه گام نامیده می‌شود.

برای رسیدن به روش اویلر، از قضیه تیلور (قضیه ۱۳.۱، ص ۷) استفاده می‌کنیم. راهی دیگر برای رسیدن به آن در تمرین ۶ ارائه می‌شود.

فرض کنیم $y(t)$ ، یعنی جواب منحصر بفرد معادله (۱۳.۶)، دو مشتق پیوسته بر $[a, b]$ داشته باشد، بطوری که بمازای هر $i = 0, 1, 2, \dots, N - 1$ می‌توان $y(t_{i+1})$ را به صورت

$$(14.6) \quad y(t_{i+1}) = y(t_i + h) = y(t_i) + hy'(t_i) + \frac{h^2}{2} y''(\xi_i)$$

بمازای نقطه‌ای مانند ξ_i ، که $t_i < \xi_i \leq t_{i+1}$ ، نوشت.

با استفاده از نماد $h = t_{i+1} - t_i$ ، نتیجه می‌شود که عدد θ_i ، که $0 < \theta_i < 1$ ، وجود دارد بطوری که

$$(15.6) \quad y(t_{i+1}) = y(t_i) + hy'(t_i) + \frac{h^2}{2} y''(t_i + \theta_i h)$$

و، چون $y(t)$ در معادله دیفرانسیل (۱۳.۶) صدق می‌کند،

$$y(t_{i+1}) = y(t_i) + hf(t_i, y(t_i)) + \frac{h^2}{2} y''(t_i + \theta_i h)$$

با دوباره نویسی این معادله، بمازای هر $i = 0, 1, \dots, N - 1$ ، نتیجه می‌شود که

$$(16.6) \quad \frac{y(t_{i+1}) - y(t_i)}{h} = f(t_i, y(t_i)) + \frac{h}{2} y''(t_i + \theta_i h)$$

وقتی h به قدر کافی کوچک باشد، بنابر پیوستگی y'' ، جمله $(h/2)y''(t_i + \theta_i h)$ نیز کوچک بوده و

$$(17.6) \quad \frac{y(t_{i+1}) - y(t_i)}{h} \approx f(t_i, y(t_i))$$

در روش اویلر از این فرض با تعریف

$$w_0 = \alpha,$$

$$(18.6)$$

$$w_{i+1} = w_i + hf(t_i, w_i)$$

و فرض $y(t_i) \approx w_i$ ، به ازای هر $i = 1, 2, \dots, N$ ، استفاده می شود .
 معادله (۱۸.۶) در ارتباط با روش اویلر معادله تفاضلی نامیده می شود . همانطور
 که بعداً در این فصل خواهیم دید ، نظریه حل معادلات تفاضلی ، به جهات زیادی ،
 شبیه نظریه حل معادلات دیفرانسیل است .
 شکل الگوریتمی روش اویلر در زیر ارائه می شود .

الگوریتم اویلر ۱۰۶

برای بدست آوردن تقریبی به جواب مسئله مقدار اولیه خوش وضع

$$y' = f(t, y); \quad a \leq t \leq b, \quad y(a) = \alpha$$

در $(N + 1)$ نقطه هم فاصله در بازه $[a, b]$ ، مراحل زیر را انجام می دهیم :
 مرحله ۱ . قرار می دهیم

$$h = \frac{b - a}{N}$$

$$t_0 = a,$$

$$w_0 = \alpha.$$

مرحله ۲ . قرار می دهیم $i = 1$.

مرحله ۳ . قرار می دهیم

$$t_i = t_0 + ih,$$

$$w_i = w_{i-1} + hf(t_{i-1}, w_{i-1})$$

مرحله ۴ . اگر $i < N$ ، ۱ را به i افزوده و به مرحله ۳ می رویم .

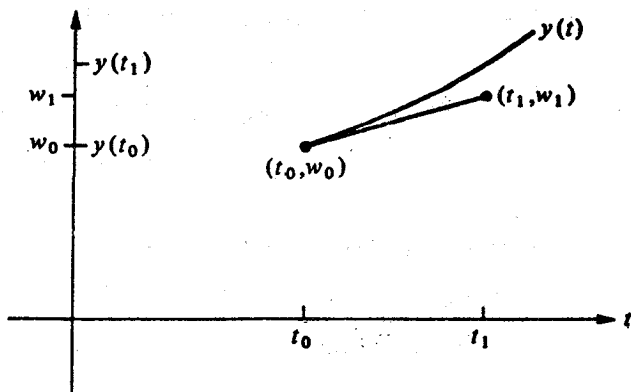
مرحله ۵ . اگر $i = N$ ، روند کامل است و w_i ، به ازای هر $i = 1, 2, \dots, N$ ، $y(t_i)$ را تقریب می کند .

برای تعبیر هندسی روش اویلر ، ابتدا به ازای هر $i = 0, 1, 2, \dots, N$ نماد $y_i = y(t_i)$ را معرفی و توجه می کنیم که ، وقتی w_i تقریب دقیقی به y_i است ، فرض خوش وضع بودن مسئله نتیجه می دهد که

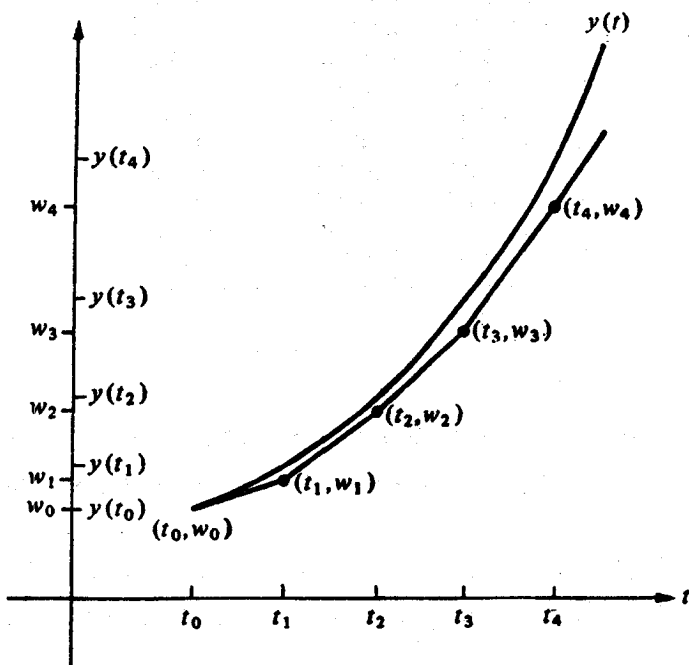
$$f(t_i, w_i) \approx y'(t_i) = f(t_i, y(t_i))$$

در نتیجه ، یک مرحله این روش مانند شکل ۱۰۶ و یک سری از مراحل مانند شکل ۲۰۶

نمایان می‌شوند



شکل ۱.۶



شکل ۲.۶

مثال ۱. برای یافتن تقریبهایی به مسئله مقدار اولیه

$$(۱۹.۶) \quad y' = -y + t + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

فرض کنیم $N = 10$ ؛ در نتیجه، $h = .1$ و $t_i = .1i$

با استفاده از نمادگذاری الگوریتم ۱۰.۶ و این مطلب که $f(t, y) = -y + t + 1$

بمازای $i = 1, 2, \dots, 10$ داریم

$$w_0 = 1$$

$$w_i = w_{i-1} + h(-w_{i-1} + t_{i-1} + 1)$$

$$= w_{i-1} + .1(-w_{i-1} + .1(i-1) + 1)$$

$$= .9w_{i-1} + .01(i-1) + .1,$$

t_i	w_i	y_i	خطا = $ w_i - y_i $
0	1.000000	1.000000	0
.1	1.000000	1.004837	.004837
.2	1.010000	1.018731	.008731
.3	1.029000	1.040818	.011818
.4	1.056100	1.070320	.014220
.5	1.090490	1.106531	.016041
.6	1.131441	1.148812	.017371
.7	1.178297	1.196585	.018288
.8	1.230467	1.249329	.018862
.9	1.287420	1.306570	.019150
1.0	1.348678	1.367879	.019201

جدول ۱۰.۶

ملاحظه می شود که جواب واقعی (۱۹.۶) برابر $y(t) = t + e^{-t}$ است، و جدول ۱۰.۶

مقایسه بین مقادیر تقریبی در t_i و مقادیر واقعی را نشان می دهد.

توجه کنید که خطا با افزایش مقدار t_i بتدریج زیاد می شود. این کنترل افزایش

خطا لازمۀ پایداری روش اویلر است، که نتیجه می دهد که، برای لخطاهای گرد کردن،

انتظار افزایشی نه بدتر از یک روش خطی وجود دارد.

برای بدست آوردن یک کران خطا در حالت کلی روش اویلر، شایسته است دو لم

محاسباتی، که در جریان کار مورد نیازند، ارائه شود.

لم ۷.۶. بهازای هر $x \geq -1$ و هر عدد صحیح مثبت m ،

$$(20.6) \quad 0 \leq (1+x)^m \leq e^{mx}$$

برهان. با استفاده از قضیه تیلور (قضیه ۱۳.۱، ص ۹) با $f(x) = e^x$ ، $x_0 = 0$ ، و $n = 1$ ، نتیجه می‌شود که

$$e^x = 1 + x + \frac{1}{2}x^2e^\xi$$

که در آن ξ بین x و صفر است. بنابراین،

$$0 \leq 1+x \leq 1+x + \frac{1}{2}x^2e^\xi = e^x$$

و، چون $1+x \geq 0$

$$0 \leq (1+x)^m \leq (e^x)^m = e^{mx}$$

لم ۸.۶. اگر m و n اعداد حقیقی مثبتی بوده و $\{a_i\}_{i=0}^k$ دنباله‌ای با $a_0 \geq 0$ باشد که در

$$(21.6) \quad a_{i+1} \leq (1+m)a_i + n \quad i = 0, 1, 2, \dots, k$$

صدق کند، آنگاه

$$(22.6) \quad a_{i+1} \leq e^{(i+1)m} \left(\frac{n}{m} + a_0 \right) - \frac{n}{m}$$

برهان. بهازای عدد صحیح و ثابت i ، از معادله (۲۱.۶) نتیجه می‌شود که

$$\begin{aligned} a_{i+1} &\leq (1+m)a_i + n \\ &\leq (1+m)[(1+m)a_{i-1} + n] + n \\ &\leq (1+m)\{(1+m)[(1+m)a_{i-2} + n] + n\} + n \\ &\vdots \\ &\leq (1+m)^{i+1}a_0 + [1 + (1+m) + (1+m)^2 + \dots + (1+m)^i]n. \end{aligned}$$

اما

$$1 + (1+m) + (1+m)^2 + \dots + (1+m)^i = \sum_{j=0}^i (1+m)^j$$

یک سری هندسی با قدر نسبت $(i + m)$ است و، در نتیجه، دارای مجموع

$$\frac{1 - (1 + m)^{i+1}}{1 - (1 + m)}$$

می باشد. لذا،

$$\begin{aligned} a_{i+1} &\leq (1 + m)^{i+1} a_0 + \frac{(1 + m)^{i+1} - 1}{m} n \\ &= (1 + m)^{i+1} \left(\frac{n}{m} + a_0 \right) - \frac{n}{m}, \end{aligned}$$

و، طبق لم ۷.۰۶،

$$a_{i+1} \leq e^{(i+1)m} \left(\frac{n}{m} + a_0 \right) - \frac{n}{m}.$$

قضیه ۹.۰۶. فرض کنیم $y(t)$ جواب منحصر بفرد مسئله مقدار اولیه خوش وضع

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

بوده و w_0, w_1, \dots, w_N تقریبهای تولید شده به وسیله روش اویلر، به ازای عدد صحیح مثبت N ، با استفاده از نمادگذاری الگوریتم ۱.۰۶، باشد. هرگاه f بر

$$D = \{(t, y) | a \leq t \leq b, -\infty < y < \infty\}$$

در شرط لیبشیتس با ثابت L صدق کرده و ثابت M با این خاصیت که

$$|y''(t)| \leq M, \quad t \in [a, b]$$

وجود داشته باشد، N نگاه به ازای هر $i = 0, 1, 2, \dots, N$

$$(۲۳.۰۶) \quad |y(t_i) - w_i| \leq \frac{hM}{2L} [e^{L(t_i - a)} - 1]$$

برهان. وقتی $i = 0$ ، درستی مطلب واضح است.

از معادله (۱۶.۰۶)، به ازای $i = 0, 1, \dots, N - 1$ ، داریم

$$y(t_{i+1}) = y(t_i) + hf(t_i, y(t_i)) + \frac{h^2}{2} y''(t_i + \theta_i h)$$

و، از معادله (۱۸.۰۶)،

$$w_{i+1} = w_i + hf(t_i, w_i)$$

در نتیجه، با استفاده از نماد $y_i = y(t_i)$ و $y_{i+1} = y(t_{i+1})$ ، خواهیم داشت

$$y_{i+1} - w_{i+1} = y_i - w_i + h[f(t_i, y_i) - f(t_i, w_i)] + \frac{h^2}{2} y''(t_i + \theta_i h)$$

و

$$|y_{i+1} - w_{i+1}| \leq |y_i - w_i| + h|f(t_i, y_i) - f(t_i, w_i)| + \frac{h^2}{2} |y''(t_i + \theta_i h)|$$

از فرض اینکه f نسبت به متغیر دوم در شرط لیبشیتس با ثابت L صدق می‌کند و اینکه $|y''(t)| \leq M$ ، نتیجه می‌شود که

$$|y_{i+1} - w_{i+1}| \leq |y_i - w_i|(1 + hL) + \frac{h^2 M}{2}$$

با توجه به لم ۸.۶ و قرار دادن $a_j = |y_j - w_j|$ ، به‌ازای هر $j = 0, 1, \dots, N$ ، در حالی که $m = hL$ و $n = h^2 M/2$ ، می‌بینیم که

$$|y_{i+1} - w_{i+1}| \leq e^{(i+1)mL} \left(|y_0 - w_0| + \frac{h^2 M}{2hL} \right) - \frac{h^2 M}{2hL}$$

یا، چون $|y_0 - w_0| = 0$ و $(i+1)h = t_{i+1} - t_0 = \tau_{i+1} - a$ ، نتیجه مطلوب

$$|y_{i+1} - w_{i+1}| \leq \frac{hM}{2L} (e^{(i+1)mL} - 1)$$

حاصل می‌شود.

ضعف قضیه ۹.۶ در نیاز به شناختن کرانی برای مشتق دوم جواب است. گرچه اغلب این شرط‌ها را از بدست آوردن یک کران خطای واقعی باز می‌دارد، باید توجه داشت که اگر $\partial f/\partial t$ و $\partial f/\partial y$ هر دو وجود داشته باشند،

$$\begin{aligned} y''(t) &= \frac{dy'}{dt}(t) = \frac{df}{dt}(t, y(t)) \\ &= \frac{\partial f}{\partial t}(t, y(t)) + \frac{\partial f}{\partial y}(t, y(t))f(t, y(t)) \end{aligned}$$

و برای مسائل معینی، بدون شناخت واقعی $y(t)$ ، احتمال بدست آوردن کرانی برای

$y''(t)$ وجود دارد.

مثال ۲. به مسئله مقدار اولیه

$$y' = -y + t + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

مطرح شده در مثال ۱ بازگشته، ملاحظه می‌کنیم که، با $f(t, y) = -y + t + 1$ داریم $\partial f / \partial y = -1$ ؛ در نتیجه، قضیه ۳.۶ ایجاب می‌کند که f در شرط لیب شیتس با $L = 1$ صدق می‌کند.

همچنین، چون در این حالت می‌دانیم که جواب دقیق $y(t) = t + e^{-t}$ است، داریم $y''(t) = e^{-t}$ و،

$$|y''(t)| \leq e^{-0} = 1, \quad t \in [0, 1]$$

با استفاده از معادله (۲۳.۶) با $h = 0.1$ و $M = L = 1$ کران خطای

$$|y_i - w_i| \leq 0.05(e^h - 1)$$

نتیجه می‌شود.

در جدول زیر این کران خطا همراه با خطای واقعی حاصل از مثال ۱ آمده است.

t_i	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
خطای واقعی	.00484	.00873	.01182	.01422	.01604	.01737	.01829	.01886	.01915	.01920
کران خطا	.00526	.01107	.01749	.02459	.03244	.04111	.05069	.06128	.07298	.08591

توجه کنید که، اگرچه خطای واقعی برای مشتق دوم جواب بکار رفته، کران خطا به طور قابل ملاحظه‌ای بیشتر از خطای واقعی است.

اهمیت اصلی فرمول کران خطا، ارائه شده در قضیه ۳.۶، این است که این کران به‌طور خطی به اندازه گام h بستگی دارد. در نتیجه، کم شدن اندازه گام متناظراً "دقت بیشتری به تقریبها می‌دهد".

آنچه در قضیه ۳.۶ نادیده گرفته شد اثر خطای گرد کردن در انتخاب اندازه گام است. واضح است که هرچه h کوچکتر باشد، محاسبات بیشتری لازم می‌شود؛ و وقتی عرفاً

ارقام حساب متناهی بکار می‌رود، باید خطای گرد کردن بیشتری را انتظار داشت. در عمل، معادله تفاضلی به شکل

$$w_0 = \alpha,$$

$$w_i = w_{i-1} + hf(t_{i-1}, w_{i-1}), \quad i = 1, 2, \dots, N$$

جهت محاسبه تقریب جواب y_i در نقطه شبکه t_i بکار نمی‌رود، بلکه به جای آن معادله‌ای به شکل

$$u_0 = \hat{\alpha},$$

$$(24.6) \quad u_i = u_{i-1} + hf(t_{i-1}, u_{i-1}) + \delta_i, \quad i = 1, 2, \dots, N$$

بکار می‌رود، که در آن $\hat{\alpha}$ تقریبی برای مقدار اولیه α است و δ_i خطای گرد کردن وابسته به محاسبه $hf(t_{i-1}, u_{i-1}) + \delta_i$ می‌باشد. با استفاده از روشهایی شبیه آنچه در برهان قضیه ۹.۶ آمد (ر.ک. تمرین ۸)، نتیجه زیر را بدست می‌آوریم، که یک کران خطا برای تقریبهای با ارقام متناهی به y_i حاصل از بکار بردن روش اوپلر است.

قضیه ۱۰.۶. فرض کنیم $y(t)$ جواب منحصر بفرد مسئله مقدار اولیه خوش وضع

$$(25.6) \quad y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

بوده و u_0, u_1, \dots, u_N تقریبهای حاصل از بکار بردن (۲۴.۶) باشد. هرگاه به ازای هر $i = 0, 1, \dots, N$ $|\delta_i| < \delta$ ، که در آن δ_0 مقدار $\hat{\alpha} - \alpha$ باشد، مفروضات قضیه ۹.۶ در مورد (۲۵.۶) برقرار باشد، آنگاه به ازای هر $i = 0, 1, \dots, N$

$$(26.6) \quad |y(t_i) - u_i| \leq \frac{1}{L} \left(\frac{hM}{2} + \frac{\delta}{h} \right) [e^{L(t_i-a)} - 1] + |\delta_0| e^{L(t_i-a)}$$

کران خطای (۲۶.۶) بهیچوجه نسبت به h خطی نیست و، در واقع، چون

$$\lim_{h \rightarrow 0} \left(\frac{hM}{2} + \frac{\delta}{h} \right) = +\infty$$

برای مقادیر به قدر کافی کوچک h انتظار می‌رود خطا بزرگ شود. این مسئله از همان نوع مسئله مذکور در بخش ۱۰.۵ در تقریب مشتق یک تابع تنها یا استفاده از مقادیر تابعی است (ر.ک. تمرین ۱۰ از بخش ۱۰.۵)، و همانند روند مورد بحث در آن بخش می‌توان، با استفاده از روشهای حساب دیفرانسیل و انتگرال برای یافتن مینیم $(hM/2) + (\delta/h)$ یک مقدار بهینه از h را بدست آورد. در واقع (ر.ک. تمرین ۷)، این مقدار برابر است

با

$$h = \sqrt{2\delta/M}$$

با تقلیل بیشتر h از این مقدار، خطای کل تقریب میل به افزایش می‌یابد. لیکن، معمولاً، مقدار δ به قدر کافی کوچک است، و این کران پایینی h در عملکرد روش اویلر اثر ندارد. این همان وضعیت بخش ۱.۵ برای تقریب مشتق نمی‌باشد.

مجموعه تمرینات ۲.۶

۱. (T) با استفاده از روش اویلر با $h = .1$ ، جواب

$$y' = \frac{2}{t}y + t^2e^t, \quad 1 \leq t \leq 2$$

$$y(1) = 0.$$

را تقریب کنید. با استفاده از معادله (۲۳.۶)، مقدار مورد لزوم h برای اینکه $|y(t_i) - w_i| \leq .1$ را حساب کنید.
(ب) برای مسئله مقدار اولیه

$$y' = \frac{1}{t^2} - \frac{y}{t} - y^2, \quad 1 \leq t \leq 2$$

$$y(1) = -1,$$

با استفاده از روش اویلر با $h = .05$ ، تقریبی برای جواب $y(t) = -1/t$ بدست آورید. این تقریب را با جواب واقعی مقایسه کنید.

(پ) با استفاده از روش اویلر با $h = .1$ و $h = .2$ ، تقریبی برای جواب $y(t) = e^{-t} + t$ از مسئله مقدار اولیه

$$y' = -y + t + 1, \quad 0 \leq t \leq 5$$

$$y(0) = 1,$$

بدست آورید.

۲. مسئله مقدار اولیه

$$y' = -10y, \quad 0 \leq t \leq 2$$

$$y(0) = 1,$$

که دارای جواب $y(t) = e^{-10t}$ است، را در نظر بگیرید. اگر روش اویلر با $h = .1$ بکار

رود، چه رخ می دهد؟ آیا این رفتار از قضیه ۹.۶ تخلف می کند؟
 ۳. با استفاده از روش اویلر با $h = 1$ ، تقریبی برای جواب

$$y' = 1 + t \sin(ty), \quad 0 \leq t \leq 2$$

$$y(0) = 0,$$

بدست آورید.

۴. در یک مدار موازی با ولتاژ \mathcal{E} ، مقاومت R ، ضریب سلف L ، و ظرفیت C ، شدت جریان i در معادله دیفرانسیل

$$\frac{di}{dt} = C \frac{d^2 \mathcal{E}}{dt^2} + \frac{1}{R} \frac{d\mathcal{E}}{dt} + \frac{1}{L} \mathcal{E}$$

صدق می کند. فرض کنید $C = 3$ فاراد، $R = 1.4$ اهم، و $L = 1.7$ هانری بوده، و ولتاژ با

$$\mathcal{E}(t) = e^{-.06t} \sin(2t - \pi)$$

داده شده باشد. با استفاده از روش اویلر، شدت جریان i را برای مقادیر $t = .1j, 0 \leq j \leq 10$ بدست آورید.

۵. در کتابی به نام "نگاهی به تاریخ از دیدگاه ریاضیات"، راشوسکی^۱ [۵۱] (صفحات ۱۱۰ - ۱۰۳) یک مدل برای مسئله‌ای مربوط به تولید مثل در یک جامعه غیر وابسته به کلیسای رسمی انگلیس را در نظر می گیرد. فرض کنید جامعه‌ای جمعیتش در زمان t ، به سال، $x(t)$ باشد، و همه افراد غیر وابسته به کلیسای رسمی انگلیس که با افرادی غیر وابسته به کلیسای رسمی انگلیس ازدواج می کنند، فرزندانشان نیز غیر وابسته به کلیسای رسمی انگلیس باشند، درحالی که یک نسبت ثابت r از تمام موالید دیگر نیز وابسته به کلیسای انگلیس نباشند. اگر میزان تولد و مرگ افراد بترتیب ثابت‌های b و d بوده، و افراد وابسته به کلیسای رسمی انگلیس به طور تصادفی با افرادی که وابسته به کلیسای رسمی انگلیس نیستند ازدواج کنند، این مسئله را می توان با معادلات دیفرانسیل

$$\frac{dx(t)}{dt} = (b - d)x(t)$$

$$\frac{dx_n(t)}{dt} = (b - r - d)x_n(t) + rb(x(t) - x_n(t))$$

بیان کرد، که در آن $x_n(t)$ تعداد افرادی از جمعیت که غیر وابسته به کلیسای رسمی انگلیس اند را در زمان t مشخص می‌کند.

(T) اگر متغیر $p(t) = x_n(t)/x(t)$ نسبت افراد غیر وابسته به کلیسای رسمی انگلیس به افراد جامعه در زمان t باشد، نشان دهید که این معادلات را می‌توان ترکیب و به صورت معادله دیفرانسیل

$$\frac{d}{dt} p(t) = rb(1 - p(t)).$$

خلاصه کرد.

(ب) با فرض $p(0) = .01$ ، $b = .02$ ، $d = .015$ ، و $r = .1$ و استفاده از الگوریتم ۱۰۶، جواب $p(t)$ از $t = 0$ تا $t = 50$ ، وقتی اندازه گام $h = 1$ سال است، را تقریب کنید.

(پ) معادله دیفرانسیل مربوط به $p(t)$ را دقیقاً "حل و نتیجه" قسمت (ب) را، وقتی $t = 50$ ، با مقدار دقیق در همان زمان مقایسه کنید.

۶. برای مسئله مقدار اولیه $y' = f(t, y)$ ، $a \leq t \leq b$ ، با $y(a) = \alpha$ ،

(T) روش اویلر را با انتگرالگیری از معادله دیفرانسیل از t_i تا t_{i+1} نتیجه‌گیری و، با استفاده از فرمول انتگرالگیری مناسبی، این انتگرال را تقریب کنید.

(ب) با استفاده از قاعده دوزنقه‌ای به عنوان فرمول انتگرالگیری، یک روش تفاضلی، موسوم به روش دوزنقه‌ای، بدست آورید.

۷. (T) نشان دهید که $h = \sqrt{2\delta/M}$ تابع

$$g(h) = \frac{hM}{2} + \frac{\delta}{h}$$

را مینیمم می‌کند.

(ب) برای مسئله مقدار اولیه

$$y' = -y + 1, \quad 0 \leq t \leq 1$$

$$y(0) = 0,$$

مقدار h برای مینیم کردن $\sqrt{g(h)}$ را محاسبه کنید. اگر در قسمت (ت) از حساب n رقمی استفاده می‌کنید، فرض کنید $\delta = 5 \times 10^{-(n+1)}$.

(پ) برای h بهینه محاسبه شده در قسمت (ب)، از معادله (۲۶.۶) استفاده کرده خطای مینیمم قابل حصول را حساب کنید.

(ت) خطای واقعی بدست آمده با $h = .1$ و $h = .01$ را با خطای مینیمم در قسمت

(پ) مقایسه کنید. آیا می‌توانید نتایج را توضیح دهید؟

۸. قضیه ۱۰.۶ را اثبات کنید.

۳۰۶ روشهای تیلور از مرتبه بالاتر

چون منظور از روشهای عددی معمولاً "تعیین تقریبهای به قدر کافی دقیق با حداقل کار است، لازم است وسیله‌ای جهت مقایسه کارآیی روشهای مختلف برحسب مورد استفاده آنها در محاسبه داشته باشیم.

در روشهای معادله تفاضلی جهت حل معادلات دیفرانسیل معمولی، نظیر روش اویلر، اولین سنجش مورد نیاز خطای برشی موضعی روش نامیده می‌شود. این خطا "موضعی" نام دارد، چون مقدار دقت روش را در یک مرحله معین، به فرض دقیق بودن آن در مرحله قبل، اندازه‌گیری می‌کند. این خطا به معادله دیفرانسیل تقریب شده، به اندازه گام، و به مرحله خاص در تقریب نیز بستگی دارد. در روش اویلر، خطای برشی موضعی در مرحله i م برای مسئله

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

عبارت است از

$$\tau_i = \frac{y_i - y_{i-1}}{h} - f(t_{i-1}, y_{i-1}), \quad i = 1, 2, \dots, N$$

که در آن، طبق معمول، $y_i = y(t_i)$ مقدار دقیق جواب در t_i است.

با در نظر گرفتن معادله (۱۶.۶) در بخش قبل، نتیجه می‌شود که به‌ازای θ_i ای

$$0 < \theta_i < 1$$

$$\tau_i = \frac{h}{2} y''(t_i + \theta_i h)$$

و وقتی $y''(t)$ به وسیله ثابت M بر $[a, b]$ کراندار باشد، نتیجه می‌شود که

$$|\tau_i| \leq \frac{h}{2} M$$

با یادآوری تعریف ۱۸.۰۱، صفحه ۲۳، نتیجه می‌گیریم که خطای برشی در روش اویلر برابر $O(h)$ است. این نتیجه به ما رهنمود می‌دهد که یک طریق انتخاب روشهای معادله تفاضلی برای حل معادلات دیفرانسیل معمولی باید چنان باشد که خطاهای برشی موضعی در آنها $O(h^p)$ ، به‌ازای یک مقدار p حتی‌الامکان بزرگ، بوده، و درعین حال تعداد و پیچیدگی محاسبات روشها درحد معقولی باشند.

پیش از بررسی روشهای انجام این منظور، دو تعریف ارائه می‌دهیم که به‌همگرایی روش معادله تفاضلی، وقتی اندازه گام h کوچک می‌شود، مربوطند.

تعریف ۱۱.۰۶. یک روش معادله تفاضلی با خطای برشی موضعی τ_i در مرحله i م را با معادله دیفرانسیلی که آن را تقریب می‌کند سازگار گوئیم اگر

$$\lim_{h \rightarrow 0} \max_{1 \leq i \leq N} |\tau_i| = 0.$$

باید توجه داشت که این تعریف نیز اساساً "تعریفی" موضعی است، زیرا به‌ازای هر یک از مقادیر τ_i ، مقدار دقیق $f(t_i, y_i)$ را با معادله تفاضلی که y را تقریب می‌کند مقایسه می‌کنیم. یک وسیله واقعی‌تر از تحلیل اثرات کوچک ساختن h تعیین اثر "کلی" این روش است که باید خطای ماکزیمم روش در تمام برد تقریب باشد، تنها با این فرض که این روش در مقدار اولیه نتیجه دقیق را ارائه می‌دهد. تعریف توصیف‌کننده روشی که با این معنی دارای همگرایی است در زیر ارائه می‌شود.

تعریف ۱۲.۰۶. یک روش معادله تفاضلی را نسبت به معادله دیفرانسیلی که آن را تقریب می‌کند همگرا گوئیم اگر

$$\lim_{h \rightarrow 0} \max_{1 \leq i \leq N} |y_i - w_i| = 0$$

که در آن $y_i = y(t_i)$ مقدار دقیق جواب معادله دیفرانسیل بوده و w_i تقریب حاصل از روش تفاضلی در مرحله i م است.

با بررسی نامعادله (۲۳.۰۶) در فرمول کران خطای روش اویلر، می‌توان دید که،

تحت مفروضات قضیه ۹.۶،

$$\max_{1 \leq i \leq N} |y_i - w_i| \leq \frac{Mh}{2L} (e^{L(b-a)} - 1)$$

در نتیجه، روش اویلر نسبت به معادله دیفرانسیلی که در شرایط این قضیه صدق می‌کند همگراست و درجه همگرایی برابر $O(h)$ است.

چون روش اویلر با استفاده از قضیه تیلور با $n = 2$ برای تقریب جواب این معادله دیفرانسیل بدست می‌آید، اولین سعی ما در بدست آوردن روشهایی با خواص همگرایی بهتر از روشهای تفاضلی، تعمیم این روش به مقادیر بزرگتر n می‌باشد. فرض کنیم جواب $y(t)$ از مسئله مقدار اولیه

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

دارای $(n + 1)$ مشتق پیوسته باشد، و جواب $y(t)$ را برحسب چند جمله‌ای تیلور درجه n م آن حول t_i بسط داده، بدست می‌آوریم که به ازای θ_i ی، که $0 < \theta_i < 1$ ،

$$y(t_{i+1}) = y(t_i) + hy'(t_i) + \frac{h^2}{2} y''(t_i) + \dots$$

(۲۷.۶)

$$+ \frac{h^n}{n!} y^{(n)}(t_i) + \frac{h^{n+1}}{(n+1)!} y^{(n+1)}(t_i + \theta_i h)$$

مشتگیری متوالی از جواب، یعنی از $y(t)$ ، نتیجه می‌دهد که

$$y'(t) = f(t, y(t));$$

$$y''(t) = \frac{\partial f}{\partial t}(t, y(t)) + \frac{\partial f}{\partial y}(t, y(t)) \frac{dy}{dt}$$

$$= \frac{\partial f}{\partial t}(t, y(t)) + \frac{\partial f}{\partial y}(t, y(t)) f(t, y(t)) \equiv f'(t, y(t))$$

$$y'''(t) = \frac{df'}{dt}(t, y(t)) \equiv f''(t, y(t));$$

و، در حالت کلی،

$$y^{(k)}(t) = \frac{dy^{(k-1)}}{dt} = \frac{df^{(k-2)}}{dt}(t, y(t)) \equiv f^{(k-1)}(t, y(t)).$$

با جایگذاری این نتایج در معادله (۲۷.۶)، به ازای θ_i ای، که $0 < \theta_i < 1$

بدست می‌آید

$$y(t_{i+1}) = y(t_i) + hf(t_i, y(t_i)) + \frac{h^2}{2} f'(t_i, y(t_i)) + \dots$$

(۲۸.۰۶)

$$+ \frac{h^n}{n!} f^{(n-1)}(t_i, y(t_i)) + \frac{h^{n+1}}{(n+1)!} f^{(n)}(t_i + \theta_i h, y(t_i + \theta_i h)) \cdot$$

روش تفاضلی متناظر با معادله (۲۸.۰۶) با صرف نظر کردن از جمله باقیمانده شامل θ_i بدست می‌آید، و روش تیلور مرتبه n نامیده می‌شود.

$$w_0 = \alpha,$$

(۲۹.۰۶)

$$w_{i+1} = w_i + hT^{(n)}(t_i, w_i) \cdot i = 0, 1, \dots, N - 1$$

که در آن

$$T^{(n)}(t_i, w_i) = f(t_i, w_i) + \frac{h}{2} f'(t_i, w_i) + \dots + \frac{h^{n-1}}{n!} f^{(n-1)}(t_i, w_i) \cdot$$

توجه کنید که، با این اصطلاح، روش اوایلر همان روش تیلور مرتبه اول است.

مثال ۱. برای بکار بردن روش تیلور مرتبه‌های دو و چهار در مسئله مقدار اولیه

$$y' = -y + t + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

که در مثال ۱ از بخش ۲.۰۶ بررسی شد، باید سه مشتق اول $f(t, y(t)) = -y + t + 1$ را بدست آوریم.

$$f'(t, y(t)) = \frac{d}{dt}(-y + t + 1) = -y' + 1 = y - t - 1 + 1 = y - t$$

$$f''(t, y(t)) = \frac{d}{dt}(y - t) = y' - 1 = -y + t + 1 - 1 = -y + t,$$

و

$$f'''(t, y(t)) = \frac{d}{dt}(-y + t) = -y' + 1 = y - t - 1 + 1 = y - t$$

و در نتیجه،

$$T^{(2)}(t_i, w_i) = f(t_i, w_i) + \frac{h}{2} f'(t_i, w_i)$$

1	مقدار دقیق	روش اولی	خطا در روش اولی	روش تیلور مرتبه دو	خطا در روش تیلور مرتبه دو	روش تیلور مرتبه چهار	خطا در روش تیلور مرتبه چهار
0	1.00000000000	1.000000	0	1.000000	0	1.00000000000	0
1.	1.0048374180	1.000000	4.837×10^{-3}	1.005000	1.626×10^{-4}	1.0048375000	8.200×10^{-8}
2.	1.0187307531	1.010000	8.731×10^{-3}	1.019025	2.942×10^{-4}	1.0187309014	1.483×10^{-7}
3.	1.0408182207	1.029000	1.182×10^{-2}	1.041218	3.998×10^{-4}	1.0408184220	2.013×10^{-7}
4.	1.0703200460	1.056100	1.422×10^{-2}	1.070802	4.820×10^{-4}	1.0703202889	2.429×10^{-7}
5.	1.1065306597	1.090490	1.604×10^{-2}	1.107076	5.453×10^{-4}	1.1065309344	2.747×10^{-7}
6.	1.1488116361	1.131441	1.737×10^{-2}	1.149404	5.924×10^{-4}	1.1488119344	2.983×10^{-7}
7.	1.1965853038	1.178297	1.829×10^{-2}	1.197211	6.257×10^{-4}	1.1965856187	3.149×10^{-7}
8.	1.2493289641	1.230467	1.887×10^{-2}	1.249976	6.470×10^{-4}	1.2493292897	3.256×10^{-7}
9.	1.3065696597	1.287420	1.915×10^{-2}	1.307228	6.583×10^{-4}	1.3065699912	3.315×10^{-7}
10	1.3678794412	1.348678	1.920×10^{-2}	1.368541	6.616×10^{-4}	1.3678797744	3.332×10^{-7}

جدول ۲-۶

$$\begin{aligned}
 &= -w_i + t_i + 1 + \frac{h}{2}(w_i - t_i) \\
 &= \left(1 - \frac{h}{2}\right)(t_i - w_i) + 1
 \end{aligned}$$

۹

$$\begin{aligned}
 T^{(4)}(t_i, w_i) &= f(t_i, w_i) + \frac{h}{2}f'(t_i, w_i) + \frac{h^2}{6}f''(t_i, w_i) + \frac{h^3}{24}f'''(t_i, w_i) \\
 &= -w_i + t_i + 1 + \frac{h}{2}(w_i - t_i) + \frac{h^2}{6}(-w_i + t_i) \\
 &\quad + \frac{h^3}{24}(w_i - t_i) \\
 &= \left(1 - \frac{h}{2} + \frac{h^2}{6} - \frac{h^3}{24}\right)(t_i - w_i) + 1.
 \end{aligned}$$

در نتیجه، روشهای تیلور مرتبه دو و چهار بترتیب عبارتند از

$$\begin{aligned}
 (۳۰.۶) \quad w_0 &= 1, \\
 w_{i+1} &= w_i + h \left[\left(1 - \frac{h}{2}\right)(t_i - w_i) + 1 \right], \quad i = 0, 1, \dots, N-1
 \end{aligned}$$

۹

$$\begin{aligned}
 (۳۱.۶) \quad w_0 &= 1, \\
 w_{i+1} &= w_i + h \left[\left(1 - \frac{h}{2} + \frac{h^2}{6} - \frac{h^3}{24}\right)(t_i - w_i) + 1 \right], \quad i = 0, 1, \dots, N-1
 \end{aligned}$$

با یکبار بردن $h = .1$ ، نتیجه می شود که $N = 10$ و، به ازای هر $i = 1, 2, \dots, 10$

$t_i = .1i$ ؛ در نتیجه، معادله (۳۰.۶) می شود

$$\begin{aligned}
 (۳۲.۶) \quad w_0 &= 1, \\
 w_{i+1} &= w_i + .1 \left[\left(1 - \frac{.1}{2}\right)(.1i - w_i) + 1 \right] \\
 &= .905w_i + .0095i + .1,
 \end{aligned}$$

و (۳۱.۶)، به ازای هر $i = 0, 1, \dots, 9$ ، خواهد شد

$$w_0 = 1,$$

$$(۳۳.۰۶) \quad w_{i+1} = w_i + .1 \left[\left(1 - \frac{.1}{2} + \frac{.01}{6} - \frac{.001}{24} \right) (1i - w_i) + 1 \right]$$

$$= .9048375w_i + .00951625i + .1.$$

مقادیر واقعی جواب $y(t) = t + e^{-t}$ همراه با تقریبهای حاصل از روش اویلر در مثال ۱ از بخش ۲.۰۶، و نتایج حاصل از روشهای تیلور مرتبه‌های دو و چهار و خطاهای واقعی مربوط به این روشها در جدول ۲.۰۶ آمده‌اند.

همانطور که از بحث ما در روش اویلر انتظار می‌رود، روش تیلور مرتبه n دارای خطای برشی موضعی $O(h^n)$ است، مشروط بر اینکه جواب معادله دیفرانسیل به قدر کافی خوشرفتار باشد. این مطلب به سادگی با توجه به اینکه معادله (۲۸.۰۶) را می‌توان به صورت

$$y_{i+1} - y_i - hf(t_i, y_i) - \frac{h^2}{2} f'(t_i, y_i) - \dots - \frac{h^n}{n!} f^{(n-1)}(t_i, y_i)$$

$$= \frac{h^{n+1}}{(n+1)!} f^{(n)}(t_i + \theta_i h, y(t_i + \theta_i h))$$

نوشت ملاحظه می‌شود. در نتیجه، خطای برشی موضعی در $(i+1)$ مین مرحله، به‌ازای هر $i = 0, 1, \dots, N-1$

$$\tau_{i+1} = \frac{y_{i+1} - y_i}{h} - T^{(n)}(t_i, y_i)$$

$$= \frac{h^n}{(n+1)!} f^{(n)}(t_i + \theta_i h, y(t_i + \theta_i h))$$

اگر $y \in C^{n+1}[a, b]$ ، نتیجه می‌شود که $y^{(n+1)}(t) = f^{(n)}(t, y(t))$ بر $[a, b]$ کراندار است و، به‌ازای هر $i = 1, 2, \dots, N$ ، $\tau_i = O(h^n)$.

مجموعه تمرینات ۳.۰۶

۱. با استفاده از روشهای تیلور مرتبه‌های دو و چهار با $h = .1$ ، جوابهای مسائل زیر را تقریب کنید:

$$y' = 1 - y, 0 \leq t \leq 2 \quad (-) \quad y(0) = 0$$

$$y' = t + y, 0 \leq t \leq 2 \quad (T) \quad y(0) = -1;$$

۲. در مسئله مقدار اولیه

$$y' = \frac{2}{t}y + t^2e^t, \quad 1 \leq t \leq 2$$

$$y(1) = 0.$$

با استفاده از روش تیلور مرتبه چهار و $h = 0.05$ ، تقریبی برای جواب بدست آورید. پاسخ خود را با مقادیر درست y مقایسه کنید.

۳. در مسئله مقدار اولیه

$$y' = -y + t + 1, \quad 0 \leq t \leq 5$$

$$y(0) = 2.$$

با استفاده از روش تیلور مرتبه چهار و $h = 0.2$ ، تقریبی برای $y(5)$ بدست آورید.

۴. با استفاده از روش تیلور مرتبه دو و $h = 0.1$ ، جواب

$$y' = 1 + t \sin(ty), \quad 0 \leq t \leq 2$$

$$y(0) = 0.$$

را تقریب کنید. جوابهای خود را با نتایج حاصل از تمرین ۳ در بخش ۲.۶ مقایسه نمایید.

۵. یک پرتابه به جرم $m = 0.11$ کیلوگرم به طور قائم با سرعت اولیه $v(0) = 8$ متر بر ثانیه

به بالا پرتاب شده و تحت اثر نیروی ثقل $F_b = -mg$ و مقاومت هوای $F_r = -kv^2$ که در آن $g = 9.8$ متر بر مجذور ثانیه و $k = 0.002$ کیلوگرم بر متر سرعتش کم می شود. معادله دیفرانسیل سرعت v به صورت زیر است:

$$mv' = -mg - kv^2$$

(آ) با استفاده از روشهای تیلور مرتبه دو و چهار، سرعت را پس از $1.0, 2.0, \dots, 1.0$ ثانیه بدست آورید.

(ب) تعیین کنید چه زمانی، به نزدیکترین ثانیه، پرتابه به بیشترین ارتفاع خود می رسد و سقوط می کند. روش تیلور مرتبه چهار را بکار ببرید.

۶. آیا روش تفاضلی

$$w_{i+1} = w_{i-1} + 2hf(t_i, w_i), \quad i = 1, 2, \dots, N-1$$

که در آن $w_0 = \alpha$ و $w_1 = y(t_1)$ ، با معادله دیفرانسیل $y' = f(t, y)$ ، $a < t < b$ ، $y(a) = \alpha$ سازگار است؟ خطای برشی موضعی را بدست آورید.

۴.۶ روشهای رونگ ۱ - کوتا ۲

روشهای تیلور خلاصه شده در بخش قبل دارای خاصیت مطلوب خطای برشی موضعی از مرتبه بالا هستند، لیکن لزوم محاسبه و ارزیابی مشتقهای $f(t, y)$ نقص آن است. این امر در بسیاری مسائل می تواند پیچیده و وقتگیر باشد و، در نتیجه، روشهای تیلور بندرت در عمل بکار می روند. روشهای رونگ - کوتا از خطای برشی موضعی مرتبه بالای روشهای تیلور استفاده می کند، در حالی که محاسبه و ارزیابی مشتقهای $f(t, y)$ را حذف می کند. پیش از ارائه مفاهیم و رای این مطالب، لازم است قضیه تیلور با دو متغیر را بیان کنیم. برهان این قضیه را می توان در هر کتاب معمولی در باب حساب دیفرانسیل و انتگرال پیشرفته یافت (مثلاً، ر. ک. فولکس [۲۹]، صفحه ۲۶۰).

قضیه ۱۳.۶. فرض کنیم $f(t, y)$ و همه مشتقهای جزئی آن از مرتبه نایبتر از $(n + 1)$ بر

$$D = \{(t, y) | a \leq t \leq b, c \leq y \leq d\}$$

پیوسته باشد. همچنین، $(t_0, y_0) \in D$. به ازای هر $(t, y) \in D$ ، نقطه ای مانند $(\xi, \eta) \in D$ وجود دارد با خاصیت

$$f(t, y) = P_n(t, y) + R_n(t, y)$$

که در آن

$$\begin{aligned} P_n(t, y) = & f(t_0, y_0) + \left[(t - t_0) \frac{\partial f}{\partial t}(t_0, y_0) + (y - y_0) \frac{\partial f}{\partial y}(t_0, y_0) \right] \\ & + \left[\frac{(t - t_0)^2}{2} \frac{\partial^2 f}{\partial t^2}(t_0, y_0) + (t - t_0)(y - y_0) \frac{\partial^2 f}{\partial t \partial y}(t_0, y_0) \right. \\ & \left. + \frac{(y - y_0)^2}{2} \frac{\partial^2 f}{\partial y^2}(t_0, y_0) \right] \\ & + \dots + \left[\frac{1}{n!} \sum_{j=0}^n \binom{n}{j} (t - t_0)^{n-j} (y - y_0)^j \frac{\partial^n f}{\partial t^{n-j} \partial y^j}(t_0, y_0) \right] \end{aligned}$$

$$R_n(t, y) = \frac{1}{(n+1)!} \sum_{j=0}^{n+1} \binom{n+1}{j} (t - t_0)^{n+1-j} (y - y_0)^j \frac{\partial^{n+1} f(\xi, \eta)}{\partial t^{n+1-j} \partial y^j}$$

P_n چند جمله‌ای تیلور درجه n دو متغیره برای تابع f حول (t_0, y_0) و خطای برشی وابسته به $P_n(t, y)$ نامیده می‌شود.

مثال ۱. چند جمله‌ای تیلور درجه سه برای $f(t, y) = \sin(ty)$ حول $(0, \pi)$ از

$$\begin{aligned} P_3(t, y) = & f(t_0, y_0) + (t - t_0) \frac{\partial f}{\partial t}(t_0, y_0) + (y - y_0) \frac{\partial f}{\partial y}(t_0, y_0) \\ & + \left[\frac{(t - t_0)^2}{2} \frac{\partial^2 f}{\partial t^2}(t_0, y_0) + (t - t_0)(y - y_0) \frac{\partial^2 f}{\partial t \partial y}(t_0, y_0) \right. \\ & \left. + \frac{(y - y_0)^2}{2} \frac{\partial^2 f}{\partial y^2}(t_0, y_0) \right] \\ & + \left[\frac{(t - t_0)^3}{6} \frac{\partial^3 f}{\partial t^3}(t_0, y_0) + \frac{(t - t_0)^2}{2} (y - y_0) \frac{\partial^3 f}{\partial t^2 \partial y}(t_0, y_0) \right. \\ & + \frac{(t - t_0)(y - y_0)^2}{2} \frac{\partial^3 f}{\partial t \partial y^2}(t_0, y_0) \\ & \left. + \frac{(y - y_0)^3}{6} \frac{\partial^3 f}{\partial y^3}(t_0, y_0) \right]. \end{aligned}$$

بدست می‌آید. با محاسبه هر یک از این مشتقهای جزئی در $(0, \pi)$ ، $P_3(t, y)$ به

$$P_3(t, y) = \pi t + t(y - \pi) - \frac{\pi^3}{6} t^3$$

تقلیل می‌یابد. این چندجمله‌ای تقریب نزدیکی به $\sin(ty)$ است مشروط بر اینکه t به صفر و y به π نزدیک باشد. مثلاً،

$$P_3(0.1, \pi + 0.1) = .03151076$$

درحالی که

$$\sin(0.1(\pi + 0.1)) = .03151071$$

اولین وسیله حصول به روش رونگ - کوتا تعیین مقادیر a_1 ، α_1 ، و β_1 با این

خاصیت است که $a_1 f(t + \alpha_1, y + \beta_1)$ عبارت

$$T^{(2)}(t, y) = f(t, y) + \frac{h}{2} f'(t, y)$$

را با خطایی نه بیشتر از $O(h^2)$ ، یعنی خطای برشی موضعی روش تیلور مرتبه دو، تقریب کند.

چون

$$\begin{aligned} f'(t, y) &= \frac{df}{dt}(t, y) \\ &= \frac{\partial f}{\partial t}(t, y) + \frac{\partial f}{\partial y}(t, y) \cdot y'(t) \end{aligned}$$

و

$$y'(t) = f(t, y)$$

نتیجه می شود که

$$(۳۴.۶) \quad T^{(2)}(t, y) = f(t, y) + \frac{h}{2} \frac{\partial f}{\partial t}(t, y) + \frac{h}{2} \frac{\partial f}{\partial y}(t, y) \cdot f(t, y)$$

باسبط $f(t + \alpha_1, y + \beta_1)$ به چند جمله ای تیلور خود از درجه یک حول (t, y) ،

داریم

$$(۳۵.۶) \quad \begin{aligned} a_1 f(t + \alpha_1, y + \beta_1) &= a_1 f(t, y) + a_1 \alpha_1 \frac{\partial f}{\partial t}(t, y) \\ &+ a_1 \beta_1 \frac{\partial f}{\partial y}(t, y) + a_1 \cdot R_1(t + \alpha_1, y + \beta_1) \end{aligned}$$

که در آن به ازای نقطه ای مانند (ξ, η) در مجموعه

$$\{ (t, y) \mid t \text{ بین } t + \alpha_1 \text{ و } t \text{ و } y \text{ بین } y + \beta_1 \text{ و } y \text{ است} \}$$

$$(۳۶.۶) \quad R_1(t + \alpha_1, y + \beta_1) = \frac{\alpha_1^2}{2} \frac{\partial^2 f}{\partial t^2}(\xi, \eta) + \alpha_1 \beta_1 \frac{\partial^2 f}{\partial t \partial y}(\xi, \eta) + \frac{\beta_1^2}{2} \frac{\partial^2 f}{\partial y^2}(\xi, \eta)$$

از مقابله ضرایب f و مشتقهای آن در معادلات (۳۴.۶) و (۳۵.۶) سه معادله

زیر بدست می آیند:

$$f(t, y): a_1 = 1$$

$$\frac{\partial f}{\partial t}(t, y): a_1 \alpha_1 = \frac{h}{2}$$

و

$$\frac{\partial f}{\partial t}(t, y): a_1 \beta_1 = \frac{h}{2} f(t, y).$$

پارامترهای a_1 ، α_1 ، و β_1 به طور منحصر بفرد به صورت

$$a_1 = 1, \quad \alpha_1 = \frac{h}{2}, \quad \beta_1 = \frac{h}{2} f(t, y)$$

تعیین می‌شوند؛ در نتیجه،

$$T^{(2)}(t, y) = f\left(t + \frac{h}{2}, y + \frac{h}{2} f(t, y)\right) - R_1\left(t + \frac{h}{2}, y + \frac{h}{2} f(t, y)\right)$$

و از (۳۶.۶)،

$$R_1\left(t + \frac{h}{2}, y + \frac{h}{2} f(t, y)\right) = \frac{h^2}{8} \frac{\partial^2 f}{\partial t^2}(\xi, \eta) + \frac{h^2}{4} f(t, y) \frac{\partial^2 f}{\partial t \partial y}(\xi, \eta) + \frac{h^2}{8} (f(t, y))^2 \frac{\partial^2 f}{\partial y^2}(\xi, \eta).$$

اگر تمام مشتقهای جزئی مرتبه دوم f کراندار باشند،

$$R_1\left(t + \frac{h}{2}, y + \frac{h}{2} f(t, y)\right)$$

برابر $O(h^2)$ ، یعنی مرتبه خطای برشی موضعی روش تیلور مرتبه دو، خواهد بود.

روش تفاضلی حاصل از جایگذاری $T^{(2)}(t, y)$ در روش تیلور مرتبه دو روش خاصی

از رونگ - کوتا است، که به روش نقطه میانی موسوم است و به صورت زیر ارائه می‌شود:

روش نقطه میانی:

$$w_0 = \alpha$$

$$(37.6) \quad w_{i+1} = w_i + hf\left(t_i + \frac{h}{2}, w_i + \frac{h}{2} f(t_i, w_i)\right), \quad i = 0, 1, \dots, N-1$$

چون نقطه پارامتر در $a_1 f(t + \alpha_1, y + \beta_1)$ حضور داشته، و همگی در رسیدن

به $T^{(2)}$ مورد نیاز بودند، انتظار لزوم شکل پیچیده‌تری در برقراری شرایط لازم برای هر

روش تیلور درجه بالاتر را باید داشت. در واقع،

مناسبتین شکل، با چهار پارامتر برای تقریب

$$T^{(3)}(t, y) = f(t, y) + \frac{h}{2} f'(t, y) + \frac{h^2}{6} f''(t, y)$$

عبارت است از

$$(38.6) \quad a_1 f(t, y) + a_2 f(t + \alpha_2, y + \delta_2 f(t, y))$$

و، حتی با این عبارت، قابلیت ناکافی جهت رسیدن به جمله^۱

$$\frac{h^2}{6} \left[\frac{\partial f}{\partial y}(t, y) \right]^2 f(t, y)$$

وجود دارد، که از بسط $(h^2/6)f''(t, y)$ نتیجه می‌شود. در نتیجه، بهترین روشی که می‌توان با استفاده از (38.6) بدست آورد روشهایی با کران خطای $O(h^2)$ است. اگرچه (38.6) دارای چهار پارامتر است، لیکن انعطافی در انتخاب آنهاست بطوری که تعدادی از روشهای $O(h^2)$ حاصل می‌شوند. دوتا از مهمترین آنها عبارتند از روش پیراسته^۱ اوپلر، که متناظر انتخاب $a_1 = a_2 = \frac{1}{2}$ و $\alpha_2 = \delta_2 = h$ بوده و دارای شکل معادله^۱ تفاضلی زیر:

روش پیراسته^۱ اوپلر:

$$w_0 = \alpha_0$$

بازای هر $i = 0, 1, 2, \dots, N-1$

$$(39.6)$$

$$w_{i+1} = w_i + \frac{h}{2} [f(t_i, w_i) + f(t_{i+1}, w_i + hf(t_i, w_i))]$$

و روش هیون^۱، که متناظر انتخاب $a_1 = \frac{1}{4}$ ، $a_2 = \frac{3}{4}$ و $\alpha_2 = \delta_2 = \frac{2}{3}h$ بوده و دارای شکل معادله^۱ تفاضلی زیر است:

روش هیون:

$$w_0 = \alpha_0$$

بازای هر $i = 0, 1, 2, \dots, N-1$

$$(40.6)$$

$$w_{i+1} = w_i + \frac{h}{4} [f(t_i, w_i) + 3f(t_i + \frac{2}{3}h, w_i + \frac{2}{3}hf(t_i, w_i))]$$

هر دو روش به‌عنوان روشهای رونگ - کوتا رده‌بندی شده‌اند.

مثال ۲. با استفاده از هر یک از روشهای رونگ - کوتای مرتبه^۱ دو در مثال همیشگی، یعنی

۱	مقادیر دقیق	روشن نقطه میانی	خطای روش	خطای روش	روشن پیکر	پیکر	روشن همون	خطای روش همون
0	1,000,000,000	1,000,000,000	0	0	1,000,000,000	0	1,000,000,000	0
1	1,000,325,252	1,000,250,000	7.52×10^{-3}	1.75×10^{-4}	1,000,500,000	1.75×10^{-4}	1,000,333,333	8.10×10^{-6}
2	1,002,538,855	1,002,426,300	1.12×10^{-4}	3.64×10^{-4}	1,002,902,255	3.64×10^{-4}	1,002,588,500	4.65×10^{-5}
3	1,008,363,636	1,008,245,800	1.18×10^{-4}	5.63×10^{-4}	1,008,926,800	5.63×10^{-4}	1,008,472,800	1.09×10^{-4}
4	1,019,935,999	1,019,262,400	9.75×10^{-5}	7.69×10^{-4}	1,020,128,800	7.69×10^{-4}	1,019,551,200	1.91×10^{-4}
5	1,036,938,877	1,036,882,200	5.62×10^{-5}	9.78×10^{-4}	1,037,916,600	9.78×10^{-4}	1,037,227,200	2.88×10^{-4}
6	1,062,376,700	1,062,378,700	2.00×10^{-6}	1.19×10^{-3}	1,063,564,500	1.19×10^{-3}	1,062,773,900	3.97×10^{-4}
7	1,096,829,400	1,096,902,700	7.33×10^{-5}	1.40×10^{-3}	1,098,225,900	1.40×10^{-3}	1,097,343,700	5.14×10^{-4}
8	1,141,342,100	1,141,496,900	1.55×10^{-4}	1.60×10^{-3}	1,142,944,400	1.60×10^{-3}	1,141,979,400	6.37×10^{-4}
9	1,196,860,700	1,197,104,700	2.44×10^{-4}	1.80×10^{-3}	1,198,664,700	1.80×10^{-3}	1,197,624,700	7.64×10^{-4}
10	1,264,241,100	1,264,579,800	3.39×10^{-4}	2.00×10^{-3}	1,266,241,600	2.00×10^{-3}	1,265,133,700	8.93×10^{-4}

جدول ۳۰۴

$$y' = -y + t + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

بخاطر سرشت این معادله دیفرانسیل، همان معادله تفاضلی حاصل از روش تیلور مرتبه دوزیر را می دهد:

$$w_0 = 1,$$

$$w_{i+1} = .905w_i + .0095i + .1$$

برای مقایسه نتایج مختلف این روشها، به جای مسئله مقدار اولیه قبلی،

$$y' = -y + t^2 + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

را بکار می بریم که $y(t) = -2e^{-t} + t^2 - 2t + 3$ جواب دقیق آن است. جدول ۳۰۶ در صفحه ۳۳۴ نتایج این محاسبات را فهرست کرده است.

گرچه $T^{(3)}(t, y)$ با خطای $O(h^3)$ را می توان با عبارتی به شکل

$$f(t + \alpha_1, y + \delta_1 f(t + \alpha_2, y + \delta_2 f(t, y)))$$

شامل چهار پارامتر تقریب کرد، لیکن جبر لازم برای تعیین $\alpha_1, \delta_1, \alpha_2, \delta_2$ کاملاً پیچیده است و ارائه نخواهد شد. در واقع، روش رونگ - کوتای مرتبه سه که از این عبارت نتیجه می شود عموماً "در عمل بکار نمی رود. در عوض، معمولترین روش رونگ - کوتا در عمل از مرتبه چهار و دارای شکل تفاضلی زیر است: به ازای هر $i = 0, 1, \dots, N - 1$

$$w_0 = \alpha,$$

$$k_1 = hf(t_i, w_i),$$

$$k_2 = hf\left(t_i + \frac{h}{2}, w_i + \frac{1}{2}k_1\right),$$

$$k_3 = hf\left(t_i + \frac{h}{2}, w_i + \frac{1}{2}k_2\right),$$

$$k_4 = hf(t_{i+1}, w_i + k_3),$$

$$w_{i+1} = w_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

این روش دارای خطای برشی موضعی $O(h^4)$ است، مشروط بر اینکه جواب $y(t)$ دارای پنج مشتق پیوسته باشد. دلیل معرفی نمادهای k_1, k_2, k_3, k_4 در این روش این است که نیاز به آشیانه ای بودن متوالی متغیر دوم تابع $f(t, y)$ را برطرف کند (ر.ک. تمرین ۱۰).

الگوریتم رونگ - کوتا (مرتبه چهار) ۲.۶

برای تقریب جواب مسئله مقدار اولیه

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

عدد صحیح مثبت N را انتخاب می‌کنیم.مرحله ۱. قرار می‌دهیم $h = (b - a)/N$ ، $t_0 = a$ ، و $w_0 = \alpha$.مرحله ۲. قرار می‌دهیم $i = 0$.

مرحله ۳. قرار می‌دهیم

$$k_1 = hf(t_i, w_i)$$

$$k_2 = hf\left(t_i + \frac{h}{2}, w_i + \frac{1}{2}k_1\right),$$

$$k_3 = hf\left(t_i + \frac{h}{2}, w_i + \frac{1}{2}k_2\right),$$

$$k_4 = hf(t_i + h, w_i + k_3),$$

$$w_{i+1} = w_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

$$t_{i+1} = a + (i + 1)h.$$

مرحله ۴. اگر $i = N - 1$ ، به مرحله ۶ می‌رویم.مرحله ۵. ۱ را به i افزوده و به مرحله ۳ می‌رویم.مرحله ۶. روند کامل است، و w_i ، به ازای هر $i = 1, 2, \dots, N$ ، $y(t_i)$ را تقریب می‌کند.

مثال ۳. استفاده از روش رونگ - کوتا مرتبه چهار در بدست آوردن تقریبهایی به

جواب مسئله مقدار اولیه

$$y' = -y + t + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

با $h = .1$ و $N = 10$ و $t_i = .i$ ، نتایج و خطاهای فهرست شده در جدول ۴.۶ را نتیجه

می‌دهد.

مهمترین عمل محاسباتی در بکار بردن روشهای رونگ - کوتا محاسبه f است. در روشهای مرتبه دوم خطای برشی $O(h^2)$ است، لیکن هزینه محاسبه دو مقدار تابعی در هر مرحله است. روش رونگ - کوتا مرتبه چهار نیاز به چهار محاسبه در هر مرحله دارد،

t_i	مقادیر حقیقی	مقادیر رونگ - کوتای مرتبه چهار	خطا
0	1.0000000000	1.0000000000	0
.1	1.0048374180	1.0048375000	8.200×10^{-8}
.2	1.0187307531	1.0187309014	1.483×10^{-7}
.3	1.0408182207	1.0408184220	2.013×10^{-7}
.4	1.0703200460	1.0703202889	2.429×10^{-7}
.5	1.1065306597	1.1065309344	2.747×10^{-7}
.6	1.1488116360	1.1488119344	2.984×10^{-7}
.7	1.1965853038	1.1965856187	3.149×10^{-7}
.8	1.2493289641	1.2493292897	3.256×10^{-7}
.9	1.3065696597	1.3065699912	3.315×10^{-7}
1.0	1.3678794412	1.3678797744	3.332×10^{-7}

جدول ۴.۶

ولی خطای برشی $O(h^4)$ است. بوجرا [۱۶] رابطه زیر را بین تعداد محاسبات در هر مرحله و مرتبه خطای برشی موضعی بیان کرده است.

تعداد محاسبه در هر مرحله	2	3	4	5	6	7	$n \geq 8$
بهترین امکان خطای برشی موضعی	$O(h^2)$	$O(h^3)$	$O(h^4)$	$O(h^4)$	$O(h^5)$	$O(h^6)$	$O(h^{n-2})$

در نتیجه، روشهای مرتبه پایین تر، با اندازه گام کوچکتر، با دادن ارجحیت به روشهای مرتبه بالاتر ممکن، که از اندازه گام بزرگتری استفاده می کنند، بکار می روند.

یک مقیاس که با آن روشهای مرتبه پایین رونگ - کوتا مقایسه می شوند را به صورت زیر شرح می دهیم: چون روش رونگ - کوتای مرتبه چهار در هر مرحله به چهار محاسبه نیاز دارد، اگر بخواهد برتر باشد، باید جوابهای دقیقتری نسبت به روش اوایلر با اندازه شبکه ای یک چهارم بدهد (که در آن منظور از اندازه شبکه ای تفاضل بین نقاط شبکه ای متوالی است). مشابهاً، اگر روش رونگ - کوتای مرتبه چهار بخواهد از روشهای

رونگ - کوتای مرتبه دو بهتر باشد، باید دقت بیشتری با اندازه گام h نسبت به روش مرتبه دو می با اندازه گام $h/2$ بدهد، زیرا روش مرتبه چهار به دو برابر محاسبه در هر مرحله نیاز دارد. یک بررسی از تفوق روش رونگ - کوتای مرتبه چهار داده شده در (۴۱.۶) با این مقیاس، در مثال زیر نشان داده می شود.

مثال ۴. برای مسئله

$$y' = -y + 1, \quad 0 \leq t \leq 1$$

$$y(0) = 0,$$

روش اویلر با $h = 0.025$ ، روش پیراسته اویلر معادله (۳۹.۶) با $h = 0.05$ ، و روش رونگ - کوتای مرتبه چهار معادله (۴۱.۶) با $h = 0.1$ در نقاط شبکه‌ای 1، 2، 3، 4، 5 مقایسه می شوند. نتایج در جدول ۵.۶ ارائه شده‌اند. در این مثال واضح است که روش مرتبه چهار بهتر است.

روش رونگ - کوتای روش پیراسته				
t	روش اویلر	آویلر	چهار	مقدار واقعی
	$h = 0.025$	$h = 0.05$	$h = 0.1$	
.1	.096312	.095123	.09516250	.095162582
.2	.183348	.181198	.18126910	.181269247
.3	.262001	.259085	.25918158	.259181779
.4	.333079	.329563	.32967971	.329679954
.5	.397312	.393337	.39346906	.393469340

جدول ۵.۶

مجموعه تمرینات ۴.۶

۱. با استفاده از روش نقطه میانی، روش پیراسته اویلر، و روش هیون با $h = 0.1$ ، جواب

$$y' = 1 + t \sin(ty), \quad 0 \leq t \leq 2$$

$$y(0) = 0.$$

را تقریب کنید. جوابهای خود را با نتایج حاصل از تمرین ۴، بخش ۳.۶، و تمرین

۳، بخش ۲.۶، مقایسه کنید.

۲. با استفاده از روش رونگ - کوتای مرتبه چهار با $h = .05$ ، جواب

$$y' = \frac{2}{t}y + t^2e^t, \quad 1 \leq t \leq 2$$

$$y(1) = 0.$$

را تقریب کنید. جوابهای خود را با نتایج حاصل از روش تیلور مرتبه چهار در تمرین ۲، بخش ۳.۶، مقایسه کنید. کدامیک از این دو روش مرتبه چهار به زمان بیشتری برای محاسبه و کدامیک به زمان اجرای بیشتر نیاز دارد؟
۳. در مسئله مقدار اولیه

$$y' = -y + t + 1, \quad 0 \leq t \leq 5$$

$$y(0) = 2;$$

با استفاده از روش رونگ - کوتای مرتبه چهار و با $h = .2$ ، $y(5)$ را تقریب و با تمرین ۳، بخش ۳.۶، مقایسه کنید.
۴. در مسئله مقدار اولیه

$$y' = 50t^2 - 50y + 2t, \quad 0 \leq t \leq 1$$

$$y(0) = \frac{1}{3},$$

با استفاده از روش رونگ - کوتای مرتبه چهار با $h = .1$ (T)؛ $h = .025$ (ب)؛ و $h = .01$ (پ)، جواب دقیق را بدست آورید. آیا نتایج با خطای برشی موضعی سازگارند؟

۵. در مسائل زیر، روش اویلر با $h = .025$ ، روش مرتبه دوم رونگ - کوتا با $h = .05$ ، و روش رونگ - کوتای مرتبه چهار با $h = .1$ را بکار برید:

$$y' = \frac{1}{t^2} - \frac{y}{t} - y^2, \quad 1 \leq t \leq 2 \quad (T)$$

$$y(1) = -1;$$

$$y' = t + y, \quad 0 \leq t \leq 2 \quad (ب)$$

$$y(0) = -1;$$

$$y' = -2y + 2t^2 + 2t, \quad 0 \leq t \leq 1 \quad (پ)$$

$$y(0) = 1.$$

- در هریک از این مسائل کدام روش تفوق دارد؟
 ۶. نشان دهید که روش نقطه میانی، روش پیراستهٔ اویلر، و روش همیون تقریبهای یکسانی از مسئلهٔ مقدار اولیهٔ

$$y' = -y + t + 1, \quad 0 \leq t \leq 1$$

$$y(0) = 1,$$

- به‌ازای هر انتخاب h ، می‌دهند. چرا؟
 ۷. مایعی به غلظت کم، نظیر آب، از بشکهای با سوراخ مستدیر به میزان

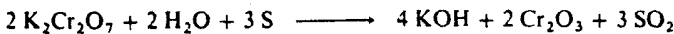
$$\frac{dx}{dt} = -6\pi r^2 \sqrt{2g} \frac{\sqrt{x}}{A(x)}$$

جریان دارد، که در آن r شعاع سوراخ، x ارتفاع سطح مایع بالای سوراخ، و $A(x)$ مساحت مقطع عرضی بشکه x واحد بالای سوراخ است. فرض کنید $r = 1$. فوت و $g = 32$ فوت بر مجذور ثانیه، و بشکه مخروطی شکل با راس رو به پایین و دارای 8 فوت ارتفاع آب اولیه و حجم $512 \pi/3$ فوت مکعب باشد.

(آ) $A(x)$ را بیابید.

(ب) با استفاده از روش رونگ-کوئای مرتبهٔ چهار با $h = 0.1$ ، سطح آب را پس از 10 ثانیه حساب کنید.

۸. واکنش شیمیایی برگشت ناپذیری که در آن دو مولکول دی کرومات پتاسیم جامد $(K_2Cr_2O_7)$ ، دو مولکول آب (H_2O) ، و سه اتم گوگرد جامد (S) ترکیب می‌شوند تا سه مولکول گاز دی‌اکسید گوگرد (SO_2) ، چهار مولکول تیدروکسید پتاسیم جامد (KOH) ، و دو مولکول اکسید کرمیک (Cr_2O_3) حاصل شوند را می‌توان با علامت به وسیلهٔ معادلهٔ شیمیایی زیر نمایش داد:



اگر n_1 مولکول $K_2Cr_2O_7$ ، n_2 مولکول H_2O ، و n_3 مولکول S در ابتدا موجود باشند، معادلهٔ دیفرانسیل زیر مقدار $x(t)$ از KOH پس از زمان t را بدست می‌دهد:

$$\frac{dx}{dt} = k \left(n_1 - \frac{x}{2} \right)^2 \left(n_2 - \frac{x}{2} \right)^2 \left(n_3 - \frac{3x}{4} \right)^3$$

که در آن k ثابت سرعت واکنش است. اگر $k = 6.22 \times 10^{-19}$ ، $n_1 = n_2 = 1000$ ، و $n_3 = 1500$ ، پس از دو ثانیه چند واحد تیدروکسید پتاسیم تشکیل می‌شود؟ روش

رونک - کوتای مرتبه چهار با $h = .1$ را بکار برید .

۹ . نشان دهید که روش تفاضلی

$$w_0 = \alpha,$$

$$w_{i+1} = w_i + a_1 f(t_i, w_i) + a_2 f(t_i + \alpha_2, w_i + \delta_2 f(t_i, w_i))$$

به ازای هر $i = 0, 1, \dots, N - 1$ و هر انتخاب از ثابتهای $a_1, a_2, \alpha_2, \delta_2$ ، نمی تواند دارای خطای برشی موضعی $O(h^3)$ باشد .

۱۰ . روش رونک - کوتای مرتبه چهار را می توان به شکل زیر نوشت :

$$w_0 = \alpha_0,$$

$$w_{i+1} = w_i + \frac{h}{6} f(t_i, w_i) + \frac{h}{3} f(t_i + \alpha_1 h, w_i + \delta_1 f(t_i, w_i))$$

$$+ \frac{h}{3} f(t_i + \alpha_2 h, w_i + \delta_2 h f(t_i + \gamma_2 h, w_i + \gamma_3 h f(t_i, w_i)))$$

$$+ \frac{h}{6} f(t_i + \alpha_3 h, w_i + \delta_3 h f(t_i + \gamma_4 h, w_i + \gamma_5 h f(t_i + \gamma_6 h, w_i$$

$$+ \gamma_7 h f(t_i, w_i))).$$

ثابتهای

$$\alpha_1, \alpha_2, \alpha_3, \delta_1, \delta_2, \delta_3, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7$$

را بدست آورید .

۵.۶ کنترل خطا و روش رونک - کوتا - فلیبرگ^۱

یک روش ایده آل معادله تفاضلی

$$(۴۲.۰۶) \quad w_{i+1} = w_i + h_i \phi(t_i, h_i, w_i), \quad i = 0, 1, \dots, N - 1$$

برای تقریب جواب $y(t)$ ، به مسئله مقدار اولیه

$$(۴۳.۰۶) \quad y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

دارای این خاصیت است که، اگر $\varepsilon > 0$ ، یعنی حداکثر خطای قابل تحمل، مفروض باشد،

باید کمترین تعداد نقاط شبکه‌ای بکار رود تا تضمین کند که خطای جامع، یعنی $|y(t_i) - w_i|$ ، به ازای هر $i = 0, 1, 2, \dots, N$ ، از ϵ تجاوز نمی‌کند. داشتن کمترین تعداد نقاط شبکه‌ای، و نیز کنترل خطای جامع یک روش تفاضلی، نه به طور غیر منتظره، با متساوی الفاصله بودن نقاط بازه ناسازگار است. در این بخش، با انتخاب مناسب نقاط شبکه‌ای، تکنیکی را جهت کنترل خطای یک روش تفاضلی، به طریقی موثر، بررسی می‌کنیم. چون عموماً "نمی‌توان خطای جامع یک روش تفاضلی را تعیین کرد، به‌جای آن با مفهومی موسوم به خطای موضعی یک روش در یک نقطه شبکه‌ای کار می‌کنیم. خطای موضعی برابر است با قدر مطلق تفاضل بین تقریب در آن نقطه و جواب درست مسئله مقدار اولیه، که با فرض تقریب در نقطه قبلی، به عنوان داده اولیه، بدست آمده است؛ یعنی، خطای موضعی عبارت است از

$$e_{i+1}(h) = |z(t_{i+1}) - w_{i+1}| \quad (44.6)$$

که در آن $z(t)$ جواب مسئله مقدار اولیه

$$z' = f(t, z), \quad t_i \leq t \leq b, \quad z(t_i) = w_i \quad (45.6)$$

و w_i جواب روش تفاضلی است.

در بخش ۹.۶ نتایجی ارائه می‌دهیم نشانگر آنکه، تحت مفروضاتی مناسب در مسئله مقدار اولیه و روش تفاضلی، خطای جامع روش به حاصل ضرب ثابتی در خطای قابل تحمل ϵ کراندار است، مشروط بر اینکه خطای موضعی به ϵh_i کراندار باشد، که در آن h_i اندازه گام وابسته به بازه $[t_i, t_{i+1}]$ است. لذا، منظور انتخاب اندازه گام در هر نقطه شبکه‌ای است، به بزرگی ممکن، که تضمین کند خطای موضعی به ϵ برابر اندازه گام کراندار است. توجه کنید که تفاوت متمایزی بین مفاهیم خطای موضعی، تعریف شده در بالا، و خطای برشی موضعی، تعریف شده در بخش ۲.۶، وجود دارد. گرچه بین این مفاهیم رابطه‌ای وجود دارد، لیکن خطای برشی موضعی جواب مسئله اصلی (۴۳.۶) را با تقریب آن در یک نقطه شبکه‌ای مقایسه می‌کند، در حالی که خطای موضعی جواب (۴۵.۶) را برای مقایسه بکار می‌برد.

برای توضیح تکنیک تخمین خطای موضعی، فرض کنیم تخمین خطای موضعی روش اولسر

$$w_0 = \alpha$$

$$w_{i+1} = w_i + hf(t_i, w_i) \quad (46.6)$$

مورد نظر باشد. چون روش اویلر دارای خطای برشی موضعی $O(h)$ است، روش دیگری، در این حالت روش پیراسته اویلر، را با خطای برشی موضعی $O(h^2)$ انتخاب می‌کنیم:

$$w_0 = \alpha, \\ (47.6) \quad \bar{w}_{i+1} = \bar{w}_i + \frac{h}{2} [f(t_i, \bar{w}_i) + f(t_{i+1}, \bar{w}_i + hf(t_i, \bar{w}_i))]$$

و فرض می‌کنیم هر دو روش در (t_i, w_i) یکی باشند؛ در نتیجه، (۴۷.۶) می‌شود

$$w_0 = \alpha, \\ (48.6) \quad \bar{w}_{i+1} = w_i + \frac{h}{2} [f(t_i, w_i) + f(t_{i+1}, w_i + hf(t_i, w_i))]$$

با نشان دادن خطای موضعی وابسته به روش اویلر در t_i با $e_i(h)$ ، یا (با فرض اینکه نقاط شبکه‌ای یکسان توزیع شده باشند) $e_i(h)$ ، داریم

$$e_{i+1}(h) = |z(t_{i+1}) - w_{i+1}|$$

چون خطای برشی موضعی روش اویلر برابر $O(h)$ است، باید انتظار داشت (ر.ک. تمرین ۲) که $e_{i+1}(h)$ برابر $O(h^2)$ باشد. مشابهاً، روش پیراسته اویلر دارای خطای برشی موضعی $O(h^2)$ است؛ در نتیجه،

$$z(t_{i+1}) = \bar{w}_{i+1} + O(h^3)$$

از اینرو،

$$z(t_{i+1}) - w_{i+1} = z(t_{i+1}) - \bar{w}_{i+1} + \bar{w}_{i+1} - w_{i+1} \\ = O(h^3) + \bar{w}_{i+1} - w_{i+1};$$

و قابل توجه‌ترین قسمت $e_{i+1}(h)$ باید به تفاضل بین \bar{w}_{i+1} و w_{i+1} اختصاص داشته باشد؛ یعنی،

$$e_{i+1}(h) \approx |\bar{w}_{i+1} - w_{i+1}| \\ = |w_i + \frac{h}{2} [f(t_i, w_i) + f(t_{i+1}, w_i + hf(t_i, w_i))] \\ - w_i - hf(t_i, w_i)| \\ = \frac{h}{2} |f(t_{i+1}, w_i + hf(t_i, w_i)) - f(t_i, w_i)|.$$

مثال ۱. خطای موضعی واقعی روش اویلر را با مقادیر تخمین زده بالا برای مسئله مقدار اولیه

$$y' = -y + t + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

با فرض اندازه گام $h = .1$ ، مقایسه می‌کنیم. جواب عمومی معادله دیفرانسیل

$$z' = -z + t + 1, \quad t_i \leq t \leq b$$

$$z(t_i) = w_i,$$

با $z(t) = t + (w_i - t_i)e^{t-t_i}$ داده می‌شود؛ در نتیجه، خطای موضعی واقعی برابر است با

$$e_{i+1}(h) = |z(t_{i+1}) - w_{i+1}| = |t_{i+1} + (w_i - t_i)e^{-h} - w_{i+1}|$$

در حالی که تخمین خطای موضعی به صورت زیر می‌باشد:

$$\frac{h}{2} |f(t_{i+1}, w_i + hf(t_i, w_i)) - f(t_i, w_i)| = \frac{h^2}{2} |w_i - t_i|$$

نتایج در جدول ۶.۶ نشان می‌دهند که در این مثال تخمین مناسب بوده است.

i	t_i	$z(t_i)$	w_i	$e_i(h)$	$\frac{h^2}{2} w_i - t_{i-1} $
0	0	—	1.000000	—	—
1	.1	1.004837	1.000000	4.837×10^{-3}	5.000×10^{-3}
2	.2	1.014354	1.010000	4.354×10^{-3}	4.500×10^{-3}
3	.3	1.032918	1.029000	3.918×10^{-3}	4.050×10^{-3}
4	.4	1.059626	1.056100	3.526×10^{-3}	3.645×10^{-3}
5	.5	1.093664	1.090490	3.174×10^{-3}	3.281×10^{-3}
6	.6	1.134297	1.131441	2.856×10^{-3}	2.952×10^{-3}
7	.7	1.180868	1.178297	2.571×10^{-3}	2.657×10^{-3}
8	.8	1.232781	1.230467	2.314×10^{-3}	2.391×10^{-3}
9	.9	1.289503	1.287420	2.083×10^{-3}	2.152×10^{-3}
10	1.0	1.350552	1.348678	1.874×10^{-3}	1.937×10^{-3}

جدول ۶.۶

حال مشخص می‌کنیم که چطور تخمین خطای موضعی یک روش تفاضلی می‌تواند در تقریب اندازه گام بهینه برای کنترل خطای جامع سودمند باشد. فرض کنیم دو روش

تفاضلی برای تقریب جواب مسئله مقدار اولیه (۴۳.۰۶) در دست باشند و یکی از روشها،

$$(49.06) \quad \begin{aligned} w_0 &= \alpha \\ w_{i+1} &= w_i + h_i \phi(t_i, h_i, w_i) \end{aligned}$$

دارای خطای برشی موضعی از مرتبه $O(h^n)$ باشد، درحالی که روش دیگر،

$$(50.06) \quad \begin{aligned} \tilde{w}_0 &= \alpha \\ \tilde{w}_{i+1} &= \tilde{w}_i + \tilde{h}_i \tilde{\phi}(\tilde{t}_i, \tilde{h}_i, \tilde{w}_i) \end{aligned}$$

دارای خطای برشی موضعی از مرتبه $O(h^{n+1})$ است.

اولین چیزی که برای تخمین خطای موضعی در (۴۹.۰۶) بکار می‌رود (۵۰.۰۶) است. برای این کار، فرض کنیم جوابها در (t_i, w_i) منطبق بوده و h انتخاب اولیه برای اندازه گام در t_i باشد. دراین صورت، همانطور که در حالت روشهای اویلر و روش پیراسته اویلر ملاحظه شد،

$$\begin{aligned} z_{i+1} - w_{i+1} &= z_{i+1} - \tilde{w}_{i+1} + \tilde{w}_{i+1} - w_{i+1} \\ &= O(h^{n+2}) + \tilde{w}_{i+1} - w_{i+1} \end{aligned}$$

$$(51.06) \quad e_{i+1}(h) = |z_{i+1} - w_{i+1}| = |\tilde{w}_{i+1} - w_{i+1}| + O(h^{n+2})$$

بعلاوه، $e_{i+1}(h)$ از مرتبه $O(h^{n+1})$ است؛ درنتیجه، طبق تعریف، ثابتی مانند

$k > 0$ ، مستقل از h ، با خاصیت

$$(52.06) \quad e_{i+1}(h) \leq kh^{n+1} + O(h^{n+2})$$

وجود دارد. از معادلات (۵۱.۰۶) و (۵۲.۰۶) نتیجه می‌شود که یک تخمین درست برای k را می‌توان با

$$kh^{n+1} = |\tilde{w}_{i+1} - w_{i+1}|$$

یا

$$(53.06) \quad k = \frac{|\tilde{w}_{i+1} - w_{i+1}|}{h^{n+1}}$$

ارائه داد.

چون طراحی این روند بخاطر کاربرد اندازه گام متغیر است، فرض کنید (۵۲.۰۶) را با تعویض h با متغیر Sh در نظر گرفته باشیم، که در آن S مثبت ولی از بالا کراندار و نیز دور از صفر کراندار باشد. دراین صورت،

$$\begin{aligned} e_{i+1}(Sh) &\leq k(Sh)^{n+1} + O((Sh)^{n+2}) \\ &= kS^{n+1}h^{n+1} + O(S^{n+2}h^{n+2}) \end{aligned}$$

و از کران‌داری S نتیجه می‌شود که

$$e_{i+1}(Sh) \leq kS^{n+1}h^{n+1} + O(h^{n+2})$$

با استفاده از تخمین (۵۳.۶) برای k بدست می‌آید

$$e_{i+1}(Sh) \leq |\tilde{w}_{i+1} - w_{i+1}|S^{n+1} + O(h^{n+2})$$

در نتیجه، قابل توجه‌ترین قسمت خطای موضعی $e_{i+1}(Sh)$ ، یعنی قسمتی که از مرتبه

$O(h^{n+1})$ است، به $|\tilde{w}_{i+1} - w_{i+1}|S^{n+1}$ کراندار است و می‌توان مطمئن بود که خطای

موضعی به ε برابر اندازه گام کراندار است؛ مشروط بر اینکه

$$|\tilde{w}_{i+1} - w_{i+1}|S^{n+1} \leq \varepsilon Sh$$

که نتیجه می‌دهد

$$(54.6) \quad S \leq \left(\frac{\varepsilon h}{|\tilde{w}_{i+1} - w_{i+1}|} \right)^{1/n}$$

یک تکنیک متداول که (۵۴.۶) را برای کنترل خطا بکار می‌برد روش رونگ - کوتا

- فلبرگ نامیده می‌شود، که به وسیله فلبرگ [۲۶] ارائه شده، و عبارت است از بکار

بستن روش رونگ - کوتا با خطای برشی موضعی مرتبه پنج،

$$\tilde{w}_{i+1} = w_i + \frac{16}{135} k_1 + \frac{6656}{12825} k_3 + \frac{28561}{56430} k_4 - \frac{9}{50} k_5 + \frac{2}{55} k_6$$

جهت تخمین خطای موضعی در روش رونگ - کوتای مرتبه چهار،

$$w_{i+1} = w_i + \frac{25}{216} k_1 + \frac{1408}{2565} k_3 + \frac{2197}{4104} k_4 - \frac{1}{5} k_5$$

که در آن

$$k_1 = hf(t_i, w_i),$$

$$k_2 = hf\left(t_i + \frac{h}{4}, w_i + \frac{1}{4} k_1\right),$$

$$k_3 = hf\left(t_i + \frac{3h}{8}, w_i + \frac{3}{32} k_1 + \frac{9}{32} k_2\right),$$

$$k_4 = hf\left(t_i + \frac{12h}{13}, w_i + \frac{1932}{2197} k_1 - \frac{7200}{2197} k_2 + \frac{7296}{2197} k_3\right),$$

$$k_5 = hf \left(t_i + h, w_i + \frac{439}{216} k_1 - 8k_2 + \frac{3680}{513} k_3 - \frac{845}{4104} k_4 \right)$$

$$k_6 = hf \left(t_i + \frac{h}{2}, w_i - \frac{8}{27} k_1 + 2k_2 - \frac{3544}{2565} k_3 \right. \\ \left. + \frac{1859}{4104} k_4 - \frac{11}{40} k_5 \right).$$

یک فایده روشن این روش آن است که در هر مرحله فقط شش محاسبه از f لازم است، در حالی که کاربرد توام روشهای رونگ - کوتای مرتبه چهار و پنج در هر گام ده محاسبه از f را نیاز دارد.

در نظریه کنترل خطا، یک مقدار اولیه h در مرحله i م جهت بدست آوردن اولین مقدار w_{i+1} و \tilde{w}_{i+1} بکار رفته که به تعیین S برای آن مرحله کشیده می شود، و سپس محاسبات تکرار می شوند. در عمل، این روند در هر مرحله به دو برابر تعداد محاسبات تابعی، که بدون کنترل خطا مورد احتیاج است، نیاز دارد؛ در نتیجه، مقدار حاصل برای S جهت ارزنده ساختن این خرج اضافی به انحاء مختلفی بکار می رود. مقدار S تعیین شده در مرحله i م را به دو منظور بکار می برند:

۱. برای طرد انتخاب اولیه h در مرحله i م و تکرار محاسبات با استفاده از Sh ،

و

۲. برای پیش بینی انتخاب اولیه مناسبی از h در مرحله $(i+1)$ م.

اگر تعدادی زیادی از مراحل تکرار شوند، به علت جریمه های (برحسب محاسبات تابعی) که باید پرداخت، انتخاب S نسبتاً "محافظة کارانه است؛ در واقع، برای روش رونگ - کوتا - فلبرگ با $n=4$ انتخاب معمولی به صورت زیر است:

$$S = \left(\frac{\epsilon h}{2|\tilde{w}_{i+1} - w_{i+1}|} \right)^{1/4} = .84 \left(\frac{\epsilon h}{|\tilde{w}_{i+1} - w_{i+1}|} \right)^{1/4}$$

الگوریتم ذیل روش رونگ - کوتا - فلبرگ با کنترل خطا را بکار می برد. در مرحله ۶، مانع تغییرات بزرگی در اندازه گام شده ایم تا از صرف وقت خیلی زیاد با اندازه گامهای بسیار کوچک در نواحی که مشتقهای y نامنظم است جلوگیری کنیم.

الگوریتم رونگ - کوتا - فلبرگ ۳.۶

برای تقریب جواب

$$y' = f(t, y), \quad a \leq t \leq b$$

$$y(a) = \alpha,$$

$\varepsilon > 0$ را حداکثر خطای قابل تحمل اختیار کرده و مراحل زیر را انجام می‌دهیم:

مرحله ۱. قرار می‌دهیم $w_0 = \alpha$ ، $t_0 = a$ ، و $h = \varepsilon^{1/4}$.

مرحله ۲. قرار می‌دهیم $i = 0$.

مرحله ۳. قرار می‌دهیم $t_{i+1} = t_i + h$.

مرحله ۴. اگر $t_{i+1} > b$ ، به مرحله ۸ می‌رویم.

مرحله ۵. قرار می‌دهیم

$$k_1 = hf(t_i, w_i),$$

$$k_2 = hf\left(t_i + \frac{1}{4}h, w_i + \frac{1}{4}k_1\right),$$

$$k_3 = hf\left(t_i + \frac{3}{8}h, w_i + \frac{3}{32}k_1 + \frac{9}{32}k_2\right),$$

$$k_4 = hf\left(t_i + \frac{12}{13}h, w_i + \frac{1932}{2197}k_1 - \frac{7200}{2197}k_2 + \frac{7296}{2197}k_3\right),$$

$$k_5 = hf\left(t_i + h, w_i + \frac{439}{216}k_1 - 8k_2 + \frac{3680}{513}k_3 - \frac{845}{4104}k_4\right),$$

$$k_6 = hf\left(t_i + \frac{1}{2}h, w_i - \frac{8}{27}k_1 + 2k_2 - \frac{3544}{2565}k_3\right.$$

$$\left. + \frac{1859}{4104}k_4 - \frac{11}{40}k_5\right),$$

$$w_{i+1} = w_i + \frac{25}{216}k_1 + \frac{1408}{2565}k_3 + \frac{2197}{4104}k_4 - \frac{1}{5}k_5,$$

$$\tilde{w}_{i+1} = w_i + \frac{16}{135}k_1 + \frac{6656}{12825}k_3 + \frac{28561}{56430}k_4 - \frac{9}{50}k_5 + \frac{2}{55}k_6$$

$$r_{i+1} = \frac{\tilde{w}_{i+1} - w_{i+1}}{h},$$

$$\delta = .84 \left(\frac{\varepsilon}{|r_{i+1}|} \right)^{1/4}.$$

مرحله ۶. اگر $\delta \leq .1$ ، h را جایگزین $.1h$ می‌کنیم.

اگر $\delta \geq 4$ ، h را جایگزین $4h$ می‌کنیم .

اگر $1 < \delta < 4$ ، h را جایگزین δh می‌کنیم .

مرحله ۷ . اگر $|r_{i+1}| > \varepsilon$ ، به مرحله ۳ (با طرد w_{i+1}) می‌رویم .

اگر $|r_{i+1}| \leq \varepsilon$ ، 1^* را به i افزوده و به مرحله ۳ (با قبول w_{i+1}) می‌رویم .

مرحله ۸ . اعداد w_i ، با حداکثر خطای موضعی نسبی ε ، $y(t_i)$ را تقریب می‌کنند ، و روند کامل خواهد بود .

مثال ۲ . برای تقریب جواب مسئله مقدار اولیه

$$y' = -y + t + 1, \quad 0 \leq y \leq 1, \quad y(0) = 1$$

با $\varepsilon = 10^{-8}$ ، الگوریتم ۳.۶ را بکار می‌بریم ، که در آن h با مقدار 2 (جهت نمایش روند طرد کردن) شروع می‌شود . جدول ۷.۶ (ص ۳۵۰) نتایج را با دقت هشت رقم نشان می‌دهد . توجه کنید که مقدار w_1 طرد شده و در نتیجه مجدداً " محاسبه شده است .

مجموعه تمرینات ۵.۶

۱ . با استفاده از الگوریتم رونگ - کوتا - فلیبرگ ۳.۶ ، جوابهای مسائل مقدار اولیه

زیر را تقریب کنید :

$$y' = 1 - y, \quad 0 \leq t \leq 1 \quad (A)$$

$y(0) = 0$ ؛ از $\varepsilon = 10^{-6}$ استفاده کنید ؛

$$y' = -y + t + 1, \quad 0 \leq t \leq 5 \quad (B)$$

$y(0) = 2$ ؛ از $\varepsilon = 10^{-4}$ استفاده کنید ؛

$$y' = \frac{2}{t}y + t^2 e^t, \quad 1 \leq t \leq 2 \quad (C)$$

$y(1) = 0$ ؛ از $\varepsilon = 10^{-6}$ استفاده کنید ؛

$$y' = 1 + y^2, \quad 0 \leq t \leq \frac{\pi}{2} \quad (D)$$

$y(0) = 0$ ؛ از $\varepsilon = 10^{-4}$ استفاده کنید ؛

$$y' = 2|t - 2|y, \quad 0 \leq t \leq 5 \quad (E)$$

$y(0) = e^{-4}$ ؛ از $\varepsilon = 10^{-4}$ استفاده کنید ؛

$$y' = 1 + t \sin(ty), \quad 0 \leq t \leq 2 \quad (F)$$

i	t_i	h_i	w_i	\hat{w}_i	r_i	S_i	$y(t_i)$	$ y(t_i) - w_i $
1	.020000000	.020000000	1.0187303	1.0187307	$0.22451282 \times 10^{-5}$.21798209	1.0187308	0.4967×10^{-6}
1	.043596418	.043596418	1.0009367	1.0009367	$0.47670754 \times 10^{-8}$	1.0141250	1.0009367	0.2114×10^{-9}
2	.087808635	.044212217	1.0037448	1.0037448	$0.47674017 \times 10^{-8}$	1.0109015	1.0037448	0.4195×10^{-9}
3	.132502283	.044694196	1.0084033	1.0084033	$0.47642404 \times 10^{-8}$	1.0110691	1.0084033	0.6206×10^{-9}
4	.17769175	.045188922	1.0148922	1.0148922	$0.47619606 \times 10^{-8}$	1.0111901	1.0148922	0.8151×10^{-9}
5	.22338635	.045694592	1.0231921	1.0231921	$0.47596204 \times 10^{-8}$	1.0113144	1.0231921	0.1003×10^{-8}
6	.26959794	.046211599	1.0332844	1.0332844	$0.47572455 \times 10^{-8}$	1.0114406	1.0332844	0.1185×10^{-8}
7	.31633823	.046740288	1.0451511	1.0451511	$0.47547942 \times 10^{-8}$	1.0115709	1.0451511	0.1360×10^{-8}
8	.36361955	.047281116	1.0587751	1.0587751	$0.47522998 \times 10^{-8}$	1.0117036	1.0587751	0.1529×10^{-8}
9	.41145383	.047834478	1.0741399	1.0741399	$0.47497640 \times 10^{-8}$	1.0118387	1.0741399	0.1692×10^{-8}
10	.45985460	.048400773	1.0912300	1.0912300	$0.47471644 \times 10^{-8}$	1.0119771	1.0912300	0.1850×10^{-8}
11	.50883508	.048980477	1.1100306	1.1100306	$0.47444912 \times 10^{-8}$	1.0121197	1.1100306	0.2002×10^{-8}
12	.55840918	.049574103	1.1305277	1.1305277	$0.47417671 \times 10^{-8}$	1.0122650	1.1305277	0.2148×10^{-8}
13	.60859131	.050182129	1.1527081	1.1527081	$0.47389777 \times 10^{-8}$	1.0124139	1.1527081	0.2289×10^{-8}
14	.65939640	.050805086	1.1765598	1.1765598	$0.47361088 \times 10^{-8}$	1.0125672	1.1765598	0.2424×10^{-8}
15	.71083996	.051443564	1.2020714	1.2020714	$0.47332021 \times 10^{-8}$	1.0127226	1.2020714	0.2555×10^{-8}
16	.76293802	.052098061	1.2292324	1.2292324	$0.47302120 \times 10^{-8}$	1.0128826	1.2292324	0.2681×10^{-8}
17	.81570724	.052769220	1.2580336	1.2580336	$0.47271416 \times 10^{-8}$	1.0130471	1.2580336	0.2801×10^{-8}
18	.86916494	.053457703	1.2884665	1.2884665	$0.47239922 \times 10^{-8}$	1.0132159	1.2884665	0.2917×10^{-8}
19	.92332914	.054164193	1.3205237	1.3205237	$0.47207759 \times 10^{-8}$	1.0133884	1.3205237	0.3029×10^{-8}
20	.97821850	.054889365	1.3541988	1.3541988	$0.47174729 \times 10^{-8}$	1.0135657	1.3541988	0.3136×10^{-8}

جدول ۷.۶

$y(0) = 0$ ؛ از $\varepsilon = 10^{-4}$ استفاده کنید؛

(چ) $y' = 50t^2 - 50y + 2t$ ، $0 \leq t \leq 1$ ؛

$y(0) = \frac{1}{3}$ ؛ از $\varepsilon = 10^{-3}$ استفاده کنید؛

(ح) $y' = -2y + 2t^2 + 2t$ ، $0 \leq t \leq 1$ ؛

$y(0) = 1$ ؛ از $\varepsilon = 10^{-6}$ استفاده کنید.

۲. نشان دهید که در روش اویلر $e_{i+1}(h)$ برابر $O(h^2)$ است.
راهنمایی.

$$\tau_{i+1} = \frac{z(t_{i+1}) - z(t_i)}{h} - f(t_i, w_i)$$

چون $z(t_i) = w_i$ ، با استفاده از تعریف w_{i+1} نشان دهید که

$$z(t_{i+1}) = w_{i+1} + h\tau_{i+1}.$$

۳. یک الگوریتم قدیمی تر از فلبرگ [۲۴]، با استفاده از روش مرتبه چهار

$$\tilde{w}_{i+1} = w_i + \frac{2}{15}k_2 + \frac{7}{90}k_3 + \frac{16}{45}k_4 + \frac{1}{45}k_5$$

خطای موضعی در روش مرتبه سوم

$$w_{i+1} = w_i + \frac{2}{3}k_2 + \frac{1}{6}k_3$$

را تخمین می‌زند، که در آن

$$k_1 = hf(t_{i+1}, w_i),$$

$$k_2 = hf\left(t_i + \frac{h}{2}, w_i + \frac{1}{8}k_1\right),$$

$$k_3 = hf\left(t_{i+1}, w_i - \frac{1}{2}k_1 + 2k_2\right)$$

$$k_4 = hf\left(t_i + \frac{h}{4}, w_i - \frac{5}{64}k_1 + \frac{5}{16}k_2 - \frac{3}{64}k_3\right),$$

$$k_5 = hf\left(t_i + \frac{3h}{4}, w_i + \frac{9}{64}k_1 - \frac{3}{16}k_2 + \frac{7}{64}k_3 + \frac{1}{2}k_4\right)$$

با استفاده از این روشها، الگوریتمی شبیه الگوریتم ۳.۶ بسازید.

۴. تمرین ۱ را، با استفاده از تمرین ۳، تکرار و نتایج را مقایسه کنید.

۵. در نظریه انتشار امراض مسری (ر.ک. بیلی^۱ [۵] یا [۶])، به شرط آنکه مفروضات

ساده سازی مناسبی در نظر گرفته شود، می توان یک معادله دیفرانسیل نسبتاً "مقدماتی برای پیشگویی تعداد افراد مبتلا در جمعیت در هر زمان بکار می رود. بخصوص، فرض کنیم امکان ابتلاء در تمامی افراد جمعیت ثابتی یکسان بوده و، هنگامی که فردی مبتلا شد، در همان حال بماند. اگر فرض کنیم $x(t)$ تعداد افراد مستعد پذیرش بیماری در لحظه t ، و $y(t)$ تعداد بیمارها باشد، معقول است فرض شود که میزان تغییر تعداد بیمارها متناسب با حاصل ضرب $x(t)$ و $y(t)$ باشد، زیرا این میزان به هر دو عامل تعداد بیمارها و تعداد افراد مستعد پذیرش بیماری در آن لحظه بستگی دارد. اگر جمعیت آنقدر انبوه باشد که بتوان $x(t)$ و $y(t)$ را متغیرهای پیوسته گرفت، مسئله به صورت

$$\frac{dy}{dt}(t) = kx(t)y(t)$$

بیان می شود، که در آن k ثابت بوده و (تعداد کل جمعیت) $x(t) + y(t) = m$ این معادله را می توان مجدداً "و فقط برحسب $y(t)$ به صورت زیر نوشت:

$$\frac{dy}{dt}(t) = ky(t)(m - y(t))$$

(آ) با فرض $k = 2 \times 10^{-6}$ ، $y(0) = 1000$ ، $m = 100,000$ و اینکه زمان برحسب روز باشد، و با استفاده از الگوریتم 3.6 و $\epsilon = 1$ ، تقریبی برای تعداد افراد بیمار در پایان روز سی ام پیدا کنید.

(ب) معادله دیفرانسیل قسمت (آ) را معادله برنولی^۱ نامیده و می توان، با قرار دادن $z(t) = (y(t))^{-1}$ ، به یک معادله دیفرانسیل خطی برحسب $z(t)$ تبدیل کرد. با استفاده از این تکنیک، و تحت مفروضات قسمت (آ)، جواب دقیق این معادله را بدست آورید، مقدار واقعی $y(t)$ را با تقریب آن در قسمت (آ) مقایسه کنید. $\lim_{t \rightarrow \infty} y(t)$ چقدر است؟ آیا این مقدار با درک شهودی شما مطابقت دارد؟ در تمرین قبل، همه افراد بیمار جهت انتشار بیماری در جمعیت باقی می ماندند. یک طرح واقع بینانه معرفی متغیر سوم $z(t)$ جهت نمایش تعداد افرادی است که از جمعیت مبتلا شده در لحظه t ، با منزوی کردن، بهبودی و مصونیت حاصله، یا

با مرگ، حذف شده‌اند. این مطلب، به‌طور طبیعی، مسئله را بهتر می‌کند؛ اما می‌توان نشان داد (ر.ک. بیلی [۶]) که یک جواب تقریبی به صورت

$$x(t) = x(0)e^{-(k_1/k_2)t}$$

و

$$y(t) = m - x(t) - z(t)$$

ارائه می‌شود، که در آن k_1 میزان ابتلا، k_2 میزان حذف، و $z(t)$ از معادله دیفرانسیل

$$\frac{dz}{dt}(t) = k_2(m - z(t) - x(0)e^{-(k_1/k_2)t})$$

تعیین می‌شود.

مولفین از هیچ روشی برای حل مستقیم این مسئله آگاه نیستند؛ در نتیجه، باید روند عددی بکار برد. با استفاده از الگوریتم ۳.۶ و فرض

$$m = 100,000, x(0) = 99,000, k_1 = 2 \times 10^{-6}, k_2 = 10^{-4}$$

تقریبی برای $x(30)$ ، $y(30)$ ، و $z(30)$ با اختیار $\varepsilon = 1$ بدست آورید.

۶.۶ روشهای چندگامی

روشهای بحث شده قبلی در این فصل روشهای تک‌گامی نامیده می‌شوند، زیرا تقریب نقطه شبکه‌ای t_{i+1} فقط مستلزم اطلاع از یکی از نقاط شبکه‌ای قبلی، یعنی t_i ، است. گرچه معمولاً این روشها اطلاعات محاسبات تابعی را در نقاط بین t_i و t_{i+1} بکار می‌برند، لیکن این مفروضات را برای استفاده مستقیم در تقریبهای آتی حفظ نمی‌کنند. در نتیجه، همه اطلاعاتی که به‌وسیله این روشها بکار می‌روند در طول بازه‌ای که تقریب جواب روی آن صورت می‌گیرد بدست می‌آیند.

چون جواب تقریبی در هر یک از نقاط شبکه‌ای t_0, t_1, \dots, t_i ، پیش از اینکه تقریب در t_{i+1} حاصل شود، در دسترس است، و چون خطای $|w_j - y(t_j)|$ با افزایش j زیاد می‌شود، ظاهراً "منطقی" است که روشهایی را که می‌توانند داده‌های دقیقتر قبلی را در تقریب جواب در t_{i+1} بکار ببرند بسط دهیم. روشهایی از این نوع که تقریب در بیشتر از یک نقطه شبکه‌ای قبلی را جهت تعیین تقریب در نقطه بعدی بکار می‌گیرند روشهای چندگامی نامیده می‌شوند. تعریف دقیق این روشها، همراه با تعریف دو نوع از روشهای

چندگامی، در زیر می آیند.

تعریف ۱۴.۶. یک روش چندگامی برای حل مسئله مقدار اولیه

$$(55.6) \quad y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

روشی است که معادله تفاضلی آن برای یافتن تقریب w_{i+1} در نقطه شبکه‌ای t_{i+1} را بتوان با معادله زیر، که در آن m عدد صحیح بزرگتر از ۱ است، نمایش داد:

$$(56.6) \quad \begin{aligned} w_{i+1} = & a_{m-1}w_i + a_{m-2}w_{i-1} + \dots + a_0w_{i+1-m} \\ & + h[b_m f(t_{i+1}, w_{i+1}) + b_{m-1}f(t_i, w_i) \\ & + \dots + b_0 f(t_{i+1-m}, w_{i+1-m})] \end{aligned}$$

بمازای $i = m-1, m, \dots, N-1$ ، که در آن مقادیر آغازی

$$(57.6) \quad w_0 = \alpha_0, w_1 = \alpha_1, w_2 = \alpha_2, \dots, w_{m-1} = \alpha_{m-1}$$

مشخص شده‌اند و، طبق معمول، $h = (b-a)/N$.

وقتی $b_m = 0$ ، این روش را روش صریح یا باز می‌نامیم، و معادله (۵۶.۶) w_{i+1}

را صریحاً "برحسب مقادیر تعیین شده قبلی بدست می‌دهد. وقتی $b_m \neq 0$ ، روش را روش ضمنی یا بسته می‌نامیم، زیرا w_{i+1} در هر دو طرف (۵۶.۶) ظاهر و فقط با روشی ضمنی تعیین می‌شود.

مثال ۱. معادلات

$$w_0 = \alpha_0, \quad w_1 = \alpha_1, \quad w_2 = \alpha_2, \quad w_3 = \alpha_3,$$

$$(58.6) \quad \begin{aligned} w_{i+1} = & w_i + \frac{h}{24} [55f(t_i, w_i) - 59f(t_{i-1}, w_{i-1}) \\ & + 37f(t_{i-2}, w_{i-2}) - 9f(t_{i-3}, w_{i-3})], \end{aligned}$$

بمازای هر $i = 3, 4, \dots, N-1$ ، یک روش صریح چهارگامی را تعریف می‌کند، که به تکنیک آدامز^۱ - بشفورث^۲ مرتبه چهار معروف است، درحالی که معادلات

$$w_0 = \alpha_0, \quad w_1 = \alpha_1, \quad w_2 = \alpha_2,$$

$$(59.6) \quad w_{i+1} = w_i + \frac{h}{24} [9f(t_{i+1}, w_{i+1}) + 19f(t_i, w_i)]$$

$$-5f(t_{i-1}, w_{i-1}) + f(t_{i-2}, w_{i-2})],$$

بمازای هر $i = 2, 3, \dots, N - 1$ یک روش ضمنی سه‌گامی را تعریف می‌کنند، که به تکنیک آدامز-مولتون مرتبه چهار موسوم است.

مقادیر آغازی در (۵۸.۰۶) یا (۵۹.۰۶) را باید معمولاً با فرض $\alpha_0 = \alpha$ و تولید کردن سایر مقادیر، با یک روش رونگ-کوتا یا با تکنیک تک‌گامی دیگر، بدست آورد. برای بکار بردن مستقیم یک روش ضمنی نظیر (۵۹.۰۶)، باید معادله ضمنی را نسبت به w_{i+1} حل کرد. همانطور که بعداً در این بخش خواهید دید، روشن نیست که در حالت کلی آیا این کار شدنی است، یا همیشه جواب منحصر بفردی برای w_{i+1} بدست می‌آید. یک روند درگیر در بدست آوردن روشهای چندگامی این است که ابتدا توجه شود که جواب مسئله مقدار اولیه (۵۵.۰۶)، اگر روی بازه $[t_i, t_{i+1}]$ انتگرالگیری شود، دارای این خاصیت است که

$$(۶۰.۰۶) \quad y(t_{i+1}) - y(t_i) = \int_{t_i}^{t_{i+1}} y'(t) dt = \int_{t_i}^{t_{i+1}} f(t, y(t)) dt$$

در نتیجه،

$$(۶۱.۰۶) \quad y(t_{i+1}) = y(t_i) + \int_{t_i}^{t_{i+1}} f(t, y(t)) dt$$

چون بدون شناختن $y(t)$ ، یعنی، جواب مسئله، نمی‌توان از $f(t, y(t))$ انتگرال گرفت، به‌جای آن از چند جمله‌ای درونیاب P که به وسیله نقاط مفروض از پیش تعیین شده $(t_0, w_0), (t_1, w_1), \dots, (t_i, w_i)$ بدست می‌آید انتگرال می‌گیریم. بعلاوه اگر فرض کنیم $y(t_i) = w_i$ ، معادله (۶۱.۰۶) می‌شود:

$$(۶۲.۰۶) \quad y(t_{i+1}) \approx w_i + \int_{t_i}^{t_{i+1}} P(t) dt$$

گرچه برای این منظور هر نوع چندجمله‌ای درونیاب را می‌توان بکار برد، در عمل چند جمله‌ای به‌نام چند جمله‌ای درونیاب پسر و نیوتن، که در بخش ۴.۳ بررسی شد، مفیدتر است. چون مشتقگیری بر پایه این چندجمله‌ای کاملاً "پر در دست راست، به‌جای آن نشان می‌دهیم چگونه یک روش دوگامی ساده را می‌توان با استفاده از یک چندجمله‌ای درونیاب لاگرانژ تعیین کرد، و، در مورد روشهای مرتبه بالاتر با استفاده از چندجمله‌ای پسر و

نیوتن، خواننده را به هنریچی [۳۳]، صفحات ۲۸۳ - ۲۷۶ ارجاع می‌دهیم. تکنیک لازم خیلی شبیه به انتگرالگیری عددی مورد بحث در فصل ۵ است. اختلاف اساسی در این است که چند جمله‌ای درونیاب برای انتگرالگیری عددی، اصولاً، نقاط را در طول بازه برای درونیابی خود بکار می‌برد، درحالی که، در این حالت، اغلب نقاط خارج بازه انتگرالگیری گرفته می‌شوند.

مثال ۲. برای بدست آوردن یک روش صریح دوگامی، ابتدا در

$$(۶۳.۰۶) \quad y(t_{i+1}) = y(t_i) + \int_{t_i}^{t_{i+1}} f(t, y(t)) dt$$

$f(t, y(t))$ را با چند جمله‌ای درونیاب لاگرانژ مرتبه یک، از نقاط $(t_{i-1}, y(t_{i-1}))$ و $(t_i, y(t_i))$ ، عوض می‌کنیم،

$$\begin{aligned} P(t) &= \frac{(t - t_i)}{(t_{i-1} - t_i)} f(t_{i-1}, y(t_{i-1})) + \frac{(t - t_{i-1})}{(t_i - t_{i-1})} f(t_i, y(t_i)) \\ &= \frac{f(t_{i-1}, y(t_{i-1}))}{-h} (t - t_i) + \frac{f(t_i, y(t_i))}{h} (t - t_{i-1}), \end{aligned}$$

و جمله خطای آن

$$R(t) = \frac{(t - t_{i-1})(t - t_i)}{2} f''(\xi_{i+1}, y(\xi_{i+1}))$$

که در آن $t_{i-1} < \xi_{i+1} < t_{i+1}$. در این صورت، معادله (۶۳.۰۶) می‌شود

$$\begin{aligned} y(t_{i+1}) &= y(t_i) + \int_{t_i}^{t_{i+1}} P(t) dt + \int_{t_i}^{t_{i+1}} R(t) dt \\ &= y(t_i) - \frac{f(t_{i-1}, y(t_{i-1}))}{h} \int_{t_i}^{t_{i+1}} (t - t_i) dt \\ &\quad + \frac{f(t_i, y(t_i))}{h} \int_{t_i}^{t_{i+1}} (t - t_{i-1}) dt \\ &\quad + \frac{1}{2} \int_{t_i}^{t_{i+1}} (t - t_i)(t - t_{i-1}) f''(\xi_{i+1}, y(\xi_{i+1})) dt, \end{aligned}$$

$$y(t_{i+1}) = y(t_i) - \frac{h}{2} f(t_{i-1}, y(t_{i-1})) + \frac{3}{2} hf(t_i, y(t_i)) \\ (۶۴.۰۶) \quad + \frac{1}{2} \int_{t_i}^{t_{i+1}} (t - t_i)(t - t_{i-1}) f''(\xi_{i+1}, y(\xi_{i+1})) dt$$

چون $(t - t_i)(t - t_{i-1})$ بر بازه $[t_i, t_{i+1}]$ تغییر علامت نمی دهد، قضیه مقدار میانگین برای انتگرالها (قضیه ۱۰.۰۱، صفحه ۹) تضمین می کند که عدد μ_{i+1} ، $t_i < \mu_{i+1} < t_{i+1}$ ، با خاصیت

$$\int_{t_i}^{t_{i+1}} (t - t_{i-1})(t - t_i) f''(\xi_{i+1}, y(\xi_{i+1})) dt \\ = f''(\mu_{i+1}, y(\mu_{i+1})) \int_{t_i}^{t_{i+1}} (t - t_{i-1})(t - t_i) dt \\ = \frac{5}{6} h^3 f''(\mu_{i+1}, y(\mu_{i+1}));$$

وجود دارد؛ و این مطلب که $y' = f(t, y)$ اجازه می دهد (۶۴.۰۶) را به صورت زیر بنویسیم:

$$y(t_{i+1}) = y(t_i) - \frac{h}{2} f(t_{i-1}, y(t_{i-1})) + \frac{3}{2} hf(t_i, y(t_i)) \\ (۶۵.۰۶) \quad + \frac{5}{12} y'''(\mu_{i+1}) h^3.$$

روش تفاضلی متناظر

$$w_0 = \alpha_0, \quad w_1 = \alpha_1,$$

$$(۶۶.۰۶) \quad w_{i+1} = w_i - \frac{h}{2} f(t_{i-1}, w_{i-1}) + \frac{3}{2} hf(t_i, w_i), \quad i = 1, 2, \dots, N-1$$

روش دوگامی آدامز - بشفورث نام دارد.

مثال ۲ یک نمونه از طرق بدست آوردن روشهای چندگامی را نشان می دهد. روندی دیگر مبتنی بر سریهای تیلور در تمرین ۶ ملاحظه می شود.

خطای برشی موضعی برای یک روش چندگامی را می توان به طریقی مشابه روشهای تک گامی تعریف کرد تا تفاوت جواب معادله دیفرانسیل با معادله تفاضلی را بدست دهد. توجه کنید که این تعریف نتیجه می دهد که خطای برشی موضعی روش دوگامی

آدامز - بشفورث داده شده در مثال ۲، در صورتی که $y \in C^3[a, b]$ ، از مرتبه دو است.

تعریف ۱۵.۶. اگر $y(t)$ جواب مسئله مقدار اولیه

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

بوده، و

$$\begin{aligned} w_{i+1} = & a_m w_i + a_{m-1} w_{i-1} + \dots + a_0 w_{i-m} \\ & + h[b_{m+1} f(t_{i+1}, w_{i+1}) + b_m f(t_i, w_i) \\ & + \dots + b_0 f(t_{i-m}, w_{i-m})] \end{aligned}$$

مرحله $(i+1)$ م در یک روش چندگامی باشد، خطای برشی موضعی در این مرحله، یعنی τ_{i+1} ، به ازای $i = m, m+1, \dots, N-1$ ، برابر است با

$$\begin{aligned} \tau_{i+1} = & \frac{y(t_{i+1}) - a_m y(t_i) - \dots - a_0 y(t_{i-m})}{h} \\ & - [b_{m+1} f(t_{i+1}, y(t_{i+1})) + \dots + b_0 f(t_{i-m}, y(t_{i-m}))]. \end{aligned}$$

بعضی از روشهای چندگامی صریح همراه با مقادیر آغازی و خطاهای برشی موضعی آنها در زیر ارائه می شوند. بدست آوردن این تکنیکها، بجز استفاده از چند جمله‌ای درونیاب پسر و نیوتن، شبیه روند موجود در مثال ۲ است. روش سه گامی آدامز - بشفورث:

$$w_0 = \alpha, \quad w_1 = \alpha_1, \quad w_2 = \alpha_2.$$

(۶۷.۶)

$$w_{i+1} = w_i + \frac{h}{12} [23f(t_i, w_i) - 16f(t_{i-1}, w_{i-1}) + 5f(t_{i-2}, w_{i-2})]$$

که در آن $i = 2, 3, \dots, N-1$ و خطای برشی موضعی برابر است با

$$\tau_{i+1} = \frac{3}{8} y^4(\mu_i) h^3$$

روش چهارگامی آدامز - بشفورث:

$$w_0 = \alpha, \quad w_1 = \alpha_1, \quad w_2 = \alpha_2, \quad w_3 = \alpha_3,$$

$$(۶۸.۶) \quad w_{i+1} = w_i + \frac{h}{24} [55f(t_i, w_i) - 59f(t_{i-1}, w_{i-1}) + 37f(t_{i-2}, w_{i-2})$$

$$- 9f(t_{i-3}, w_{i-3})],$$

که در آن $i = 3, 4, \dots, N - 1$ و خطای برشی موضعی برابر است با

$$\tau_{i+1} = \frac{251}{720} y^{(5)}(\mu_i) h^4$$

روش پنج گامی آدامز - بشفورث :

$$w_0 = \alpha, \quad w_1 = \alpha_1, \quad w_2 = \alpha_2, \quad w_3 = \alpha_3, \quad w_4 = \alpha_4$$

$$(۶۹.۰۶) \quad w_{i+1} = w_i + \frac{h}{720} [1901f(t_i, w_i) - 2774f(t_{i-1}, w_{i-1}) \\ + 2616f(t_{i-2}, w_{i-2}) - 1274f(t_{i-3}, w_{i-3}) \\ + 251f(t_{i-4}, w_{i-4})],$$

که در آن $i = 4, 5, \dots, N - 1$ و خطای برشی موضعی برابر است با

$$\tau_{i+1} = \frac{95}{288} y^{(6)}(\mu_i) h^5.$$

اگر نقطه $(t_{i+1}, f(t_{i+1}, y(t_{i+1})))$ نیز به عنوان یک گره درونیاب در تقریب انتگرال

$$\int_{t_i}^{t_{i+1}} f(t, y(t)) dt$$

بکار رود، روشهای ضمنی حاصل می شوند (ر. ک. تمرین ۷). بعضی از روشهای ضمنی معمولتر در زیر آمده اند.

روش دوگامی آدامز - مولتون :

$$w_0 = \alpha_0, \quad w_1 = \alpha_1,$$

$$(۷۰.۰۶) \quad w_{i+1} = w_i + \frac{h}{12} [5f(t_{i+1}, w_{i+1}) + 8f(t_i, w_i) - f(t_{i-1}, w_{i-1})]$$

که در آن $i = 1, 2, \dots, N - 1$ و خطای برشی موضعی برابر است با

$$\tau_{i+1} = -\frac{1}{24} y^{(4)}(\mu_i) h^3$$

روش سهگامی آدامز - مولتون :

$$w_0 = \alpha_0, \quad w_1 = \alpha_1, \quad w_2 = \alpha_2,$$

$$(۷۱.۰۶) \quad w_{i+1} = w_i + \frac{h}{24} [9f(t_{i+1}, w_{i+1}) + 19f(t_i, w_i) - 5f(t_{i-1}, w_{i-1}) \\ + f(t_{i-2}, w_{i-2})],$$

که در آن $i = 2, 3, \dots, N - 1$ و خطای برشی موضعی برابر است با

$$\tau_{i+1} = -\frac{19}{720}y^{(5)}(\mu_i)h^4.$$

روش چهارگامی آدامز - مولتون:

$$w_0 = \alpha_0, \quad w_1 = \alpha_1, \quad w_2 = \alpha_2, \quad w_3 = \alpha_3,$$

$$w_{i+1} = w_i + \frac{h}{720} [251f(t_{i+1}, w_{i+1}) + 646f(t_i, w_i)$$

(۷۲.۶)

$$- 264f(t_{i-1}, w_{i-1}) + 106f(t_{i-2}, w_{i-2})$$

$$- 19f(t_{i-3}, w_{i-3})],$$

که در آن $i = 3, 4, \dots, N - 1$ و خطای برشی موضعی برابر است با

$$\tau_{i+1} = -\frac{3}{160}y^{(6)}(\mu_i)h^5$$

مقایسه یک روش صریح m گامی آدامز - بشفورث با یک روش ضمنی $(m-1)$ گامی آدامز - مولتون جالب است. در هر مرحله هر دو به m محاسبه f نیاز دارند، و هر دو دارای جملات $y^{(m+1)}(\mu_i)h^m$ در خطاهای برشی موضعی خود هستند. در حالت کلی، ضرایب جملات درگیر با f و ضرایب موجود در خطای برشی موضعی در روشهای آدامز مولتون کوچکتراند. این مطلب به پایداری بیشتر و خطاهای گرد کردن کوچکتری در روشهای ضمنی منجر می شود. این امر در مثال بعد توضیح داده می شود.

مثال ۳. مسئله مقدار اولیه

$$(۷۳.۶) \quad y' = -y + t + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

را در نظر می گیریم. تقریبهایی که داده خواهند شد با روش چهارگامی آدامز - بشفورث، یعنی معادله (۶۸.۶)، و با روش سه گامی آدامز - مولتون، یعنی معادله (۷۱.۶)، بوده و در هر دوازده $h = 0.1$ و مقادیر جواب دقیق $y(t) = e^{-t} + t$ به عنوان مقادیر آغازی آنها استفاده می کنند.

روش آدامز - بشفورث دارای معادله تفاضلی

$$w_{i+1} = w_i + \frac{h}{24} [55f(t_i, w_i) - 59f(t_{i-1}, w_{i-1}) + 37f(t_{i-2}, w_{i-2})$$

$$- 9f(t_{i-3}, w_{i-3})],$$

به‌ازای $i = 3, 4, \dots, 9$ ، است که، در صورت ساده شدن با استفاده از معادله (۷۳.۰۶) و $h = .1$ و $t_i = .1i$ ، می‌شود

$$w_{i+1} = \frac{1}{24}[18.5w_i + 5.9w_{i-1} - 3.7w_{i-2} + .9w_{i-3} + .24i + 2.52]$$

به‌ازای $i = 3, 4, \dots, 9$.

روش آدامز - مولتون دارای معادله تفاضلی

$$w_{i+1} = w_i + \frac{h}{24} [9f(t_{i+1}, w_{i+1}) + 19f(t_i, w_i) - 5f(t_{i-1}, w_{i-1}) + f(t_{i-2}, w_{i-2})],$$

به‌ازای $i = 2, 3, \dots, 9$ ، است که به

$$w_{i+1} = \frac{1}{24}[-.9w_{i+1} + 22.1w_i + .5w_{i-1} - .1w_{i-2} + .24i + 2.52]$$

به‌ازای $i = 2, 3, \dots, 9$ ، ساده می‌شود. معیناً، برای استفاده از این روش می‌توان آن را صریحاً "نسبت به w_{i+1} حل کرد، که می‌دهد

$$w_{i+1} = \frac{1}{24.9} [22.1w_i + .5w_{i-1} - .1w_{i-2} + .24i + 2.52],$$

به‌ازای $i = 2, 3, \dots, 9$.

با استفاده از مقادیر دقیق $y(t) = e^{-t} + t$ برای α_0 ، α_1 ، و α_2 در حالت آدامز - بشفورث و برای α_0 و α_1 در حالت آدامز - مولتون، نتایج جدول ۸.۰۶ حاصل می‌شوند.

t_i	آدامز - بشفورث w_i	خطا	آدامز - مولتون w_i	خطا
.3	starting value		1.0408180061	2.146×10^{-7}
.4	1.0703229200	2.874×10^{-6}	1.0703196614	3.846×10^{-7}
.5	1.1065354755	4.816×10^{-6}	1.1065301384	5.213×10^{-7}
.6	1.1488184077	6.772×10^{-6}	1.1488110076	6.285×10^{-7}
.7	1.1965933934	8.090×10^{-6}	1.1965845932	7.106×10^{-7}
.8	1.2493381564	9.192×10^{-6}	1.2493281927	7.714×10^{-7}
.9	1.3065796139	9.954×10^{-6}	1.3065688456	8.141×10^{-7}
1.0	1.3678899580	1.052×10^{-5}	1.3678785994	8.418×10^{-7}

جدول ۸.۰۶

در مثال ۳، روش آدامز - مولتون به‌طور قابل ملاحظه‌ای نتایج بهتری نسبت به روش

آدامز - بشفورث از همان مرتبه بدست می‌دهد. گرچه این امر کلی است، لیکن روشهای ضمنی این ضعف ذاتی را دارند که ابتدا روش را به‌طور جبری به‌نمایش صریحی برای w_{i+1} برمی‌گردانند. اینکه این روند می‌تواند مشکل باشد، اگر غیر ممکن نباشد، با توجه به مسئله مقدار اولیه نسبتاً "مقدماتی زیر دیده می‌شود:

$$y' = e^y, \quad 0 \leq t \leq .25, \quad y(0) = 1$$

چون $f(t, y) = e^y$ ، روش سه‌گامی آدامز - مولتون دارای معادله تفاضلی

$$w_{i+1} = w_i + \frac{h}{24} [9e^{w_{i+1}} + 19e^{w_i} - 5e^{w_{i-1}} + e^{w_{i-2}}]$$

می‌باشد، که دقیقاً "نسبت به w_{i+1} غیر قابل حل است.

در عمل، روشهای چندگامی ضمنی به‌صورت توصیف شده در فوق بکار نمی‌روند، بلکه برای بهتر کردن تقریبهای حاصل از روشهای صریح بکار می‌روند. روند معمول، که ترکیبی از یک تکنیک صریح و ضمنی است، روش پیشگو - اصلاحگر نامیده می‌شود.

اگر یک روش مرتبه چهار برای حل یک مسئله مقدار اولیه لازم باشد، اولین مرحله محاسبه مقادیر آغازی w_0 ، w_1 ، w_2 ، و w_3 برای روش آدامز - بشفورث مرتبه چهار با استفاده از روش رونگ - کوتای مرتبه چهار است، که در الگوریتم ۲.۶ ارائه شد. مرحله بعد، با استفاده از روش آدامز - بشفورث، محاسبه یک تقریب $w_4^{(0)}$ به $y(t_4)$ است.

$$w_4^{(0)} = w_3 + \frac{h}{24} [55f(t_3, w_3) - 59f(t_2, w_2) + 37f(t_1, w_1) - 9f(t_0, w_0)].$$

سپس این تقریب را با استفاده از روش سه‌گامی آدامز - مولتون به

$$w_4^{(1)} = w_3 + \frac{h}{24} [9f(t_4, w_4^{(0)}) + 19f(t_3, w_3) - 5f(t_2, w_2) + f(t_1, w_1)]$$

اصلاح می‌کنند.

مقدار $w_4^{(1)}$ به عنوان تقریبی به $y(t_4)$ بکار می‌رود و تکنیک استفاده از روش آدامز - بشفورث به عنوان یک پیشگو و استفاده از روش آدامز - مولتون به عنوان یک اصلاحگر را

جهت تعیین $w_3^{(0)}$ و $w_3^{(1)}$ ، یعنی تقریبهای ابتدایی و انتهایی $y(t_0)$ ، و غیره تکرار می‌کنند.

اگرچه، در تئوری، تقریبهای اصلاح شده‌ای برای $y(t_{i+1})$ با تکرار فرمول آدامز-مولتون

$$w_{i+1}^{(k+1)} = w_i + \frac{h}{24} [9f(t_{i+1}, w_{i+1}^{(k)}) + 19f(t_i, w_i) - 5f(t_{i-1}, w_{i-1}) + f(t_{i-2}, w_{i-2})],$$

بدست می‌آیند، لیکن، در عمل، چون $\{w_{i+1}^{(k+1)}\}$ به تقریب واقعی ارائه شده به‌وسیله فرمول ضمنی همگراست تا به جواب $y(t_{i+1})$ ، اگر اصلاح دقیقی لازم باشد، استفاده از کاهش اندازه گام موثرتر است.

الگوریتم ارائه شده در زیر براساس روش آدامز-بشفورث مرتبه چهار به‌عنوان پیشگو و یک تکرار روش آدامز-مولتون به‌عنوان اصلاحگر، با مقادیر آغازی حاصل از بکار بردن روش رونگ-کوتای مرتبه چهار می‌باشد.

الگوریتم پیشگو-اصلاحگر مرتبه چهار آدامز ۴.۶

برای تقریب جواب مسئله مقدار اولیه

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

عدد صحیح مثبت N را اختیار می‌کنیم.

مرحله ۱. قرار می‌دهیم $h = \frac{b-a}{N}$ ، $t_0 = a$ ، $w_0 = \alpha$.

مرحله ۲. قرار می‌دهیم $i = 0$.

مرحله ۳. قرار می‌دهیم

$$k_1 = hf(t_i, w_i),$$

$$k_2 = hf\left(t_i + \frac{h}{2}, w_i + \frac{k_1}{2}\right)$$

$$k_3 = hf\left(t_i + \frac{h}{2}, w_i + \frac{k_2}{2}\right)$$

$$k_4 = hf(t_i + h, w_i + k_3),$$

$$w_{i+1} = w_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

$$t_{i+1} = a + (i + 1)h.$$

مرحله ۴. ۱ را به i می‌افزاییم .

مرحله ۵. اگر $i \leq 2$ ، به مرحله ۳ می‌رویم .

مرحله ۶. قرار می‌دهیم

$$w_{i+1}^{(0)} = w_i + \frac{h}{24} [55f(t_i, w_i) - 59f(t_{i-1}, w_{i-1})$$

$$+ 37f(t_{i-2}, w_{i-2}) - 9f(t_{i-3}, w_{i-3})].$$

مرحله ۷. قرار می‌دهیم $t_{i+1} = a + (i + 1)h$

مرحله ۸. قرار می‌دهیم

$$w_{i+1} = w_i + \frac{h}{24} [9f(t_{i+1}, w_{i+1}^{(0)}) + 19f(t_i, w_i)$$

$$- 5f(t_{i-1}, w_{i-1}) + f(t_{i-2}, w_{i-2})].$$

مرحله ۹. ۱ را به i می‌افزاییم .

مرحله ۱۰. اگر $i \leq N - 1$ ، به مرحله ۶ می‌رویم .

مرحله ۱۱. روند تکمیل خواهد بود؛ w_i به‌ازای هر $i = 1, 2, \dots, N$ ، $y(t_i)$ را تقریب می‌کند .

مثال ۴. در جدول ۹.۶ نتایج حاصل برای مسئله مقدار اولیه

$$y' = -y + t + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

با $N = 10$ و با استفاده از الگوریتم ۴.۶ آمده است . گرچه در اینجا نتایج به دقت نتایج مثال ۳، که فقط از اصلاحگر (یعنی، روش آدامز - مولتون استفاده می‌کند) ، نیست ، ولی ساد یادآور شد که برای بکار بردن این روش در آ. مثال ابتدا لازم بود که آن را به یک نمایش صریح برای w_{i+1} تغییر دهیم و ، بعلاوه ، از مقادیر آغازی دقیقی که در آن حالت بکار رفته بودند استفاده کنیم .

t_i	w_i	$ y(t_i) - w_i $
.4	1.0703199182	1.278×10^{-7}
.5	1.1065302684	3.923×10^{-7}
.6	1.1488110326	6.035×10^{-7}
.7	1.1965845314	7.724×10^{-7}
.8	1.2493280604	9.043×10^{-7}
.9	1.3065686568	1.003×10^{-6}
1.0	1.3678783660	1.075×10^{-6}

جدول ۹.۶

همانطور که در حالت روشهای تک مرحله‌ای ملاحظه شد، وقتی برای یک مقدار دو تقریب وجود دارد، تقریب خطا در یکی از روشها امکان پذیر است. در این صورت، این خطای تقریبی را می‌توان جهت تعدیل اندازه‌گام بکار برد تا خطای موضعی کنترل شود. چون روش چهارگامی آدامز - بشفورث از رابطه

$$y(t_{i+1}) = y(t_i) + \frac{h}{24} [55f(t_i, y(t_i)) - 59f(t_{i-1}, y(t_{i-1})) + 37f(t_{i-2}, y(t_{i-2})) - 9f(t_{i-3}, y(t_{i-3}))] + \frac{251}{720} y^{(5)}(\mu_i)h^5$$

به‌ازای یک $t_{i-3} < \mu_i < t_{i+1}$ ، بدست می‌آید، از فرض اینکه تقریبهای w_0, w_1, \dots, w_i همگی دقیق‌اند نتیجه می‌شود که

$$(۷۴.۶) \quad y(t_{i+1}) = w_{i+1}^{(0)} + \frac{251}{720} y^{(5)}(\mu_i)h^5$$

تحلیل مشابهی از روش سه‌گامی آدامز - مولتون به فرمول زیر منجر می‌شود:

$$(۷۵.۶) \quad y(t_{i+1}) = w_{i+1} - \frac{19}{720} y^{(5)}(\mu_i)h^5, \quad t_{i-2} < \mu_i < t_{i+1}$$

اگر معادله (۷۴.۶) را از معادله (۷۵.۶) کم و فرض کنیم، به‌ازای h کوچکی، بتوانیم $y^{(5)}$ را در هر دو معادله در نقطه یکسان μ حساب کرد، نتیجه می‌شود که

$$w_{i+1} - w_{i+1}^{(0)} = \frac{h^5}{720} [251y^{(5)}(\mu) + 19y^{(5)}(\mu)]$$

$$= \frac{3}{8} h^5 y^{(5)}(\mu).$$

با استفاده از این نتیجه در حذف عبارت شامل $h^5 y^{(5)}$ از (۷۵.۶) ، تقریب خطای زیر حاصل می‌گردد:

$$(۷۶.۶) \quad |y(t_{i+1}) - w_{i+1}| \approx \frac{19}{270} |w_{i+1} - w_{i+1}^{(0)}|.$$

چون برای تعیین تقریب موجود در (۷۶.۶) چند فرض بکار رفته‌اند ، معمولاً " بکار بردن تخمین محافظه کارانه‌ای چون σ برای خطای موضعی عملی است ، که در آن

$$\sigma = \frac{1}{10} |w_{i+1} - w_{i+1}^{(0)}| \approx |y(t_{i+1}) - w_{i+1}|.$$

برای حداکثر خطای قابل تحمل $\varepsilon > 0$ و اندازه گام اولیه h ، اگر $\varepsilon/10 < \sigma < \varepsilon$ ، مقدار w_{i+1} قابل قبول است و روند با محاسبه $w_{i+2}^{(0)}$ ، با استفاده از همان اندازه گام h ، ادامه می‌یابد . اما ، وقتی در این محدوده نباشد ، اندازه گام با ضرب در ثابت $q > 0$ تغییر خواهد کرد . برای تعیین مقدار مناسب q ، از معادله (۷۵.۶) ، با فرض اندازه گام h ، یادآور می‌شویم که

$$\sigma \approx |y(t_{i+1}) - w_{i+1}| \approx \frac{19}{720} y^{(5)}(\mu) h^5$$

اگر اندازه گام به qh تغییر یافته و \hat{w}_{i+1} تقریب جدید در $t_i + qh$ را مشخص کند ،

$$|y(t_i + qh) - \hat{w}_{i+1}| \approx \frac{19}{720} |y^{(5)}(\hat{\mu}) h^5 q^5| \approx \sigma q^5$$

برای کنترل خطای $|y(t_i + qh) - \hat{w}_{i+1}|$ ، انتخاب مناسبی از q تضمین می‌کند که σq^5 از ε تجاوز ننماید . روند مطرح شده در الگوریتم بعدی رابطه

$$\sigma q^5 = \frac{\varepsilon}{2}$$

را برای این منظور بکار می‌برد ، که انتخاب q به صورت زیر را نتیجه می‌دهد :

$$q = \sqrt[5]{\frac{\varepsilon}{2\sigma}}$$

این انتخاب q تضمین می‌کند که خطا در هر مرحله بین $\varepsilon/10$ و ε باشد . در عمل ،

معمولاً به مقدار q یک کران بالایی داده می‌شود، در این حالت 4؛ در نتیجه، اندازه گام خیلی بزرگ نمی‌شود. باید توجه داشت که، چون روشهای چندگامی اندازه‌های گام یکسانی را برای مقادیر آغازی لازم دارند، هر تغییر در اندازه گام محاسبه مجدد مقادیر آغازی جدید در آن نقطه را ایجاب می‌کند. این امر با زیر الگوریتمی به نام رونگ - کوتا، نظیر الگوریتم ۲.۶، انجام خواهد شد.

الگوریتم پیشگو - اصلاحگر با اندازه گام متغیر آدامز ۵.۶
برای تقریب جواب مسئله مقدار اولیه

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

با خطای موضعی کمتر از $\varepsilon > 0$ ، اندازه گام اولیه $h = 1.5\varepsilon^{1/5}$ را انتخاب می‌کنیم.
مرحله ۱. یک زیر الگوریتم مرتبه چهار رونگ - کوتا، به نام RK4(H, W, T) را، با اعمال زیر، وضع می‌کنیم:
قرار می‌دهیم

$$k_1 = H \cdot f(T, W),$$

$$k_2 = H \cdot f(T + H/2, W + \frac{1}{2}k_1),$$

$$k_3 = H \cdot f(T + H/2, W + \frac{1}{2}k_2),$$

$$k_4 = H \cdot f(T + H, W + k_3),$$

$$RK4(H, W, T) = W + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

و مقدار RK4(H, W, T) را به نقطه مربوطه برگشت می‌دهیم.

مرحله ۲. قرار می‌دهیم $t_0 = a$ و $w_0 = \alpha$.

مرحله ۳. به ازای هر $i = 0, 1, 2$ ، قرار می‌دهیم

$$t_{i+1} = a + (i + 1)h$$

$$w_{i+1} = RK4(h, w_i, t_i).$$

مرحله ۴. قرار می‌دهیم $i = 3$ و $L = 0$.

مرحله ۵. قرار می‌دهیم $t_{i+1} = t_i + h$.

مرحله ۶. اگر $t_{i+1} > b$ ، به مرحله ۱۸ می‌رویم.

مرحله ۷. قرار می دهیم

$$w_{i+1}^{(0)} = w_i + \frac{h}{24} [55f(t_i, w_i) - 59f(t_{i-1}, w_{i-1}) + 37f(t_{i-2}, w_{i-2}) - 9f(t_{i-3}, w_{i-3})].$$

مرحله ۸. قرار می دهیم

$$w_{i+1} = w_i + \frac{h}{24} [9f(t_{i+1}, w_{i+1}^{(0)}) + 19f(t_i, w_i) - 5f(t_{i-1}, w_{i-1}) + f(t_{i-2}, w_{i-2})].$$

مرحله ۹. قرار می دهیم $\sigma = .1 |w_{i+1} - w_{i+1}^{(0)}|$

مرحله ۱۰. اگر $1 \leq \sigma \leq \varepsilon$ ، i را به $i + 1$ می افزاییم ، و قرار می دهیم $L = 0$ و به

مرحله ۵ می رویم . (w_{i+1} را می پذیریم .)

مرحله ۱۱. قرار می دهیم $q = (\varepsilon/2\sigma)^{1/5}$.

مرحله ۱۲. اگر $\sigma > \varepsilon$ ، به مرحله ۱۷ می رویم . (w_{i+1} طرد می شود .)

مرحله ۱۳. اگر $q > 4$ ، h را با $4h$ عوض می کنیم . در غیر این صورت ، h را با qh

عوض می کنیم . i را به $i + 1$ می افزاییم .

مرحله ۱۴. قرار می دهیم $j = i$ و $L = 1$.

مرحله ۱۵. قرار می دهیم $t_{j+1} = t_j + h$.

اگر $t_{j+1} > b$ ، به مرحله ۱۸ می رویم .

قرار می دهیم $w_{j+1} = \text{RK4}(h, w_j, t_j)$.

i را به j می افزاییم .

مرحله ۱۶. اگر $3 \leq j = i + 3$ ، i را به $i + 3$ می افزاییم و به مرحله ۵ می رویم . در غیر

این صورت ، به مرحله ۱۵ می رویم .

مرحله ۱۷. اگر $q < .1$ ، h را با $1h$ عوض می کنیم . در غیر این صورت ، h را با qh

عوض می کنیم .

اگر $i = 3$ ، به مرحله ۳ می رویم .

اگر $L = 1$ ، 3 را از i کم می کنیم .

به مرحله ۱۴ می رویم .

مرحله ۱۸. روند کامل خواهد بود . اعداد w_i ، $y(t_i)$ را با خطای موضعی حداکثر

ε تقریب می‌کنند.

مثال ۵. در جدول ۱۰.۶ بعضی از نتایج حاصل از استفاده الگوریتم ۵.۶ در تعیین تقریبهایی برای جواب مسئله مقدار اولیه

$$y' = -y + t + 1, \quad 0 \leq t, \quad y(0) = 1$$

با حداکثر امید خطای موضعی 10^{-8} آمده است. چون تغییری در اندازه گام برای

t_i	w_i	$ y(t_i) - w_i $	h_{i+1}
.0376783047	1.0007009963	6.59×10^{-10}	.0376783047
.0753566095	1.0027693139	1.23×10^{-9}	.0376783047
.1130349142	1.0061543930	1.78×10^{-9}	.0376783047
.1507132190	1.0108075405	5.02×10^{-10}	.0376783047
.1883915237	1.0166818696	2.54×10^{-9}	.0376783047
.2260698285	1.0237322242	4.45×10^{-9}	.0376783047
.2637481332	1.0319151181	6.16×10^{-9}	.0376783047
.3014264380	1.0411886729	7.84×10^{-9}	.0376783047
.3391047427	1.0515125591	9.27×10^{-9}	.0376783047
.3767830475	1.0628479381	1.06×10^{-8}	.0376783047
.4144613522	1.0751574077	1.19×10^{-8}	.0376783047
.4521396570	1.0884049485	1.31×10^{-8}	.0376783047
.4898179617	1.1025558734	1.41×10^{-8}	.0376783047
.5274962665	1.1175767775	1.51×10^{-8}	.0376783047
⋮	⋮	⋮	⋮
1.1680274471	1.4790071867	2.11×10^{-8}	.0376783047
1.2226601238	1.5171059668	1.87×10^{-8}	.0546326766
1.2772928004	1.5560838065	1.66×10^{-8}	.0546326766
⋮	⋮	⋮	⋮
3.0801711288	3.1261224928	2.84×10^{-8}	.0546326766
3.1596056397	3.2020480901	2.50×10^{-8}	.0794345109
3.2390401505	3.2782416331	2.20×10^{-8}	.0794345109
⋮	⋮	⋮	⋮
4.9865993895	4.9934282146	2.22×10^{-8}	.0794345109
5.1015242266	5.1076116686	1.88×10^{-8}	.1149248370
5.2164490636	5.2218756122	1.57×10^{-8}	.1149248370
⋮	⋮	⋮	⋮
6.9403216188	6.9412895611	1.59×10^{-8}	.1149248370
7.1072944343	7.1081835300	1.25×10^{-8}	.1669728155
7.2742672498	7.2749603880	9.67×10^{-9}	.1669728155

$0 \leq t \leq 1$ ظاهر نمی‌شود، چند نتیجه^۶ نمونه که به‌ازای $t < 1$ بدست آمده‌اند نیز ارائه می‌شود. توجه کنید که خطای مندرج در این جدول، اگر با جواب $y(t) = t + e^{-t}$ مقایسه شود، خطای واقعی تقریب است، و نیز این امر که این مقدار در بعضی از موارد از مقدار قابل تحمل 10^{-8} بیشتر می‌شود البته این نتیجه را نمی‌دهد که خطای موضعی از این مقدار تجاوز می‌نماید.

روشهای چندگامی دیگری را می‌توان با استفاده از انتگرالگیری چند جمله‌ایهای درونیاب روی بازه‌هایی به شکل $[t_j, t_{j+1}]$ ، به‌ازای $j \leq i-1$ ، برای تعیین تقریب $y(t_{i+1})$ نتیجه گرفت. یک روش مخصوص که از انتگرالگیری چند جمله‌ای پسر نیوتن روی $[t_{i-3}, t_{i+1}]$ حاصل می‌شود، روشی صریح به‌نام روش میلن^۱ است:

$$(۷۷.۶) \quad w_{i+1} = w_{i-3} + \frac{4h}{3} [2f(t_i, w_i) - f(t_{i-1}, w_{i-1}) + 2f(t_{i-2}, w_{i-2})],$$

که دارای خطای برشی موضعی $(14/45)h^4 y^{(5)}(\xi_i)$ است، که در آن $t_{i-3} < \xi_i < t_{i+1}$. این روش گاهی به‌عنوان پیشگو برای روشی ضمنی، به‌نام روش سیمپسون،

$$(۷۸.۶) \quad w_{i+1} = w_{i-1} + \frac{h}{3} [f(t_{i+1}, w_{i+1}) + 4f(t_i, w_i) + f(t_{i-1}, w_{i-1})]$$

بکار می‌رود، که دارای خطای برشی موضعی $-(h^4/90)y^{(5)}(\xi_i)$ به‌ازای $t_{i-1} < \xi_i < t_{i+1}$ است، و با انتگرالگیری یک چند جمله‌ای پسر نیوتن روی $[t_{i-1}, t_{i+1}]$ حاصل می‌شود.

گرچه خطای برشی موضعی مربوط به روشی پیشگو - اصلاح‌گر از نوع میلن - سیمپسون معمولاً "کوچکتر از روش آدامز - بشفورث - مولتون است، این تکنیک به علت مسائل پایداری، که در روند آدامز ظاهر نمی‌شوند، کاربرد کمی دارد. جزئیات بیشتر در زمینه^۶ این مشکل در بخش ۹.۶ مطرح می‌شود.

مجموعه تمرینات ۶.۶

- با استفاده از روش آدامز - بشفورث (۵۸.۶) و روش آدامز - مولتون (۵۹.۶)، جوابهای مسائل مقدار اولیه^۶ زیر را تقریب کنید:

$$y' = 1 - y, \quad 0 \leq t \leq 1 \quad (\text{ا})$$

$y(0) = 0$ ؛ از $h = .1$ استفاده کنید؛

$$y' = -y + t + 1, \quad 0 \leq t \leq 5 \quad (\text{ب})$$

$y(0) = 2$ ؛ از $h = .2$ استفاده کرده، و نتایج تمرین ۳، بخش ۴.۰۶، را مقایسه کنید؛

$$y' = \frac{2}{t}y + t^2e^t, \quad 1 \leq t \leq 2 \quad (\text{پ})$$

$y(1) = 0$ ؛ از $h = .05$ استفاده کرده، و نتایج تمرین ۲، بخش ۴.۰۶، را مقایسه کنید؛

$$y' = 50t^2 - 50y + 2t, \quad 0 \leq t \leq 1 \quad (\text{ت})$$

$y(0) = \frac{1}{3}$ ؛ از $h = .1, .025, .01$ استفاده کرده، و نتایج تمرین ۴، بخش ۴.۰۶، را مقایسه کنید.

۲. الگوریتم ۵.۰۶ را در مورد مسائل زیر اجرا، و نتایج خود را با نتایج حاصل از تمرین

۱، بخش ۵.۰۶، مقایسه کنید؛

$$y' = 1 - y, \quad 0 \leq t \leq 1 \quad (\text{ا})$$

$y(0) = 0$ ؛ از $\varepsilon = 10^{-6}$ استفاده کنید؛

$$y' = -y + t + 1, \quad 0 \leq t \leq 5 \quad (\text{ب})$$

$y(0) = 2$ ؛ از $\varepsilon = 10^{-4}$ استفاده کنید؛

$$y' = \frac{2}{t}y + t^2e^t, \quad 1 \leq t \leq 2 \quad (\text{پ})$$

$y(1) = 0$ ؛ از $\varepsilon = 10^{-6}$ استفاده کنید؛

$$y' = 1 + y^2, \quad 0 \leq t \leq \frac{\pi}{2} \quad (\text{ت})$$

$y(0) = 0$ ؛ از $\varepsilon = 10^{-4}$ استفاده کنید؛

$$y' = 2|t - 2|y, \quad 0 \leq t \leq 5 \quad (\text{ث})$$

$y(0) = e^{-4}$ ؛ از $\varepsilon = 10^{-4}$ استفاده کنید؛

$$y' = 1 + t \sin(ty), \quad 0 \leq t \leq 2 \quad (\text{ج})$$

$y(0) = 0$ ؛ از $\varepsilon = 10^{-4}$ استفاده کنید؛

$$y' = 50t^2 - 50y + 2t, \quad 0 \leq t \leq 1 \quad (\text{چ})$$

$y(0) = \frac{1}{3}$ ؛ از $\varepsilon = 10^{-3}$ استفاده کنید،

$$y' = -2y + 2t^2 + 2t, \quad 0 \leq t \leq 1 \quad (\text{ح})$$

$y(0) = 1$ ؛ از $\varepsilon = 10^{-6}$ استفاده کنید.

۳. روش سهگامی آدامز - مولتون برای معادله دیفرانسیل

$$y' = e^y, \quad 0 \leq t \leq .20$$

$$y(0) = 1,$$

معادله تفاضلی

$$w_{i+1} = w_i + \frac{h}{24} [9e^{w_{i+1}} + 19e^{w_i} - 5e^{w_{i-1}} + e^{w_{i-2}}]$$

را نتیجه می‌دهد.

(آ) نشان دهید این معادله تفاضلی دارای جواب منحصر بفردی برای مقادیر

معلومی از $h > 0$ است.

(ب) w_i را به وسیله تکرار تابعی برای $t = 3, \dots, 20$ ، با $h = .01$ ، و استفاده از

مقادیر آغازی دقیق w_0, w_1, w_2 بدست آورید. در هر مرحله برای تقریب ابتدایی

w_{i+1} از w_i استفاده کنید. جواب برابر است با

$$y(t) = \ln \left| \frac{e}{1 - et} \right|$$

(پ) آیا روش نیوتن به همگرایی روی تکرار تابعی سرعت می‌دهد؟

۴. روش پیشگو - اصلاحگر میلن - سیمپسون را در مورد جواب

$$y' = -5y, \quad 0 < t < 2$$

$$y(0) = 1,$$

با $h = .1$ بکار برید. روند را با $h = .05$ تکرار کنید. آیا جوابها با خطای برشی

موضعی سازگارند؟

۵. یک مدار الکتریکی متشکل از یک خازن با ظرفیت ثابت $C = 1.1$ فاراد، و یک سیم

به مقاومت ثابت $R_0 = 2.1$ اهم به‌طور سری است. یک ولتاژ $\mathcal{E}(t) = 110 \sin t$ در زمان

t بکاررفته است. وقتی سیم گرم می‌شود، مقاومت تابعی از شدت جریان i می‌گردد؛

یعنی،

$$R(t) = R_0 + ki$$

و معادله دیفرانسیل برای i به صورت زیر درمی‌آید:

$$\left(1 + \frac{2k}{R_0} i\right) \frac{di}{dt} + \frac{1}{R_0 C} i = \frac{1}{R_0} \frac{d\mathcal{E}}{dt}$$

با استفاده از الگوریتم ۵.۰۶ با $\varepsilon = 10^{-4}$ ، شدت جریان i را پس از 2 ثانیه بدست آورید .

۶ . با استفاده از روش زیر ، فرمول (۶۷.۰۶) را نتیجه بگیرید . قرار دهید

$$y(t_{i+1}) = y(t_i) + ahf(t_i, y(t_i)) + bhf(t_{i-1}, y(t_{i-1})) \\ + chf(t_{i-2}, y(t_{i-2})).$$

سری تیلور $y(t_{i+1})$ ، $f(t_{i-2}, y(t_{i-2}))$ و $f(t_{i-1}, y(t_{i-1}))$ را در $(t_i, y(t_i))$ تعیین و برای بدست آوردن a ، b ، c ، ضرایب h ، h^2 ، و h^3 را برابر بگیرید .

۷ . فرمول (۷۰.۰۶) و خطای برشی موضعی آن را با استفاده از شکل مناسبی از یک چند جمله‌ای درونیاب نتیجه بگیرید .

۸ . الگوریتمی برای روش آدامز - بشفورث و روش پیشگو - اصلاحگر آدامز - مولتون بنویسید ، که با تغییر اندازه گام با نصف یا دو برابر کردن h ، بسته به تخمین خطا ، توام باشد .

۹ . خطاهای برشی موضعی روشهای میلن و سیمپسون را نتیجه بگیرید .

۱۰ . روش سیمپسون را با بکار بردن قاعده سیمپسون در انتگرال زیر نتیجه بگیرید :

$$y(t_{i+1}) - y(t_{i-1}) = \int_{t_{i-1}}^{t_{i+1}} f(t, y(t)) dt$$

۷.۰۶ روشهای برونابی

تکنیک برونابی برای تعیین تقریبهای دقیقی از فرمولهای مرتبه پایین در بخشهای ۲.۰۵ و ۶.۰۵ به عنوان کاربرد دیفرانسیلگیری و انتگرالگیری عددی ارائه شد . همچنین ، این تکنیک را می توان جهت توسعه روشهای موثر برای تقریب جواب مسائل مقدار اولیه مربوط به معادلات دیفرانسیل معمولی بکار برد . برای تجدید نظر در تکنیک برونابی ، ابتدا روش اویلر را بکار می بریم . سپس برونابی با طرحی متفاوت جهت حصول به روند برونابی ، که ایده آن به گراگ^۱ [۳۲] نسبت داده می شود ، داخل می گردد .

برای مسئله مقدار اولیه

$$(۷۹.۶) \quad y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

روش اویلر با اندازه گام $h > 0$ به وسیله

$$w_0 = \alpha$$

و، به ازای هر $i = 0, 1, \dots, N - 1$

$$(۸۰.۶) \quad w_{i+1} = w_i + hf(t_i, w_i)$$

ارائه می شود، که در آن $t_i = a + ih$ و $N = (b - a)/h$. یک تحلیل دقیق خطای $y(t_i) - w_i$ (ر.ک. گیر^۱ [۳۰]، صفحه ۱۵) به وجود تابع $\delta(t)$ با این خاصیت که به ازای هر $i = 1, 2, \dots, N$

$$(۸۱.۶) \quad y(t_i) = w_i + h\delta(t_i) + O(h^2)$$

منجر می شود، نکته مهم عدم وابستگی δ به h است.

چون روند برونمایی شامل اندازه های گام مختلف است، نماد $w(t_i, h)$ را برای نمایش مقدار w_i ، یعنی تقریب $y(t_i)$ منتج از (۸۰.۶) با اندازه گام h ، بکار می بریم. با استفاده از (۸۰.۶) با $h_0 = h$ و $h_1 = h/2$ ، تقریبهای $w(t_i, h_0)$ به ازای $i = 0, 1, \dots, N$ و $w(t_i, h_1)$ به ازای $i = 0, 1, \dots, 2N$ بدست می آیند. نقاط مشترک در هر دو اندازه گام با $t = t_{2i} = a + 2ih_1$ ، به ازای $i = 0, 1, \dots, N$ ، ارائه می شوند. در این نقاط از معادله (۸۱.۶) نتیجه می شود که

$$(۸۲.۶) \quad y(t) = w(t, h) + h\delta(t) + O(h^2)$$

$$(۸۳.۶) \quad y(t) = w\left(t, \frac{h}{2}\right) + \frac{h}{2}\delta(t) + O(h^2)$$

با ضرب (۸۳.۶) در دو و تفریق آن از (۸۲.۶) نتیجه می شود که

$$y(t) = 2w\left(t, \frac{h}{2}\right) - w(t, h) + O(h^2)$$

بطوری که روند برونمایی تقریب

$$2w\left(t, \frac{h}{2}\right) - w(t, h)$$

را نتیجه می دهد، که یک تقریب $O(h^2)$ است.

مثال ۱. روش اویلر با برونمایی را در مسئله مقدار اولیه

$$y' = -y + t + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

با استفاده از $h_0 = .1$ و $h_1 = .05$ ، بکار می بریم. فرمول تفاضلی به ازای $h = h_0$ برابر است با

$$w_0 = 1$$

$$w_{i+1} = .9w_i + .1t_i + .1, \quad i = 0, 1, \dots, 9$$

و، به ازای $h = h_1$ برابر است با

$$w_0 = 1,$$

$$w_{i+1} = .95w_i + .05t_i + .05, \quad i = 0, 1, \dots, 19$$

مقادیر موجود در جدول ۱۱.۶ فقط در نقاط $t_i = .1i$ ، به ازای $i = 0, 1, \dots, 10$ ، درج شده اند، و جواب دقیق $y(t) = t + e^{-t}$ برای مقایسه داده می شود.

$t = t_i$	$w(t, .1)$	$w(t, .05)$	$2w(t, .05) - w(t, .1)$	$y(t)$	Error
0	1.000000	1.000000	1.000000	1.000000	0
.1	1.000000	1.002500	1.005000	1.004837	1.63×10^{-4}
.2	1.010000	1.014506	1.019012	1.018731	2.81×10^{-4}
.3	1.029000	1.035092	1.041184	1.040818	3.66×10^{-4}
.4	1.056100	1.063420	1.070740	1.070320	4.20×10^{-4}
.5	1.090490	1.098737	1.106984	1.106531	4.53×10^{-4}
.6	1.131441	1.140360	1.149279	1.148812	4.67×10^{-4}
.7	1.178297	1.187675	1.197053	1.196585	4.68×10^{-4}
.8	1.230467	1.240126	1.249785	1.249329	4.56×10^{-4}
.9	1.287420	1.297214	1.307008	1.306570	4.38×10^{-4}
1.0	1.348678	1.358485	1.368292	1.367879	4.13×10^{-4}

جدول ۱۱.۶

برای بکار بردن روند برونمایی ریچاردسون در یک روش تفاضلی، لازم است روشی با نوع خاصی بسط خطا در نظر گرفته شود. فرض کنیم یک روش تفاضلی به شکل

$$(۱۴.۶)$$

$$w_0 = \alpha$$

به ازای هر $i = 0, 1, \dots, N - 1$ ، $w_{i+1} = w_i + \phi(t_i, w_i, h)$ ،
 با خاصیت زیر در دسترس باشد :

هرگاه $y \in C^{2p+2}[a, b]$ ، آنگاه به ازای p ای

$$(۸۵.۰۶) \quad y(t) = w(t, h) + \sum_{k=1}^p \delta_k(t)h^{2k} + O(h^{2p+2})$$

که در آن توابع δ_k به ازای هر $k = 1, 2, \dots, p$ مستقل از h اند . روند برونمایی را می توان به صورت زیر بکار برد :

$h > 0$ را ثابت گرفته و قرار می دهیم $h_0 = h$. یک نقطه شبکه ای t را اختیار و $w(t, h_0)$ را از (۸۴.۰۶) محاسبه می کنیم . طبق (۸۵.۰۶) ، داریم

$$(۸۶.۰۶) \quad y(t) = w(t, h_0) + \delta_1(t)h^2 + \delta_2(t)h^4 + \dots + \delta_p(t)h^{2p} + O(h^{2p+2})$$

سپس قرار می دهیم $h_1 = h/2$ ، و ، با استفاده از (۸۴.۰۶) ، $w(t, h_1)$ را محاسبه می کنیم . باید توجه داشت که این روند برای تقریب $y(t)$ ، با استفاده از h_1 ، به دو برابر تعداد گامهای لازم برای h_0 نیاز است ، و نتیجه خواهد بود

$$(۸۷.۰۶) \quad y(t) = w(t, h_1) + \frac{1}{4} \delta_1(t)h^2 + \frac{1}{16} \delta_2(t)h^4 + \dots + \frac{1}{4^p} \delta_p(t)h^{2p} + O(h^{2p+2}).$$

با ضرب (۸۷.۰۶) در چهار و تفریق آن از (۸۶.۰۶) و تقسیم نتیجه بر سه ، تقریب $O(h^4)$ از عبارت زیر حاصل می شود :

$$(۸۸.۰۶) \quad y(t) = \frac{4w(t, h_1) - w(t, h_0)}{3} - \frac{1}{4} \delta_2(t)h^4 - \dots - \left(\frac{4^p - 1}{3 \cdot 4^{p-1}} \right) \delta_p(t)h^{2p} + O(h^{2p+2})$$

اولین مرحله این روند کامل است . برای ادامه کار ، قرار می دهیم $h_2 = h/4$ و $w(t, h_2)$ را محاسبه می کنیم ، مجدداً از (۸۴.۰۶) استفاده می کنیم . نتیجه می شود که

$$(۸۹.۰۶) \quad y(t) = w(t, h_2) + \frac{1}{16} \delta_1(t)h^2 + \frac{1}{256} \delta_2(t)h^4 + \dots + \frac{1}{4^{2p}} \delta_p(t)h^{2p} + O(h^{2p+2}).$$

با حذف جمله h^2 از (۸۷.۶) و (۸۹.۶) ،

$$(90.6) \quad y(t) = \frac{4w(t, h_2) - w(t, h_1)}{3} - \frac{1}{64} \delta_2(t)h^4 \\ - \dots - \left(\frac{4^{p-1} - 1}{3 \cdot 4^{2p-1}} \right) \delta_p(t)h^{2p} + O(h^{2p+2})$$

چون نمادگذاری کاملاً "بفرنج می‌شود" ، آن را با قرار دادن

$$y_{i,1} = w(t, h_{i-1}) = w\left(t, \frac{h}{2^{i-1}}\right) \quad , \quad i = 1, 2, 3$$

و

$$y_{i,2} = \frac{4y_{i,1} - y_{i-1,1}}{3} \quad , \quad i = 2, 3$$

ساده می‌کنیم . حال ، با ضرب (۹۰.۶) در ۱۶ ، تفریق (۸۸.۶) و تقسیم حاصل بر ۱۵ ، یک تقریب $O(h^6)$ برای $y(t)$ بدست می‌آید . با استفاده از نمادگذاری جدید ، این اعمال نتیجه می‌دهند که

$$y(t) = \frac{16y_{3,2} - y_{2,2}}{15} + \frac{1}{64} \delta_3(t)h^6 \\ + \dots + \frac{(4^{p-1} - 1)(4^{p-2} - 1)}{45 \cdot 4^{2p-3}} \delta_p(t)h^{2p} + O(h^{2p+2})$$

اگر $y_{3,3}$ را با

$$y_{3,3} = \frac{16y_{3,2} - y_{2,2}}{15}$$

تعریف کنیم ، جدولی که نتایج

$$y_{1,1} = w(t, h),$$

$$y_{2,1} = w\left(t, \frac{h}{2}\right), \quad y_{2,2} = \frac{4y_{2,1} - y_{1,1}}{3},$$

$$y_{3,1} = w\left(t, \frac{h}{4}\right), \quad y_{3,2} = \frac{4y_{3,1} - y_{2,1}}{3}, \quad y_{3,3} = \frac{16y_{3,2} - y_{2,2}}{15}$$

را می‌دهد شبیه جدول انتگرالگیری رامبرگ، ارائه شده در بخش ۶.۵، است. یک روند تفاضلی که از روش نقطهء میانی

$$w_{i+1} = w_{i-1} + 2hf(t_i, w_i)$$

و روش اویلر برای شروع استفاده می‌کند را می‌توان، برای ارائهء بسط خطای مطلوب (۸۵.۶)، تعدیل کرد. در این تعدیل، قاعدهء نقطهء میانی را در افراز مناسبی از $[t_i, t_{i+1}]$ بکار می‌بریم؛ در نتیجه، دنباله‌ای از مراحل لازم است تا w_{i+1} از w_i بدست آید. روش تفاضلی را به گراگ نسبت می‌دهند، و تحقیق اینکه این روش معادلهء بسط خطای (۸۵.۶) را برقرار می‌کند در [۳۲] آمده است.

الگوریتم زیر روند برونابی را بکار می‌برد، و با شروع از روش نقطهء میانی پیراسته، برای تولید تقریبها و کنترل خطای موضعی ایجاب می‌کند که عناصر قطری محاسبه شوند مگر اینکه اختلاف آنها کمتر از یک مقدار معین قابل تحمل باشد. مرحلهء ۱۱ الگوریتم تضمین می‌کند که، جدول برونابی هرگز شامل بیش از شش سطر نباشد. این نوع برش الزاما" تضمین می‌کند که خطاهای گرد کردن بر محاسبات چیره نمی‌شوند. اگر، در هر مرحله، درایهء آخر در سطر ششم، یعنی $y_{6,6}$ ، به اندازهء کافی دقیق نباشد، در مرحلهء ۱۲ لازم است که محاسبات برای این تقریب با اندازهء گام تعدیل شده‌ای تکرار شود.

الگوریتم برونابی ۶.۶

برای تعیین یک تقریب عددی به جواب مسئلهء مقدار اولیهء

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

یک اندازهء گام اولیهء $h > 0$ و یک حداکثر خطای قابل تحمل $\epsilon > 0$ را انتخاب می‌کنیم.

مرحلهء ۱. قرار می‌دهیم $t_0 = a$ ، $w_0 = \alpha$ ، $p = 6$ ، و $i = 0$.

مرحلهء ۲. قرار می‌دهیم $m = 1$.

مرحلهء ۳. قرار می‌دهیم $k = 1$.

مرحلهء ۴. قرار می‌دهیم

$$h_k = 2^{-k}h,$$

$$n_k = 2^k m,$$

$$t_{i,0}^{(k)} = t_i,$$

$$\begin{aligned} w_{i,0}^{(k)} &= w_i, \\ w_{i,j+1}^{(k)} &= w_{i,j}^{(k)} + h_k f(t_{i,j}^{(k)}, w_{i,j}^{(k)}) \\ t_{i,j+1}^{(k)} &= t_{i,j}^{(k)} + h_k. \end{aligned}$$

مرحله ۵. به ازای هر $j = 1, 2, \dots, n_k - 1$ ، قرار می‌دهیم

$$\begin{aligned} w_{i,j+1}^{(k)} &= w_{i,j-1}^{(k)} + 2h_k f(t_{i,j}^{(k)}, w_{i,j}^{(k)}) \\ t_{i,j+1}^{(k)} &= t_i + (j+1)h_k. \end{aligned}$$

مرحله ۶. قرار می‌دهیم

$$y_{k,1} = [w_{i,n_k}^{(k)} + w_{i,n_k-1}^{(k)} + h_k f(t_{i,n_k}^{(k)}, w_{i,n_k}^{(k)})]/2$$

مرحله ۷. 1 را به k می‌افزاییم.

مرحله ۸. اگر $k \leq 2$ ، به مرحله ۴ می‌رویم.

مرحله ۹. به ازای هر $l = 2, \dots, k - 1$ ، قرار می‌دهیم

$$y_{k-1,l} = \frac{2^{2(l-1)} y_{k-1,l-1} - y_{k-2,l-1}}{2^{2(l-1)} - 1}$$

مرحله ۱۰. اگر $|y_{k-1,k-1} - y_{k-2,k-2}| < \varepsilon$ ، به مرحله ۱۳ می‌رویم.

مرحله ۱۱. اگر $k \leq 6$ ، به مرحله ۴ می‌رویم.

مرحله ۱۲. قرار می‌دهیم $h = h/2$ و به مرحله ۱ می‌رویم.

مرحله ۱۳. قرار می‌دهیم $w_{i+1} = y_{k-1,k-1}$.

مرحله ۱۴. 1 را به i می‌افزاییم.

مرحله ۱۵. اگر $t_i < b$ ، به مرحله ۲ می‌رویم.

مرحله ۱۶. روند کامل خواهد بود؛ به ازای هر i ، w_i تقریبی از $y(t_i)$ با خطای حداکثر

ε است.

مثال ۲. مسئله مقدار اولیه

$$(91.6) \quad y' = -y + t + 1, \quad 0 \leq t \leq 1, \quad y(0) = 1$$

که دارای جواب $y(t) = t + e^{-t}$ است، را در نظر می‌گیریم.

الگوریتم برونمایی ۶.۶ را در (۹۱.۶) با $h = .1$ و $\varepsilon = 10^{-10}$ بکار می‌بریم. در محاسبه

w_1 ، جدول زیر حاصل می‌شود:

$$y_{1,1} = 1.0048750000,$$

$$y_{2,1} = 1.0048468359, \quad y_{2,2} = 1.0048374479,$$

$$y_{3,1} = 1.0048397739, \quad y_{3,2} = 1.0048374199, \quad y_{3,3} = 1.0048374180$$

$$y_{4,1} = 1.0048380071, \quad y_{4,2} = 1.0048374182, \quad y_{4,3} = 1.0048374180,$$

$$y_{4,4} = 1.0048374180$$

چون $|y_{3,3} - y_{4,4}| \leq 10^{-10}$ ، محاسبه با $w_1 = y_{4,4}$ متوقف می‌شود. مجموعه کامل تقریبها در جدول ۱۲.۶ ارائه شده است، و در هر حال $y_{4,4}$ قابل قبول است. علاوه، هر w_i تا ده رقم اعشار دقیق است.

i	t_i	w_i
0	0	1.0000000000
1	.1	1.0048374180
2	.2	1.0187307531
3	.3	1.0408182207
4	.4	1.0703200460
5	.5	1.1065306597
6	.6	1.1488116361
7	.7	1.1965853038
8	.8	1.2493289641
9	.9	1.3065696597
10	1.0	1.3678794412

جدول ۱۲.۶

اثبات همگرایی روش ارائه شده در الگوریتم ۶.۶ مستلزم نتایجی از نظریه مجموعه‌پذیری است، و می‌توان آن را در مقاله^۶ اولیه^۶ گراگ [۳۲] یافت. چند روند برونمایی در دسترس‌اند، که در بعضی از آنها از تکنیکهای اندازه‌گام متغیر استفاده می‌شود، و تحقیق در این زمینه بسیار دایر است. برای روندهای دیگر مبتنی بر روش برونمایی، به بولیرش^۱ و استور^۲ [۱۳]، [۱۴]، [۱۵]، یا کتبا^۳ [۶۱] رجوع کنید.

1. Bulirsch

2. Stoer

3. Stetter

مجموعه تمرینات ۷.۶

۱. الگوریتم برونیاپی ۶.۶ را در مسائل مقدار اولیه زیر بکار برید:

$$y' = 1 - y, \quad 0 \leq t \leq 1 \quad (\text{آ})$$

$y(0) = 0$ ؛ از $h = .1$ استفاده کنید؛

$$y' = -y + t + 1, \quad 0 \leq t \leq 5 \quad (\text{ب})$$

$y(0) = 2$ ؛ از $h = .2$ استفاده کرده، و نتایج را با تمرین ۱، بخش ۵.۶، و تمرین

۲، بخش ۶.۶، مقایسه نمایید؛

$$y' = \frac{2}{t}y + t^2e^t, \quad 1 \leq t \leq 2 \quad (\text{پ})$$

$y(1) = 0$ ؛ از $h = .1$ استفاده کرده، و نتایج را با تمرین ۱، بخش ۵.۶، و تمرین

۲، بخش ۶.۶، مقایسه نمایید؛

$$y' = 1 + y^2, \quad 0 \leq t \leq \frac{\pi}{2} \quad (\text{ت})$$

$y(0) = 0$ ؛ از $h = \frac{\pi}{20}$ استفاده کرده، و نتایج را با تمرین ۱، بخش ۵.۶، و تمرین

۲، بخش ۶.۶، مقایسه نمایید؛

$$y' = 2|t - 2|y, \quad 0 \leq t \leq 5 \quad (\text{ث})$$

$y(0) = e^{-4}$ ؛ از $h = .2$ استفاده کرده، و نتایج را با تمرین ۱، بخش ۵.۶، و تمرین

۲، بخش ۶.۶، مقایسه نمایید؛

$$y' = 1 + t \sin(ty), \quad 0 \leq t \leq 2 \quad (\text{ج})$$

$y(0) = 0$ ؛ از $h = .1$ استفاده کرده، و نتایج را با تمرین ۱، بخش ۵.۶، و تمرین

۲، بخش ۶.۶، مقایسه نمایید؛

$$y' = 50t^2 - 50y + 2t, \quad 0 \leq t \leq 1 \quad (\text{چ})$$

$y(0) = \frac{1}{3}$ ؛ از $h = .1$ استفاده کرده، و نتایج را با تمرین ۱، بخش ۵.۶، و تمرین

۲، بخش ۶.۶، مقایسه نمایید؛

$$y' = -2y + 2t^2 + 2t, \quad 0 \leq t \leq 1 \quad (\text{ح})$$

$y(0) = 1$ ؛ از $h = .1$ استفاده کرده، و نتایج را با تمرین ۱، بخش ۵.۶، و تمرین

۲، بخش ۶.۶، مقایسه نمایید.

۲. فرض کنید $P(t)$ تعداد افراد یک جمعیت در زمان t ، برحسب سال، باشد. اگر

b میزان متوسط تولد ثابت، و d میزان متوسط مرگ متناسب با انبوهی جمعیت (ناشی از ازدحام) باشد، میزان رشد جمعیت با معادله لوجستیکی

$$\frac{dP(t)}{dt} = bP(t) - k[P(t)]^2,$$

داده می‌شود، که در آن $d = kP(t)$ فرض کنید $P(0) = 50.976$ ، $b = 2.9 \times 10^{-2}$ ، و $k = 1.4 \times 10^{-7}$. با استفاده از الگوریتم ۶.۶، جمعیت را بعد از پنج سال تعیین کنید.

۸.۶ معادلات مرتبه بالاتر و دستگاههای معادلات دیفرانسیل

این بخش مشتمل بر مقدمه‌ای از حل عددی معادلات دیفرانسیل مرتبه بالاتر تحت شرایط اولیه است. روشهای مورد بحث محدود به آنهایی است که یک معادله مرتبه بالاتر از یک را به دستگاهی از معادلات دیفرانسیل مرتبه اول تبدیل می‌کند. پیش از بحث درباره روند تبدیل، نکاتی چند در باب دستگاههای معادلات دیفرانسیل مرتبه اول لازم است. یک دستگاه مرتبه m م از مسائل مقدار اولیه مرتبه اول را می‌توان به صورت

$$\begin{aligned} \frac{du_1}{dt} &= f_1(t, u_1, u_2, \dots, u_m), \\ \frac{du_2}{dt} &= f_2(t, u_1, u_2, \dots, u_m), \\ &\vdots \\ \frac{du_m}{dt} &= f_m(t, u_1, u_2, \dots, u_m), \end{aligned} \quad (92.6)$$

به‌ازای $a \leq t \leq b$ ، با شرایط اولیه زیر بیان کرد:

$$\begin{aligned} u_1(a) &= \alpha_1, \\ u_2(a) &= \alpha_2, \\ &\vdots \\ u_m(a) &= \alpha_m. \end{aligned} \quad (93.6)$$

هدف تعیین m تابع $u_1(t), u_2(t), \dots, u_m(t)$ است که در دستگاه معادلات دیفرانسیل و شرایط اولیه صدق کنند.

برای بحث وجود و یکتایی جوابهای دستگاه معادلات، ابتدا لازم است تعریف شرط لیب شیتس را به توابع چند متغیره تعمیم دهیم.

۱۶.۶ تعریف. گوئیم تابع $f(t, y_1, \dots, y_m)$ تعریف شده بر مجموعه^۱

$$D = \{(t, u_1, \dots, u_m) \mid a \leq t \leq b, -\infty < u_i < \infty, i = 1, 2, \dots, m\}$$

در شرط لیب شیتس بر D با متغیرهای u_1, u_2, \dots, u_m صدق می‌کند اگر ثابتی مانند $L > 0$ با خاصیت زیر وجود داشته باشد:

به‌ازای هر (t, z_1, \dots, z_m) و $(t, u_1, u_2, \dots, u_m)$ در D .

$$(۹۴.۶) \quad |f(t, u_1, \dots, u_m) - f(t, z_1, \dots, z_m)| \leq L \sum_{j=1}^m |u_j - z_j|.$$

با استفاده از قضیه مقدار میانگین، می‌توان نشان داد که هرگاه f و مشتقات جزئی

اول آن در D پیوسته بوده و به‌ازای هر $i = 1, 2, \dots, m$ و هر (t, u_1, \dots, u_m) در D

$$\left| \frac{\partial f(t, u_1, \dots, u_m)}{\partial u_i} \right| \leq L$$

آنگاه f در شرط لیب شیتس بر D با ثابت لیب شیتس L صدق می‌کند (ر. ک. بیرکف - روتا [۱۱]، صفحه ۱۰۲). یک قضیه وجودی و یکتایی اساسی در زیر ارائه می‌شود. اثباتش را می‌توان در صفحات ۱۱۲ - ۱۱۳ بیرکف - روتا [۱۱] یافت.

قضیه ۱۷.۶. فرض کنیم

$$D = \{(t, u_1, u_2, \dots, u_m) \mid a \leq t \leq b, -\infty < u_i < \infty, i = 1, 2, \dots, m\}$$

و نیز $f_i(t, u_1, \dots, u_m)$ ، به‌ازای هر $i = 1, 2, \dots, m$ ، در D پیوسته و بر آن در شرط لیب شیتس صدق کنند. در این صورت، دستگاه معادلات دیفرانسیل مرتبه اول (۹۳.۶)، تحت شرایط اولیه^۲ (۹۳.۶)، دارای جواب منحصر بفرد $u_1(t), \dots, u_m(t)$ ، به‌ازای $a \leq t \leq b$ است.

برای تحویل یک معادله دیفرانسیل مرتبه m کلی به شکل

$$y^{(m)}(t) = f(t, y, y', \dots, y^{(m-1)}), \quad a \leq t \leq b,$$

با شرایط اولیه^۳ $y(a) = \alpha_1, y'(a) = \alpha_2, \dots, y^{(m-1)}(a) = \alpha_m$ به یک دستگاه معادلات به

شکل (۹۲.۶) و (۹۳.۶) ، فرض می‌کنیم

$$u_1(t) = y(t), u_2(t) = y'(t), \dots, u_m(t) = y^{(m-1)}(t).$$

با این نمادها، دستگاه مرتبه اول

$$\frac{du_1}{dt} = \frac{dy}{dt} = u_2,$$

$$\frac{du_2}{dt} = \frac{dy'}{dt} = u_3,$$

⋮

$$\frac{du_m}{dt} = \frac{dy^{(m-1)}}{dt} = y^{(m)} = f(t, y, y', \dots, y^{(m-1)})$$

$$= f(t, u_1, u_2, \dots, u_m),$$

با شرایط اولیه زیر بدست می‌آید:

$$u_1(a) = y(a) = \alpha_1,$$

$$u_2(a) = y'(a) = \alpha_2,$$

⋮

$$u_m(a) = y^{(m-1)}(a) = \alpha_m.$$

مثال ۱. معادله دیفرانسیل مرتبه دوم

$$(95.6) \quad y'' - 2y' + 2y = e^{2t} \sin t, \quad 0 \leq t \leq 1$$

با شرایط اولیه $y(0) = -4, y'(0) = -6$ را در نظر می‌گیریم. با فرض $u_1(t) = y(t)$ و

$$u_2(t) = y'(t), \quad (95.6) \quad \text{به دستگاه}$$

$$u_1'(t) = u_2(t),$$

(96.6)

$$u_2'(t) = e^{2t} \sin t - 2u_1(t) + 2u_2(t),$$

با شرایط اولیه زیر تبدیل می‌شود:

$$u_1(0) = -4,$$

(97.6)

$$u_2(0) = -6.$$

واضح است که $f_1(t, u_1, u_2) = u_2$ بر $D = \{(t, u_1, u_2) | 0 \leq t \leq 1, -\infty < u_1, u_2 < \infty\}$ در شرط لیبشیتس صدق می‌کند، و بر D پیوسته است. با فرض

$$f_2(t, u_1, u_2) = e^{2t} \sin t - 2u_1 + 2u_2$$

$$\begin{aligned} & |f_2(t, u_1, u_2) - f_2(t, z_1, z_2)| \\ &= |e^{2t} \sin t - 2u_1 + 2u_2 - e^{2t} \sin t + 2z_1 - 2z_2| \\ &\leq 2|u_1 - z_1| + 2|u_2 - z_2|; \end{aligned}$$

در نتیجه، f_2 نیز به D در شرط لیبشیتس، لیکن با $L = 2$ ، صدق می‌کند. بعلاوه، f_2 نیز بر D پیوسته است؛ در نتیجه، دستگاه به‌ازای $0 \leq t \leq 1$ دارای جواب منحصر بفرد است. به سادگی می‌توان تحقیق کرد که این جواب برابر است با

$$\begin{aligned} u_1(t) &= .2e^{2t} (\sin t - 2 \cos t), \\ u_2(t) &= .2e^{2t} (4 \sin t - 3 \cos t); \end{aligned}$$

در نتیجه، $y(t) = .2e^{2t}(\sin t - 2 \cos t)$ ، جواب منحصر بفرد معادله مرتبه دوم است که در $y(0) = -.4$ و $y'(0) = -.6$ صدق می‌کند.

روشهای حل دستگاههای معادلات دیفرانسیل مرتبه اول چیزی جز تعمیم روشهای حل یک معادله مرتبه اول که پیشتر در این فصل ارائه شد نیست. مثلاً، روش کلاسیک مرتبه چهار رونگ - کوتای ارائه شده با

$$\begin{aligned} w_0 &= \alpha, \\ k_1 &= hf(t_i, w_i), \\ k_2 &= hf\left(t_i + \frac{h}{2}, w_i + \frac{1}{2}k_1\right), \\ k_3 &= hf\left(t_i + \frac{h}{2}, w_i + \frac{1}{2}k_2\right), \\ k_4 &= hf(t_{i+1}, w_i + k_3), \end{aligned}$$

و، به‌ازای هر $i = 0, 1, \dots, N - 1$ ،

$$w_{i+1} = w_i + \frac{1}{6} [k_1 + 2k_2 + 2k_3 + k_4]$$

برای حل مسئله مقدار اولیه مرتبه اول

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

را می‌توان به صورت زیر تعمیم داد: عدد صحیح $N > 0$ را اختیار و قرار می‌دهیم $h = (b - a)/N$. بازه $[a, b]$ را به N زیر بازه با نقاط شبکه‌ای به‌ازای هر $j = 0, 1, \dots, N$ ، $t_j = a + jh$

افراز می‌کنیم. نمادگذاری w_{ij} را جهت نمایش تقریب $u_i(t_j)$ به‌ازای هر $j = 0, 1, \dots, N$ و $i = 1, 2, \dots, m$ بکار می‌بریم، به‌این معنی که w_{ij} تقریب جواب i م $u_i(t)$ از (۹۲.۶) در t_j ، یعنی نقطه شبکه‌ای z م، است. برای شرایط اولیه، قرار می‌دهیم

$$(98.6) \quad \begin{aligned} w_{1,0} &= \alpha_1, \\ w_{2,0} &= \alpha_2, \\ &\vdots \\ w_{m,0} &= \alpha_m \end{aligned}$$

اگر فرض کنیم مقادیر $w_{1,j}, w_{2,j}, \dots, w_{m,j}$ محاسبه شده باشند، با اولین محاسبه $w_{1,j+1}, w_{2,j+1}, \dots, w_{m,j+1}$ را بدست می‌آوریم:

$$(99.6) \quad k_{1,i} = hf_i(t_j, w_{1,j}, w_{2,j}, \dots, w_{m,j}), \quad i = 1, 2, \dots, m$$

به‌ازای هر $i = 1, 2, \dots, m$

$$(100.6) \quad k_{2,i} = hf_i\left(t_j + \frac{h}{2}, w_{1,j} + \frac{1}{2}k_{1,1}, w_{2,j} + \frac{1}{2}k_{1,2}, \dots, w_{m,j} + \frac{1}{2}k_{1,m}\right)$$

به‌ازای هر $i = 1, 2, \dots, m$

$$(101.6) \quad k_{3,i} = hf_i\left(t_j + \frac{h}{2}, w_{1,j} + \frac{1}{2}k_{2,1}, w_{2,j} + \frac{1}{2}k_{2,2}, \dots, w_{m,j} + \frac{1}{2}k_{2,m}\right)$$

به‌ازای هر $i = 1, 2, \dots, m$

$$(102.6) \quad k_{4,i} = hf_i(t_j + h, w_{1,j} + k_{3,1}, w_{2,j} + k_{3,2}, \dots, w_{m,j} + k_{3,m})$$

و، در این صورت،
به‌ازای هر $i = 1, 2, \dots, m$

$$(103.6) \quad w_{i,j+1} = w_{i,j} + \frac{1}{6}[k_{1,i} + 2k_{2,i} + 2k_{3,i} + k_{4,i}]$$

باید توجه داشت که لازم است تعام $k_{1,1}, k_{1,2}, \dots, k_{1,m}$ ، پیش از آنکه بتوان $k_{2,1}$ را تعیین کرد، محاسبه شوند. درحالت کلی، باید هر یک از $k_{1,1}, k_{1,2}, \dots, k_{1,m}$ پیش از هر یک از عبارات $k_{l+1,i}$ محاسبه شوند.

مثال ۲. دستگاه مرتبه اول مثال ۱، داده شده با

$$(104.6) \quad \begin{aligned} u'_1 &= u_2, \\ u'_2 &= e^{2t} \sin t - 2u_1 + 2u_2 \end{aligned}$$

بهازی $0 \leq t \leq 1$ ، با شرایط اولیه^۱ زیر را در نظر می‌گیریم:

$$(105.6) \quad \begin{aligned} u_1(0) &= -4, \\ u_2(0) &= -6 \end{aligned}$$

روش کلاسیک مرتبه^۲ چهار رונک - کوتا را برای تقریب جواب این مسئله با استفاده از $h = .1$ بکار می‌بریم. از شرایط اولیه نتیجه می‌شود که $w_{1,0} = -4$ و $w_{2,0} = -6$. بنابراین معادلات (۹۹.۶) تا (۱۰۲.۶) با $j = 1$

$$k_{1,1} = hf_1(t_0, w_{1,0}, w_{2,0}) = hw_{2,0} = -.06,$$

$$\begin{aligned} k_{1,2} &= hf_2(t_0, w_{1,0}, w_{2,0}) \\ &= h[e^{2t_0} \sin t_0 - 2w_{1,0} + 2w_{2,0}] = -.04, \end{aligned}$$

$$k_{2,1} = hf_1\left(t_0 + \frac{h}{2}, w_{1,0} + \frac{1}{2}k_{1,1}, w_{2,0} + \frac{1}{2}k_{1,2}\right)$$

$$= h\left[w_{2,0} + \frac{1}{2}k_{1,2}\right] = -.062,$$

$$k_{2,2} = hf_2\left(t_0 + \frac{h}{2}, w_{1,0} + \frac{1}{2}k_{1,1}, w_{2,0} + \frac{1}{2}k_{1,2}\right)$$

$$\begin{aligned} &= h[e^{2(t_0+.05)} \sin(t_0 + .05) - 2(w_{1,0} + \frac{1}{2}k_{1,1}) \\ &\quad + 2(w_{2,0} + \frac{1}{2}k_{1,2})] \\ &= -.03247644757, \end{aligned}$$

$$k_{3,1} = h[w_{2,0} + \frac{1}{2}k_{2,2}] = -.06162382238,$$

$$\begin{aligned} k_{3,2} &= h[e^{2(t_0+.05)} \sin(t_0 + .05) - 2(w_{1,0} + \frac{1}{2}k_{2,1}) \\ &\quad + 2(w_{2,0} + \frac{1}{2}k_{2,2})] \\ &= -.03152409237, \end{aligned}$$

$$k_{4,1} = h[w_{2,0} + k_{3,2}] = -.06315240924,$$

$$k_{4,2} = h[e^{2(t_0+.1)} \sin(t_0 + .1) - 2(w_{1,0} + k_{3,1})]$$

$$+ 2(w_{2,0} + k_{3,2})] \\ = -.02178637298;$$

در نتیجه ،

$$w_{1,1} = w_{1,0} + \frac{1}{6}[k_{1,1} + 2k_{2,1} + 2k_{3,1} + k_{4,1}] = -.4617333423,$$

و

$$w_{2,1} = w_{2,0} + \frac{1}{6}[k_{1,2} + 2k_{2,2} + 2k_{3,2} + k_{4,2}] = -.6316312421$$

مقدار $w_{1,1}$ تقریب $u_1(1) = y(1) = .2e^{2(1)}[\sin .1 - 2 \cos .1]$ و $w_{2,1}$ تقریب $u_2(1) = y(1) = .2e^{2(1)}[4 \sin .1 - 3 \cos .1]$ می باشد .

مقادیر $w_{1,j}$ و $w_{2,j}$ ، به ازای $j = 0, 1, \dots, 10$ ، در جدول ۱۳.۶ درج شده اند ، و مقدار واقعی $u_1(t) = .2e^{2t}[\sin t - 2 \cos t]$ نیز داده شده است . چون $u_2 = u_1'$ ، مقادیر $u_2(t)$ داده نشده اند .

t_j	$w_{1,j}$	$w_{2,j}$	$y(t_j) = u_1(t_j)$	$ y(t_j) - w_{1,j} $
0	-.40000000	-.60000000	-.40000000	0
.1	-.46173334	-.63163124	-.46173297	3.7×10^{-7}
.2	-.52555988	-.64014895	-.52555905	8.3×10^{-7}
.3	-.58860144	-.61356381	-.58860005	1.39×10^{-6}
.4	-.64661231	-.53658203	-.64661028	2.03×10^{-6}
.5	-.69356666	-.38873810	-.69356395	2.71×10^{-6}
.6	-.72115190	-.14438087	-.72114849	3.41×10^{-6}
.7	-.71815295	.22899702	-.71814890	4.05×10^{-6}
.8	-.66971133	.77199180	-.66970677	4.56×10^{-6}
.9	-.55644290	.15347815	-.55643814	4.76×10^{-6}
1.0	-.35339886	.25787663	-.35339436	4.50×10^{-6}

جدول ۱۳.۶

روشهای دیگری را می توان مشابه "به دستگامها تعمیم داد . اگر روشهایی نظیر روش رونگ - کوتا - فلبرگ با کنترل خطا تعمیم یافته باشند ، هر مولفهء جواب عددی $(w_{1,j}, w_{2,j}, \dots, w_{m,j})$ را باید از حیث دقت بررسی کرد . اگر مولفه ای به قدر کافی دقیق نباشد ، جواب عددی $(w_{1,j}, w_{2,j}, \dots, w_{m,j})$ را باید مجدداً محاسبه کرد . روشهای چندمرحله ای و تکنیکهای پیشگو - اصلاحگر را نیز می توان به سادگی به

دستگاهها تعمیم داد. مجدداً، اگر کنترل خطا بکار می‌رود، هر مولفه باید دقیق باشد.

تکنیک برون‌یابی را می‌توان به دستگاهها نیز تعمیم داد، لیکن نمادگذاری کاملاً پیچیده می‌شود. تعمیمی از این نوع در تمرین ۴ ملاحظه می‌شود.

قضایای همگرایی و تخمینهای خطا برای دستگاهها خیلی شبیه آنهاست که برای تک معادلات در بخش ۹.۶ بررسی خواهد شد، جز آنکه کرانها برحسب نرمهای برداری داده شده‌اند، بحثی که در فصل ۸ مطرح می‌شود. [مرجع خوبی برای این قضایا، گیر [۳۰] صفحات ۷۲-۴۵ است.]

مجموعه تمرینات ۸.۶

۱. با تغییر الگوریتم ۴.۶ مرتبه چهار پیشگو - اصلاحگر آدامز، تقریب جوابهای دستگاههای معادلات دیفرانسیل مرتبه اول را بدست آورده، و مسائل مقدار اولیه زیر را حل کنید:

$$\begin{aligned} u_1'(t) &= 3u_1(t) + 2u_2(t), & 0 \leq t \leq 1, & u_1(0) = 0, \\ u_2'(t) &= 4u_1(t) + u_2(t), & 0 \leq t \leq 1, & u_2(0) = 1. \end{aligned} \quad (\text{آ})$$

۱. h را بکار برده و جواب را با جواب واقعی

$$u_1(t) = \frac{1}{3}(e^{5t} - e^{-t}) \quad \text{و} \quad u_2(t) = \frac{1}{3}(e^{5t} + 2e^{-t})$$

مقایسه کنید؛

$$\begin{aligned} y''(t) + 2ty'(t) + t^2y(t) &= e^t, & 0 \leq t \leq 2, & \\ y(0) = 1, y'(0) &= -1. \end{aligned} \quad (\text{ب})$$

۱. h را بکار برده و جواب را با جواب حاصل از روش به صورت سری توانی مقایسه کنید؛

$$\begin{aligned} y''(t) - 2y'(t) + 2y(t) &= e^{2t} \sin t, & 0 \leq t \leq 1, & \\ y(0) = -.4, y'(0) &= -6. \end{aligned} \quad (\text{پ})$$

۱. h را بکار برده و جواب را با مثال ۲ مقایسه کنید؛

$$\begin{aligned} t^2y'' - 2ty' + 2y &= t^3 \ln t, & 1 \leq t \leq 2, & \\ y(1) = 1, y'(1) &= 0. \end{aligned} \quad (\text{ت})$$

۰.۵. h را بکار برده و جواب را با جواب واقعی

$$y(t) = \frac{1}{2}t + \frac{t^3}{2} \ln t - \frac{3}{2}t^3$$

مقایسه کنید :

$$y''' + 2y'' - y' - 2y = e^t, \quad 0 \leq t \leq 3, \quad (ث)$$

$$y(0) = 1, \quad y'(0) = 2, \quad y''(0) = 0$$

$h = 2$ را بکار برده و جواب را با جواب واقعی مقایسه کنید :

$$u_1' + 4u_1 + 2u_2 = \cos t + 4 \sin t, \quad 0 \leq t \leq 5, \quad u_1(0) = 0, \quad (ج)$$

$$u_2' - 3u_1 - u_2 = -3 \sin t, \quad 0 \leq t \leq 5, \quad u_2(0) = -1$$

$h = 1$ را بکار برده و جواب را با جواب واقعی

$$u_1(t) = 2e^{-t} - 2e^{-2t} + \sin t, \quad u_2(t) = -3e^{-t} + 2e^{-2t}.$$

مقایسه کنید :

$$u_1' = u_2, \quad 0 \leq t \leq 1, \quad u_1(0) = 3, \quad (چ)$$

$$u_2' = -u_1 + 2e^{-t} + 1, \quad 0 \leq t \leq 1, \quad u_2(0) = 0.$$

$h = 0.5$ را بکار برده و جواب را با جواب واقعی

$$u_1(t) = \cos t + \sin t + e^{-t} + 1, \quad u_2(t) = -\sin t + \cos t - e^{-t}.$$

مقایسه کنید :

$$y'' = 2y^3, \quad 1 < t < 1.9, \quad (ح)$$

$$y(1) = -1, \quad y'(1) = -1$$

$h = 0.5$ را بکار برده و جواب را با جواب واقعی

$$y(t) = \frac{t}{t-2}$$

مقایسه نمایید .

۲ . با استفاده از روش مرتبه چهار رونگ - کوتا ، تمرین ۱ را تکرار کنید . قسمت (پ)

را حذف نمایید .

۳ . با تغییر الگوریتم رونگ - کوتا - فلیبرگ ۳.۶ ، جواب دستگاهها را تقریب و تمرین

۱ را تکرار کنید .

۴ . با تغییر الگوریتم برونیابی ۶.۶ ، جواب دستگاهها را تقریب و تمرین ۱ را تکرار

کنید .

۵ . منشاء مطالعه مدل‌های ریاضی جهت پیشگویی دینامیک جمعیتی نمونه‌های رقابتی ،

آثار مستقل ۱. ج. لوتکا^۱ و وولتر^۲، که اوایل قرن اخیر منتشر شده، می باشد. مسئله پیشگویی دو نمونه از جمعیت را در نظر بگیرید، که یکی شکارچی است و در زمان t دارای جمعیت $x_2(t)$ است، و از دیگری (به نام شکار)، با جمعیت $x_1(t)$ ، تغذیه می شود. فرض کنید شکار همیشه ذخیره غذایی کافی دارد و میزان موالید آن در هر زمان متناسب با تعداد شکار زنده در آن زمان است؛ یعنی، میزان موالید (شکار) برابر $k_1 x_1(t)$ است، درحالی که میزان مرگ و میر آن هم به تعداد شکار و هم به تعداد شکارچی زنده در آن زمان بستگی دارد. برای سادگی، فرض کنید میزان مرگ و میر (شکار) برابر $k_2 x_1(t)x_2(t)$ است. از آن سو، میزان موالید شکارچی، همانطور که به تعداد شکارچیهای موجود برای تولید مثل وابسته است، به ذخیره غذایی آن، یعنی $x_1(t)$ ، نیز بستگی دارد. به این دلیل، فرض می شود که میزان موالید (شکارچی) برابر $k_3 x_1(t)x_2(t)$ است. میزان مرگ و میر شکارچی متناسب با تعداد شکارچیهای زنده در آن زمان گرفته می شود؛ یعنی، میزان مرگ و میر شکارچی برابر $k_4 x_2(t)$ است. چون

$$\frac{dx_1(t)}{dt} \quad \text{و} \quad \frac{dx_2(t)}{dt}$$

بترتیب تغییرات جمعیت شکار و شکارچی را، نسبت به زمان، نشان می دهند، مسئله با دستگاه معادلات دیفرانسیل غیر خطی

$$\frac{dx_1(t)}{dt} = k_1 x_1(t) - k_2 x_1(t)x_2(t)$$

و

$$\frac{dx_2(t)}{dt} = k_3 x_1(t)x_2(t) - k_4 x_2(t).$$

بیان می شود. با استفاده از روش مرتبه چهار رانگ - کوتا، و فرض اینکه جمعیت اولیه شکار ۱۰۰۰ و شکارچی ۲۰۰ است و اینکه مقادیر ثابت عبارتند از $k_1 = 3$ ، $k_2 = .002$ ، $k_3 = .0006$ ، و $k_4 = .5$ ، این دستگاه را حل کنید. نمودار جوابهای این مسئله را، با در نظر گرفتن هر دو جمعیت نسبت به زمان، بکشید و پدیده

فیزیکی بیان شده را شرح دهید.

۶. در تمرین ۵، مسئله پیشگویی جمعیت را در طرح شکارچی - شکار در نظر گرفتیم. مسئله دیگر از این نوع رقابت دو نمونه به ازای ذخیره غذایی یکسان می باشد. اگر تعداد افراد زنده دو نمونه در زمان t با $x_1(t)$ و $x_2(t)$ نموده شوند، اغلب فرض می شود که، درحالی که میزان مولید هر یک از نمونه ها متناسب با تعداد افراد زنده نمونه در آن زمان است، میزان مرگ و میر هر یک از نمونه ها به جمعیت هر دو نمونه بستگی دارد. فرض کنید جمعیت یک زوج خاص از نمونه ها با معادلات

$$\frac{dx_1(t)}{dt} = x_1(t)(4 - .0003x_1(t) - .0004x_2(t))$$

و

$$\frac{dx_2(t)}{dt} = x_2(t)(2 - .0002x_1(t) - .0001x_2(t)).$$

بیان شوند. اگر جمعیت اولیه هر نمونه را 10,000 نفر بدانیم، جواب این دستگاه را بدست آورید.

۹.۶ پایداری و انتخاب روشها

در این فصل روشهای متعددی جهت تقریب جواب یک مسئله مقدار اولیه ارائه شده است. گرچه برای این منظور روشهای بسیار دیگری نیز موجود است، روشهای مطرح شده در اینجا را به این علت انتخاب کرده ایم که عموماً سه معیار را برآورده می کنند: اول، تعمیم آنها آنقدر روشن است که دانشجوی سال اول آنالیز عددی چگونگی و دلیل عملکرد آنها را می فهمد؛ دوم، اغلب روشهای پیچیده و پیشرفته تر بر یک یا چند روند مطرح شده در اینجا استوارند؛ و سوم، یک یا چند تا از این روشها برای اغلب مسائلی که دانشجویان دوره لیسانس علوم و مهندسی با آن مواجه اند نتایج رضایت بخشی بدست می دهد. مطالب مورد بحث در این بخش دلیل این است که چرا از این روشها انتظار حصول نتایج رضایت بخش می رود درحالی که روشهای مشابه دیگر چنین نیستند.

در بحث روش تک گامی معروف به روش اویلر، در بخش ۲.۶، تعاریف کاملاً کلی برای همگرایی و سازگار بودن آن ارائه شد. برای روشهای تک گامی تعریف سازگاری یک روش دقیقاً وقتی برقرار است که معادله تفاضلی روش با میل کردن اندازه گام به صفر

به معادله دیفرانسیل نزدیک شود؛ یعنی، اگر اندازه گام به صفر نزدیک شود، خطای برشی موضعی به صفر نزدیک گردد. تعریف همگرایی دارای مفهوم مشابهی است، چون یک روش دقیقاً وقتی همگراست که جواب معادله تفاضلی با نزدیک شدن اندازه گام به صفر به جواب معادله دیفرانسیل نزدیک شود.

نوع دیگری از کران خطای مسئله، که وقتی موجود است که برای تقریب جوابهای معادلات دیفرانسیل از روشهای تفاضلی استفاده می‌شود، از عدم استفاده از نتایج دقیق منتج می‌شوند. در عمل، نه شرایط اولیه و نه حسابی که متعاقباً انجام می‌شود، به علت خطای گرد کردن وابسته به حساب ارقام متناهی، دقیقاً "ارائه نمی‌شوند. در بخش ۲.۰۶ دیدیم که این امر حتی در مورد روشهای سازگار و همگرا به مشکلاتی منجر می‌شود. برای آنکه لااقل این مطلب جزئاً مورد تحلیل قرار گیرد، سعی می‌کنیم روشهای پایدار، به این مفهوم که تغییرات یا انحرافهای جزئی در شرایط اولیه متناظراً به تغییرات کوچکی در تقریبهای بعدی منجر می‌شود، را مشخص کنیم؛ یعنی، یک روش پایدار روشی است که به‌طور پیوسته به داده‌های اولیه بستگی دارد.

چون این مفهوم از پایداری یک معادله تفاضلی تک گامی بنحوی مشابه شرط خوش وضعی یک معادله دیفرانسیل است، تعجب آور نیست که در اینجا شرط لیب شیتس ظاهر شود، همانطور که در قضیه متناظر برای معادلات دیفرانسیل، یعنی قضیه ۶.۶ (صفحه ۲۰۶)، آمد. اثبات این نتیجه مشکل نیست و در تمرین ۱ ملاحظه می‌شود.

قضیه ۱۸.۰۶. هرگاه مسئله مقدار اولیه

$$(106.06) \quad y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

با روش تفاضلی تک گامی به شکل

$$(107.06) \quad \begin{aligned} w_0 &= \alpha, \\ w_{i+1} &= w_i + h\phi(t_i, w_i, h) \end{aligned}$$

تقریب شود و $\phi(t, w, h)$ با متغیر w بر مجموعه

$$D = \{(t, w, h) \mid a \leq t \leq b, \quad -\infty < w < \infty, \quad 0 \leq h \leq h_0\}$$

در شرط لیب شیتس صدق کند، آنگاه این روش پایدار است.

قضایای مشابهی نیز وجود دارند که شرایط کافی برای همگرایی و سازگار بودن روش

تک گامی را بدست می دهند. اثبات قضیه زیر را می توان در صفحات ۵۸-۵۷ از مطالب ارائه شده توسط گیر [۳۰] یافت.

قضیه ۱۹.۶. هرگاه مسئله مقدار اولیه

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha,$$

با روش تفاضلی تک گامی به شکل

$$w_0 = \alpha,$$

$$w_{i+1} = w_i + h\phi(t_i, w_i, h)$$

تقریب شود و $\phi(t, w, h)$ با متغیر w بر مجموعه

$$D = \{(t, w, h) | a \leq t \leq b, \quad -\infty < w < \infty, \quad 0 \leq h \leq h_0\},$$

پیوسته و در شرط لیب شیتس صدق کند، آنگاه

(یک) روش تفاضلی همگراست اگر و فقط اگر سازگار باشد، یعنی، اگر و فقط اگر

$$\phi(t, y, 0) = f(t, y), \quad a \leq t \leq b$$

(دو) اگر خطای برشی موضعی τ_i به ازای هر $i = 1, 2, \dots, N$ در $|\tau_i| \leq \tau(h)$ صدق کند،

که در آن $0 \leq h \leq h_0$ ، آنگاه به ازای هر $i = 1, 2, \dots, N$

$$|y(t_i) - w_i| \leq \frac{\tau(h)}{L} e^{L(t_i - a)}$$

که در آن L ثابت لیب شیتس است.

توجه کنید که قسمت (دو) این قضیه تبصره داده شده در بخش ۵.۶ در باب کنترل

خطای جامع یک روش با کنترل خطای موضعی آن را توجیه می کند و نتیجه می دهد که،

وقتی خطای برشی موضعی به میزان همگرایی $O(h^n)$ باشد، خطای جامع همان میزان

همگرایی را دارد. به عنوان روشی از نوع روشهای تک گامی مطرح شده، روش پیراسته

اویلر را از حیث همگرایی و پایداری مورد تحلیل قرار می دهیم.

مثال ۱. روش پیراسته اویلر به صورت زیر را در نظر می گیریم:

$$w_0 = \alpha,$$

$$w_{i+1} = w_i + \frac{h}{2} [f(t_i, w_i) + f(t_{i+1}, w_i + hf(t_i, w_i))], \quad i = 0, 1, \dots, N-1$$

برای این روش

$$^*\phi(t, w, h) = \frac{1}{2}f(t, w) + \frac{1}{2}f(t + h, w + hf(t, w)).$$

با قرار دادن $h = 0$ ،

$$\phi(t, w, 0) = \frac{1}{2}f(t, w) + \frac{1}{2}f(t + 0, w + 0 \cdot f(t, w)) = f(t, w),$$

در نتیجه ، شرط سازگاری در قسمت (یک) قضیه ۱۹.۶ برقرار است .

هرگاه f در شرط لیب شیتس بر $\{(t, w) | a \leq t \leq b, -\infty < w < \infty\}$ با ثابت

L صدق کند ، آنگاه ، چون

$$\begin{aligned} \phi(t, w, h) - \phi(t, \bar{w}, h) &= \frac{1}{2}f(t, w) + \frac{1}{2}f(t + h, w + hf(t, w)) \\ &\quad - \frac{1}{2}f(t, \bar{w}) - \frac{1}{2}f(t + h, \bar{w} + hf(t, \bar{w})), \end{aligned}$$

شرط لیب شیتس بر f منجر می شود به

$$\begin{aligned} |\phi(t, w, h) - \phi(t, \bar{w}, h)| &\leq \frac{1}{2}L|w - \bar{w}| \\ &\quad + \frac{1}{2}L|w + hf(t, w) - \bar{w} - hf(t, \bar{w})| \\ &\leq L|w - \bar{w}| + \frac{1}{2}L|hf(t, w) - hf(t, \bar{w})| \\ &\leq L|w - \bar{w}| + \frac{1}{2}hL^2|w - \bar{w}| \\ &= (L + \frac{1}{2}hL^2)|w - \bar{w}|. \end{aligned}$$

از اینرو ، ϕ با متغیر w بر مجموعه

$$\{(t, w, h) | a \leq t \leq b, -\infty < w < \infty, 0 \leq h \leq h_0\}$$

بمازای هر $h_0 > 0$ با ثابت

$$L' = (L + \frac{1}{2}h_0L^2)$$

در شرط لیب شیتس صدق می کند .

بالاخره هرگاه f بر $\{(t, w) | a \leq t \leq b, -\infty < w < \infty\}$ پیوسته باشد ، آنگاه ϕ بر

$$\{(t, w, h) | a \leq t \leq b, -\infty < w < \infty, 0 \leq h \leq h_0\}$$

پیوسته است ؛ لذا ، از قضایای ۱۸.۶ و ۱۹.۶ همگرایی و پایداری روش پیراسته ، اوپلر

نتیجه می شود . بعلاوه ، دیده ایم که برای این روش ، خطای برشی موضعی برابر $O(h^2)$

است . بنابراین ، همگرایی روش پیراسته ، اوپلر نیز به میزان $O(h^2)$ می باشد .

مسائل مربوط به سازگاری ، همگرایی ، و پایداری روشهای چندگامی ، به علت تعدد

تقریبهای مربوط به هرگام ترکیب می شوند . در روشهای تک گامی ، تقریب w_{i+1} مستقیماً

فقط به تقریب قبلی w_i بستگی دارد، در حالی که در روشهای چندگامی لااقل دو تقریب قبلی بکار می‌روند و در روشهایی که عموماً " بکار می‌روند بیشتر. پیش از بحث در ماهیت مسائل روشهای چندگامی، مناسب است که آن روشها در زمینهٔ مختصر متفاوتی ارائه شوند.

روش چندگامی عمومی برای تقریب جواب مسئلهٔ مقدار اولیهٔ

$$(108.06) \quad y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha,$$

را می‌توان به شکل

$$w_0 = \alpha, \quad w_1 = \alpha_1, \quad \dots, \quad w_{m-1} = \alpha_{m-1}.$$

$$(109.06) \quad w_{i+1} + a_{m-1}w_i + a_{m-2}w_{i-1} + \dots + a_0w_{i+1-m}$$

$$= hF(t_i, h, w_{i+1}, w_i, \dots, w_{i+1-m})$$

به‌ازای هر $i = m-1, m, \dots, N-1$ نوشت، که در آن a_0, a_1, \dots, a_{m-1} مقادیر ثابت و، مثل همیشه، $t_i = a + ih$ و $h = (b-a)/N$.

اگر F شامل w_{i+1} نباشد، این روش صریح است؛ در غیر این صورت، روش ضمنی است. روش مرتبهٔ چهار آدامز-بشفورث را می‌توان با انتخاب $a_0 = 0, a_1 = 0, a_2 = 0, a_3 = -1$ و

$$\begin{aligned} F(t_i, h, w_{i+1}, w_i, \dots, w_{i-3}) &= \frac{1}{24}[55f(t_i, w_i) - 59f(t_{i-1}, w_{i-1}) \\ &+ 37f(t_{i-2}, w_{i-2}) \\ &- 9f(t_{i-3}, w_{i-3})] \end{aligned}$$

به شکل معادلهٔ (109.06) بیان کرد، زیرا این روش عبارت است از

$$\begin{aligned} w_{i+1} - w_i &= \frac{h}{24} [55f(t_i, w_i) - 59f(t_{i-1}, w_{i-1}) \\ &+ 37f(t_{i-2}, w_{i-2}) - 9f(t_{i-3}, w_{i-3})]. \end{aligned}$$

به‌همین نحو، روش مرتبهٔ چهار آدامز-مولتون را می‌توان با انتخاب $a_0 = 0, a_1 = 0, a_2 = -1$ و

$$\begin{aligned} F(t_i, h, w_{i+1}, \dots, w_{i-2}) &= \frac{1}{24}[9f(t_{i+1}, w_{i+1}) \\ &+ 19f(t_i, w_i) - 5f(t_{i-1}, w_{i-1}) \\ &+ f(t_{i-2}, w_{i-2})]. \end{aligned}$$

به شکل معادله (۱۰۹.۶) بیان کرد.

در طول این تحلیل دو فرض در مورد تابع F می توان کرد: اول اینکه، هرگاه $f \equiv 0$ (یعنی، معادله دیفرانسیل همگن باشد)، آنگاه نیز $F \equiv 0$ ؛ و دوم اینکه، F با در نظر گرفتن دنباله $\{w_j\}_{j=i+1-m}^{i+1}$ در نوعی شرط لیبشیتس صدق می کند، به این معنی که، برای تابع معین f مقدار ثابت C وجود دارد بطوری که به ازای هر $i = m - 1, m, \dots, N$

$$\begin{aligned} & \{v_j\}_{j=0}^N \text{ و } \{w_j\}_{j=0}^N \\ & |F(t_i, h, w_{i+1}, \dots, w_{i+1-m}) - F(t_i, h, v_{i+1}, \dots, v_{i+1-m})| \\ & \leq C \sum_{j=0}^m |w_{i+1-j} - v_{i+1-j}| \end{aligned}$$

اینکه روشهای آدامز - بشفورث و آدامز - مولتون، مشروط بر اینکه f در شرط لیبشیتس صدق کند، در هر دو شرط صدق می کنند در تعریف آ۲ آمده است.

خطای برشی موضعی روش چندگامی ارائه شده به شکل معادله (۱۹.۶) به ازای هر $i = m - 1; m, \dots, N - 1$ برابر

$$\begin{aligned} \tau_{i+1} = & \frac{y(t_{i+1}) + a_{m-1}y(t_i) + \dots + a_0y(t_{i+1-m})}{h} \\ & - F(t_i, h, y(t_{i+1}), y(t_i), \dots, y(t_{i+1-m})) \end{aligned}$$

است و، مانند روشهای تک گامی، با چگونگی عدم صدق جواب y از معادله دیفرانسیل در معادله تفاضلی اندازه گیری می شود.

در مورد روش چهارگامی آدامز - بشفورث، دیده ایم که

$$\tau_{i+1} = \frac{251}{720} y^{(5)}(\mu_i) h^4, \quad t_{i-3} < \mu_i < t_{i+1} \text{ به ازای}$$

درحالی که خطای برشی موضعی ملاحظه شده برای روش آدامز - مولتون برابر

$$\tau_{i+1} = \frac{-19}{720} y^{(5)}(\mu_i) h^4, \quad t_{i-2} < \mu_i < t_{i+1}$$

است، البته به شرط اینکه $y \in C^5[a, b]$.

مفهوم همگرایی روشهای چندگامی نظیر روشهای تک گامی است؛ یعنی، یک روش چندگامی همگراست اگر جواب معادله تفاضلی، وقتی اندازه گام به صفر نزدیک می شود،

به جواب معادله دیفرانسیل نزدیک شود. این یعنی

$$\lim_{h \rightarrow 0} \max_{0 \leq i \leq N} |w_i - y(t_i)| = 0.$$

لیکن، در مورد سازگاری، وضعیت تقریباً متفاوتی رخ می‌دهد. مجدداً، روشی مورد نظر است که سازگار باشد، مشروط بر اینکه با نزدیک شدن اندازه گام به صفر معادله تفاضلی به معادله دیفرانسیل نزدیک شود؛ یعنی، خطای برشی موضعی با نزدیک شدن اندازه گام به صفر باید در هر گام به صفر میل کند. به علت تعدد مقادیر اولیه لازم برای روشهای چندگامی، شرط اضافی پیش می‌آید. چون انتظار می‌رود که حداکثر اولین مقدار اولیه، یعنی w_0 ، دقیق باشد، علاوه بر آن لازم است خواسته شود که خطاها در تمام مقادیر اولیه، با نزدیک شدن اندازه گام به صفر، به صفر نزدیک شوند؛ یعنی، برای آنکه یک روش چندگامی به شکل معادله (۱۰۹.۶) سازگار باشد، باید

$$\lim_{h \rightarrow 0} |\tau_i| = 0, \quad i = m, m+1, \dots, N$$

و

$$\lim_{h \rightarrow 0} |\alpha_i - y(t_i)| = 0, \quad i = 0, 1, \dots, m-1$$

درست باشند.

پیش از بحث درباره ارتباط بین سازگاری، همگرایی، و پایداری روشهای چندگامی، باید معادله تفاضلی یک روش چندگامی را با تفصیل بیشتری در نظر گرفت.

به معادله تفاضلی ارائه شده در (۱۰۹.۶)

$$w_0 = \alpha, \quad w_1 = \alpha_1, \quad \dots, \quad w_{m-1} = \alpha_{m-1},$$

$$w_{i+1} + a_{m-1}w_i + a_{m-2}w_{i-1} + \dots + a_0w_{i+1-m}$$

$$= hF(t_i, h, w_{i+1}, w_i, \dots, w_{i+1-m}),$$

یک چندجمله‌ای، به نام چندجمله‌ای مشخص آن روش، وابسته می‌شود، که با

$$(110.6) \quad p(\lambda) = \lambda^m + a_{m-1}\lambda^{m-1} + a_{m-2}\lambda^{m-2} + \dots + a_1\lambda + a_0.$$

داده می‌شود.

مشاهده اینک که p دارای یک ریشه در β است دقیقاً اگر به‌زای هر n

یک جواب معادله تفاضلی، در حالت $F \equiv 0$ ، باشد، مشکل نیست (تمرین ۳). به این دلیل است که ریشه‌های p جالب توجه‌اند.

تعریف ۲۰۰۶. فرض کنیم $\lambda_1, \lambda_2, \dots, \lambda_m$ ریشه‌های (نه الزاماً متمایز) چندجمله‌ای

مشخص

$$p(\lambda) = \lambda^m + a_{m-1}\lambda^{m-1} + \dots + a_1\lambda + a_0$$

وابسته به روش تفاضلی چندگامی

$$w_0 = \alpha, w_1 = \alpha_1, \dots, w_{m-1} = \alpha_{m-1}$$

$$w_{i+1} + a_{m-1}w_i + a_{m-2}w_{i-1} + \dots + a_0w_{i+1-m}$$

$$= hF(t_i, h, w_{i+1}, w_i, \dots, w_{i+1-m})$$

را مشخص کنند. هرگاه به ازای هر $i = 1, 2, \dots, m$ ، $|\lambda_i| \leq 1$ و ریشه‌های با قدرمطلق i ریشه‌های ساده باشند، آنگاه گوییم روش تفاضلی در شرط ریشه‌های صدق می‌کند.

مثال ۲. ملاحظه شد که روش مرتبه چهار آدامز - بشفورث را می‌توان به شکل

$$w_{i+1} - w_i = hF(t_i, h, w_{i+1}, w_i, \dots, w_{i-3})$$

بیان کرد، که در آن

$$F(t_i, h, w_{i+1}, w_i, \dots, w_{i-3}) = \frac{1}{24}[55f(t_i, w_i) - 59f(t_{i-1}, w_{i-1}) \\ + 37f(t_{i-2}, w_{i-2}) \\ - 9f(t_{i-3}, w_{i-3})];$$

بطوری که $a_3 = -1$ و $a_2 = 0$ ، $a_1 = 0$ ، $a_0 = 0$ ، $m = 4$

در نتیجه، چند جمله‌ای مشخص روش آدامز - بشفورث

$$p(\lambda) = \lambda^4 - \lambda^3 = \lambda^3(\lambda - 1),$$

است، که دارای ریشه‌های $\lambda_4 = 1$ و $\lambda_3 = 0$ ، $\lambda_2 = 0$ ، $\lambda_1 = 0$ می‌باشد و در شرط ریشه‌های صدق می‌کند.

روش آدامز - مولتون دارای چند جمله‌ای مشخص مشابهی است، یعنی $P(\lambda) = \lambda^3 - \lambda^2$

و نیز در شرط ریشه‌های صدق می‌کند.

اهمیت شرط ریشه‌ای در ارتباط آن با سازگاری، همگرایی، و پایداری یک روش است.

پایداری یک روش چندگامی به این معنی است که، به ازای اندازه‌های گام به قدر کافی کوچک، انحرافات کوچکی در یک یا همه مقادیر اولیه فقط منجر به انحرافات کوچکی در مقادیر بعدی می‌شود. برای اثبات این مطلب و نظریه متکی بر آن، خواننده را مصرا"

به ایزاکسون و کالر [۳۹] ارجاع می‌دهیم.

قضیه ۲۱.۶. یک روش چندگامی به شکل

$$w_0 = \alpha, \quad w_1 = \alpha_1, \quad \dots, \quad w_{m-1} = \alpha_{m-1}$$

و

$$w_{i+1} + a_{m-1}w_i + a_{m-2}w_{i-1} + \dots + a_0w_{i+1-m} = hF(t_i, h, w_{i+1}, w_i, \dots, w_{i+1-m})$$

پایدار است اگر و فقط اگر در شرط ریشه‌های صدق کند. بعلاوه، هرگاه روش تفاضلی با معادله دیفرانسیل سازگار باشد، آنگاه این روش پایدار است اگر و فقط اگر همگرا باشد.

مثال ۳. روش چندگامی ضمنی ارائه شده با

$$w_{i+1} - w_{i-1} = \frac{h}{3} [f(t_{i+1}, w_{i+1}) + 4f(t_i, w_i) + f(t_{i-1}, w_{i-1})]$$

در بخش ۶.۶ به عنوان روش مرتبه چهار سیمپسون معرفی شد. چون چند جمله‌ای مشخص این روش، یعنی $p(\lambda) = \lambda^2 - 1$ ، دارای ریشه‌های $\lambda_1 = 1$ و $\lambda_2 = -1$ است، این روش در شرط ریشه‌های صدق می‌کند و روشی پایدار است. لیکن، همانطور که با در نظر گرفتن مسئله مقدار اولیه

$$(111.06) \quad y' = -6y + 6, \quad 0 \leq t \leq 1.5, \quad y(0) = 2$$

ملاحظه می‌شود، این امر ایجاب نمی‌کند که این روش همیشه تقریبهای دقیقی بدست دهد. با $h = 0.1$ و $t_i = 0.1i$ ، این روش را می‌توان به‌طور صریح به شکل

$$w_{i+1} = \frac{-0.8w_i + 0.8w_{i-1} + 1.2}{1.2}$$

به‌ازای $i = 1, 2, \dots, 9$ بیان کرد. با استفاده از مقادیر اولیه $w_0 = 2$ و $w_1 = e^{-0.6} + 1$ ، مقادیر مربوطه و مقادیر دقیق جواب $y = e^{-6t} + 1$ در جدول ۱۴.۶ ارائه شده‌اند.

یکی از دلایل خطای بیشتر در این مثال این است که، درحالی که روش سیمپسون پایدار است، معادله مشخص آن دارای دو ریشه با بزرگی ۱ است. این امر سبب می‌شود که این روش برای معادلات دیفرانسیل از روشهایی که فقط دارای یک ریشه مشخص به

t_i	w_i	y_i	$ w_i - y_i $
0	—	2.0000000	—
.1	—	1.5488116	—
.2	1.3007923	1.3011942	4.019×10^{-4}
.3	1.1653462	1.1652989	4.731×10^{-5}
.4	1.0902974	1.0907۸80	4.206×10^{-4}
.5	1.0500325	1.0497871	2.454×10^{-4}
.6	1.0268433	1.0273237	4.864×10^{-4}
.7	1.0154595	1.0149956	4.639×10^{-4}
.8	1.0075892	1.0082298	6.405×10^{-4}
.9	1.0052469	1.0045166	7.303×10^{-4}
1.0	1.0015615	1.0024788	9.173×10^{-4}
1.1	1.0024569	1.0013604	1.097×10^{-3}
1.2	.9994031	1.0007466	1.343×10^{-3}
1.3	1.0020359	1.0004097	1.626×10^{-3}
1.4	.9978060	1.0002249	2.419×10^{-3}
1.5	1.0028199	1.0001234	2.696×10^{-3}

جدول ۱۴.۶

بزرگی 1 است کمتر پایدار باشد. به این دلیل، روشهایی که در شرط ریشه‌های صدق می‌کنند فقط دارای یک ریشه به بزرگی 1 هستند، یعنی $\lambda = 1$ ، را به طور قوی پایدار گوئیم، در حالی که روشهایی با بیشتر از یک ریشه به بزرگی 1 را به طور ضعیف پایدار می‌نامیم. دلیل انتخاب روش آدامز - بشفورث - مولتون، در بخش ۶.۶، به عنوان روش پیشگو - اصلاحگر مرتبه چهار به جای روش میلن - سیمپسون از همان مرتبه، این است که هر دو روش آدامز - بشفورث و آدامز - مولتون روشهایی به طور قوی پایدارند؛ و در نتیجه، تقریبهای دقیقتری را به دسته گسترده‌تری از مسائل، نسبت به پیشگو - اصلاحگرهای منکی بر روشهای سیمپسون ر میلن که هر دو به طور ضعیف پایدارند، ارائه می‌دهند.

معادله دیفرانسیل موجود در مثال ۳ از نوع معادله "سخت" است، معادله‌ای که جواب دقیق آن شامل تابعی به شکل e^{2t} ، به ازای مقدار ثابت λ ، است. معادلاتی از این نوع اغلب در عمل، بخصوص در تحلیل دستگاههای کنترل معین و در مسائل جنبشی شیمی، ظاهر می‌شوند، و مخصوصاً وقتی λ منفی و بزرگی آن زیاد باشد باعث دردسر هستند، زیرا بخش e^{2t} از جواب به سرعت به صفر تنزل می‌کند. جهت بررسی مشکل مورد

بحث، وضعیتی را که با استفاده از روش تک گامی ساده معروف به روش اویلر در مسئله مقدار اولیه

$$(112.06) \quad y' = \lambda y, \quad y(0) = \alpha$$

پیش می آید در نظر می گیریم. این معادله را اغلب معادله "آزمونی" می نامیم، چونکه ساده ترین نوع معادله دیفرانسیل با جوابی به شکل $e^{\lambda t}$ است و، اگر آنچه را که روشهای متعدد تفاضلی در این مسئله انجام می دهند مشخص کنیم، معیاری خوب از وضعیتی که برای حالت کلی معادله "سخت" ظاهر می شود نتیجه خواهد شد.

برای روش اویلر داریم

$$w_0 = \alpha$$

و، به ازای $i = 0, 1, 2, \dots$

$$w_{i+1} = w_i + hf(t_i, w_i)$$

$$= w_i + h\lambda w_i$$

$$= (1 + h\lambda)w_i,$$

فرض کنیم یک خطای δ در مرحله z م روش تفاضلی حادث شود. احتمالاً گرد کردن به ارقام متناهی موجب این خطاست که به تعویض w_j با $w_j + \delta$ در محاسبه بعدی منجر می شود. از اینرو،

$$w_{j+1} = (1 + h\lambda)(w_j + \delta),$$

$$w_{j+2} = (1 + h\lambda)^2(w_j + \delta);$$

و در حالت کلی،

$$w_{j+m} = (1 + h\lambda)^m(w_j + \delta)$$

در نتیجه، خطای حاصل از خطا در w_j برابر $(1 + h\lambda)^m \delta$ است.

اگر $\lambda > 0$ ، جواب معادله (۱۱۲.۰۶)، یعنی $y(t) = \alpha e^{\lambda t}$ ، به طور نمایی افزایش می یابد و احتمالاً این خطا قابل توجه نیست. اما، اگر $\lambda < 0$ ، $y(t) = \alpha e^{\lambda t}$ به صفر نزدیک می شود و، بخصوص، اگر $|1 + \lambda h| > 1$ (یعنی، اگر $h > -2/\lambda$)، خطا به طور نمایی افزایش یافته و بر تقریب چیره می شود.

در مورد سایر روشهای تک گامی وضعیت مشابهی برقرار است که در آن یک تابع $Q(h)$ با این خاصیت وجود دارد که معادله تفاضلی، بکار رفته در معادله آزمونی، عبارت

است از

$$w_{i+1} = Q(\lambda h)w_i$$

و، مجدداً، وقوع تقریبهای نادقیقی، بجز وقتی $|Q(\lambda h)| < 1$ ، انتظار می‌رود. اکثراً در مسائل مورد بحث دستگامهای معادلات دیفرانسیل "سختی" ظاهر می‌شود، و در این وضعیت در نظر گرفتن معادله آزمون که در آن λ عددی مختلط با قسمت حقیقی منفی است طبیعی‌تر می‌باشد. در این حالت، ناحیه‌ای از صفحه مختلط که برای آن $|Q(z)| < 1$ ، که $z = \lambda h$ ، ناحیه پایداری مطلق روش نامیده می‌شود. با در نظر گرفتن مقداری از λ ، این ناحیه مقادیری از h را تعیین می‌کند که برای آنها انتظار می‌رود روش تقریبهای دقیقی را بدست دهد. روشی برای رسیدگی به مسائل شامل معادلات سخت مناسبتر است که ناحیه پایداری مطلق آن وسیعتر باشد. وضعیت روشهای چندگامی مشابه ولی پیچیده‌تر است، زیرا روش تفاضلی را نمی‌توان به سادگی روشی چون معادله (۱۱۳.۶) بیان کرد.

برای یک مسئله مقدار اولیه کلی

$$(114.6) \quad y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha$$

وقتی انتظار "سختی" می‌رود که $\partial f / \partial y$ در زیربازه‌ای از $[a, b]$ منفی باشد. در این حالت، باید روش تفاضلی انتخاب شود که دارای وسیعترین ناحیه پایداری مطلق است. بخصوص، روش ذوزنقه‌ای

$$w_0 = \alpha,$$

$$(115.6) \quad w_{i+1} = w_i + \frac{h}{2} [f(t_{i+1}, w_{i+1}) + f(t_i, w_i)],$$

روشی مرتبه دوم است که در تمام نیم صفحه چپ به طور مطلق پایدار است، و معمولاً چنین روشی مناسب است.

مثال ۴. اگر روش ذوزنقه‌ای (۱۱۵.۶) را در مورد مسئله مقدار اولیه

$$y' = -6y + 6, \quad 0 \leq t \leq 1.5, \quad y(0) = 1$$

که در مثال ۳ با $h = 1$ آمده بود، بکار ببریم، نتایج موجود در جدول ۱۵.۶ حاصل می‌شود.

گرچه این تقریبها کاملاً دقیق نیستند، ولی همان است که با استفاده از $h = 1$ و یک روش مرتبه دوم انتظار می‌رود، و واضح است که این روش پایدار می‌باشد. می‌توان

t_i	w_i	y_i	قدر مطلق خطا
0	—	2.0000000	—
.1	1.5384615	1.5488116	1.04×10^{-2}
.2	1.2899408	1.3011942	1.13×10^{-2}
.3	1.1561220	1.1652989	9.18×10^{-3}
.4	1.0840657	1.0907180	6.65×10^{-3}
.5	1.0452661	1.0497871	4.52×10^{-3}
.6	1.0243741	1.0273237	2.95×10^{-3}
.7	1.0131245	1.0149956	1.87×10^{-3}
.8	1.0070670	1.0082298	1.16×10^{-3}
.9	1.0038053	1.0045166	7.11×10^{-4}
1.0	1.0020490	1.0024788	4.30×10^{-4}
1.1	1.0011033	1.0013604	2.57×10^{-4}
1.2	1.0005941	1.0007466	1.53×10^{-4}
1.3	1.0003199	1.0004097	8.98×10^{-5}
1.4	1.0001723	1.0002249	5.26×10^{-5}
1.5	1.0000928	1.0001234	3.06×10^{-5}

جدول ۱۵-۶

نشان داد که روشهای چندگامی با مرتبه بزرگتر از ۲ در تمام نیم صفحه چپ به طور مطلق پایدار نیستند.

در اینجا برای خواننده‌های که اغلب انتظار مواجهه با معادلات "سخت" را دارند فقط مقدار کمی از آنچه باید بدانند ارائه شده است. در این صورت، توصیه می‌شود که برای بحث بیشتری در این مطلب گیر [۳۵] مورد استفاده قرار گیرد.

در خاتمه این فصل، به نظر مناسب می‌آید که توصیه‌هایی در باب استفاده از روشها جهت حل مسائل مقدار اولیه ارائه شود. اغلب این توصیه‌ها متکی بر نتایج ارائه شده در مقالات هول^۱، انرایت^۲، فلن^۳، و سجویک^۴ [۳۸]، منتشر شده در سال ۱۹۷۲، و در مقالات در مقالات انرایت، هول، و لیندبرگ^۵ [۲۳]، منتشر شده در سال ۱۹۷۵، و در مقالات هول و انرایت [۳۲]، منتشر شده در سال ۱۹۷۶ است. محتمل است که هر مطلب جدیدی

1. Hull

2. Enright

3. Fellen

4. Sedgwick

5. Lindberg

در مورد این نتایج در ژورنال آنالیز عددی SIAM ظاهر شود.

وقتی تابعی که باید حساب شود نسبتاً ساده باشد؛ یعنی، نیاز به محاسبات خیلی وقت گیر نداشته باشد، اغلب یک روند برونمایی نظیر الگوریتم ۶.۶ موثرتر است، در حالی که اگر محاسبه تابع پیچیده باشد، روش پیشگو - اصلاحگر آدامز مطلوب است. کاربرد روندهای رونگ - کوتا را باید به تعیین مقادیر اولیه در روش آدامز، یا به مسائلی که در آنها تابع به سادگی محاسبه می شود و دقت مورد نیاز کم، در حدود 10^{-4} ، است محدود کرد. در مورد مسائل مشکوک به "سخت بودن"، روش ذوزنقه ای مناسب است، اگرچه روشهای ماهرانه تری نظیر آنچه درگیر [۳۰] آمده است (بخصوص، مطالب مندرج در صفحات ۱۶۶ - ۱۵۸) وجود دارند.

مجموعه تمرینات ۹.۶

۱. برای اثبات قضیه ۱۸.۶، نشان دهید مفروضات وجود یک مقدار ثابت $K > 0$ را ایجاب می کنند بطوری که هر وقت $\{u_i\}_{i=1}^N$ و $\{v_i\}_{i=1}^N$ در (۱۰۷.۶) صدق کنند،

$$|u_i - v_i| \leq K |u_0 - v_0|, \quad 1 \leq i \leq N$$

۲. نشان دهید که برای روشهای آدامز - بشفورث و آدامز - مولتون یک مقدار ثابت وجود دارد بطوری که

$$F(t_i, h, w_{i+1}, \dots, w_{i+1-m}) = 0, \quad f \equiv 0$$

و

$$|F(t_i, h, w_{i+1}, \dots, w_{i+1-m}) - F(t_i, h, v_{i+1}, \dots, v_{i+1-m})|$$

$$\leq C \sum_{j=0}^m |w_{i+1-j} - v_{i+1-j}|$$

۳. نشان دهید که $w_n = \beta^n$ در

$$w_{i+1} + a_{m-1}w_i + \dots + a_0w_{i+1-m} = 0, \quad i = m-1, \dots, N$$

صدق می کند اگر و فقط اگر $p(\beta) = 0$ ، که در آن

$$p(\lambda) = \lambda^m + a_{m-1}\lambda^{m-1} + \dots + a_0.$$

۴. نشان دهید که روش مرتبه چهار رونگ - کوتا پایدار است.

۵. معادله دیفرانسیل

$$y' = f(t, y), \quad a \leq t \leq b,$$

$$y(a) = \alpha.$$

را در نظر بگیرید

(T) نشان دهید به ازای ξ_i ی، که $t_i < \xi_i < t_{i+2}$ ،

$$y'(t_i) = \frac{-3y(t_i) + 4y(t_{i+1}) - y(t_{i+2}))}{2h} + \frac{h^2}{3} y'''(\xi_i)$$

(ب) قسمت (T) معادله تفاضلی

$$w_{j+2} = 4w_{j+1} - 3w_j - 2hf(t_j, w_j) \quad j = 0, 1, \dots, N-2$$

را پیشنهاد می کند. با استفاده از این روش،

$$y' = 1 - y, \quad 0 \leq t \leq 1$$

$$y(0) = 0,$$

را وقتی $h = 1$ حل کنید. مقادیر اولیه $w_0 = 0$ و $w_1 = y(t_1) = 1 - e^{-1}$ را بکار برید.

(پ) قسمت (ب) را با $h = 0.01$ و $w_1 = 1 - e^{-0.01}$ تکرار کنید.

(ت) این روش را از حیث سازگاری، پایداری، و همگرایی تحلیل کنید.

۶. روش چندگامی

به ازای هر $i = 2, \dots, N-1$ ، $w_{i+1} = -\frac{3}{2}w_i + 3w_{i-1} - \frac{1}{2}w_{i-2} + 3hf(t_i, w_i)$ ، مقادیر اولیه w_0, w_1, w_2 مفروض است.

(T) خطای برشی موضعی را بیابید.

(ب) سازگاری، پایداری، و همگرایی را مورد تفسیر قرار دهید.

۷. یک جواب تقریبی معادله دیفرانسیل

$$y' = -y, \quad 0 \leq t \leq 10, \quad y(0) = 1$$

را با استفاده از روش میلن، ابتدا با $h = 1$ و سپس با $h = 0.01$ و مقادیر اولیه

$w_0 = 1$ و $w_1 = e^{-h}$ در هر دو حالت، بدست آورید. آیا تنزل h از $h = 1$ به $h = 0.01$

در تعداد ارقام درست جوابهای تقریبی در $t = 1$ و $t = 10$ اثر دارد؟

۸. در سازگاری، پایداری، و همگرایی روش ذوزنقه‌ای زیر بحث کنید:

$$w_{i+1} = w_i + \frac{h}{2} [f(t_{i+1}, w_{i+1}) + f(t_i, w_i)], \quad i = 0, 1, \dots, N-1$$

با $w_0 = \alpha$ اعمال شده بر معادله دیفرانسیل

$$y' = f(t, y), \quad a \leq t \leq b, \quad y(a) = \alpha.$$

۹. پایداری روش تفاضلی زیر را بررسی کنید:

$$w_{i+1} + 4w_i - 5w_{i-1} = 2h[f(t_i, w_i) + 2hf(t_{i-1}, w_{i-1})], \quad i = 1, 2, \dots, N-1$$

بمازای w_0, w_1 با مقادیر اولیه^۶

۱۰. مسئله^۶

$$y' = 0, \quad 0 \leq t \leq 10, \quad y(0) = 0$$

را در نظر بگیرید، که جواب $y \equiv 0$ را دارد.

(آ) هرگاه روش تفاضلی تمرین ۵ بر مسئله اعمال شود، آنگاه

$$w_{i+1} = 4w_i - 3w_{i-1}, \quad i = 1, 2, \dots, N-1$$

$$w_0 = 0,$$

$$w_1 = \alpha_1.$$

فرض کنید $\varepsilon = w_1 = \alpha_1$ ، که در آن ε یک خطای گرد کردن کوچک است. با محاسبه^۶

دقیق w_i بمازای $i = 2, 3, \dots, 6$ ، نحوه^۶ افزایش ε را پیدا کنید.

(ب) قسمت (آ) را با استفاده از قاعده^۶ دوزنقه‌ای تکرار کنید.

۱۱. روش مرتبه^۶ چهار آدامز - مولتون بکار رفته در معادله^۶ دیفرانسیل

$$y' = \lambda y, \quad t \geq 0, \quad y(0) = 1$$

با

$$w_{i+1} = w_i + \frac{9h\lambda}{24} w_{i+1} + \frac{19h\lambda}{24} w_i - \frac{5h\lambda}{24} w_{i-1} + \frac{h\lambda}{24} w_{i-2}$$

ارائه می‌شود.

(آ) چند جمله‌ای مشخص این معادله^۶ تفاضلی را تعیین کنید.

(ب) هرگاه $\beta_1, \beta_2, \beta_3$ ریشه‌های این چندجمله‌ای مشخص باشند، آنگاه

جواب عمومی معادله^۶ تفاضلی عبارت است از

$$w_i = c_1 \beta_1^i + c_2 \beta_2^i + c_3 \beta_3^i$$

هرگاه بمازای z ای $|\beta_j| > 1$ ، آنگاه β_j^z با صعود i به‌طور نمایی رشد می‌کند، بطوری

که روش آدامز - مولتون ناپایدار می‌شود. ریشه‌های چندجمله‌ای مشخص را برای

مقادیر متعدد $h\lambda < 0$ تعیین کنید تا به‌طور آزمایشی یک مقدار h مشخص شود که

بمازای مقادیر بزرگتر از آن یک ریشه^۶ β_j با $|\beta_j| > 1$ وجود داشته باشد.

۱۲. تمرین ۱۱ را با استفاده از روش سیمپسون تکرار کنید.

۱۳. مسائل مقدار اولیه^۶ "سخت" زیر را با استفاده از روش اویلر، روش مرتبه^۶ چهار

رونک - کوتا، روش پیشگو - اصلاحگر آدامز - بشفورث و آدامز - مولتون، روش

دوزنقهای، و روش برونیاپی حل کنید.

$$y' = -20(y - t^2) + 2t, \quad 0 \leq t \leq 1, \quad (T)$$

$$y(0) = 0.$$

$h = .075$ و $h = .01$ را برای $0 \leq t \leq 2$ و $h = .075$ را برای $1 \leq t \leq 2$ بکار ببرید. جواب دقیق به صورت $y(t) = t^2 + \frac{1}{3}e^{-20t}$ ارائه می شود.

$$u_1'(t) = 32u_1 + 66u_2 + \frac{2}{3}t + \frac{2}{3}, \quad (-)$$

$$u_2'(t) = -66u_1 - 133u_2 - \frac{1}{3}t - \frac{1}{3}, \quad 0 \leq t \leq 1,$$

$$u_1(0) = \frac{1}{3}, \quad u_2(0) = \frac{1}{3}.$$

$h = .005$ را برای $0 \leq t \leq 2$ و $h = .1$ را برای $1 \leq t \leq 2$ بکار ببرید. جواب دقیق عبارت است از

$$u_1(t) = \frac{2}{3}t + \frac{2}{3}e^{-t} - \frac{1}{3}e^{-100t} \quad \text{و} \quad u_2(t) = -\frac{1}{3}t - \frac{1}{3}e^{-t} + \frac{2}{3}e^{-100t}.$$

$$y' = -20y + 20 \sin t + \cos t, \quad 0 \leq t \leq 1 \quad (P)$$

$$y(0) = 1.$$

$h = .01$ را برای $0 \leq t \leq 2$ و $h = .075$ را برای $1 \leq t \leq 2$ بکار ببرید. جواب دقیق عبارت است از

$$y(t) = e^{-20t} + \sin t.$$

$$y' = (50/y) - 50y, \quad 0 \leq t \leq 1, \quad (T)$$

$$y(0) = \sqrt{2}.$$

$h = .01$ را برای $0 \leq t \leq 2$ و $h = .1$ را برای $1 \leq t \leq 2$ بکار ببرید. جواب دقیق عبارت است از

$$y(t) = \sqrt{1 + e^{-100t}}.$$

در تمرین ۵ بخش ۲۰۶، معادله دیفرانسیل

$$\frac{dp(t)}{dt} = rb(1 - p(t))$$

به عنوان نمونه‌ای جهت مطالعه نسبت $p(t)$ از مخالفین کلیسای رسمی، در اجتماعی که میزان تولد آن h است، بدست آمد که در آن r نمایش میزان مولیدی است که مخالف کلیسا می‌شوند، در حالی که لااقل یکی از والدین آنها موافق است. در آن تمرین لازم شده که تقریبی برای $p(t)$ با استفاده از روش اویلر، یعنی الگوریتم ۱۰۶،

برای مقادیر صحیح t وقتی $p(0) = .01$ و $b = .02$ و $r = .1$ بدست‌آید و سپس تقریب $p(50)$ با مقدار واقعی مقایسه شد. علت خطای زیاد این تقریب را شرح دهید. با استفاده از روش دوزنقه‌ای، که در (۱۱۵.۰۶) بیان شد، تقریب دیگری برای $p(50)$ با فرض مجدد 1 سال $h =$ بدست‌آورید.

روشهای مستقیم حل دستگاههای خطی^۷

قوانین کیرشف درباره مدارهای الکتریکی بیان می‌کنند که برآیند فلوی جریان عبور کرده از هر اتصال یک مدار صفر است، و کل افت ولتاژ در طول هر حلقه بسته مدار صفر می‌باشد. اگر پتانسیل V ولت بین نقاط A و G در مدار شکل ۱۰۷ اعمال شود و v_b ، v_c ، v_d ، v_e ، v_f پتانسیلهایی بترتیب در نقاط B ، C ، D ، E ، F باشند، با بکار بردن G به‌عنوان نقطه بازگشت، قوانین کیرشف ایجاب می‌کنند که این پتانسیلها در دستگاه معادلات خطی زیر صدق کنند. حل دستگاههایی از این نوع در این فصل مورد بحث قرار می‌گیرد.

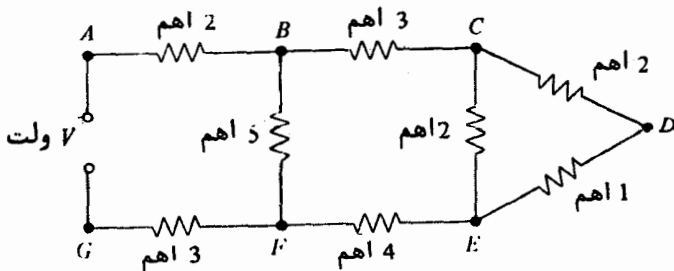
$$-31v_b + 10v_c + 6v_f = 15 \text{ V,}$$

$$2v_b - 8v_c + 3v_d + 3v_e = 0,$$

$$v_c - 3v_d + 2v_e = 0,$$

$$2v_c + 4v_d - 9v_e + v_f = 0,$$

$$12v_b + 15v_d - 47v_f = 0.$$



شکل ۱۰۷

حل دستگاههایی از این نوع در این فصل مورد بحث قرار می‌گیرد.

در این فصل، روش‌های مستقیم برای حل دستگاه‌های خطی را بررسی می‌کنیم. در این روش‌ها، ماتریس ضرایب را به یک ماتریس مثلثی تبدیل می‌کنیم و سپس با استفاده از روش حذف به عقب، مقادیر متغیرها را به دست می‌آوریم. این روش‌ها برای سیستم‌های خطی با ضرایب عددی یا نمادین قابل استفاده است.

در این فصل، روش‌های مستقیم برای حل دستگاه‌های خطی را بررسی می‌کنیم. در این روش‌ها، ماتریس ضرایب را به یک ماتریس مثلثی تبدیل می‌کنیم و سپس با استفاده از روش حذف به عقب، مقادیر متغیرها را به دست می‌آوریم. این روش‌ها برای سیستم‌های خطی با ضرایب عددی یا نمادین قابل استفاده است.

در این فصل، روش‌های مستقیم برای حل دستگاه‌های خطی را بررسی می‌کنیم. در این روش‌ها، ماتریس ضرایب را به یک ماتریس مثلثی تبدیل می‌کنیم و سپس با استفاده از روش حذف به عقب، مقادیر متغیرها را به دست می‌آوریم. این روش‌ها برای سیستم‌های خطی با ضرایب عددی یا نمادین قابل استفاده است.

در این فصل، روش‌های مستقیم برای حل دستگاه‌های خطی را بررسی می‌کنیم. در این روش‌ها، ماتریس ضرایب را به یک ماتریس مثلثی تبدیل می‌کنیم و سپس با استفاده از روش حذف به عقب، مقادیر متغیرها را به دست می‌آوریم. این روش‌ها برای سیستم‌های خطی با ضرایب عددی یا نمادین قابل استفاده است.

در این فصل، روش‌های مستقیم برای حل دستگاه‌های خطی را بررسی می‌کنیم. در این روش‌ها، ماتریس ضرایب را به یک ماتریس مثلثی تبدیل می‌کنیم و سپس با استفاده از روش حذف به عقب، مقادیر متغیرها را به دست می‌آوریم. این روش‌ها برای سیستم‌های خطی با ضرایب عددی یا نمادین قابل استفاده است.

در این فصل، روش‌های مستقیم برای حل دستگاه‌های خطی را بررسی می‌کنیم. در این روش‌ها، ماتریس ضرایب را به یک ماتریس مثلثی تبدیل می‌کنیم و سپس با استفاده از روش حذف به عقب، مقادیر متغیرها را به دست می‌آوریم. این روش‌ها برای سیستم‌های خطی با ضرایب عددی یا نمادین قابل استفاده است.

در این فصل، روش‌های مستقیم برای حل دستگاه‌های خطی را بررسی می‌کنیم. در این روش‌ها، ماتریس ضرایب را به یک ماتریس مثلثی تبدیل می‌کنیم و سپس با استفاده از روش حذف به عقب، مقادیر متغیرها را به دست می‌آوریم. این روش‌ها برای سیستم‌های خطی با ضرایب عددی یا نمادین قابل استفاده است.

دستگاههای معادلات خطی با اکثر مسائل مهندسی و علوم در ارتباطند، همانطور که به کاربردهای ریاضیات در علوم اجتماعی و مطالعه کمی تجارت و مسائل اقتصادی ارتباط دارند. در این فصل، روشهای مستقیم حل دستگاه خطی

$$(1.7) \begin{matrix} E_1: & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = h_1 \\ E_2: & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = h_2 \\ & \vdots \\ E_n: & a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = h_n \end{matrix}$$

نسبت به x_1, \dots, x_n با مقادیر مفروض $a_{i,j}$ ، به ازای هر $i, j = 1, 2, \dots, n$ ، و h_i ، به ازای هر $i = 1, 2, \dots, n$ بررسی می شود. روشهای مستقیم روشهایی هستند که در تعداد ثابتی مرحله جوابی، فقط تابع خطاهای گرد کردن، بدست می دهند. برای ارائه مطلب، معرفی چند مفهوم مقدماتی از جبر خطی نیز لازم است.

روشهای تقریب جواب دستگاههای خطی با روشهای تکراری در فصل ۸ بررسی می شوند.

۱.۷ دستگاههای معادلات خطی

برای حل یک دستگاه خطی، مانند (۱.۷)، انجام سه عمل روی معادلات مجاز است:

- (۱) می توان معادله E_i را در ثابت ناصفر λ ضرب کرد و نتیجه را به جای معادله E_i قرار داد. این عمل را با $(E_i) \rightarrow (\lambda E_i)$ مشخص می کنیم.
- (۲) می توان با ضرب معادله E_j در ثابت λ و افزودن آن به معادله E_i نتیجه را به

جای معادله E_i قرار داد. این عمل را با $(E_i) \rightarrow (E_i + \lambda E_j)$ مشخص می‌کنیم. (۳) می‌توان ترتیب معادلات E_i و E_j را تعویض کرد. این عمل را با $(E_j) \leftrightarrow (E_i)$ مشخص می‌کنیم.

با انجام یک رشته از اعمال فوق، می‌توان یک دستگاه خطی را به دستگاهی خطی، با همان مجموعه جواب، که خیلی ساده‌تر حل می‌شود تبدیل کرد. در مثال بعد این رشته از اعمال تشریح می‌شوند.

مثال ۱. چهار معادله

$$\begin{aligned} E_1: & x_1 + x_2 + 3x_4 = 4, \\ E_2: & 2x_1 + x_2 - x_3 + x_4 = 1, \\ E_3: & 3x_1 - x_2 - x_3 + 2x_4 = -3, \\ E_4: & -x_1 + 2x_2 + 3x_3 - x_4 = 4, \end{aligned} \quad (2.7)$$

رأنسبت به مجهولات x_1 ، x_2 ، x_3 و x_4 حل می‌کنیم. اولین مرحله بکارگیری معادله E_1 جهت حذف مجهول x_1 از معادلات E_2 ، E_3 ، و E_4 بوسیله اعمال $(E_2 - 2E_1) \rightarrow (E_2)$ ، $(E_3 - 3E_1) \rightarrow (E_3)$ ، و $(E_4 + E_1) \rightarrow (E_4)$ است. دستگاه

$$\begin{aligned} E_1: & x_1 + x_2 + 3x_4 = 4, \\ E_2: & -x_2 - x_3 - 5x_4 = -7, \\ E_3: & -4x_2 - x_3 - 7x_4 = -15, \\ E_4: & 3x_2 + 3x_3 + 2x_4 = 8, \end{aligned} \quad (3.7)$$

حاصل می‌شود، که مجدداً "معادلات جدید را برای سادگی با E_1 ، E_2 ، E_3 ، و E_4 نشان داده‌ایم.

دراین دستگاه، E_2 را جهت حذف x_2 از E_3 و E_4 بوسیله اعمال $(E_3 - 4E_2) \rightarrow (E_3)$ و $(E_4 + 3E_2) \rightarrow (E_4)$ بکار می‌بریم؛ دستگاه

$$\begin{aligned} E_1: & x_1 + x_2 + 3x_4 = 4, \\ E_2: & -x_2 - x_3 - 5x_4 = -7, \\ E_3: & 3x_3 + 13x_4 = 13, \\ E_4: & -13x_4 = -13, \end{aligned} \quad (4.7)$$

حاصل می‌شود.

حال دستگاه معادلات (۴.۷) دروضعیتی است که فرم مثلثی یا تحویل یافته نامیده

$$a_{i1}x'_1 + a_{i2}x'_2 + \dots + a_{in}x'_n = b_i$$

با ضرب این معادله در λ داریم

$$\lambda a_{i1}x'_1 + \lambda a_{i2}x'_2 + \dots + \lambda a_{in}x'_n = \lambda b_i$$

در نتیجه، x'_1, \dots, x'_n در E_i نیز صدق می‌کنند.

از طرف دیگر، فرض کنیم x'_1, \dots, x'_n یک جواب (۶.۷) باشد. برای اینکه نشان دهیم یک جواب (۵.۷) نیز هست فقط باید برقرار بودن معادله E_i از (۵.۷) را تحقیق کنیم. چون x'_1, \dots, x'_n جواب (۶.۷) است،

$$\lambda a_{i1}x'_1 + \lambda a_{i2}x'_2 + \dots + \lambda a_{in}x'_n = \lambda b_i$$

اما $\lambda \neq 0$ ؛ در نتیجه، تقسیم بر λ جایز است، و نتیجه می‌دهد

$$a_{i1}x'_1 + a_{i2}x'_2 + \dots + a_{in}x'_n = b_i.$$

در اجرای محاسبات مثال ۱، واقعا" لازم نیست همهء معادلات در هر مرحله به تفصیل نوشته شوند یا متغیرهای x_1 ، x_2 ، x_3 ، و x_4 در طول محاسبات تکرار شوند، زیرا آنها همیشه در همان ستون باقی می‌مانند. تنها تغییری که از دستگاهی به دستگاهی رخ می‌دهد در ضرایب مجهولات و مقادیر طرف راست معادلات است. به این دلیل، اغلب یک دستگاه خطی با یک ماتریس که شامل همهء اطلاعات لازم، منتها فشرده، جهت تعیین جوابهای آن است جایگزین می‌شود.

تعریف ۲.۷. یک ماتریس n در m آرایه‌ای مستطیلی از عناصر با n سطرها و m ستون است که در آن نه تنها مقدار یک عنصر، بلکه محل آن در آرایه نیز، حائز اهمیت است.

برای نمایش یک ماتریس $m \times n$ (n در m) از حرف بزرگی نظیر A ، و برای نمایش درایهء واقع در محل تلاقی سطر i و ستون m آن از حروف کوچک با زیرنویس مضاعف، نظیر a_{ij} ، استفاده می‌شود؛ یعنی،

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

و

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

و سپس از تلفیق این دو ماتریسی به نام ماتریس افزوده تشکیل می‌شود:

$$[A, b] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array} \right]$$

که در آن خط چین اغلب برای تفکیک ضرایب مجهولات از مقادیر طرف راست بکار می‌رود.

مثال ۳. تکرار اعمال موجود در مثال ۱ با نماد ماتریس، ابتدا بررسی ماتریس افزوده^۲ وابسته به دستگاه (۲.۷)، یعنی

$$(۷.۷) \quad \left[\begin{array}{cccc|c} 1 & 1 & 0 & 3 & 4 \\ 2 & 1 & -1 & 1 & 1 \\ 3 & -1 & -1 & 2 & -3 \\ -1 & 2 & 3 & -1 & 4 \end{array} \right]$$

را موجب می‌شود. اجرای اعمال مربوطه $(E_2 - 2E_1) \rightarrow (E_2)$ ، $(E_3 - 3E_1) \rightarrow (E_3)$ ، و $(E_4 + E_1) \rightarrow (E_4)$ در (۲.۷) با تغییرات لازم روی سطرهای مربوطه از ماتریس افزوده^۲ (۷.۷)، ماتریس متناظر دستگاه (۳.۷) را بدست می‌دهد:

$$(۸.۷) \quad \left[\begin{array}{cccc|c} 1 & 1 & 0 & 3 & 4 \\ 0 & -1 & -1 & -5 & -7 \\ 0 & -4 & -1 & -7 & -15 \\ 0 & 3 & 3 & 2 & 8 \end{array} \right]$$

با اجرای آخرین تغییرات، ماتریس افزودهٔ متناظر دستگاه (۴.۷) حاصل می‌شود:

$$(۹.۷) \quad \left[\begin{array}{cccc|c} 1 & 1 & 0 & 3 & 4 \\ 0 & -1 & -1 & -5 & -7 \\ 0 & 0 & 3 & 13 & 13 \\ 0 & 0 & 0 & -13 & -13 \end{array} \right].$$

حال می‌توان این ماتریس را به دستگاه خطی (۴.۷) متناظر آن تبدیل و جوابهای x_1, x_2, x_3 ، و x_4 را بدست آورد. روند موجود در این روش حذف گاوسی با جایگذاری پسرو نامیده می‌شود. در بخشهای بعد، شرایطی برای دستگاه خطی که بتوان تحت آنها این روش را بکار برد بررسی می‌شود.

مجموعه تمرینات ۱.۷

۱. دستگاههای خطی زیر را با استفاده از روش حذف حل کنید.

$$x_1 + 3x_2 - x_3 + x_4 = 5, \quad (A)$$

$$2x_2 + x_3 + x_4 = 7,$$

$$x_3 + 2x_4 = 6,$$

$$3x_4 = 9.$$

$$x_1 - 2x_2 = 3, \quad (B)$$

$$2x_1 + x_2 = 4.$$

$$x_1 + x_2 - x_3 = 3, \quad (C)$$

$$2x_1 - x_2 + 3x_3 = 0,$$

$$-x_1 - 2x_2 + x_3 = -5.$$

$$3x - y + 2z = -3, \quad (D)$$

$$x + y + z = -4,$$

$$2x + y - z = -3.$$

$$x_1 + \frac{1}{2}x_2 + \frac{1}{3}x_3 = \frac{11}{6}, \quad (E)$$

$$\frac{1}{2}x_1 + \frac{1}{3}x_2 + \frac{1}{4}x_3 = \frac{13}{12},$$

$$\frac{1}{3}x_1 + \frac{1}{4}x_2 + \frac{1}{5}x_3 = \frac{47}{60}.$$

۲. برای هر یک از دستگاههای ذیل، در صورت امکان، یک جواب موداری بدست آورید، و سپس سعی کنید آنها را با روشهای این بخش حل نمایید.

$$x_1 + 2x_2 = 3, \quad (T)$$

$$x_1 - x_2 = 0.$$

$$x_1 + 2x_2 = 3, \quad (ب)$$

$$-2x_1 - 4x_2 = 6.$$

$$x_1 + 2x_2 = 3, \quad (پ)$$

$$2x_1 + 4x_2 = 6.$$

$$0 \cdot x_1 + x_2 = 3, \quad (ت)$$

$$2x_1 - x_2 = 7.$$

۳. با استفاده از روشهای این بخش سعی کنید دستگاههای خطی زیر را حل کنید. شرح دهید چرا این روش در هر حالت که جوابی نمی‌توان بدست آورد ناموفق است.

$$x_1 + x_2 + x_3 = 2, \quad (T)$$

$$2x_1 + x_2 - x_3 + x_4 = 1,$$

$$-x_1 + 2x_2 + 3x_3 - x_4 = 4,$$

$$3x_1 - x_2 - x_3 + 2x_4 = -3.$$

$$x_1 + x_2 - x_3 = 3, \quad (ب)$$

$$-x_1 + x_2 + x_3 = 2,$$

$$x_1 + 3x_2 - x_3 = 8.$$

$$x_1 + x_2 - x_3 = 3, \quad (پ)$$

$$-x_1 + x_2 + x_3 = 2,$$

$$x_1 + 3x_2 - x_3 = 6.$$

۴. درباره هر دسته از معادلات زیر، از دیدگاه هندسی، چه نظری می‌توان ارائه داد:

$$2x + y = -1, \quad (T)$$

$$x + y = 2,$$

$$x - 3y = 5.$$

$$2x + y + z = 1, \quad (ب)$$

$$2x + 4y - z = -1.$$

۵. اثبات قضیه ۱.۷ را کامل کنید.

۶. فرض کنید در یک دستگاه زیست‌شناسی n نوع حیوان و m منبع غذایی وجود داشته باشد. همچنین، x_j جمعیت نوع j م، به‌ازای هر $j = 1, \dots, n$ ، و h_i ذخیره روزانه قابل استفاده غذای i م را نشان دهد؛ و نیز a_{ij} متوسط مقدار غذای i م

تشکیل می‌دهیم، که در آن A ماتریس متشکل از ضرایب و درایه‌های موجود در ستون $(n+1)$ م مقادیر b را مشخص می‌کند؛ یعنی، به‌ازای هر $i=1, 2, \dots, n$ ، $a_{i,n+1} = b_i$ ، $a_{11} \neq 0$ ، اعمال متناظر با $(E_j) \rightarrow (E_j - (a_{j1}/a_{11})E_1)$ ، به‌ازای هر $j=2, 3, \dots, n$ جهت حذف ضریب x_1 در هریک از این سطرها انجام می‌شود. گرچه انتظار می‌رود درایه‌های سطرها $2, 3, \dots, n$ تغییر یابند، ولی برای سادگی نمادگذاری مجدداً "درایه" موجود در سطر i م و ستون j م را با a_{ij} نشان می‌دهیم. با بخاطر سپردن این مطلب، یک روند متوالی را به‌ازای $i=2, 3, \dots, n-1$ دنبال می‌کنیم و، مشروط بر اینکه $a_{ii} \neq 0$ ، عمل $(E_j) \rightarrow (E_j - (a_{ji}/a_{ii})E_i)$ را به‌ازای هر $j=i+1, i+2, \dots, n$ انجام می‌دهیم. این امر x_i را در هر سطر زیر سطر i م به‌ازای همه مقادیر $i=1, 2, \dots, n-1$ حذف می‌کند (یعنی، ضریب را به صفر تبدیل می‌کند). ماتریس حاصل به شکل

$$\bar{A} = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & a_{1,n+1} \\ 0 & a_{22} & \cdots & a_{2n} & a_{2,n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & a_{nn} & a_{n,n+1} \end{array} \right]$$

خواهد بود که در آن، همانطور که در بالا گفته شد، انتظار نمی‌رود مقادیر a_{ii} با مقادیر نظیر در ماتریس اصلی \bar{A} مطابقت داشته باشد. این ماتریس دستگاهی خطی با همان مجموعه جواب (۱۰.۷) ارائه می‌دهد. چون این دستگاه خطی معادل مثلثی شکل است، یعنی به شکل

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= a_{1,n+1} \\ a_{22}x_2 + \cdots + a_{2n}x_n &= a_{2,n+1} \\ &\vdots \\ a_{nn}x_n &= a_{n,n+1} \end{aligned}$$

است، روش جایگذاری پسرورا می‌توان اجرا کرد. با حل معادله n م نسبت به x_n نتیجه می‌شود که

$$x_n = \frac{a_{n,n+1}}{a_{nn}}$$

با حل معادله $(n-1)$ م نسبت به x_{n-1} و استفاده از x_n نتیجه می‌شود

$$x_{n-1} = \frac{[a_{n-1,n+1} - a_{n-1,n}x_n]}{a_{n-1,n-1}}$$

و با ادامه این روند، بدست می‌آید:

بمازای هر $i = n-1, n-2, \dots, 2, 1$

$$\begin{aligned} x_i &= \frac{[a_{i,n+1} - a_{in}x_n - a_{i,n-1}x_{n-1} - \dots - a_{i,i+1}x_{i+1}]}{a_{ii}} \\ &= \frac{[a_{i,n+1} - \sum_{j=i+1}^n a_{ij}x_j]}{a_{ii}} \end{aligned}$$

روند حذف گاوسی را می‌توان دقیقتر، ولی پیچیده‌تر، ارائه داد. با تشکیل دنباله‌ای از ماتریسهای افزوده $\vec{A}^{(1)}, \vec{A}^{(2)}, \dots, \vec{A}^{(n)}$ ، که در آن $\vec{A}^{(1)}$ ماتریس \vec{A} ارائه شده در (۱۱.۷) بوده و $\vec{A}^{(k)}$ ، بمازای هر $k = 2, 3, \dots, n$ ، دارای درایه‌های $a_{ij}^{(k)}$ است بطوری که

$$a_{ij}^{(k)} = \begin{cases} a_{ij}^{(k-1)} & , j = 1, 2, \dots, n+1 \text{ و } i = 1, 2, \dots, k-1 \\ 0 & , j = 1, 2, \dots, k-1 \text{ و } i = k, k+1, \dots, n \\ a_{ij}^{(k-1)} - \frac{a_{i,k-1}^{(k-1)}}{a_{k-1,k-1}^{(k-1)}} a_{k-1,j}^{(k-1)} & , j = k, k+1, \dots, n+1 \text{ و } i = k, k+1, \dots, n \end{cases}$$

بنابراین،

$$\vec{A}^{(k)} = \left[\begin{array}{ccccccc|c} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \dots & a_{1,k-1}^{(1)} & a_{1k}^{(1)} & \dots & a_{1n}^{(1)} & a_{1,n+1}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \dots & a_{2,k-1}^{(2)} & a_{2k}^{(2)} & \dots & a_{2n}^{(2)} & a_{2,n+1}^{(2)} \\ & & & & \vdots & \vdots & & \vdots & \vdots \\ & & & & & \vdots & & \vdots & \vdots \\ & & & & & a_{k-1,k-1}^{(k-1)} & a_{k-1,k}^{(k-1)} & \dots & a_{k-1,n}^{(k-1)} & a_{k-1,n+1}^{(k-1)} \\ & & & & & & 0 & a_{kk}^{(k)} & \dots & a_{kn}^{(k)} & a_{k,n+1}^{(k)} \\ & & & & & & & \vdots & & \vdots & \vdots \\ 0 & & & & & & & & & a_{nk}^{(k)} & \dots & a_{nn}^{(k)} & a_{n,n+1}^{(k)} \end{array} \right]$$

دستگاه خطی معادلی ارائه می‌دهد که در آن متغیر x_{k-1} از معادلات E_k, E_{k+1}, \dots, E_n حذف شده است.

اگر یکی از عناصر $a_{11}^{(1)}, a_{22}^{(2)}, a_{33}^{(3)}, \dots, a_{n-1,n-1}^{(n-1)}, a_{nn}^{(n)}$ صفر باشد، این روند کارگر

نیست، زیرا در این صورت مرحلهٔ

$$\left(E_i - \frac{a_{i,k}^{(k)}}{a_{kk}^{(k)}} E_k \right) \rightarrow E_i$$

یا غیر قابل اجراست (این در صورتی است که یکی از $a_{11}^{(1)}, \dots, a_{n-1,n-1}^{(n-1)}$ صفر باشد) یا روند جایگذاری پسرو را نمیتوان انجام داد (در حالتی که $a_{nn}^{(n)} = 0$). این مطلب الزاما "به معنی حل ناپذیری دستگاه خطی نیست، بلکه روش حل باید تغییر یابد. این مطلب در مثال زیر تشریح می‌شود.

مثال ۱. دستگاه خطی

$$\begin{aligned} E_1: & x_1 - x_2 + 2x_3 - x_4 = -8, \\ E_2: & 2x_1 - 2x_2 + 3x_3 - 3x_4 = -20, \\ E_3: & x_1 + x_2 + x_3 = -2, \\ E_4: & x_1 - x_2 + 4x_3 + 3x_4 = 4 \end{aligned}$$

را در نظر می‌گیریم. ماتریس افزوده عبارت است از

$$\tilde{A} = \tilde{A}^{(1)} = \left[\begin{array}{cccc|c} 1 & -1 & 2 & -1 & -8 \\ 2 & -2 & 3 & -3 & -20 \\ 1 & 1 & 1 & 0 & -2 \\ 1 & -1 & 4 & 3 & 4 \end{array} \right]$$

و، با اجرای اعمال

$$(E_2 - 2E_1) \rightarrow (E_2), \quad (E_3 - E_1) \rightarrow (E_3),$$

و

$$(E_4 - E_1) \rightarrow (E_4),$$

می‌نویسیم

$$\tilde{A}^{(2)} = \left[\begin{array}{cccc|c} 1 & -1 & 2 & -1 & -8 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & 2 & 4 & 12 \end{array} \right]$$

چون عنصر $a_{22}^{(2)}$ ، به نام عنصر محوری، صفر است، روند را به شکل معمول نمی‌توان ادامه داد. لیکن عمل $(E_i) \leftrightarrow (E_j)$ مجاز است، در نتیجه، برای تعیین اولین عنصر ناصفر،

عناصر $a_{42}^{(2)}$ و $a_{32}^{(2)}$ بررسی می‌شوند. چون $a_{32}^{(2)} \neq 0$ ، با انجام عمل $(E_3) \leftrightarrow (E_2)$ ماتریس جدیدی که با

$$\bar{A}^{(2)} = \left[\begin{array}{cccc|c} 1 & -1 & 2 & -1 & -8 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 0 & 2 & 4 & 12 \end{array} \right]$$

نموده می‌شود بدست می‌آید. چون قبلاً x_2 از E_3 و E_4 حذف شده است، $\bar{A}^{(3)}$ همان $\bar{A}^{(2)}$ است و محاسبه با عمل $(E_4 + 2E_3) \rightarrow (E_4)$ ادامه می‌یابد، که نتیجه می‌دهد

$$\bar{A}^{(4)} = \left[\begin{array}{cccc|c} 1 & -1 & 2 & -1 & -8 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 0 & 0 & 2 & 4 \end{array} \right]$$

بالاخره، با اعمال جایگذاری پسرو، بدست می‌آوریم

$$x_4 = \frac{4}{2} = 2,$$

$$x_3 = \frac{[-4 - (-1)x_4]}{(-1)} = 2,$$

$$x_2 = \frac{[6 - x_4 - (-1)x_3]}{2} = 3,$$

$$x_1 = \frac{[-8 - (-1)x_4 - 2x_3 - (-1)x_2]}{(1)} = -7.$$

مثال ۱ روشن می‌کند که اگر به ازای $k = 1, 2, \dots, n-1$ ای $a_{kk}^{(k)} = 0$ ، چه باید کرد. در ستون k ام $\bar{A}^{(k-1)}$ از سطر k ام تا سطر n م باید در جستجوی اولین درایهٔ منصفربود. هرگاه به ازای p ای که $k+1 \leq p \leq n$ ، $a_{pk}^{(k)} \neq 0$ ، آنگاه باید عمل $(E_k) \leftrightarrow (E_p)$ را جهت تعیین $\bar{A}^{(k-1)}$ انجام داد. سپس روند را جهت تشکیل $\bar{A}^{(k)}$ ادامه می‌دهیم، و غیره. هرگاه به ازای $n, n-1, \dots, k+1$ ، $a_{pk}^{(k)} = 0$ ، می‌توان نشان داد (ر. ک. قضیهٔ ۱۶.۷، صفحهٔ ۴۵۲) که این دستگاه خطی فاقد جواب منحصر بفرد است و روند متوقف می‌شود. بالاخره، هرگاه $a_{nn}^{(n)} = 0$ ، این دستگاه خطی فاقد جواب منحصر

بفرد است و مجدداً " روند متوقف می شود. برای خلاصه کردن تمام روند حذف گاوسی با جایگذاری پسرو، الگوریتم زیر ارائه می شود.

الگوریتم حذف گاوسی با جایگذاری پسرو ۱.۷ برای حل دستگاه خطی

$$\begin{aligned} E_1: & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \equiv a_{1,n+1} \\ E_2: & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \equiv a_{2,n+1} \\ & \vdots \\ E_n: & a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \equiv a_{n,n+1} \end{aligned}$$

ماتریس افزوده^۶ زیر را می سازیم :

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & a_{1,n+1} \\ a_{21} & a_{22} & \cdots & a_{2n} & a_{2,n+1} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & a_{n,n+1} \end{array} \right]$$

- مرحله^۱ . قرار می دهیم $i = 1$.
- مرحله^۲ . اگر $a_{ii} \neq 0$ ، به مرحله^۴ می رویم .
- اگر $a_{ii} = 0$ ، به مرحله^۳ می رویم .
- مرحله^۳ . فرض کنیم p عددی صحیح باشد بطوری که $a_{p-1,i} = 0, \dots, a_{i+1,i} = 0, a_{ii} = 0$ ولی $a_{pi} \neq 0$.
- اگر p وجود داشته باشد، عمل $(E_i) \leftrightarrow (E_p)$ را انجام داده و به مرحله^۴ می رویم .
- اگر چنین p ای موجود نباشد، دستگاه فاقد جواب منحصر بفرد است و روند متوقف می شود.
- مرحله^۴ . فرض کنیم به ازای $j = i + 1, i + 2, \dots, n$ ، $m_{ji} = a_{ji}/a_{ii}$ ، و اعمال $(E_j - m_{ji}E_i) \rightarrow (E_j)$ را انجام می دهیم .
- مرحله^۵ . i را به $i + 1$ می افزاییم .
- مرحله^۶ . اگر $i < n$ ، به مرحله^۲ می رویم .
- اگر $i = n$ ، به مرحله^۷ می رویم .
- مرحله^۷ . اگر $a_{nn} = 0$ ، دستگاه فاقد جواب منحصر بفرد است و روند متوقف می شود .
- اگر $a_{nn} \neq 0$ ،

(۲) قرار می‌دهیم $x_n = a_{n,n+1}/a_{nn}$ ؛

(۳) رابطه‌ای برای $x_i = [a_{i,n+1} - \sum_{j=i+1}^n a_{ij}x_j]/a_{ii}$ محاسبه می‌کنیم.

مرحله ۸. روند تمام خواهد بود.

مثال ۲. هدف از این مثال این است که نشان دهیم اگر الگوریتم ۱.۷ موفق نباشد چه پیش می‌آید. محاسبات روی دو دستگاه زیر همزمان انجام می‌شود.

$$\begin{aligned}x_1 + x_2 + x_3 + x_4 &= 7, \\x_1 + x_2 + 2x_4 &= 8, \\2x_1 + 2x_2 + 3x_3 &= 10, \\-x_1 - x_2 - 2x_3 + 2x_4 &= 0.\end{aligned}$$

و

$$\begin{aligned}x_1 + x_2 + x_3 + x_4 &= 7, \\x_1 + x_2 + 2x_4 &= 5, \\2x_1 + 2x_2 + 3x_3 &= 10, \\-x_1 - x_2 - 2x_3 + 2x_4 &= 0.\end{aligned}$$

از این دستگاهها ماتریسهای زیر حاصل می‌شوند:

$$\bar{A} = \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 7 \\ 1 & 1 & 0 & 2 & 8 \\ 2 & 2 & 3 & 0 & 10 \\ -1 & -1 & -2 & 2 & 0 \end{array} \right]$$

و

$$\bar{A} = \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 7 \\ 1 & 1 & 0 & 2 & 5 \\ 2 & 2 & 3 & 0 & 10 \\ -1 & -1 & -2 & 2 & 0 \end{array} \right]$$

مرحله ۱.۱ $i = 1$.

مرحله ۲. $a_{ii} = a_{11} = 1 \neq 0$.

مرحله ۴.

$$m_{ji} = m_{j1} = a_{1j}/1 = a_{1j}.$$

$$j = 2, \quad m_{12} = 1,$$

$$j = 3, \quad m_{13} = 2,$$

$$j = 4, \quad m_{14} = -1.$$

$$(E_2 - E_1) \rightarrow (E_2),$$

$$(E_3 - 2E_1) \rightarrow (E_3),$$

$$(E_4 + E_1) \rightarrow (E_4).$$

ماتریسها می شوند:

$$\tilde{A} = \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 7 \\ 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & -2 & -4 \\ 0 & 0 & -1 & 3 & 7 \end{array} \right]$$

و

$$\tilde{A} = \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 7 \\ 0 & 0 & -1 & 1 & -2 \\ 0 & 0 & 1 & -2 & -4 \\ 0 & 0 & -1 & 3 & 7 \end{array} \right]$$

مرحله ۵. $i = 2$.

مرحله ۶. $i = 2 < 4 = n$.

مرحله ۲. $a_{ii} = a_{22} = 0$.

مرحله ۳. $a_{33} = a_{44} = 0$.

الگوریتم ایجاب می کند که روند متوقف و هیچ جوابی برای هر یک از دستگاهها بدست نیاید.

جهت بررسی دقیقتر اشکال کار، $(E_4 + E_3) \rightarrow (E_4)$ را انجام می دهیم، بدست می آید.

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 7 \\ 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & -2 & -4 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right]$$

و

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 7 \\ 0 & 0 & -1 & 1 & -2 \\ 0 & 0 & 1 & -2 & -4 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right]$$

با نوشتن معادلات هر دستگاه نتیجه می شود

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 7, \\ -x_3 + x_4 &= 1, \\ x_3 - 2x_4 &= -4, \\ x_4 &= 3. \end{aligned}$$

و

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 7, \\ -x_3 + x_4 &= -2, \\ x_3 - 2x_4 &= -4, \\ x_4 &= 3. \end{aligned}$$

با انجام روش جایگذاری پسرو در دستگاه اول نتیجه می شود

$$\begin{aligned} x_4 &= 3, & x_4 &= 3, \\ x_3 &= -4 + 2x_4 = 2 & x_3 &= -4 + 2x_4 = 2 \end{aligned}$$

اگر روش جایگذاری پسرو روی معادله دوم هر دستگاه ادامه یابد، اختلاف این دو دستگاه معلوم می شود، چون

$$-x_3 + x_4 = 1$$

درحالی که

$$-x_3 + x_4 = -2$$

دستگاه خطی اول دارای بی نهایت جواب $x_4 = 3$ ، $x_3 = 2$ ، x_2 دلخواه، و $x_1 = 2 - x_2$ است، درحالی که دستگاه دیگر به تناقض منجر می شود، و هیچ جوابی ندارد. لیکن، در هر دو حالت، همانطور که از الگوریتم ۱.۷ نتیجه گرفتیم، جوابهای منحصر بفرود وجود ندارند.

گرچه الگوریتم ۱.۷ را می توان مانند ساختمان ماتریسهای افزوده $\bar{A}^{(1)}, \dots, \bar{A}^{(n)}$ نگریست، لیکن محاسبات در یک ماشین حساب فقط با بکار بردن یک آرایه n در $(n+1)$

برای ذخیره، با جایگزینی ساده مقدار قبلی a_{ij} با مقدار جدید آن در هر مرحله، انجام می‌شود. همچنین، ذخیره‌سازی مضارب m_{ji} در محل a_{ji} سودمند است، چون a_{ji} ، به ازای هر $i = 1, 2, \dots, n-1$ و $j = i + 1, i + 2, \dots, n$ ، دارای مقدار صفر خواهد بود. بنابراین، A را می‌توان بوسیله مضارب زیر قطر اصلی و بوسیله درایه‌های ناصفروو بالای قطر اصلی $\bar{A}^{(m)}$ نوشت. این مقادیر را می‌توان جهت حل دیگر دستگاههای خطی شامل ماتریس اصلی A بکار برد. (ر.ک. بخش ۶.۷).

هنگام مقایسه تکنیکهای حل دستگاههای خطی، علاوه بر مقدار حافظه لازم، به بررسی مفاهیم دیگر نیز نیاز داریم. یکی از اینها اثر خطای گردگردن، و دیگری مقدار زمان لازم جهت انجام کامل محاسبات است. این دو مطلب به تعداد اعمال حسابی لازم جهت حل کردن یک مسئله عادی بستگی دارند. در حالت کلی، مقدار وقت لازم جهت ضرب یا تقسیم در یک ماشین حساب تقریباً یکسان بوده، و به‌طور قابل ملاحظه‌ای بیشتر از وقت لازم جهت جمع یا تفریق است. لیکن، تفاوت‌های واقعی در زمان اجرا به دستگاه محاسبه بخصوصی که بکار می‌رود بستگی دارد. برای اینکه روند شمارش اعمال را برای یک روش مفروض تشریح کنیم، اعمال لازم جهت حل یک دستگاه خطی نمونه با n معادله و n مجهول را با استفاده از الگوریتم ۱.۷ می‌شماریم. به علت اختلاف زمان، تعداد جمع و تفریقها را جدا از تعداد ضرب و تقسیمها می‌شماریم.

در این الگوریتم تا مرحله چهار هیچ عمل حسابی انجام نمی‌شود. اولین خطدر این مرحله نیاز به انجام $(n - i)$ تقسیم دارد. جایگزینی معادله (E_j) با معادله $(E_j - m_{ji}E_i)$ نیاز به ضرب m_{ji} در هر جمله معادله (E_i) دارد؛ در نتیجه، جمعاً $(n - i)(n - i + 1)$ ضرب لازم می‌شود. پس از اتمام این امر، هر جمله معادله حاصل از جمله نظیر آن در (E_j) تفریق می‌شود، که $(n - i)(n - i + 1)$ تفریق را سبب می‌شود. به‌مازای هر $i = 1, 2, \dots, n - 1$ ، اعمال لازم در مرحله چهار عبارت است از

$$\text{تقسیمها/ضربها} : (n - i) + (n - i)(n - i + 1)$$

$$\text{تفریقها/جمعها} : (n - i)(n - i + 1)$$

نتیجتاً، تعداد کل اعمال لازم در این مرحله عبارتند از

تقسیمها/ضربها :

$$\sum_{i=1}^{n-1} (n - i)(n - i + 2) = (n^2 + 2n) \sum_{i=1}^{n-1} 1 - 2(n + 1) \sum_{i=1}^{n-1} i + \sum_{i=1}^{n-1} i^2$$

$$\begin{aligned}
 &= (n^2 + 2n)(n - 1) - 2(n + 1) \frac{(n)(n - 1)}{2} \\
 &\quad + \frac{(n - 1)(n)(2n - 1)}{6} \\
 &= \frac{2n^3 + 3n^2 - 5n}{6}
 \end{aligned}$$

تفریقها/جمعها :

$$\begin{aligned}
 \sum_{i=1}^{n-1} (n - i)(n - i + 1) &= (n^2 + n) \sum_{i=1}^{n-1} 1 - (2n + 1) \sum_{i=1}^{n-1} i + \sum_{i=1}^{n-1} i^2 \\
 &= (n^2 + n)(n - 1) - (2n + 1) \frac{(n)(n - 1)}{2} \\
 &\quad + \frac{(n - 1)(n)(2n - 1)}{6} \\
 &= \frac{n^3 - n}{3}
 \end{aligned}$$

تنها مرحله دیگر الگوریتم ۱.۷ که به اعمال حسابی نیازمند است مرحله ۷ می باشد، که در آن (T) یک تقسیم لازم دارد و (ب) (n - i) ضرب و برای هر جمله مجموع (n - i - 1) جمع، و سپس به یک تفریق و یک تقسیم نیازمند است. تعداد کل اعمال مرحله ۷ عبارت است از

$$1 + \sum_{i=1}^{n-1} ((n - i) + 1) = \frac{n^2 + n}{2}$$

تقسیمها/ضربها

$$\sum_{i=1}^{n-1} ((n - i - 1) + 1) = \frac{n^2 - n}{2}$$

تفریقها/جمعها

در نتیجه، مجموع اعمال حسابی در الگوریتم ۱.۷ عبارت است از

$$\frac{2n^3 + 3n^2 - 5n}{6} + \frac{n^2 + n}{2} = \frac{n^3 + 3n^2 - n}{3}$$

تقسیمها/ضربها

$$\frac{n^3 - n}{3} + \frac{n^2 - n}{2} = \frac{2n^3 + 3n^2 - 5n}{6}$$

تفریقها/جمعها

چون تعداد کل ضرب و تقسیمها، و همین طور برای جمع و تفریقها، تقریباً $\frac{1}{3}n^3$ است، مقدار محاسبه و زمان لازم با افزایش n به نسبت n^3 افزایش می یابد. جدول زیر نشانگر افزایش محاسبات با افزایش n می باشد.

"	تقسیمها / ضربها	تفریقها / جمعها
3	17	11
10	430	375
50	44,150	42,875
100	343,300	338,250

مجموعه تمرینات ۲.۷

۱. دستگاههای خطی زیر را، در صورت امکان، با استفاده از الگوریتم ۱.۷ حل کنید، و تعیین کنید آیا تعویض سطرها لازم است.

$$x_1 - x_2 + 3x_3 = 2, \quad (\text{A})$$

$$3x_1 - 3x_2 + x_3 = -1,$$

$$x_1 + x_2 = 3.$$

$$\frac{1}{4}x_1 + \frac{1}{4}x_2 + \frac{1}{6}x_3 = 9, \quad (\text{B})$$

$$\frac{1}{4}x_1 + \frac{1}{4}x_2 + \frac{1}{4}x_3 = 8,$$

$$\frac{1}{2}x_1 + x_2 + 2x_3 = 8.$$

$$x_1 + \frac{1}{2}x_2 + \frac{1}{3}x_3 + \frac{1}{4}x_4 = \frac{1}{6}, \quad (\text{C})$$

$$\frac{1}{2}x_1 + \frac{1}{3}x_2 + \frac{1}{4}x_3 + \frac{1}{4}x_4 = \frac{1}{6},$$

$$\frac{1}{3}x_1 + \frac{1}{4}x_2 + \frac{1}{5}x_3 + \frac{1}{6}x_4 = \frac{1}{8},$$

$$\frac{1}{4}x_1 + \frac{1}{5}x_2 + \frac{1}{6}x_3 + \frac{1}{7}x_4 = \frac{1}{6}.$$

۲. دستگاههای خطی زیر را، در صورت امکان، با استفاده از الگوریتم ۱.۷ حل کنید، و تعیین کنید آیا تعویض سطرها لازم است.

$$4.01x_1 + 1.23x_2 + 1.43x_3 - .73x_4 = 5.94, \quad (\text{A})$$

$$1.23x_1 + 7.41x_2 + 2.41x_3 + 3.02x_4 = 14.07$$

$$1.43x_1 + 2.41x_2 + 5.79x_3 - 1.11x_4 = 8.52,$$

$$-.73x_1 + 3.02x_2 - 1.11x_3 + 6.41x_4 = 7.59.$$

$$3.333x_1 + 15920x_2 - 10.333x_3 = 15913, \quad (\text{B})$$

$$2.222x_1 + 16.71x_2 + 9.612x_3 = 28.544,$$

$$1.5611x_1 + 5.1791x_2 + 1.6852x_3 = 8.4254.$$

۳.۷ جبر خطی و معکوس کردن ماتریس

مفهوم ماتریس در دو بخش قبل به عنوان روش مناسبی برای بیان و حل یک دستگاه خطی

معرفی شد. مطالب این بخش مربوط به جبر ماتریسها و هر طریقه‌ای است که می‌توان آن را برای حل مسائل مربوط به دستگاههای خطی بکار برد.

تعریف ۳.۷. دو ماتریس A و B را مساوی گوئیم اگر هر دو از یک اندازه، مثلا " $m \times n$ " باشند و، به ازای هر $i = 1, 2, \dots, m$ و $j = 1, 2, \dots, n$ ، $a_{ij} = b_{ij}$.

مثال ۱. فرض کنیم

$$A = \begin{bmatrix} 2 & -1 & 7 \\ 3 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -1 & 7 \\ 3 & 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 3 \\ -1 & 1 \\ 7 & 0 \end{bmatrix}$$

آنگاه $A = B$ ولی $A \neq C$.

دو عمل مهم که روی ماتریسها انجام می‌شود جمع دو ماتریس و ضرب یک عدد حقیقی در یک ماتریس است.

تعریف ۴.۷. هرگاه A و B هر دو ماتریسهایی $m \times n$ باشند، آنگاه مجموع A و B که با $A + B$ نموده می‌شود، ماتریس $m \times n$ می‌است که به ازای هر $i = 1, 2, \dots, m$ و $j = 1, 2, \dots, n$ درایه‌های آن $a_{ij} + b_{ij}$ است.

تعریف ۵.۷. هرگاه A ماتریسی $m \times n$ و λ عددی حقیقی باشد، آنگاه حاصل ضرب λA و A ، که با λA نموده می‌شود، ماتریس $m \times n$ می‌است که به ازای هر $i = 1, 2, \dots, m$ و $j = 1, 2, \dots, n$ درایه‌های آن λa_{ij} است.

مثال ۲. فرض کنیم

$$A = \begin{bmatrix} 2 & -1 & 7 \\ 3 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 2 & -8 \\ 0 & 1 & 6 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

و $\lambda = -2, \mu = 3$ در این صورت،

$$A + B = \begin{bmatrix} 2+4 & -1+2 & 7-8 \\ 3+0 & 1+1 & 0+6 \end{bmatrix} = \begin{bmatrix} 6 & 1 & -1 \\ 3 & 2 & 6 \end{bmatrix};$$

$$B + A = \begin{bmatrix} 6 & 1 & -1 \\ 3 & 2 & 6 \end{bmatrix};$$

$$(A + B) + C = \begin{bmatrix} (2+4)+1 & (-1+2)+2 & (7-8)+3 \\ (3+0)+4 & (1+1)+5 & (0+6)+6 \end{bmatrix} = \begin{bmatrix} 7 & 3 & 2 \\ 7 & 7 & 12 \end{bmatrix};$$

$$A + (B + C) = \begin{bmatrix} 7 & 3 & 2 \\ 7 & 7 & 12 \end{bmatrix};$$

$$\lambda(A + B) = \begin{bmatrix} -2(6) & -2(1) & -2(-1) \\ -2(3) & -2(2) & -2(6) \end{bmatrix} = \begin{bmatrix} -12 & -2 & 2 \\ -6 & -4 & -12 \end{bmatrix};$$

$$\lambda A + \lambda B = \begin{bmatrix} -4 & 2 & -14 \\ -6 & -2 & 0 \end{bmatrix} + \begin{bmatrix} -8 & -4 & 16 \\ 0 & -2 & -12 \end{bmatrix} = \begin{bmatrix} -12 & -2 & 2 \\ -6 & -4 & -12 \end{bmatrix};$$

$$(\lambda + \mu)A = (-2 + 3)A = \begin{bmatrix} 2 & -1 & 7 \\ 3 & 1 & 0 \end{bmatrix};$$

$$\begin{aligned} \lambda A + \mu A &= -2A + 3A = \begin{bmatrix} -4 & 2 & -14 \\ -6 & -2 & 0 \end{bmatrix} + \begin{bmatrix} 6 & -3 & 21 \\ 9 & 3 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -1 & 7 \\ 3 & 1 & 0 \end{bmatrix}; \end{aligned}$$

$$\lambda(\mu A) = (-2)(3A) = \begin{bmatrix} -12 & 6 & -42 \\ -18 & -6 & 0 \end{bmatrix};$$

$$(\lambda\mu)A = [(-2)(3)]A = \begin{bmatrix} -12 & 6 & -42 \\ -18 & -6 & 0 \end{bmatrix};$$

$$1A = \begin{bmatrix} 2 & -1 & 7 \\ 3 & 1 & 0 \end{bmatrix} = A.$$

فرض کنیم ماتریسی که همه درایه‌های آن صفر است به‌طور خلاصه با O نموده شود، و $-A$ ماتریسی باشد که درایه‌های آن $-a_{ij}$ است. می‌توانیم خواص کلی جمع ماتریسی و ضرب اسکالر را در زیر فهرست کنیم. این خواص برای اینکه مجموعه همه

ماتریسهای $m \times n$ با درایه‌های حقیقی را به‌عنوان یک فضای برداری روی میدان اعداد حقیقی طبقه‌بندی کند کافی است. (ر.ک. نوبل^۱، [۴۵]، صفحات ۴۶۳ - ۴۶۰).

قضیه ۶.۷. فرض کنیم A ، B ، C ماتریسهایی $m \times n$ و λ و μ اعدادی حقیقی باشند. خواص زیر در مورد جمع و ضرب اسکالر برقرارند:

$$: A + B = B + A \quad (ت)$$

$$: (A + B) + C = A + (B + C) \quad (ب)$$

$$: A + O = O + A = A \quad (پ)$$

$$: A + (-A) = -A + A = O \quad (ت)$$

$$: \lambda(A + B) = \lambda A + \lambda B \quad (ث)$$

$$: (\lambda + \mu)A = \lambda A + \mu A \quad (ج)$$

$$: \lambda(\mu A) = (\lambda\mu)A \quad (چ)$$

$$: 1A = A \quad (ح)$$

در مثال ۲، خواص $(ت)$ ، $(ب)$ ، $(ث)$ ، $(ج)$ ، و $(ح)$ برای ماتریسهای خاص داده شده در آنجا بررسی شدند. درستی همه این خواص به‌سادگی تحقیق می‌شود. (ر.ک. تمرین ۰۵)

تعریف ۷.۷. فرض کنیم A یک ماتریس $m \times n$ و B یک ماتریس $n \times p$ باشد. حاصل ضرب A و B ، که با AB نموده می‌شود، ماتریس $m \times p$ ای مانند C است که درایه‌های آن c_{ij} به‌صورت زیرند:

بمازای هر $i = 1, 2, \dots, m$ و $j = 1, 2, \dots, p$

$$(۱۲۰۷) \quad c_{ij} = \sum_{k=1}^n a_{ik} b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj}$$

محاسبه c_{ij} با ضرب درایه‌های سطر i م A در درایه‌های نظیر آنها در ستون

j م B و با جمع این حاصل ضربها تعیین می شود؛ یعنی،

$$[a_{i1} \quad a_{i2} \quad \cdots \quad a_{in}] \begin{bmatrix} h_{1j} \\ h_{2j} \\ \vdots \\ h_{nj} \end{bmatrix} = c_{ij}$$

که در آن

$$c_{ij} = a_{i1}h_{1j} + a_{i2}h_{2j} + \cdots + a_{in}h_{nj} = \sum_{k=1}^n a_{ik}h_{kj}$$

این مطلب نشان می دهد که چرا باید تعداد ستونهای A برابر تعداد سطرهای B باشد تا حاصل ضرب AB تعریف شود.

مثال زیر روند ضرب ماتریسی را روشنتر می کند.

مثال ۳. فرض کنیم

$$A = \begin{bmatrix} 2 & 1 & -1 \\ 3 & 1 & 2 \\ 0 & -2 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 2 \\ -1 & 1 \\ 6 & 4 \end{bmatrix}$$

$$C = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 3 & 2 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & -1 & 1 \\ 2 & -1 & 2 \\ 3 & 0 & 3 \end{bmatrix}$$

در این صورت،

$$AD = \begin{bmatrix} 2 & 1 & -1 \\ 3 & 1 & 2 \\ 0 & -2 & -3 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 2 & -1 & 2 \\ 3 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & -3 & 1 \\ 11 & -4 & 11 \\ -13 & 2 & -13 \end{bmatrix}$$

$$DA = \begin{bmatrix} 1 & -1 & 1 \\ 2 & -1 & 2 \\ 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 & -1 \\ 3 & 1 & 2 \\ 0 & -2 & -3 \end{bmatrix} = \begin{bmatrix} -1 & -2 & -6 \\ 1 & -5 & -10 \\ 6 & -3 & -12 \end{bmatrix}$$

که نشان می‌دهد $AD \neq DA$.
بعلاوه،

$$BC = \begin{bmatrix} 3 & 2 \\ -1 & 1 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ -1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 9 & 4 \\ -3 & 2 & 2 \\ 8 & 18 & 8 \end{bmatrix}$$

و

$$CB = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 3 & 2 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ -1 & 1 \\ 6 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 5 \\ 6 & 9 \end{bmatrix}$$

در اینجا BC و CB حتی از یک اندازه هم نیستند.
بالاخره،

$$AB = \begin{bmatrix} 2 & 1 & -1 \\ 3 & 1 & 2 \\ 0 & -2 & -3 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ -1 & 1 \\ 6 & 4 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 20 & 15 \\ -16 & -14 \end{bmatrix}$$

درحالی که

$$BA = \begin{bmatrix} 3 & 2 \\ -1 & 1 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 & -1 \\ 3 & 1 & 2 \\ 0 & -2 & -3 \end{bmatrix}$$

را نمی‌توان محاسبه کرد.

تعریف ۸.۷. ماتریس همانی از مرتبه n ، $I_n = (\delta_{ij})$ ، ماتریسی با درایه‌های

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

است. وقتی اندازه I_n معلوم است، گاهی ماتریس را فقط به صورت I می‌نویسیم.

مثال ۴. ماتریس همانی از مرتبه ۳ عبارت است از

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

هرگاه A یک ماتریس 3×3 دیگر باشد، آنگاه

$$AI_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = A.$$

ملاحظه می شود که ماتریس همانی می تواند با A جابجا شود؛ یعنی، در ضرب ترتیب اهمیت ندارد؛ در نتیجه، $I_3 A = A$.

در مثال ۳ دیدیم که خاصیت $AB = BA$ برای ضرب ماتریسی معمولاً برقرار نیست. بعضی از خواص ضرب ماتریسی که برقرارند در قضیه بعد ارائه می شوند.

قضیه ۹.۷. فرض کنیم A یک ماتریس $n \times m$ ، B یک ماتریس $m \times k$ ، C یک ماتریس $k \times p$ ، D یک ماتریس $m \times k$ و λ عددی حقیقی باشد. در این صورت خواص زیر برقرارند.

$$A(BC) = (AB)C \quad (\text{ت})$$

$$I_m B = B \text{ and } BI_k = B \quad (\text{ب})$$

$$A(B + D) = AB + AD \quad (\text{پ})$$

$$\lambda(AB) = (\lambda A)B = A(\lambda B) \quad (\text{ت})$$

برهان. تحقیق درستی قسمت (ت) جهت نشان دادن روش بکار رفته ارائه می شود. تحقیق درستی سایر قسمتها در تمرین ۷ خواسته شده است. برای اثبات اینکه $A(BC) = (AB)C$ ، درایه i, j هر دو طرف معادله را حساب می کنیم. BC یک ماتریس $m \times p$ با درایه i, j .

$$(BC)_{ij} = \sum_{l=1}^k b_{il}c_{lj}.$$

است. بنابراین، $A(BC)$ ماتریسی $n \times p$ با درایه های

$$\begin{aligned} [A(BC)]_{ij} &= \sum_{s=1}^m a_{is}(BC)_{sj} \\ &= \sum_{s=1}^m a_{is} \left(\sum_{l=1}^k b_{sl}c_{lj} \right) \\ &= \sum_{s=1}^m \sum_{l=1}^k a_{is}b_{sl}c_{lj}. \end{aligned}$$

می‌باشد. مشابهاً، AB ماتریسی $n \times k$ با درایه‌های

$$(AB)_{ij} = \sum_{s=1}^m a_{is}b_{sj}$$

و از آنجا $(AB)C$ ماتریسی $n \times p$ با درایه‌های

$$\begin{aligned} [(AB)C]_{ij} &= \sum_{l=1}^k (AB)_{il}c_{lj} \\ &= \sum_{l=1}^k \left(\sum_{s=1}^m a_{is}b_{sl} \right) c_{lj} \\ &= \sum_{l=1}^k \sum_{s=1}^m a_{is}b_{sl}c_{lj} \\ &= \sum_{s=1}^m \sum_{l=1}^k a_{is}b_{sl}c_{lj} \end{aligned}$$

است. در نتیجه، برای هر $i = 1, 2, \dots, n$ و $j = 1, 2, \dots, p$ ، $[A(BC)]_{ij} = [(AB)C]_{ij}$ و $A(BC) = (AB)C$.

با تعریف ضرب ماتریسی، رابطه دستگاه‌های خطی با جبر خطی را می‌توان مطرح کرد. دستگاه خطی

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n. \end{aligned}$$

را می‌توان به عنوان معادله ماتریسی

$$Ax = b$$

در نظر گرفت، که در آن-

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

حال نشان می‌دهیم که حذف گاوسی بدون تغییرات سطرها ماتریس A را به حاصل ضرب دو ماتریس از نوع بسیار خاص تجزیه می‌کند.

تعریف ۱۰.۷. یک ماتریس $n \times n$ بالا مثلثی U دارای درایه‌های

$$u_{ij} = 0, \quad i = j + 1, j + 2, \dots, n$$

و یک ماتریس پایین مثلثی L دارای درایه‌های

$$l_{ij} = 0, \quad i = 1, 2, \dots, j - 1$$

است. یک ماتریس قطری D ماتریسی است که هم بالا مثلثی و هم پایین مثلثی باشد؛ یعنی، درایه‌های ناصفر فقط روی قطرند.

مثال ۵. مثالهای ۱ و ۳ بخش ۱۰.۷ را مجدداً در نظر می‌گیریم، که در آن‌ها دستگاه خطی

$$\begin{aligned} x_1 + x_2 + 3x_4 &= 4, \\ 2x_1 + x_2 - x_3 + x_4 &= 1, \\ 3x_1 - x_2 - x_3 + 2x_4 &= -3, \\ -x_1 + 2x_2 + 3x_3 - x_4 &= 4, \end{aligned}$$

به دستگاه معادل

$$\left[\begin{array}{cccc|c} 1 & 1 & 0 & 3 & 4 \\ 0 & -1 & -1 & -5 & -7 \\ 0 & 0 & 3 & 13 & 13 \\ 0 & 0 & 0 & -13 & -13 \end{array} \right]$$

تحویل شد. فرض کنیم U ماتریس 4×4 بالا مثلثی

$$U = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix}$$

باشد، که نتیجه اجرای حذف گاوسی روی A است. به‌ازای $i = 1, 2, 3$ ، به‌ازای m_{ji} را به‌ازای هر $j = i + 1, i + 2, \dots, 4$ برابر عدد بکار رفته در مرحله حذفی $(E_j - m_{ji}E_i) \rightarrow E_j$ تعریف کنید؛ یعنی، $m_{21} = 2$ ، $m_{31} = 3$ ، $m_{41} = -1$ ، $m_{32} = 4$ ، $m_{42} = -3$ ، و $m_{43} = 0$. هرگاه L ماتریس 4×4 پایین مثلثی با درایه‌های l_{ij} داده شده به صورت

$$l_{ij} = \begin{cases} 0, & \text{اگر } i = 1, 2, \dots, j - 1 \\ 1, & \text{اگر } i = j \\ m_{ij}, & \text{اگر } i = j + 1, j + 2, \dots, n \end{cases}$$

تعریف شود، آنگاه

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix};$$

و به سادگی تحقیق می‌شود که

$$LU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix}$$

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

نتایج حاصل در مثال ۵ در حالت کلی درست و در قضیه زیر ارائه می‌شوند.

قضیه ۱۱.۷. هرگاه روش حذف گاوسی (الگوریتم ۱۰.۷) را بتوان روی دستگاه $Ax = b$ بدون تعویض سطرها انجام داد، آنگاه ماتریس A به حاصل ضرب یک ماتریس پایین مثلثی در یک ماتریس بالا مثلثی تجزیه می‌شود:

$$A = LU,$$

که در آن $U = (u_{ij})$ و $L = (l_{ij})$ به صورت

$$u_{ij} = \begin{cases} a_{ij}^{(i)} & \text{اگر } i = 1, 2, \dots, j \\ 0 & \text{اگر } i = j + 1, j + 2, \dots, n \end{cases}$$

و

$$l_{ij} = \begin{cases} 0 & \text{اگر } i = 1, 2, \dots, j-1 \\ 1 & \text{اگر } i = j \\ m_{ij} & \text{اگر } i = j + 1, j + 2, \dots, n \end{cases}$$

تعریف می‌شوند، که $a_{ij}^{(i)}$ عنصر i, j از $A^{(i)}$ ، آخرین ماتریس حاصل از روش حذف گاوسی،

و m_{ij} مضرب است.

برهان. ر.ک. تمرین ۱۱.

اگر برای ادامه این روند تعویض سطرها لازم باشد، A را می توان به LU تجزیه کرد، که در آن U همان ماتریس موجود در قضیه ۱۱.۷ است، ولی در حالت کلی، L پایین مثلثی نیست. به این موضوع در تمرینات بیشتر پرداخته شده است (بخصوص، ر.ک. تمرینات ۱۲ و ۱۳).

همچنین، مفهوم معکوس یک ماتریس با دستگاههای خطی در ارتباط است.

تعریف ۱۲.۷. معکوس ماتریس $n \times n$ ، A ماتریسی $n \times n$ است، که با A^{-1} نموده می شود، و در خاصیت $AA^{-1} = A^{-1}A = I_n$ صدق می کند. ماتریسی که دارای معکوس است نامنفرد و ماتریس بدون معکوس منفرد نامیده می شود.

مثال ۶. فرض کنیم

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} \quad \text{و} \quad B = \frac{1}{9} \begin{bmatrix} -2 & 5 & -1 \\ 4 & -1 & 2 \\ -3 & 3 & 3 \end{bmatrix}$$

چون

$$AB = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} \frac{1}{9} \begin{bmatrix} -2 & 5 & -1 \\ 4 & -1 & 2 \\ -3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

و

$$BA = \frac{1}{9} \begin{bmatrix} -2 & 5 & -1 \\ 4 & -1 & 2 \\ -3 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

A و B نامنفردند، $B = A^{-1}$ و $A = B^{-1}$.

فایده معکوس یک ماتریس را با در نظر گرفتن دستگاه خطی

$$\begin{aligned}x_1 + 2x_2 - x_3 &= 2, \\2x_1 + x_2 &= 3, \\-x_1 + x_2 + 2x_3 &= 4\end{aligned}$$

می‌توان نشان داد. ابتدا دستگاه را به صورت ماتریسی برمی‌گردانیم:

$$\begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}.$$

و سپس طرفین را در معکوس داده شده در بالا ضرب می‌کنیم:

$$\begin{aligned}\frac{1}{9} \begin{bmatrix} -2 & 5 & -1 \\ 4 & -1 & 2 \\ -3 & 3 & 3 \end{bmatrix} \left(\begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) \\ = \frac{1}{9} \begin{bmatrix} -2 & 5 & -1 \\ 4 & -1 & 2 \\ -3 & 3 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix};\end{aligned}$$

در نتیجه،

$$\left(\begin{bmatrix} -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ -\frac{3}{9} & \frac{3}{9} & \frac{3}{9} \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) = \frac{1}{9} \begin{bmatrix} 7 \\ 13 \\ 15 \end{bmatrix}$$

و

$$I_3 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{7}{9} \\ \frac{13}{9} \\ \frac{15}{9} \end{bmatrix}.$$

که جواب $x_1 = \frac{7}{9}$ ، $x_2 = \frac{13}{9}$ و $x_3 = \frac{15}{9}$ را بدست می‌دهد.

در حالت کلی، این روش برای حل $Ax = b$ به علت تعدد اعمال لازم توصیه نمی‌شود.

(ر.ک. تمرینهای ۹ (ج)، ۹ (چ)، و ۹ (ح).)

جهت یافتن روشی برای محاسبه A^{-1} ، با فرض وجود A^{-1} ، مجدداً مناسب است

که ضرب ماتریسی ملاحظه شود. فرض کنیم B_j ستون j م ماتریس $n \times n$ ، B باشد،

$$B_j = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}$$

بنابر ضرب ،

$$AB_j = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n a_{1k} b_{kj} \\ \sum_{k=1}^n a_{2k} b_{kj} \\ \vdots \\ \sum_{k=1}^n a_{nk} b_{kj} \end{bmatrix}$$

هرگاه $AB = C$ ، آنگاه ستون j م C با

$$C_j = \begin{bmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{nj} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n a_{1k} b_{kj} \\ \sum_{k=1}^n a_{2k} b_{kj} \\ \vdots \\ \sum_{k=1}^n a_{nk} b_{kj} \end{bmatrix}$$

داده می‌شود. از اینرو، ستون j م حاصل ضرب AB برابر حاصل ضرب A در ستون j م B است.

فرض کنیم A^{-1} وجود داشته باشد و $A^{-1} = B = (b_{ij})$ ؛ در این صورت، $AB = I$ و

$$AB_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

که در آن مقدار ۱ در سطر j م ظاهر می‌شود. برای تعیین دقیق B ، باید n دستگاه خطی را حل کنیم ، که در آن ستون j م ماتریس معکوس جواب دستگاه خطی است که طرف راست آن ستون j م I است. مثال بعد این روش را نشان می‌دهد.

مثال ۷. فرض کنیم

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

برای محاسبه A^{-1} ، باید سه دستگاه خطی

$$\begin{aligned} x_1 + 2x_2 - x_3 &= 1, & x_1 + 2x_2 - x_3 &= 0, \\ 2x_1 + x_2 &= 0, & 2x_1 + x_2 &= 1, \\ -x_1 + x_2 + 2x_3 &= 0; & -x_1 + x_2 + 2x_3 &= 0; \\ x_1 + 2x_2 - x_3 &= 0, \\ 2x_1 + x_2 &= 0, \\ -x_1 + x_2 + 2x_3 &= 1 \end{aligned}$$

را حل کنیم. با استفاده از حذف گاوسی، محاسبات براحتی روی ماتریس افزوده بزرگتر زیر، که از الحاق ماتریسها تشکیل شده، انجام می‌شود:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 2 & 0 & 0 & 1 \end{array} \right]$$

برای اتمام، چون ماتریس ضرایب تغییر نمی‌کند، باید رشته اعمال سطری یکسان برای هر دستگاه انجام داد. ابتدا با انجام $(E_2 - 2E_1) \rightarrow (E_2)$ و $(E_3 + E_1) \rightarrow (E_3)$ ، داریم

$$\left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -3 & 2 & -2 & 1 & 0 \\ 0 & 3 & 1 & 1 & 0 & 1 \end{array} \right]$$

سپس، با انجام $(E_3 + E_2) \rightarrow (E_3)$ بدست می‌آوریم

$$\left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -3 & 2 & -2 & 1 & 0 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array} \right]$$

با انجام جایگذاری پسر و روی هریک از سه ماتریس افزوده

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 0 & -3 & 2 & -2 \\ 0 & 0 & 3 & -1 \end{array} \right], \quad \left[\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 0 & -3 & 2 & 1 \\ 0 & 0 & 3 & 1 \end{array} \right], \quad \left[\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 0 & -3 & 2 & 0 \\ 0 & 0 & 3 & 1 \end{array} \right]$$

همه درایه‌های ۱. بدست می‌آیند؛ لیکن اغلب مناسبتر است که تحویل سطری بیشتری

مورد استفاده قرار گیرد. بخصوص، عمل $(E_3) \rightarrow (\frac{1}{3}E_3)$ نتیجه می دهد

$$\left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -3 & 2 & -2 & 1 & 0 \\ 0 & 0 & 1 & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right]$$

و
 $(E_2 - 2E_3) \rightarrow (E_2)$ و $(E_1 + E_3) \rightarrow (E_1)$ ماتریس

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & -3 & 0 & -\frac{4}{3} & \frac{1}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right]$$

را موجب می شود. با انجام $(E_2) \rightarrow (-\frac{1}{3}E_2)$ بدست می آید

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 1 & 0 & \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ 0 & 0 & 1 & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right];$$

و، بالاخره، از $(E_1) \rightarrow (E_1 - 2E_2)$ نتیجه می شود

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ 0 & 1 & 0 & \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ 0 & 0 & 1 & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right].$$

آخرین ماتریس افزوده جواب این سه دستگاه خطی را ارائه می دهد:

$$\begin{array}{lll} x_1 = -\frac{2}{9}, & x_1 = \frac{5}{9}, & x_1 = -\frac{1}{9}, \\ x_2 = \frac{4}{9}, & x_2 = -\frac{1}{9}, & x_2 = \frac{2}{9}, \\ x_3 = -\frac{1}{3}, & x_3 = \frac{1}{3}, & x_3 = \frac{1}{3}. \end{array}$$

در نتیجه،

$$A^{-1} = \begin{bmatrix} -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \frac{1}{9} \begin{bmatrix} -2 & 5 & -1 \\ 4 & -1 & 2 \\ -3 & 3 & 3 \end{bmatrix}$$

در آخرین مثال، محاسبه A^{-1} بررسی شد. همانطور که در آن مثال دیدیم، شایسته است ماتریس افزوده بزرگتری تشکیل دهیم:

$$[A : I].$$

با انجام حذف، مطابق الگوریتم ۱.۷، ماتریس

$$[U : Y]$$

بدست می‌آید، که در آن Y ماتریس $n \times n$ حاصل از انجام همان اعمال روی ماتریس همانی I است که A را به U تبدیل می‌کند. در این لحظه انتخابی بین n کاربرد الگوریتم جایگذاری پسرو یا تحویل به

$$[I : A^{-1}]$$

وجود دارد. اگر جایگذاری پسرو مستقیماً بکار رود، $\frac{1}{3}n - \frac{1}{3}n^3$ ضرب و تقسیم و $\frac{1}{6}n + \frac{2}{3}n^2 - \frac{1}{3}n^3$ جمع و تفریق لازم است (ر.ک. تمرین ۹ (پ)). همچنین، در آن تمرین نشان داده می‌شود که تحویل بیشتر، که حالت خاصی از روش گاوس-ژردان است، نیازه $\frac{1}{2}n - \frac{1}{3}n^3$ ضرب و تقسیم و $\frac{1}{2}n + 2n^2 - \frac{1}{3}n^3$ جمع و تفریق دارد. جدول زیر مقدار نسبی محاسبات به‌ازای تغییرات اندازه n را نشان می‌دهد.

n	حذف گاوس		گاوس-ژردان	
	تقسیم / ضرب	تفریق / جمع	تقسیم / ضرب	تفریق / جمع
	$(4n^3 - n)/3$	$(8n^3 - 9n^2 + n)/6$	$(3n^3 - n)/2$	$(3n^3 - 4n^2 + n)/2$
3	35	23	39	24
10	1330	1185	1495	1305
50	166,650	162,925	187,475	182,525
100	1,333,300	1,318,350	1,499,950	1,480,050

در اجرای واقعی هر دو روش، می‌توان با بذل توجه خاص، اعمالی که احتیاج نیست انجام شوند را شناسایی کرد؛ به‌عنوان مثال، یک ضرب هنگام 1 بودن مضرب، یا یک تفریق وقتی مفروق صفر است. برای هر دو روش، تعداد ضربها و تقسیمهای لازم به n^3

n	هر دو روش، با حذف اعمال غیر لازم	
	تقسیم / ضرب	تفریق / جمع
3	27	12
10	1000	810
50	125,000	120,050
100	1,000,000	980,100

و تعداد جمعها و تفریقها به $n^3 - 2n^2 + n$ کاهش می یابد (ر. ک. تمرینهای ۹ (ت) و ۹ (ث)).
 یک مفهوم اساسی جبر خطی که در تعیین وجود و یکتایی جوابهای دستگاههای خطی بسیار مفید است دترمینان یک ماتریس $n \times n$ است. در اینجا تنها راه ارائه شده برای تعریف و محاسبه دترمینان تعریف بازگشتی است. دترمینان ماتریس A با $\det A$ ، و نیز به طور مرسوم با $|A|$ ، مشخص می شود. زیرماتریس یک ماتریس A ماتریسی "استخراج شده" از آن با حذف سطرها یا ستونهای معینی از A است.

تعریف ۱۳.۷ (آ). هرگاه $A = [a]$ یک ماتریس 1×1 باشد، آنگاه $\det A = a$.

(ب) مینور، M_{ij} ، عبارت است از دترمینان زیرماتریس $(n-1) \times (n-1)$ از ماتریس $n \times n$ ، A ، که با حذف سطر i م و ستون j م حاصل می شود.

(پ) همعامل، A_{ij} ، وابسته به M_{ij} ، بنابه تعریف عبارت است از $A_{ij} = (-1)^{i+j} M_{ij}$.

(ت) دترمینان ماتریس $n \times n$ ، A ، که در آن $n > 1$ ، با

$$\det A = \sum_{j=1}^n a_{ij} A_{ij}, \quad i = 1, 2, \dots, n \quad \text{بهازای هر } i \quad (13.7)$$

یا با

$$\det A = \sum_{i=1}^n a_{ij} A_{ij}, \quad j = 1, 2, \dots, n \quad \text{بهازای هر } j \quad (14.7)$$

ارائه می شود.

گرچه بنظر می رسد $2n$ تعریف مختلف برای $\det A$ وجود دارد، ولی می توان نشان داد که همه آنها نتیجه عددی یکسان بدست می دهند. (ر. ک. نوبل [۴۵]، صفحات ۲۰۳ - ۲۰۱).

برای حالت ماتریسهای 2×2 ، مثال زیر معادل بودن این تعاریف را نشان می دهد.

مثال ۸

(آ) فرض کنیم $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. با استفاده از معادله (۱۳.۷) با $i = 1$ ،

$$\begin{aligned} \det A &= a_{11}A_{11} + a_{12}A_{12} \\ &= a_{11}[(-1)^{1+1} \det(a_{22})] + a_{12}[(-1)^{1+2} \det(a_{21})] \\ &= a_{22}a_{11} - a_{21}a_{12}. \end{aligned}$$

با استفاده از (۱۳.۷) با $i = 2$ ،

$$\begin{aligned} \det A &= a_{21}A_{21} + a_{22}A_{22} \\ &= -a_{21}a_{12} + a_{22}a_{11}. \end{aligned}$$

با استفاده از (۱۴.۷) با $j = 1$ ،

$$\begin{aligned} \det A &= a_{11}A_{11} + a_{21}A_{21} \\ &= a_{11}a_{22} - a_{21}a_{12}. \end{aligned}$$

با استفاده از (۱۴.۷) به‌ازای $j = 2$ ،

$$\begin{aligned} \det A &= a_{12}A_{12} + a_{22}A_{22} \\ &= -a_{12}a_{21} + a_{22}a_{11}. \end{aligned}$$

در تمام حالات ،

$$\det A = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

(ب) فرض کنیم

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 3 & -1 & 1 \\ 4 & 6 & -2 \end{bmatrix}$$

با استفاده از (۱۳.۷) به‌ازای $i = 3$ ،

$$\begin{aligned} \det A &= a_{31}A_{31} + a_{32}A_{32} + a_{33}A_{33} \\ &= 4A_{31} + 6A_{32} - 2A_{33} \\ &= 4(-1)^{3+1} \det \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix} + 6(-1)^{3+2} \det \begin{bmatrix} 2 & 0 \\ 3 & 1 \end{bmatrix} \\ &\quad - 2(-1)^{3+3} \det \begin{bmatrix} 2 & 0 \\ 3 & -1 \end{bmatrix} \\ &= -8. \end{aligned}$$

واضح است که مناسبترین کار محاسبه $\det A$ نسبت به سطر یا ستون با بیشترین صفر است .

(پ) فرض کنیم

$$A = \begin{bmatrix} 2 & -1 & 3 & 0 \\ 4 & -2 & 7 & 0 \\ -3 & -4 & 1 & 5 \\ 6 & -6 & 8 & 0 \end{bmatrix}$$

برای محاسبه $\det A$ ، T ، سادترین کار استفاده از ستون چهارم است، (۱۴.۷) به ازای $j = 4$ خواهد شد

$$\begin{aligned} \det A &= a_{14}A_{14} + a_{24}A_{24} + a_{34}A_{34} + a_{44}A_{44} \\ &= 5A_{34} \\ &= -5 \det \begin{bmatrix} 2 & -1 & 3 \\ 4 & -2 & 7 \\ 6 & -6 & 8 \end{bmatrix} \\ &= -5 \left\{ 2 \det \begin{bmatrix} -2 & 7 \\ -6 & 8 \end{bmatrix} - (-1) \det \begin{bmatrix} 4 & 7 \\ 6 & 8 \end{bmatrix} \right. \\ &\quad \left. + 3 \det \begin{bmatrix} 4 & -2 \\ 6 & -6 \end{bmatrix} \right\} \\ &= -30. \end{aligned}$$

خواص زیر از دترمینانها در ارتباط با دستگاههای خطی و حذف گاوسی با دترمینان مفیدند.

قضیه ۱۴.۷. فرض کنیم A یک ماتریس $n \times n$ باشد.

- (ت) هرگاه تمام درایه‌های سطری یا ستونی از A برابر صفر باشد، آنگاه $\det A = 0$.
- (ب) هرگاه \tilde{A} بوسیله عمل $(E_i) \leftrightarrow (E_j)$ ، با $i \neq j$ ، از A حاصل شود، آنگاه $\det \tilde{A} = -\det A$.
- (پ) هرگاه A دارای دو سطر یکسان باشد، آنگاه $\det A = 0$.
- (ت) هرگاه \tilde{A} بوسیله عمل $(\lambda E_i) \rightarrow (E_i)$ از A حاصل شود، آنگاه $\det \tilde{A} = \lambda \det A$.
- (ث) هرگاه \tilde{A} بوسیله عمل $(E_i + \lambda E_j) \rightarrow (E_i)$ ، با $i \neq j$ ، از A حاصل شود، آنگاه $\det \tilde{A} = \det A$.
- (ج) هرگاه B نیز ماتریسی $n \times n$ باشد، آنگاه $\det AB = \det A \det B$.

برهان

(آ) فرض کنیم سطر i م A صفر باشد. با بسط نسبت به سطر i م، می‌نویسیم

$$\det A = \sum_{j=1}^n a_{ij} A_{ij} = \sum_{j=1}^n 0 \cdot A_{ij} = 0.$$

(ب) برهان را در هر کتاب جبر خطی متعارف می‌توان یافت. (مثلاً، ر.ک. نوبل

[۴۵]، صفحه ۲۰۳.) حالتی که n برابر سه است در تعریف آمده است.

(پ) فرض کنیم سطرهای i م و j م یکسان باشند. \bar{A} را با تعویض سطرهای i م و j م

تشکیل می‌دهیم. بنا به (ب)، $\det \bar{A} = -\det A$ ، چون عمل $(E_i) \leftrightarrow (E_j)$ ماتریس A را

تغییر نمی‌دهد، در این حالت $\det \bar{A} = \det A$ ، که نتیجه می‌دهد $\det A = -\det A$ و از آنجا

$$\det A = 0$$

(ت) با بسط نسبت به سطر i م \bar{A} نتیجه می‌شود

$$\det \bar{A} = \sum_{k=1}^n \lambda a_{ik} A_{ik} = \lambda \sum_{k=1}^n a_{ik} A_{ik} = \lambda \det A.$$

(ث) با بسط نسبت به سطر i م \bar{A} نتیجه می‌شود

$$\begin{aligned} \det A &= \sum_{k=1}^n (a_{ik} + \lambda a_{jk}) A_{ik} \\ &= \sum_{k=1}^n a_{ik} A_{ik} + \lambda \sum_{k=1}^n a_{jk} A_{ik} = \det A + \lambda \det B, \end{aligned}$$

که در آن ماتریس B دقیقاً همان ماتریس A است جز اینکه سطر i م B سطر j م A است.

پس، بنا به (پ)، $\det B = 0$ و از آنجا $\det \bar{A} = \det A$.

(ج) اثبات این حکم را می‌توان در صفحات ۲۰۷ - ۲۰۶ نوبل [۴۵] پیدا کرد.

قضیه ۱۵۰۷. هرگاه $A = (a_{ij})$ ماتریسی $n \times n$ بالا مثلثی یا پایین مثلثی یا قطری

$$\text{باشد، آنگاه } \det A = \prod_{i=1}^n a_{ii}.$$

اثبات این حکم متشکل از بسط ماتریس و هر زیرماتریس آن نسبت به سطر اول یا

ستون اول است، و به خواننده واگذار می‌شود.

مسئله محاسبه دترمینان یک ماتریس را می‌توان ابتدا با استفاده از خواص ارائه

شده در قضیه ۱۴.۷ برای تحویل ماتریس به صورت مثلثی و سپس استفاده از قضیه ۱۵.۷ جهت تعیین دترمینان ماتریس مثلثی خلاصه کرد. مثال زیر این روش را تشریح می‌کند.

مثال ۹. فرض کنیم

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

بوسیله سه نوع عمل مجاز، ماتریس A را به یک ماتریس بالا مثلثی تحویل می‌کنیم. ابتدا $(E_2 - 2E_1) \rightarrow (E_2)$ ، $(E_3 + E_1) \rightarrow (E_3)$ ، و $(E_4 - 3E_1) \rightarrow (E_4)$ را انجام می‌دهیم تا

$$\tilde{A}_1 = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 3 & 3 & 2 \\ 0 & -4 & -1 & -7 \end{bmatrix}$$

بدست آید. بنا بر قضیه ۱۴.۷ (۵)، $\det A = \det \tilde{A}_1$ ، \tilde{A}_2 را از \tilde{A}_1 بوسیله اعمال $(E_3 + 3E_2) \rightarrow (E_3)$ و $(E_4 - 4E_2) \rightarrow (E_4)$ تشکیل می‌دهیم،

$$\tilde{A}_2 = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 0 & -13 \\ 0 & 0 & 3 & 13 \end{bmatrix}$$

و مجدداً " $\det \tilde{A}_2 = \det \tilde{A}_1 = \det A$ ". فرض کنیم \tilde{A}_3 بوسیله $(E_3) \leftrightarrow (E_4)$ از \tilde{A}_2 تشکیل شده باشد،

$$\tilde{A}_3 = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix}$$

بنابر قضیه ۱۵.۷، $\det \bar{A}_3 = (1)(-1)(3)(-13) = 39$ ، و چون \bar{A}_3 از \bar{A}_2 با تعویض سطر بدست آمده است،

$$\det A = \det \bar{A}_2 = -\det \bar{A}_3 = -39.$$

حال کلیدی ارائه می‌شود که نامنفردی، حذف گاوسی، دستگاه‌های خطی، و دترمینانها را بهم ربط می‌دهد. اثبات این قضیه مشکل نیست ولی پرزحمت است، و در اینجا ارائه نمی‌شود. (ر.ک. نوبل [۴۵].)

قضیه ۱۶.۷. عبارات زیر برای هر ماتریس $n \times n$ ، A معادلند:

(آ) معادله $Ax = 0$ دارای جواب منحصر بفرد $x = 0$ است؛

(ب) دستگاه خطی $Ax = b$ به‌ازای هر بردار ستونی b بعدی دارای جواب منحصر بفرد است؛

(پ) ماتریس A نامنفرد است؛ یعنی، A^{-1} وجود دارد.

(ت) $\det A \neq 0$ ؛

(ث) الگوریتم ۱۰.۷ (حذف گاوسی با تعویض سطرها) را روی دستگاه خطی $Ax = b$ به‌ازای هر بردار ستونی b بعدی می‌توان انجام داد.

مجموعه تمرینات ۳.۷

۱. دستگاه خطی

$$2x + \quad \quad z = 2,$$

$$x - y - z = 0,$$

$$4x + 5y + 3z = 7.$$

را به‌صورت معادله بردار - ماتریسی بنویسید.

۲. تعیین کنید از ماتریسهای زیر کدامها نامنفردند، و معکوس آنها را مشخص کنید.

$$A = \begin{bmatrix} 4 & 2 & 6 \\ 3 & 0 & 7 \\ -2 & -1 & -3 \end{bmatrix} \quad (\text{آ})$$

$$B = \begin{bmatrix} 2 & 3 & 1 & 2 \\ -2 & 4 & -1 & 5 \\ 3 & 7 & \frac{1}{2} & 1 \\ 6 & 9 & 3 & 7 \end{bmatrix} \quad (\text{ب})$$

$$C = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 6 & 7 & 0 & 0 \\ 9 & 11 & 1 & 0 \\ 5 & 4 & 1 & 1 \end{bmatrix} \quad (\text{پ})$$

۳. برای ماتریسهای

$$A = \begin{bmatrix} 4 & 6 & 1 & -1 \\ 2 & 1 & 0 & \frac{1}{2} \\ 3 & 0 & 0 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix} \quad \text{و} \quad B = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & -1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

، $\det A$ ، $\det B$ ، $\det AB$ ، و $\det BA$ را محاسبه کنید.

۴. نشان دهید هرگاه A و B ماتریسهایی $n \times n$ بوده و $\bar{A}B = I$ ، آنگاه $B = A^{-1}$.

۵. قضیه ۶.۷ را ثابت کنید.

۶. فرض کنید A ماتریسی 3×3 باشد. نشان دهید هرگاه \bar{A} ماتریس حاصل از A با استفاده از هریک از اعمال

$$(E_2) \leftrightarrow (E_3) \text{ یا } (E_1) \leftrightarrow (E_3) \text{ ، } (E_1) \leftrightarrow (E_2)$$

باشد، آنگاه $\det \bar{A} = -\det A$.

۷. درستی قسمتهای (ب) ، (پ) ، و (ت) قضیه ۹.۷ را تحقیق کنید.

۸. روش گاوس - ژردان برای حل معادله خطی $Ax = b$ به صورت زیر توصیف می شود:

معادله i را جهت حذف x_i نه فقط از معادلات $E_{i+1}, E_{i+2}, \dots, E_n$ ، همانطور که در روش حذف گاوس انجام شد، بلکه از E_1, E_2, \dots, E_{i-1} بکار می بریم. تا $[A : b]$ به

$$\left[\begin{array}{cccc|c} a_{11}^{(1)} & 0 & \cdots & 0 & a_{1,n+1}^{(1)} \\ 0 & a_{22}^{(2)} & \cdots & \vdots & a_{2,n+1}^{(2)} \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & a_{nn}^{(n)} & a_{n,n+1}^{(n)} \end{array} \right]$$

تحویل شود. با قرار دادن

$$x_i = \frac{a_{i,n+1}^{(i)}}{a_{ii}^{(i)}}$$

به ازای هر $i = 1, 2, \dots, n$ ، جواب بدست می آید. این روند بر جایگذاری پسرو در حذف گاوسی مزیت دارد.

(آ) الگوریتمی برای روند گاوس-ژردان با تقلید از الگوریتم ۱۰۷ بسازید.

(ب) دستگاههای خطی تمرین ۱ از بخش ۲۰۷ را بوسیله روش گاوس-ژردان حل کنید.

(پ) دستگاههای خطی تمرین ۲ از بخش ۲۰۷ را بوسیله روش گاوس-ژردان حل کنید.

(ت) نشان دهید که روش گاوس-ژردان به $\frac{n^3}{2} + n^2 - \frac{n}{2}$ ضرب و تقسیم و $\frac{n^3}{2} - \frac{n}{2}$ جمع و تفریق نیاز دارد.

(ث) یک جدول مقایسه اعمال لازم برای زوشهای گاوس-ژردان و حذف گاوسی به ازای $n = 3, 10, 50, 100$ بسازید. کدام روش به محاسبه کمتری نیاز دارد؟

۹. فرض کنید m دستگاه خطی

$$AX^{(p)} = b^{(p)}, \quad p = 1, 2, \dots, m.$$

هریک با ماتریس ضرایب A ، برای حل موجود باشند.

(آ) نشان دهید بکار بردن حذف گاوسی روی ماتریس افزوده

$$[A: b^{(1)} b^{(2)} \dots b^{(m)}]$$

به $\frac{n^3}{3} + mm^2 - \frac{n}{3}$ ضرب و تقسیم و به $\frac{n^3}{3} + mm^2 - \frac{n}{2} - mm + \frac{n}{6}$ جمع و تفریق نیاز

دارد.

(ب) نشان دهید بکار بردن روش گاوس-ژردان روی ماتریس افزوده

$$[A: b^{(1)} b^{(2)} \dots b^{(m)}]$$

به $\frac{1}{2}n^3 + mm^2 - \frac{1}{2}n$ ضرب و تقسیم و به $(\frac{1}{2} - m)n + (m-1)n^2 + \frac{1}{2}n^3$ جمع و تفریق نیاز

دارد.

(پ) برای حالت خاص

$$b^{(p)} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \text{، سطر } p \text{ ام}$$

به‌ازای هر $p = 1, \dots, m$ با $m = n$ ، جواب $x^{(p)}$ ستون p ام A^{-1} است. نشان دهید در این مورد روش حذف گاوسی به $\frac{1}{3}n^3 - \frac{1}{3}n^2 + \frac{1}{6}n$ ضرب و تقسیم و به $\frac{1}{3}n^3 - \frac{1}{2}n^2 + \frac{1}{6}n$ جمع و تفریق نیاز دارد و روش گاوس-ژردان به $\frac{1}{3}n^3 - \frac{1}{2}n^2 + \frac{1}{6}n$ ضرب و تقسیم و به $\frac{1}{3}n^3 - 2n^2 + \frac{1}{2}n$ جمع و تفریق نیاز دارد.

(ت) با استفاده از حذف گاوسی، الگوریتمی بسازید که A^{-1} را بدست دهد ولی وقتی یکی از مضارب واحد است عمل ضرب انجام نشود، و وقتی یکی از عناصر درگیر جمع یا تفریق صفر است عمل جمع و تفریق انجام نشود. نشان دهید که محاسبات لازم‌رانی‌تان به n^3 ضرب و تقسیم و $n^3 - 2n^2 + n$ جمع و تفریق تقلیل داد.

(ث) با استفاده از روش گاوس-ژردان، قسمت (ت) را تکرار کنید.

(ج) نشان دهید که حل دستگاه خطی $AX = b$ ، با معلوم بودن A^{-1} ، هنوز به n^2 ضرب و تقسیم و $n^2 - n$ جمع و تفریق نیاز دارد.

(چ) نشان دهید که حل m دستگاه خطی $AX^{(m)} = b^{(m)}$ ، به‌ازای $p = 1, 2, \dots, m$ ، بوسیله روش $x^{(m)} = A^{-1}b^{(m)}$ ، اگر A^{-1} معلوم باشد، نیاز به mn^2 ضرب و $m(n^2 - n)$ جمع دارد.

(ح) فرض کنید A ماتریسی $n \times n$ باشد. به‌ازای $n = 3, 10, 50, 100$ ، تعداد اعمال لازم جهت حل m دستگاه خطی شامل A بوسیله حذف گاوسی و بوسیله ابتدا معکوس کردن A و سپس ضرب $AX = b$ در A^{-1} را مقایسه کنید. آیا به‌منظور حل دستگاههای خطی، محاسبه A^{-1} مفید است؟

۱۰. با استفاده از الگوریتم پرورده شده در تمرین ۹، معکوس ماتریسهای نامفرد تمرین ۲ را بیابید.

۱۱. فرض کنید به‌ازای $j = i + 1, i + 2, \dots, n$ ، $m_{ji} = a_{ji}^{(i)} / a_{ii}^{(i)}$ ، و به‌ازای هر $i = 1, 2, \dots, n - 1$ تعریف کنید

$$(M^{(1)}) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ & \ddots & & \\ 0 & 0 & 1 & \dots \\ & & 0 & -m_{i+1,i} & 1 & \dots \\ \vdots & & \vdots & -m_{i+2,i} & 0 & \dots & 0 \\ & & \vdots & \vdots & \vdots & \dots & \\ 0 & \dots & 0 & -m_{n,i} & 0 & \dots & 0 & 1 \end{bmatrix}$$

- (ت) نشان دهید $\bar{A}^{(2)} = M^{(1)}\bar{A}^{(1)}$ ؛ یعنی به ازای هر $j = 2, 3, \dots, n$ اعمال $(E_i - m_{i+1,i}E_i) \rightarrow (E_i)$ معادل ضربهای ماتریسی $M^{(1)}A$ و $M^{(1)}b$ است.
- (ب) فرض کنید تعویض سطرها لازم نباشد؛ نشان دهید به ازای هر $k = 2, \dots, n$ ،
- $$\bar{A}^{(k)} = M^{(k-1)}M^{(k-2)}\dots M^{(1)}\bar{A}^{(1)}.$$
- (پ) نشان دهید به ازای هر i ، $M^{(i)}$ نامنفرد است، و

$$(M^{(i)})^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ & \ddots & & \\ 0 & 0 & 1 & \dots \\ & & 0 & m_{i+1,i} & 1 & \dots \\ \vdots & & \vdots & m_{i+2,i} & 0 & \dots & 0 \\ & & \vdots & \vdots & \vdots & \dots & \\ 0 & \dots & 0 & m_{n,i} & 0 & \dots & 0 & 1 \end{bmatrix}$$

- (ت) فرض کنید $M = M^{(n-1)}M^{(n-2)}\dots M^{(1)}$. نشان دهید M پایین مثلثی و نامنفرد است و $U = MA$ ، که در آن U ماتریس بالامثلثی تشکیل شده در حذف گاوسی است.

(ث) نشان دهید

$$L = M^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ m_{21} & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ m_{n1} & \dots & m_{n,n-1} & 1 \end{bmatrix}$$

و A را می‌توان به حاصل ضرب یک ماتریس پایین مثلثی و یک ماتریس بالا مثلثی تجزیه کرد.

۱۲. یک ماتریس جایگشت ماتریسی است که در هر ستون و هر سطر یک درایه با مقدار ۱ دارد و همه درایه‌های دیگر صفرند. مثلاً،

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

یک ماتریس جایگشت 3×3 است.

(T) نشان دهید عمل $(E_i) \leftrightarrow (E_j)$ که روی یک ماتریس $n \times n$ انجام شده معادل است با حاصل ضرب PA ، که در آن P یک ماتریس جایگشت است که از بکاربردن $(E_i) \leftrightarrow (E_j)$ روی I_n تشکیل یافته است.

(ب) ضرب AP ، که در آن P ماتریس جایگشت حاصل در قسمت (T) است، دارای چه اثری است؟

(پ) نشان دهید برای هر ماتریس جایگشت P می‌توان P^{-1} را با تغییر سطرها و ستونها بدست آورد؛ یعنی، $(P^{-1})_{ij} = (P)_{ji}$.

(ت) اگر تعویض سطرها در الگوریتم ۱۰.۷ جهت تحویل A لازم باشد و روی I بکار رود تا ماتریس P حاصل شود، نشان دهید $LU = PA$. از آنجا نتیجه می‌شود $A = (P^{-1}L)U$ ، لیکن $P^{-1}L$ لزوماً پایین مثلثی نیست.

۱۳. فرض کنید

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

با استفاده از بحث موجود در تمرین ۱۲، تجزیه A به صورت $A = (P^{-1}L)U$ را بدست آورید.

۱۴. افزاز ماتریسها به دسته‌ای از زیرماتریسها، بوسیله رسم خطوط افقی و عمودی، اغلب مفید است. مثلاً، ماتریسهای

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 3 & -4 & -3 \\ 6 & 5 & 0 \end{bmatrix} \quad \text{و} \quad B = \begin{bmatrix} 2 & -1 & 7 & 0 \\ 3 & 0 & 4 & 5 \\ -2 & 1 & -3 & 1 \end{bmatrix}$$

را می‌توان به

$$\left[\begin{array}{cc|c} 1 & 2 & -1 \\ 3 & -4 & -3 \\ \hline 6 & 5 & 0 \end{array} \right] = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array} \right]$$

و

$$\left[\begin{array}{ccc|c} 2 & -1 & 7 & 0 \\ 3 & 0 & 4 & 5 \\ \hline -2 & 1 & -3 & 1 \end{array} \right] = \left[\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array} \right]$$

افراز کرد.

(T) نشان دهید در این حالت حاصل ضرب A در B برابر است با

$$AB = \left[\begin{array}{ccc|ccc} A_{11} & B_{11} + A_{12} & B_{21} & A_{11} & B_{12} + A_{12} & B_{22} \\ \hline A_{21} & B_{11} + A_{22} & B_{21} & A_{21} & B_{12} + A_{22} & B_{22} \end{array} \right]$$

(ب) اگر B به صورت

$$B = \left[\begin{array}{ccc|c} 2 & -1 & 7 & 0 \\ \hline 3 & 0 & 4 & 5 \\ -2 & 1 & -3 & 1 \end{array} \right] = \left[\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array} \right]$$

افراز می‌شد، نتیجه قسمت (T) برقرار می‌بود؟

(پ) برای برقراری قسمت (T) در حالت کلی، شرایط لازم را حدس بزنید.

۱۵. A یک ماتریس $n \times n$ است. یک زیرماتریس اصلی پیشرو A به صورت یک زیرماتریس

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix}$$

تعریف می‌شود، که در آن $1 \leq k \leq n$. نشان دهید حذف گاوسی بدون تعویض سطرها را می‌توان روی A انجام داد اگر و فقط اگر همه زیرماتریسهای اصلی پیشرو A نامنفرد باشند.

[راهنمایی. $1 \leq k \leq n$ را ثابت و از تمرین ۱۱ به یاد آورید که $A^{(k)} = M^{(k-1)} \cdots M^{(1)} A$]

هریک از این ماتریسها را به طور عمودی بین ستونهای k ام و $(k + 1)$ م و به طور افقی بین سطرهای k ام و $(k + 1)$ م افراز کنید، بدست می آید

$$\begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ A_{21}^{(k)} & A_{22}^{(k)} \end{bmatrix} = \begin{bmatrix} M_{11}^{(k-1)} & 0 \\ M_{21}^{(k-1)} & M_{22}^{(k-1)} \end{bmatrix} \begin{bmatrix} M_{11}^{(k-2)} & 0 \\ M_{21}^{(k-2)} & M_{22}^{(k-2)} \end{bmatrix} \cdots \begin{bmatrix} M_{11}^{(1)} & 0 \\ M_{21}^{(1)} & M_{22}^{(1)} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

نشان دهید که نامنفرد بودن $A_{11}^{(k)}$ نامنفردن $A_{11}^{(k)}$ را نتیجه می دهد. هرگاه حذف گاوسی را بتوان بدون تعویض سطرها انجام داد، آنگاه $A = LU$ با استفاده از این مطلب و ملاحظه حاصل ضرب افراز شده، نتیجه را بدست آورید.

۱۶. درباب زندگی، بحث زنجیره های غذایی یک موضوع مهم در تعیین گسترش و تجمع آلودگیهای محیطی است. فرض کنید یک زنجیره غذایی دارای سه حلقه باشد. اولین حلقه متشکل از انواع گیاهی v_1, v_2, \dots, v_n است، که همه غذای لازم برای گیاه خواران از نوع h_1, h_2, \dots, h_m ، موجود در حلقه دوم، را تهیه می کند. حلقه سوم متشکل از حیوانات گوشتخوار c_1, c_2, \dots, c_k است، که برای تهیه غذایشان کاملاً به گیاهخواران حلقه دوم وابسته اند. درایه a_{ij} از ماتریس

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

تعداد کل گیاهان نوع v_i خورده شده توسط گیاهخواران نوع h_j را نشان می دهد، درحالی که b_{ij} در

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mk} \end{bmatrix}$$

تعداد گیاهخواران نوع h_i خورده شده توسط حیوانات نوع c_j را توصیف می کند. (T) نشان دهید که تعداد گیاهان نوع v_i که سرانجام توسط حیوانات نوع c_j پایان می یابند با درایه محل سطر i م و ستون j م ماتریس AB ارائه می شود.

(ب) چه معنی فیزیکی به ماتریسهای $B^{-1}A^{-1} = (AB)^{-1}$ ، A^{-1} ، B^{-1} ، وابسته است؟
 ۱۷. در مقاله‌ای با عنوان "امواج جمعیت"، برنادلی^۱ [۱۰] (ر.ک. سیرل^۲ [۵۷]) نوعی سوسک ساده که دارای عمر طبیعی حدود سه سال است را در نظر گرفت. ماده^۳ این نوع در سال اول حیات دارای میزان بقای $\frac{1}{2}$ ، از سال دوم به سال سوم میزان بقای $\frac{1}{3}$ ، و قبل از مردن در انتهای سال سوم به طور متوسط 6 ماده^۴ جدید بدنیا می‌آورد. برای نشان دادن سهم یک سوسک ماده در ساخت جمعیت ماده^۴ نوع خود، به مفهوم احتمالی، می‌توان ماتریسی بکار برد، با قرار دادن a_{ij} در ماتریس $A = (a_{ij})$ ، که سهم یک سوسک ماده^۴ تک به سن j را در ساخت جمعیت ماده^۴ بعدی به سن i مشخص می‌کند؛ یعنی،

$$A = \begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix}$$

(ت) سهمی که یک سوسک ماده در ساخت جمعیت دوساله دارد باید از درایه‌های A^2 ، و لذا در جمعیت سه‌ساله از A^3 و غیره، مشخص شود. A^2 و A^3 را بسازید و سعی کنید برای هر عدد صحیح مثبت n ، عبارتی کلی درباره^۴ سهمی که یک سوسک ماده در ساخت جمعیت n ساله دارد بسازید.

(ب) از نتیجه^۴ حاصل در قسمت (ت) برای توصیف آنچه در سالهای آینده برای جمعیتی از این ماده‌ها، که به طور آغازی متشکل از 6000 سوسک ماده هریک در گروههای سه‌ساله، اتفاق می‌افتد استفاده کنید.

(پ) A^{-1} را بسازید و اهمیت آن را در مورد این نوع جمعیت توصیف کنید.

۱۸. حل دستگاه خطی

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1,$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2,$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3.$$

بوسیله^۴ قاعده^۴ گرامر^۳ به صورت

$$x_1 = \frac{1}{D} \det \begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix} \equiv \frac{D_1}{D}.$$

$$x_2 = \frac{1}{D} \det \begin{bmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{bmatrix} \equiv \frac{D_2}{D},$$

و

$$x_3 = \frac{1}{D} \det \begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{bmatrix} \equiv \frac{D_3}{D}.$$

است، که در آن

$$D = \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

(ت) جواب دستگاه خطی

$$2x_1 + 3x_2 - x_3 = 4,$$

$$x_1 - 2x_2 + x_3 = 6,$$

$$x_1 - 12x_2 + 5x_3 = 10,$$

را بوسیلهٔ قاعدهٔ کرامر بیابید.

(ب) نشان دهید که دستگاه خطی

$$2x_1 + 3x_2 - x_3 = 4,$$

$$x_1 - 2x_2 + x_3 = 6,$$

$$-x_1 - 12x_2 + 5x_3 = 9,$$

فاقد جواب است. D_1 ، D_2 ، و D_3 را محاسبه کنید.

(پ) نشان دهید که دستگاه خطی

$$2x_1 + 3x_2 - x_3 = 4,$$

$$x_1 - 2x_2 + x_3 = 6,$$

$$-x_1 - 12x_2 + 5x_3 = 10,$$

دارای بی‌نهایت جواب است. D_1 ، D_2 ، و D_3 را محاسبه کنید.

(ت) ثابت کنید هرگاه یک دستگاه خطی 3×3 با $D=0$ جوابهایی داشته باشد،

$$\text{آنگاه } D_1 = D_2 = D_3 = 0.$$

۴.۷ استراتژیهای محورگیری

دراثنای بدست آوردن الگوریتم ۱۰۷، دریافتیم که بدست آوردن یک صفر برای عنصر

محوری $a_{k,k}^{(k)}$ ایجاب می‌کند که یک تعویض سطر به صورت $(i_k) \leftrightarrow (i_n)$ انجام گیرد، که در آن $n+1 \leq k \leq n$ کوچکترین عدد صحیح با خاصیت $a_{p,k}^{(k)} \neq 0$ بود. در عمل اغلب انجام تعویض سطرها، حتی وقتی عناصر محوری صفر نباشند، مطلوب است. علت این امر این است که وقتی محاسبات با استفاده از حساب ارقام متناهی انجام می‌شود، مانند حالت ماشین حساب یا جوابهای تولید شده با کامپیوتر، یک عنصر محوری که در مقایسه با درایه‌های زیر آن در همان ستون کوچکتر است می‌تواند به خطای گرد کردن قابل توجهی منجر شود. یک بررسی از این اشکال در مثال زیر ارائه می‌شود.

مثال ۱. دستگاه خطی

$$E_1: .003000x_1 + 59.14x_2 = 59.17,$$

$$E_2: 5.291x_1 - 6.130x_2 = 46.78,$$

دارای جواب دقیق $x_1 = 10.00$ و $x_2 = 1.000$ است.

جهت بررسی اشکالات خطای گرد کردن، انجام حذف گاوسی روی این دستگاه، و استفاده از حساب چهاررقمی با گرد کردن بررسی می‌شوند. اولین عنصر محوری $a_{11}^{(1)} = .003000$ ، و مضرب وابسته به آن

$$m_{21} = \frac{5.291}{.003000} = 1763.6\bar{6},$$

است که به 1764 گرد می‌شود. با انجام عمل $(E_2 - m_{21}E_1) \rightarrow (E_2)$ و گرد کردن مناسب، دستگاه جدید

$$.003000x_1 + 59.14x_2 = 59.17$$

$$-104300x_2 = -104400.$$

حاصل می‌شود. از جایگذاری پسر نتیجه می‌شود

$$x_2 = 1.001$$

و

$$x_1 = \frac{59.17 - (59.14)(1.001)}{.003000}$$

$$= -10.00.$$

این خطای بزرگ در جواب عددی x_1 از خطای کوچک 001 حاصل از حل برای

x_2 نتیجه شده است. این خطا، به علت ترتیبی که محاسبات انجام شده‌اند، بوسیلهٔ عامل 20,000 در جواب x_1 بزرگ شده است.

این مثال مسلماً "اشکالاتی را که در بعضی حالات بهنگام کوچک بودن عنصر محوری $a_{k,k}^{(k)}$ نسبت به درایه‌های $a_{i,j}^{(k)}$ به‌ازای $k \leq i \leq n$ ، $k \leq j \leq n$ است روشن می‌کند. استراتژیهای محورگیری درحالت کلی با انتخاب عنصر جدید به‌عنوان محور $a_{p,q}^{(k)}$ و سپس تعویض سطرها k ام و q ام، و در صورت لزوم، تعویض ستونهای k ام و p ام، انجام می‌شوند. شاید ساده‌ترین استراتژی از این نوع، انتخاب عنصری در همان‌ستون و زیرقطر با بیشترین قدرمطلق باشد؛ یعنی p ، را چنان تعیین کنید که

$$|a_{p,k}^{(k)}| = \max_{k \leq i \leq n} |a_{i,k}^{(k)}|,$$

و $(E_k) \leftrightarrow (E_p)$ را انجام دهید. در این حالت هیچ تعویض ستونی در نظر گرفته نمی‌شود.

مثال ۲. مجدداً " دستگاه

$$E_1: .003000x_1 + 59.14x_2 = 59.17,$$

$$E_2: 5.291x_1 - 6.130x_2 = 46.78.$$

را در نظر می‌گیریم و با استفاده از روند محورگیری ستونی که در بالا گفته شد ابتدا

$$\begin{aligned} \max\{|a_{11}^{(1)}|, |a_{21}^{(1)}|\} &= \max\{.003000, |5.291|\} \\ &= |5.291| \\ &= |a_{21}^{(1)}|; \end{aligned}$$

را تعیین می‌کنیم. با انجام عمل $(E_1) \leftrightarrow (E_2)$ دستگاه

$$E_1: 5.291x_1 - 6.130x_2 = 46.78,$$

$$E_2: .003000x_1 + 59.14x_2 = 59.17$$

بدست می‌آید. ضرب لازم برای این دستگاه عبارت است از

$$m_{21} = \frac{a_{21}^{(1)}}{a_{11}^{(1)}} = .0005670,$$

و عمل $(E_2 - m_{21}E_1) \rightarrow (E_2)$ دستگاه را به

اگر $i = n$ ، به مرحله ۶ می‌رویم .

مرحله ۶ . اگر $a_{nn} = 0$ ، دستگاه دارای جواب منحصر بفرد نیست و روند متوقف می‌شود .
اگر $a_{nn} \neq 0$ ،

(T) قرار می‌دهیم $x_n = \frac{a_{n,n+1}}{a_{nn}}$

(ب) $x_i = \frac{[a_{i,n+1} - \sum_{j=i+1}^n a_{ij}x_j]}{a_{ii}}$ را به‌ازای هر $i = n-1, n-2, \dots, 1$ محاسبه می‌کنیم .

مرحله ۷ . روند کامل خواهد بود .

روندهای شرح داده شده در این الگوریتم جهت تضمین اینکه هر ضرب m_{ji} بیشتر از یک نباشد کافی بوده و استراتژی محورگیری ستونی ماکزیمال برای اغلب دستگاههای خطی کفایت می‌کند . لیکن ، مواردی هست که این استراتژی برای آنها کافی نیست .

مثال ۳ . دستگاه خطی

$$E_1: 30.00x_1 + 591.400x_2 = 591,700,$$

$$E_2: 5.291x_1 - 6.130x_2 = 46.78,$$

همان دستگاهی است که در مثال قبل ارائه شد جز آنکه تمام درایه‌های معادله اول آن در 10^4 ضرب شده است . روند گفته شده در الگوریتم ۲.۷ با حساب چهاررقمی همان نتایجی را که در مثال ۱ بدست آمد می‌دهد ، زیرا مقدار ماکزیم در اولین ستون برابر 30.00 است و ضریب

$$m_{21} = \frac{5.291}{30.00} = .1764$$

به دستگاه

$$30.00x_1 + 591,400x_2 = 591,700,$$

$$-104300x_2 = -104,400,$$

منجر می‌شود که دارای جواب $x_1 = -10.00$ و $x_2 = 1.001$ است .

یک روند محورگیری که از پس این وضعیت و اکثر دستگاههایی که بهنگام استفاده از

الگوریتم ۲.۷ با مسائلی روبرو می‌شوند برمی‌آید به محورگیری ستونی مقیاس شده مشهور است. روند محورگیری ستونی مقیاس شده، به سادگی بجای اینکه بزرگترین عنصر زیر قطر را برای استفاده محور انتخاب کند، ابتدا مقادیر را نسبت به بزرگترین درایه در سطر مربوطه مقیاس می‌کند، و متعاقباً به عنوان عنصر محوری یکی را با بزرگترین مقدار مقیاس شده انتخاب می‌کند. ملاحظه می‌شود که بکار بردن این روش در مثال ۳ نتایج زیر را با مقادیر مقیاس شده بدست می‌دهد:

$$\frac{5.291}{6.130} = .8631 \quad \text{و} \quad \frac{30.00}{591400} = .00005073$$

با بکار بردن استراتژی محورگیری، E_1 و E_2 تعویض خواهند شد و نتایج همانند مثال ۲ خواهند بود، یعنی مقادیر درست $x_1 = 10.00$ و $x_2 = 1.000$.

الگوریتم حذف گاوسی با محورگیری ستونی مقیاس شده ۳.۷

تنها مرحله‌هایی که در این الگوریتم با مراحل الگوریتم ۲.۷ متفاوتند عبارتند از:

مرحله ۱. به ازای $i = 1, 2, \dots, n$ تعریف می‌کنیم $s_i = \max_{1 \leq j \leq n} |a_{ij}|$. اگر به ازای i $s_i = 0$ ، دستگاه جواب منحصر بفرد ندارد و روند متوقف می‌شود. اگر به ازای هر $i = 1, 2, \dots$ ، $s_i \neq 0$ ، قرار می‌دهیم $i = 1$ و به مرحله ۲ می‌رویم.

مرحله ۲. فرض کنیم p اولین عدد صحیحی باشد که $i \leq p \leq n$ و

$$\frac{|a_{p,i}|}{s_p} = \max_{i \leq j \leq n} \frac{|a_{ji}|}{s_j}$$

عمل $(i; i) \leftrightarrow (i; p)$ را انجام می‌دهیم و به مرحله ۳ می‌رویم.

کاملترین استراتژی محورگیری موسوم به محورگیری ماکزیمال (یا کامل) است و مستلزم انتخاب عنصر محوری $(i; i)$ برابر $\max_{k \leq i, j \leq n} |a_{ij}^{(k)}|$ است، که بوسیله تعویض سطرها و ستونها انجام می‌شود. به علت تعداد مقایسه‌های لازم، این استراتژی فقط برای دستگاههای خیلی سخت، که در آنها زمان اجرای زیاد برای این روش قابل توجیه باشد، توصیه می‌شود. مسائل مستلزم محورگیری ماکزیمال را در تمرینات ۴ و ۵ ملاحظه کنید.

مجموعه تعریفات ۴.۷

۱. الگوریتمهای ۱.۷، ۲.۷ و ۳.۷ را روی دستگاههای خطی ذیل با حساب گرد کردن

تا چهار رقم اعمال کنید. کدام الگوریتم دقیقترین نتایج را می دهد؟

$$58.09x_1 + 1.003x_2 = 68.12, \quad (T)$$

$$321.8x_1 + 5.550x_2 = 377.3;$$

$$1.003x_1 + 58.09x_2 = 68.12, \quad (ب)$$

$$5.550x_1 + 321.8x_2 = 377.3;$$

$$1.003x_1 + 58.09x_2 = 68.12, \quad (پ)$$

$$321.8x_1 + 5.550x_2 = 377.3.$$

۲. الگوریتمهای ۱.۰۷، ۲.۰۷، و ۳.۰۷ را برای حل دستگاههای خطی زیر، با حساب گرد کردن تا سه رقم، بکار برید.

$$.03x_1 + 58.9x_2 = 59.2, \quad (T)$$

$$5.31x_1 - 6.10x_2 = 47.0;$$

$$58.9x_1 + .03x_2 = 59.2, \quad (ب)$$

$$-6.10x_1 + 5.31x_2 = 47.0;$$

$$x_1 + \frac{1}{2}x_2 + \frac{1}{3}x_3 = \frac{11}{6}, \quad (پ)$$

$$5x_1 + \frac{1}{3}x_2 + \frac{1}{2}x_3 = \frac{95}{6},$$

$$\frac{100}{3}x_1 + 25x_2 + 20x_3 = \frac{235}{3}.$$

۳. در الگوریتمهای ۲.۰۷ و ۳.۰۷ در واقع لازم نیست تعویض سطرها انجام شود اگر، در عوض، ترتیب محاسبات اصلاح شود. برای انجام این منظور، با بردار $XROW(i) = i$ ، به ازای هر $i = 1, 2, \dots, n$ ، شروع و هروقت مناسب بود، $XROW(p)$ و $XROW(i)$ را تعویض کنید. عمل لازم به ازای هر $j = i + 1, \dots, n$ به صورت

$$\left(E_{XROW(j)} - \frac{a_{XROW(j),i}}{a_{XROW(i),i}} E_{XROW(i)} \right) \rightarrow (E_{XROW(j)})$$

می باشد. الگوریتمهای ۲.۰۷ و ۳.۰۷ را با منظور کردن این تعویض مجدداً بنویسید.

۴. برای روند محورگیری ماکزیمال مورد بحث بعد از الگوریتم ۳.۰۷ الگوریتمی بسازید.

۵. با استفاده از الگوریتم محورگیری ماکزیمال ساخته شده در تمرین ۴، جوابهای تمرینات ۱ و ۲ را بدست آورده و جوابهای حاصل را با نتایج این دو تمرین مقایسه کنید.

۶. جهت برازش یک مجموعه از داده های $\{(x_i, y_i) | i = 1, 2, \dots, M\}$ با یک چندجمله ای کمترین مربعات گسسته P از درجه n مانند

$$P(x) = a_0 + a_1x + \dots + a_nx^n, \quad n < M - 1,$$

استفاده از معادلات نرمال، موجود در صفحه ۱۹۶، ایجاب می‌کند که یک دستگاه خطی

$$A \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = b,$$

را حل کنیم، که در آن

$$A_{ij} = \sum_{k=1}^M x_k^{i+j-2}, \quad i, j = 1, 2, \dots, n+1$$

و

$$b_i = \sum_{k=1}^M y_k x_k^{i-1}, \quad i = 1, 2, \dots, n+1$$

(T) با استفاده از الگوریتمهای ۱۰۷ و ۲۰۷، برای داده‌های زیر

i	1	2	3	4
x_i	0	1/3	2/3	1
y_i	1/9	25/27	41/27	19/9

چند جمله‌ای کمترین مربعات گسسته درجه سه بدست آورید.
(ب) برای هر چند جمله‌ای بدست آمده در قسمت (T) خطای

$$\left(\sum_{i=1}^4 |y_i - P(x_i)|^2 \right)^{1/2}$$

را مقایسه کنید.

(پ) داده‌ها از تابع $f(x) = x^3 - 2x^2 + 3x + 1/9$ گرفته شده‌اند. آیا چند جمله‌ایها مناسب هستند؟

۷. تعیین چند جمله‌ای کمترین مربعات پیوسته

$$P(x) = a_0 + a_1x + a_2x^2 + a_3x^3$$

از درجه سه که $f(x) = e^{-1x}$ را تقریب می‌کند، با استفاده از معادلات نرمال موجود در صفحه ۲۰۴، حل دستگاه خطی

$$\begin{bmatrix} \int_0^1 dx & \int_0^1 x dx & \int_0^1 x^2 dx & \int_0^1 x^3 dx \\ \int_0^1 x dx & \int_0^1 x^2 dx & \int_0^1 x^3 dx & \int_0^1 x^4 dx \\ \int_0^1 x^2 dx & \int_0^1 x^3 dx & \int_0^1 x^4 dx & \int_0^1 x^5 dx \\ \int_0^1 x^3 dx & \int_0^1 x^4 dx & \int_0^1 x^5 dx & \int_0^1 x^6 dx \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \int_0^1 e^{-1x} dx \\ \int_0^1 xe^{-1x} dx \\ \int_0^1 x^2 e^{-1x} dx \\ \int_0^1 x^3 e^{-1x} dx \end{bmatrix}$$

را ایجاب می‌کند. انتگرالها را محاسبه و دستگاه خطی را بوسیله الگوریتمهای ۱۰.۷ و ۲.۷ حل کنید. جواب خود را با جواب تمرین ۳ قسمت (پ) بخش ۴.۴ مقایسه نمایید.

۵.۷ انواع ویژه ماتریسها

در این بخش، مطالبی اضافی در مورد ماتریسها که در فصلهای بعد مفیدند ارائه می‌شود.

تعریف ۱۷.۷. **ترانسهاده** یک ماتریس $m \times n$ ، A که با A' نموده می‌شود، ماتریسی $n \times m$ است که درایه‌های آن عبارتند از $(A')_{ij} = (A)_{ji}$. ماتریسی که ترانسهاده آن خودش است **متقارن** نامیده می‌شود.

مثال ۱. ماتریسهای

$$A = \begin{bmatrix} 7 & 2 & 0 \\ 3 & 5 & -1 \\ 0 & 5 & -6 \end{bmatrix} \quad B = \begin{bmatrix} 2 & 4 & 7 \\ 3 & -5 & -1 \end{bmatrix} \quad C = \begin{bmatrix} 6 & 4 & 3 \\ 4 & -2 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

دارای ترانسهاده

$$A' = \begin{bmatrix} 7 & 3 & 0 \\ 2 & 5 & 5 \\ 0 & -1 & -6 \end{bmatrix} \quad B' = \begin{bmatrix} 2 & 3 \\ 4 & -5 \\ 7 & -1 \end{bmatrix} \quad C' = \begin{bmatrix} 6 & 4 & 3 \\ 4 & -2 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

هستند. چون C' و C دارای درایه‌های یکسانند، C متقارن است.

قضیه ۱۸.۷. احکام زیر، در صورت امکان، در مورد ترانسهاده ماتریسها برقرارند:

$$1. (A')' = A$$

$$2. (A + B)' = A' + B'$$

$$3. (AB)' = B'A'$$

$$4. \text{اگر } A^{-1} \text{ موجود باشد، } (A^{-1})' = (A')^{-1}$$

$$5. \det A' = \det A$$

اثبات این احکام ساده است و در تمرینات می‌آید.

تعریف ۱۹.۷. یک ماتریس $n \times n$ را ماتریس نواری گوئیم اگر اعداد صحیح p و q ، که $1 < p, q < n$ ، با خاصیت $a_{ij} = 0$ هرگاه $a_{ij} = 0$ هرگاه $i + p \leq j$ یا $j + q \leq i$ ، وجود داشته باشند. برای چنین ماتریسی، عرض نواری با $w = p + q - 1$ تعریف می‌شود.

ماتریس A تعریف شده در مثال ۱ یک ماتریس نواری با $p = q = 2$ و عرض نواری

۳ است.

تعریف ماتریس نواری سبب می‌شود که چنین ماتریسهایی همهء درایه‌های ناصفرشان را در اطراف قطر متمرکز کنند. دو حالت خاص از ماتریسهای نواری که اغلب در عمل ظاهر می‌شوند، دارای $p = q = 2$ و $p = q = 4$ هستند. با ماتریسی به عرض نواری ۳، که وقتی $p = q = 2$ رخ می‌دهد، قبلاً در رابطه با بحث تقریب با اسپلاین مکعبی در بخش ۶.۳ مواجه شده‌ایم. چنین ماتریسهایی اغلب سه قطری نامیده می‌شوند. زیرا به صورت

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & \ddots & \vdots \\ 0 & a_{32} & a_{33} & a_{34} & 0 \\ \vdots & \ddots & \ddots & \ddots & a_{n-1,n} \\ 0 & \cdots & 0 & a_{n,n-1} & a_{nn} \end{bmatrix}$$

هستند. چنین ماتریسهایی را در فصل ۱۰، در رابطه با بحث تقریب قطعه قطعه خطی به مسائل با مقدار مرزی دوباره مطرح خواهیم کرد. حالت $p = q = 4$ ، وقتی تابع تقریب ساز یک اسپلاین مکعبی فرض شود، برای حل مسائل با مقدار مرزی نیز بکار می‌رود. دستهء دیگری از ماتریسها که مفید می‌باشند ذیلاً تعریف می‌شوند.

روشهای مستقیم حل دستگاههای خطی ۴۷۱

تعریف ۲۰.۷. ماتریس $n \times n$ ، A را "اکیدا" قطر غالب گوئیم هرگاه به ازای هر $i = 1, 2, \dots, n$ نامساوی

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \quad (15.7)$$

برقرار باشد.

ملاحظه می شود که ماتریس A در مثال ۱ "اکیدا" قطر غالب است، زیرا $|7| > |2| + |0|$ ، $|5| > |3| + |-1|$ ، و $|6| > |0| + |5|$ ، اما جالب توجه است که A' "اکیدا" قطر غالب نیست.

قضیه ۲۱.۷. هرگاه A یک ماتریس $n \times n$ "اکیدا" قطر غالب باشد، آنگاه A نامفرد است. علاوه، حذف گاوسی را می توان روی هر دستگاه خطی به صورت $Ax = b$ ، جهت تعیین جوابهای منحصر بفرد آن بدون تعویض سطر و ستون، اعمال کرد و محاسبات نسبت به رشد خطاهای گرد کردن پایدارند.

اثبات اینکه حذف گاوسی را می توان بدون تعویض سطر و ستون انجام داد بستگی به تمرین ۱۵ بخش ۳.۷ دارد و (به عنوان تمرین ۷) به خواننده محول شده است. اثبات پایداری این روش را می توان در وندروف [۶۷] یافت.

برای نشان دادن اینکه A نامفرد است، دستگاه خطی توصیف شده با $Ax = 0$ را در نظر می گیریم و فرض می کنیم یک جواب ناصفر برای این دستگاه وجود داشته باشد. در این صورت، به ازای k ای، $|x_k| = \max_{1 \leq j \leq n} |x_j|$ ، و چون به ازای هر $i = 1, 2, \dots, n$ داریم

$$a_{kk}x_k = - \sum_{\substack{j=1 \\ j \neq k}}^n a_{kj}x_j.$$

این نتیجه می دهد که

$$|a_{kk}||x_k| \leq \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}||x_j|$$

یا

$$|a_{kk}| \leq \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}| \frac{|x_j|}{|x_k|}$$

$$\leq \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}|,$$

که متناقض با اکیدا " قطر غالب بودن A است. در نتیجه، تنها جواب $Ax = 0$ برابر $x = 0$ است، شرطی که در قضیه ۱۶.۷ جهت معادل بودن بانامفردی A مطرح شده است. توجه کنید که قضیه ۲۱.۷ را جهت ارائه اثبات مقدماتی قضیه ۹.۳ و قضیه ۱۰.۳ بخش ۶.۳، با اعمال روی ماتریسهای داده شده در این قضایا، می توان بکار برد. آخرین رده از ماتریسهای مورد بحث در این بخش معین مثبت نامیده می شود.

تعریف ۲۲.۷. یک ماتریس $n \times n$ متقارن A را معین مثبت گوئیم اگر به ازای هر بردار ستونی n بعدی $x \neq 0$ ، $x'Ax > 0$.

جهت دقیق بودن، تعریف باید مشخص کند که تنها درایه ماتریس 1×1 تولید شده با عمل $x'Ax$ دارای مقداری مثبت است، زیرا این عمل در واقع به صورت زیر انجام می شود:

$$x'Ax = [x_1, x_2, \dots, x_n] \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= [x_1, x_2, \dots, x_n] \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j \end{bmatrix}$$

$$= [\sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j].$$

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

یک ماتریس معین مثبت است، چرا که اگر x یک بردار ستونی سه بعدی دلخواه باشد،

$$\begin{aligned} x'Ax &= [x_1, x_2, x_3] \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= [x_1, x_2, x_3] \begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 - x_3 \\ -x_2 + 2x_3 \end{bmatrix} \\ &= [2x_1^2 - 2x_1x_2 + 2x_2^2 - 2x_2x_3 + 2x_3^2] \\ &= [x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2] \end{aligned}$$

و

$$x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2 > 0,$$

مگر اینکه $x_1 = x_2 = x_3 = 0$.

از این مثال باید آشکار شود که بکار بردن تعریف در تعیین معین مثبت بودن یک ماتریس فوق العاده کسالت‌آور است. خوشبختانه، محک ساده‌تری برای تعیین عناصر این دسته مهم از ماتریسها وجود دارد، که در فصل بعد ارائه می‌شود. شرایطی که می‌توان جهت حذف ماتریسهای مشخصی از این جرگه بکار برد در تمرین ۵ مطرح می‌شوند. قضیه بعد در راستای قضیه ۲۱.۷ در مورد ماتریسهای اکیدا "قطر غالب" است.

قضیه ۲۳.۷. هرگاه A یک ماتریس $n \times n$ معین مثبت باشد، آنگاه A نامنفرد است. بعلاوه، حذف گاوسی را می‌توان روی هر دستگاه خطی به صورت $Ax = b$ جهت تعیین جوابهای منحصر بفرد آن بدون تعویض سطر و ستون، اعمال کرد و محاسبات نسبت به رشد خطای گرد کردن پایدارند.

برهان. هرگاه $x \neq 0$ برداری باشد که در $Ax = 0$ صدق کند، آنگاه $x'Ax = 0$ ، که با

معین مثبت بودن A در تناقض است. در نتیجه، $Ax = 0$ فقط دارای جواب صفر است، و از قضیه ۱۶.۷ نتیجه می‌شود که A نامنفرد است. برای اثبات برقراری حکم اول در مورد حذف گاوسی، فرض کنیم A ، به‌ازای k ای دلخواه که $k = 1, 2, \dots, n$ ، یک زیر ماتریس اصلی پیشرواز A باشد، که تعریف زیرماتریس اصلی پیشرو در تمرین ۱۵ بخش ۳.۷ ارائه شده است. هرگاه A معین مثبت نباشد، آنگاه یک بردار ستونی k بعدی مانند $x \neq 0$ که $x'Ax \leq 0$ وجود دارد. از x یک بردار ستونی n بعدی $x \neq 0$ با گذاشتن صفر در $(n - k)$ مولفه آخر می‌سازیم. چون

$$x'Ax = \hat{x}'A\hat{x} \leq 0$$

A معین مثبت نیست، و این تناقض است. در نتیجه، هر زیرماتریس اصلی A معین مثبت و از آنجا نامنفرد است، و نتیجه از مسئله ۱۵ بخش ۷.۳ حاصل می‌شود. اثبات پایداری روند حذف گاوسی را می‌توان در وندروف [۶۷]، در آغاز صفحه ۱۲۰، یافت.

مجموعه تمرینات ۵.۷

۱. احکام زیر را ثابت کنید یا برای نشان دادن نادرستی آنها مثال نقض بیاورید.
 - (ت) حاصل ضرب دو ماتریس متقارن متقارن است.
 - (ب) معکوس یک ماتریس متقارن نامنفرد ماتریس متقارن نامنفرد است.
 - (پ) هرگاه A و B ماتریس‌هایی $n \times n$ باشند، آنگاه $(AB)' = A'B'$.
۲. احکام ارائه شده در قضیه ۱۸.۷ را ثابت کنید.
۳. ماتریس نامتقارن A را طوری بسازید که به‌ازای هر $x \neq 0$ ، $x'Ax > 0$.
۴. ماتریس $n \times n$ ، A را قطر غالب گوئیم اگر

$$|a_{ii}| \geq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, 2, \dots, n$$

- (ت) یک ماتریس قطر غالب منفرد بدون درایه صفر پیدا کنید.
- (ب) آیا یک ماتریس قطر غالب منفرد بدون درایه صفر وجود دارد که برای آن نامساوی اکید در معادله (۱۵.۷) برای تمام سطرها جز یکی برقرار باشد؟
۵. فرض کنید A یک ماتریس $n \times n$ معین مثبت باشد. نشان دهید که
 - (ت) به‌ازای هر $i = 1, 2, \dots, n$ ، $a_{ii} > 0$ ؛

$$: \max_{1 \leq i \leq n} |a_{ii}| \geq \max_{1 \leq i, j \leq n} |a_{ij}| \quad (\beta)$$

$$\cdot (a_{ij})^2 < a_{ii}a_{jj}, \quad i \neq j \quad \text{با } i, j = 1, 2, \dots, n \quad (\rho)$$

۶. نشان دهید که ماتریس

$$A = \begin{bmatrix} 4 & -1 & 1 \\ -1 & 4.25 & 2.75 \\ 1 & 2.75 & 3.5 \end{bmatrix}$$

معین مثبت است.

۷. با استفاده از تمرین ۵ بخش ۳۰۷ و استقرای ریاضی، ثابت کنید روند حذف گاوسی

را برای تعیین جواب منحصر بفرد $Ax = b$ ، که در آن A یک ماتریس اکیدا "قطر

غالب است، بدون تعویض سطر یا ستون می‌توان اعمال کرد.

۸. نشان دهید که معادلات نرمال مطرح شده در صفحه ۱۹۶، حاصل از تقریب کمترین

مربعات گسسته، یک ماتریس متقارن و نامنفرد را بدست می‌دهد؛ و در نتیجه، دارای

جواب منحصر بفرد است.

[راهنمایی. فرض کنید $A = (a_{ij})$ ، که در آن

$$a_{ij} = \sum_{k=1}^M x_k^{i+j-2},$$

و x_1, x_2, \dots, x_n نقاط متمایز با $n < M$ است. همچنین، A منفرد و $c \neq 0$ چنان

باشد که $c^T A c = 0$. نشان دهید که چندجمله‌ای درجه n که ضرایب آن مولفه‌های

هستند دارای بیش از n صفر است، و با استفاده از این مطلب یک تناقض بدست

آورید.]

۹. در مقاله‌ای توسط دورن^۱ و بورردیک^۲ [۲۱]، گزارش شده است که متوسط طول

بال بدست آمده از آمیزش سه نوع متفاوت از حشرات میوه‌جات را می‌توان با ماتریس

متقارن

$$A = \begin{bmatrix} 1.59 & 1.69 & 2.13 \\ 1.69 & 1.31 & 1.72 \\ 2.13 & 1.72 & 1.85 \end{bmatrix}$$

بیان کرد، که در آن a_{ij} متوسط طول بال یک مولود حاصل از آمیزش یک مذکر نوع i با یک مؤنث نوع j می باشد.

(آ) چه مفهوم فیزیکی وابسته به این ماتریس متقارن است؟

(ب) آیا این ماتریس معین مثبت است؟ اگر پاسخ مثبت است، آن را اثبات کنید؛

در غیر این صورت، بردار ناصفر x را طوری بیابید که به ازای آن $x^T A x \leq 0$.

۶.۱ تجزیه مستقیم ماتریسها

حت متمرکز حول قضیه ۱۱.۷ صفحه ۴۴۰ درباره تجزیه ماتریس A بر حسب یک اتریس پایین مثلثی L و یک ماتریس بالا مثلثی U بود. وجود این تجزیه، اگر دستگاه خطی $Ax = b$ با استفاده از حذف گاوسی بدون تعویض سطر یا ستون دارای جواب منحصر فرد باشد، نشان داده شد. دستگاه $LUX = Ax = b$ را می توان به دستگاه $UX = L^{-1}b$ تبدیل کرد، چون U بالا مثلثی است، جایگذاری پسر را بکار برد. گرچه مورویژه L و U را می توان از روند حذف گاوسی بدست آورد، ولی حصول روش مستقیم تری برای تعیین آنها مورد نظر است بطوری که، اگر دستگاههای زیادی با استفاده از A باید حل شوند، فقط یک جایگذاری پسر یا پیشرو مورد نیاز باشد (ر.ک. مراحل ۱۵ - ۱۲ لگوریتم ۵.۷). برای بررسی یک روند جهت محاسبه درایه های این ماتریسها، مثالی در نظر می گیریم.

مثال ۱. ماتریس 4×4 اکیدا " قطر غالب زیر را در نظر می گیریم:

$$A = \begin{bmatrix} 6 & 2 & 1 & -1 \\ 2 & 4 & 1 & 0 \\ 1 & 1 & 4 & -1 \\ -1 & 0 & -1 & 3 \end{bmatrix}$$

قضایای ۲۱.۷ (صفحه ۴۷۱) و ۱۱.۷ (صفحه ۴۴۰) تضمین می کنند که A را به

صورت $A = LU$ می توان تجزیه کرد، که در آن

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{bmatrix} \quad \text{و} \quad L = \begin{bmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{bmatrix}$$

16 درایه معلوم A را می توان برای تعیین 10 درایه نامعلوم L و همین تعداد درایه U بکار برد. اگر روندی برای تعیین جوابی منحصر بفرد مورد نظر باشد، چهار شرط اضافی روی درایه های L و U لازم است. روش بکار رفته در این مثال به طور دلخواهی طلب می کند که $l_{11} = l_{22} = l_{33} = l_{44} = 1$ این روش به روش دولیتل¹ موسوم است. بعداً² در این بخش روشهایی، روش کروت³ که ایجاب می کند همه عناصر قطری U برابر یک باشد، و روش چولسکی³ که ایجاب می کند به ازای هر یک از مقادیر i ، $l_{ii} = u_{ii}$ ، ملاحظه خواهند شد. بخشی از ضرب L در U ، یعنی

$$L \cdot U = \begin{bmatrix} 1 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 \\ l_{41} & l_{42} & l_{43} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{bmatrix}$$

که سطر اول A را مشخص می کند چهار معادله زیر را بدست می دهد:

$$u_{11} = 6, \quad u_{12} = 2, \quad u_{13} = 1, \quad u_{14} = -1$$

بخشی از این ضرب که بقیه دریه های ستون اول A را مشخص می کند.

معادلات

$$\begin{aligned} l_{21} &= \frac{1}{3}, & l_{21}u_{11} &= 2, \\ l_{31} &= \frac{1}{6}, & l_{31}u_{11} &= 1, \\ l_{41} &= -\frac{1}{6}, & l_{41}u_{11} &= -1, \end{aligned}$$

را فراهم می آورد.

در این مرحله L و U به صورت زیر می باشند:

$$U = \begin{bmatrix} 6 & 2 & 1 & -1 \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{bmatrix} \quad \text{و} \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & 1 & 0 & 0 \\ \frac{1}{6} & l_{32} & 1 & 0 \\ -\frac{1}{6} & l_{42} & l_{43} & 1 \end{bmatrix}$$

بخشی از این ضرب که بقیه درایه های سطر دوم A را مشخص می کند منجر به معادلات

ذیل می شود:

$$\begin{aligned} u_{22} &= \frac{10}{3}, & \frac{2}{3} + u_{22} &= 4, \\ u_{23} &= \frac{2}{3}, & \text{در نتیجه} \quad \frac{1}{3} + u_{23} &= 1, \\ u_{24} &= \frac{1}{3}; & -\frac{1}{3} + u_{24} &= 0, \end{aligned}$$

و بخشی که بقیه درایه‌های ستون A را مشخص می‌کند عبارتند از

$$\begin{aligned} l_{32} &= \frac{1}{3}, & \frac{2}{6} + \frac{10}{3}l_{32} &= 1, \\ l_{42} &= \frac{1}{10}, & -\frac{2}{6} + \frac{10}{3}l_{42} &= 0, \end{aligned}$$

این روند ادامه می‌یابد، با تغییر تناوبی ستونها و سطرها به‌طور قطری، تا بالاخره

نتیجه می‌شود

$$U = \begin{bmatrix} 6 & 2 & 1 & -1 \\ 0 & \frac{10}{3} & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & \frac{37}{10} & -\frac{9}{10} \\ 0 & 0 & 0 & \frac{191}{74} \end{bmatrix} \quad \text{و} \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & 1 & 0 & 0 \\ \frac{1}{6} & \frac{1}{5} & 1 & 0 \\ -\frac{1}{6} & \frac{1}{10} & -\frac{9}{37} & 1 \end{bmatrix}$$

الگوریتم زیر شامل روند کلی تجزیه ماتریسها به حاصل ضرب ماتریسهای مثلثی است. گرچه ماتریسهای جدید L و U ساخته می‌شوند، مقادیر واقعی تولید شده ممکن است جایگزین درایه‌های نظیر A ، که دیگر مورد نیاز نیستند، شوند. بنابراین، نتیجه عبارت می‌شود از:

بمازای هر $i = 2, 3, \dots, n$ و $j = 1, 2, \dots, i - 1$ ؛ $a_{ij} = l_{ij}$ و بمازای هر $i = 1, 2, \dots, n$ ؛ $l_{ii} = 1$ و بمازای هر $i = 1, 2, \dots, n$ و $j = i, i + 1, \dots, n$ ؛ $a_{ij} = u_{ij}$.

۴.۷ الگوریتم تجزیه مستقیم

برای تجزیه ماتریس

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

به ماتریسهای مثلثی

$$L = \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix}$$

و

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{nn} \end{bmatrix}$$

- مرحله ۰۱. l_{11} و u_{11} را چنان انتخاب می‌کنیم که در $l_{11}u_{11} = a_{11}$ صدق کنند.
 مرحله ۰۲. درایه‌های اولین ستون L را بوسیله شرایط

$$l_{j1} = \frac{a_{j1}}{u_{11}}$$

بمازای هر $j = 2, 3, \dots, n$ تولید می‌کنیم.

- مرحله ۰۳. درایه‌های اولین سطر U را با شرایط

$$u_{1j} = \frac{a_{1j}}{l_{11}}$$

بمازای هر $j = 2, 3, \dots, n$ تولید می‌کنیم.

- مرحله ۰۴. قرار می‌دهیم $i = 2$.

- مرحله ۰۵. l_{ii} و u_{ii} ی صادق در

$$l_{ii}u_{ii} = a_{ii} - \sum_{k=1}^{i-1} l_{ik}u_{ki}$$

انتخاب می‌کنیم.

- مرحله ۰۶. اگر $i < n$ ، به مرحله ۰۷ می‌رویم.

- اگر $i = n$ ، به مرحله ۰۱۰ می‌رویم.

- مرحله ۰۷. درایه‌های ستون i م L را با شرایط

$$l_{ji} = \frac{1}{u_{ii}} \left[a_{ji} - \sum_{k=1}^{i-1} l_{jk}u_{ki} \right],$$

بمازای هر $j = i + 1, i + 2, \dots, n$ تولید می‌کنیم.

مرحله ۸. درایه‌های سطر i م U را با شرایط

$$u_{ij} = \frac{1}{l_{ii}} \left[a_{ij} - \sum_{k=1}^{i-1} l_{ik} u_{kj} \right]$$

به‌ازای هر $n, i+1, i+2, \dots, n$ z تولید می‌کنیم .

مرحله ۹. ۱ را به i می‌افزاییم و به مرحله می‌رویم .

مرحله ۱۰. روند تکمیل است؛ همه درایه‌های L و U مشخص شده‌اند .

اشکال ناشی از بکار بردن این الگوریتم جهت تعیین تجزیه ماتریس ضرایب یک دستگاه خطی از معادلات به‌این علت است که از هیچ محورگیری برای تقلیل اثر خطای گردکردن استفاده نمی‌شود. در محاسبات قبلی ملاحظه شد که اگر حساب ارقام متناهی بکار رود، خطای گرد کردن کاملاً قابل توجه است و در هر الگوریتم با ارزشی این امر باید در نظر گرفته شود.

گرچه وارد کردن تعویض ستون در الگوریتم تجزیه مشکل است، این الگوریتم را می‌توان به‌سادگی تغییر داد تا شامل یک روش تعویض سطر، که با روند محورگیری ستونی ماکزیمال ذکر شده در الگوریتم ۲.۷ معادل و در اغلب حالات پسنده است، باشد. الگوریتم زیر روند الگوریتم تجزیه ۴.۷ و محورگیری ماکزیمال و جایگذاری پسرو و پیشرو را، برای حل یک دستگاه خطی از معادلات، بهم می‌پیوندد. روند واقعی مستلزم نوشتن دستگاه خطی $Ax = b$ به صورت $LUx = b$ است. جایگذاری پیشرو دستگاه $Lz = b$ و جایگذاری پسرو دستگاه $Ux = z = L^{-1}b$ را حل می‌کند. باید توجه داشت که درایه‌های ناصفر L و U را می‌توان در محل درایه‌های متناظر A ، جز برای قطر L که همگی می‌توانند یک باشند، ذخیره کرد.

الگوریتم تجزیه مستقیم با محورگیری ستونی ماکزیمال ۵.۷

برای حل دستگاه خطی

$$E_1: a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \equiv a_{1,n+1},$$

$$E_2: a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \equiv a_{2,n+1},$$

$$\vdots$$

$$E_n: a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n \equiv a_{n,n+1},$$

ماتریس افزوده زیر را تشکیل می‌دهیم :

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & a_{1,n+1} \\ a_{21} & a_{22} & \cdots & a_{2n} & a_{2,n+1} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & a_{n,n+1} \end{array} \right]$$

مرحله ۱. کوچکترین عدد صحیح p را چنان اختیار می‌کنیم که

$$|a_{p1}| = \max_{1 \leq j \leq n} |a_{j1}|$$

و سطرهای p و ۱ را با هم عوض می‌کنیم.

مرحله ۲. l_{11} و $u_{11} = a_{11}$ صادق در $l_{11}u_{11} = a_{11}$ را اختیار می‌کنیم.

مرحله ۳. درایه‌های ستون اول L را با شرایط

$$l_{j1} = \frac{a_{j1}}{u_{11}}$$

به‌ازای هر $j = 2, 3, \dots, n$ تولید می‌کنیم.

مرحله ۴. درایه‌های سطر اول U را با شرایط

$$u_{1j} = \frac{a_{1j}}{l_{11}}$$

به‌ازای هر $j = 2, 3, \dots, n$ تولید می‌کنیم.

مرحله ۵. قرار می‌دهیم $i = 2$.

مرحله ۶. کوچکترین عدد صحیح $p \geq i$ را چنان اختیار می‌کنیم که

$$\left| a_{pi} - \sum_{k=1}^{i-1} l_{pk}u_{ki} \right| = \max_{i \leq j \leq n} \left| a_{ji} - \sum_{k=1}^{i-1} l_{jk}u_{ki} \right|$$

و سطرهای p و i را، در هر دو ماتریس A و L ، با هم عوض می‌کنیم.

مرحله ۷. l_{ii} و u_{ii} ی صادق در

$$l_{ii}u_{ii} = a_{ii} - \sum_{k=1}^{i-1} l_{ik}u_{ki}$$

را اختیار می‌کنیم.

مرحله ۸. اگر $i < n$ ، به مرحله ۹ می‌رویم.

اگر $i = n$ ، به مرحله ۱۲ می‌رویم.

مرحله ۹. درایه‌های ستون i م L را با شرایط

$$l_{ji} = \frac{1}{u_{ii}} \left[a_{ji} - \sum_{k=1}^{i-1} l_{jk} u_{ki} \right]$$

بازای هر $j = i + 1, i + 2, \dots, n$ تولید می‌کنیم.

مرحله ۱۰. درایه‌های سطر i م U را با شرایط

$$u_{ij} = \frac{1}{l_{ii}} \left[a_{ij} - \sum_{k=1}^{i-1} l_{ik} u_{kj} \right]$$

بازای هر $j = i + 1, i + 2, \dots, n$ تولید می‌کنیم.

مرحله ۱۱. ۱ را به i افزوده و به مرحله ۹ می‌رویم.

مرحله ۱۲. قرار می‌دهیم $z_1 = \frac{a_{1,n+1}}{l_{11}}$.

مرحله ۱۳. قرار می‌دهیم $z_i = \frac{1}{l_{ii}} \left[a_{i,n+1} - \sum_{j=1}^{i-1} l_{ij} z_j \right]$ ، بازای هر $i = 2, 3, \dots, n$.

مرحله ۱۴. قرار می‌دهیم $x_n = \frac{z_n}{u_{nn}}$.

مرحله ۱۵. قرار می‌دهیم $x_i = \frac{1}{u_{ii}} \left[z_i - \sum_{j=i+1}^n u_{ij} x_j \right]$ ، بازای هر

$i = n - 1, n - 2, \dots, 1$.

مرحله ۱۶. روند کامل خواهد بود.

اگر ماتریس معین مثبت باشد، می‌توان پیشرفت قابل توجهی در رابطه با تعداد ا

اعمال حسابی مورد نیاز در روند تجزیه ماتریس ایجاد کرد.

قضیه ۲۴.۷. هرگاه A یک ماتریس $n \times n$ معین مثبت باشد، آنگاه A دارای تجزیه‌ای

به صورت $A = LL'$ است، که در آن L یک ماتریس پایین مثلثی است. الگوریتم ۴.۷ با

$l_{ii} = u_{ii}$ ، بازای هر $i = 1, 2, \dots, n$ ، را برای حصول این تجزیه می‌توان بکار برد.

برهان. چون A معین مثبت است، از قضایای ۲۳.۷ و ۱۱.۷ نتیجه می‌شود که می‌توان

$A = LU$ را به صورت $A = LU$ تجزیه کرد، که در آن L یک ماتریس پایین مثلثی و U یک ماتریس

بالامثلثی است. همچنین، از معین مثبت بودن A نتیجه می‌شود (ر.ک. تمرین ۵ بخش ۵.۷) که به‌ازای هر $i = 1, 2, \dots, n$ ، $u_{ii} > 0$ ، و با نمادگذاری الگوریتم ۴.۷، این روند با انتخاب $l_{11} = u_{11} = \sqrt{a_{11}}$ آغاز می‌شود.

چون A متقارن است، به‌ازای هر $j = 2, 3, \dots, n$ ،

$$l_{j1} = \frac{a_{j1}}{u_{11}} = \frac{a_{1j}}{l_{11}} = u_{1j}$$

بطوری که درایه‌های سطر اول U با درایه‌های متناظر ستون اول L مطابقت دارند. اثبات به‌استقرای ریاضی پیش می‌رود. فرض کنیم $k < n$ و درایه‌های k سطر اول U با درایه‌های نظیر آنها در k ستون اول مطابقت داشته‌باشد. اثبات قضیه با برقراری تطابق درایه‌های سطر $(k+1)$ U با درایه‌های ستون $(k+1)$ L کامل خواهد بود.

چون l پایین مثلثی و U بالا مثلثی است، واضح است که اگر $j = 1, 2, \dots, k$ ،

$$l_{j,k+1} = 0 = u_{k+1,j}.$$

انتخاب می‌کنیم $l_{k+1,k+1} = u_{k+1,k+1} = (a_{k+1,k+1} - \sum_{j=1}^k l_{k+1,j}^2)^{1/2}$ که می‌توان نشان داد عددی حقیقی است (ر.ک. تمرین ۱۵).

چون A متقارن است و به‌ازای هر $i = 1, 2, \dots, k$ و $j = 1, 2, \dots, n$ ، $l_{ji} = u_{ij}$ ، پس به‌ازای هر $j = k+2, k+3, \dots, n$ ،

$$\begin{aligned} l_{j,k+1} &= \frac{1}{u_{k+1,k+1}} \left[a_{j,k+1} - \sum_{i=1}^k l_{ji} u_{i,k+1} \right] \\ &= \frac{1}{l_{k+1,k+1}} \left[a_{k+1,j} - \sum_{i=1}^k u_{ij} l_{k+1,i} \right] \\ &= u_{k+1,j}. \end{aligned}$$

این قضیه برای ساده‌کردن الگوریتم تجزیه ۴.۷، در مورد یک ماتریس معین مثبت بکار می‌رود. اگر حل یک دستگاه خطی شامل یک ماتریس معین مثبت مورد نظر باشد، مراحل ۹ - الگوریتم ذیل را می‌توان، برای بهره‌گیری از ساده‌سازی حاصل شده، جایگزین مراحل ۱۱ - الگوریتم ۵.۷ نمود، و درایه‌های تولید شده L را در درایه‌های متناظر A ذخیره کرد.

الگوریتم چولسکی ۶.۷

برای تجزیه ماتریس معین مثبت

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

به ماتریسهای مثلثی

$$L = \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix}$$

و

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{nn} \end{bmatrix}$$

که $U = U'$

مرحله ۱. قرار می‌دهیم $l_{11} = \sqrt{a_{11}}$

مرحله ۲. ستون اول L را با شرایط

$$l_{j1} = \frac{a_{j1}}{l_{11}}$$

بمازای هر $j = 2, 3, \dots, n$ تولید می‌کنیم.

مرحله ۳. قرار می‌دهیم $i = 2$

مرحله ۴. قرار می‌دهیم $l_{ii} = \left[a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2 \right]^{1/2}$

مرحله ۵. اگر $i < n$ ، به مرحله ۶ می‌رویم.

اگر $i = n$ ، به مرحله ۸ می‌رویم.

مرحله ۶. ستون i م L را با شرایط

$$l_{ji} = \frac{1}{l_{ii}} \left[a_{ji} - \sum_{k=1}^{i-1} l_{jk} l_{ik} \right]$$

بمازای هر $j = i + 1, i + 2, \dots, n$ تولید می‌کنیم.

مرحله ۷. ۱ را به i افزوده به مرحله ۴ می‌رویم.

مرحله ۰۸. به ازای $i = 1, 2, \dots, n-1$ یا $i = 1, 2, \dots, n$ ، $j = i + 1, i + 2, \dots, n$ ، قرار می‌دهیم
 $l_{ij} = 0$.

مرحله ۰۹. به ازای هر $i = 1, 2, \dots, n$ و $j = 1, 2, \dots, n$ ، قرار می‌دهیم $u_{ij} = l_{ji}$.

مرحله ۰۱۰. روند کامل خواهد بود.

تحقیق اینکه حل یک نوع دستگاه خطی شامل ماتریس معین مثبت با استفاده از الگوریتم چولسکی نیاز به n جذر و $\frac{n^3 + 9n^2 + 2n}{6}$ ضرب و تقسیم و $\frac{n^3 + 6n^2 - 7n}{6}$ جمع و تفریق دارد سخت نیست (ر.ک. تمرین ۱۳). این مقادیر تقریباً " نصف اعمال حسابی لازم در الگوریتم حذف گاوسی ۱۰۷، روشی که در رابطه با تعداد اعمال حسابی مورد نیاز الگو است، می‌باشند. مزیت محاسباتی روش چولسکی بستگی به تعداد اعمال مورد نیاز برای تعیین مقادیر n جذر دارد، که چون عاملی خطی از n است، با افزایش n به طور قابل ملاحظه‌ای کاهش می‌یابد.

مثال ۰۲. برای بررسی مراحل موجود در روش بیکار رفته چولسکی، ماتریس

$$A = \begin{bmatrix} 4 & -1 & 1 \\ -1 & 4.25 & 2.75 \\ 1 & 2.75 & 3.5 \end{bmatrix}$$

را در نظر می‌گیریم، که می‌توان نشان داد معین مثبت است (ر.ک. تمرین ۶ بخش ۵۰۷).
 با دنبال کردن مراحل الگوریتم ۰۶۰۷،

مرحله ۰۱. $l_{11} = \sqrt{a_{11}} = \sqrt{4} = 2$

مرحله ۰۲. $l_{21} = \frac{a_{21}}{l_{11}} = \frac{-1}{2} = -.5$

$l_{31} = \frac{a_{31}}{l_{11}} = \frac{1}{2} = .5$.

مرحله ۰۳. $i = 2$

مرحله ۰۴. $l_{22} = (a_{22} - l_{21}^2)^{1/2}$
 $= (4.25 - (-.5)^2)^{1/2}$
 $= 2$.

مرحله ۵. $i = 2$ ؛ به مرحله ۶ می‌رویم.

$$\begin{aligned} l_{32} &= \frac{1}{l_{22}} (a_{32} - l_{31}l_{21}) \quad \text{مرحله ۶} \\ &= \frac{1}{2} (2.75 - .5(-.5)) \\ &= 1.5. \end{aligned}$$

مرحله ۷. $i = 3$ ؛ به مرحله ۴ می‌رویم.

$$\begin{aligned} l_{33} &= (a_{33} - l_{31}^2 - l_{32}^2)^{1/2} \quad \text{مرحله ۴} \\ &= (3.5 - (.5)^2 - (1.5)^2)^{1/2} \\ &= 1. \end{aligned}$$

مرحله ۵. $i = 3$ ؛ به مرحله ۸ می‌رویم.

$$l_{12} = l_{13} = l_{23} = 0 \quad \text{مرحله ۸}$$

$$u_{11} = l_{11} = 2, \quad u_{12} = l_{21} = -.5, \quad u_{13} = l_{31} = .5, \quad \text{مرحله ۹}$$

$$u_{21} = l_{12} = 0, \quad u_{22} = l_{22} = 2, \quad u_{23} = l_{32} = 1.5,$$

$$u_{31} = l_{13} = 0, \quad u_{32} = l_{23} = 0, \quad u_{33} = l_{33} = 1.$$

$$\text{مرحله ۱۰.} \quad L = \begin{bmatrix} 2 & 0 & 0 \\ -.5 & 2 & 0 \\ .5 & 1.5 & 1 \end{bmatrix}, \quad U = L' = \begin{bmatrix} 2 & -.5 & .5 \\ 0 & 2 & 1.5 \\ 0 & 0 & 1 \end{bmatrix}.$$

الگوریتمهای تجزیه را در حالت ماتریسهای نواری، به علت تعداد زیاد صفرهای موجود در آنها در الگوهای منظم، می‌توان به طور قابل ملاحظه‌ای ساده کرد. بخصوص، بررسی صورتی را که در این مورد روش کروت یا دولیتل در نظر می‌گیرند جالب توجه است. برای توضیح در این زمینه، ماتریس سه قطری

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & \cdots & 0 \\ & a_{22} & a_{23} & \cdots & 0 \\ 0 & \cdots & a_{32} & \cdots & a_{n-1,n} \\ \vdots & & & \ddots & \\ 0 & \cdots & 0 & a_{n,n-1} & a_{nn} \end{bmatrix}.$$

که به ماتریسهای زیر تجزیه می شود، را در نظر می گیریم :

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{nn} \end{bmatrix} \quad \text{و} \quad L = \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix}$$

چون A فقط $(3n - 2)$ درایه ناصفر دارد، فقط $(3n - 2)$ شرط جهت تعیین درایه های L و U ، البته بشرطی که درایه های صفر A نیز بدست آیند، اعمال می شود. فرض کنیم این ماتریسها را بتوان عملاً" به شکل زیر بدست آورد:

$$U = \begin{bmatrix} 1 & u_{12} & 0 & \cdots & 0 \\ 0 & 1 & & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ & & & & u_{n-1,n} \\ 0 & \cdots & 0 & & 1 \end{bmatrix} \quad \text{و} \quad L = \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \ddots & \vdots \\ 0 & & \ddots & 0 \\ \vdots & & & \ddots \\ 0 & \cdots & 0 & l_{n,n-1} & l_{nn} \end{bmatrix}$$

در این صورت، $(2n - 1)$ درایه از L و $(n - 1)$ درایه از U نامعلومند، که جمعا" برابر تعداد شرایط ذکر شده در بالا هستند، و درایه های صفر A خود بخود بدست آمده اند. ضرب $A = LU$ ، علاوه بر درایه های صفر، معادلات زیر را نتیجه می دهد:

$$(16.7) \quad a_{11} = l_{11};$$

$$(17.7) \quad \text{بازای هر } i = 2, 3, \dots, n \quad a_{i,i-1} = l_{i,i-1};$$

$$(18.7) \quad \text{بازای هر } i = 2, 3, \dots, n \quad a_{ii} = l_{i,i-1}u_{i-1,i} + l_{ii};$$

و

$$(19.7) \quad \text{بازای هر } i = 1, 2, \dots, n - 1 \quad a_{i,i+1} = l_{ij}u_{i,i+1};$$

یک جواب این دستگاه معادلات را می توان ابتدا با استفاده از معادله (17.7)، تعیین همه جملات ناصفر غیر بر قطر L ، و سپس با بکار بردن معادلات (19.7) و (18.7)، متناوبا"، برای تعیین سایر درایه های U و L ، که در درایه های متناظر A ذخیره می شوند، بدست آورد.

الگوریتم کاملی برای حل یک دستگاه $n \times n$ از معادلات خطی که ماتریس ضرایب آن سه قطری است ذیلا" ارائه می شود.

الگوریتم تحویل کروت برای دستگاههای خطی سه‌قطری ۷.۷
برای حل دستگاه معادلات خطی

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \equiv a_{1,n+1}, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \equiv a_{2,n+1}, \\ a_{32}x_2 + a_{33}x_3 + a_{34}x_4 &= b_3 \equiv a_{3,n+1}, \\ &\vdots \\ a_{n,n-1}x_{n-1} + a_{nn}x_n &= b_n \equiv a_{n,n+1}. \end{aligned}$$

ماتریس افزوده، زیر را تشکیل می‌دهیم:

$$\left[\begin{array}{cccccc|c} a_{11} & a_{12} & & 0 & \cdots & 0 & a_{1,n+1} \\ & & & & & \vdots & \\ a_{21} & a_{22} & a_{23} & & & 0 & a_{2,n+1} \\ & & & & & \vdots & \\ 0 & a_{32} & \cdots & & & a_{n-1,n} & \\ \vdots & \ddots & \ddots & & & & \\ 0 & \cdots & 0 & a_{n,n-1} & & a_{nn} & a_{n,n+1} \end{array} \right]$$

مرحله ۱. قرار می‌دهیم $l_{11} = a_{11}$

$$u_{12} = \frac{a_{12}}{l_{11}}.$$

مرحله ۲. قرار می‌دهیم $i = 2$.

مرحله ۳. قرار می‌دهیم $l_{i,i-1} = a_{i,i-1}$.

$$l_{ii} = a_{ii} - l_{i,i-1}u_{i-1,i}.$$

$$u_{i,i+1} = \frac{a_{i,i+1}}{l_{ii}}.$$

مرحله ۴. i را به ۱ می‌افزاییم.

مرحله ۵. اگر $i < n$ ، به مرحله ۳ می‌رویم.

اگر $i = n$ ، به مرحله ۶ می‌رویم.

مرحله ۶. قرار می‌دهیم $l_{n,n-1} = a_{n,n-1}$.

$$l_{nn} = a_{nn} - l_{n,n-1}u_{n-1,n}.$$

مرحله ۷. قرار می‌دهیم $z_1 = \frac{a_{1,n+1}}{l_{11}}$

بهازای هر $i = 2, 3, \dots, n$ ، $z_i = \frac{1}{l_{ii}} [a_{i,n+1} - l_{i,i-1}z_{i-1}]$.

مرحله ۰۸. قرار می‌دهیم $x_n = z_n$ ،

$$x_i = z_i - u_{i,i+1}x_{i+1} \quad , \quad i = n-1, n-2, \dots, 1$$

مرحله ۰۹. روند کامل خواهد بود .

این الگوریتم فقط به $(5n - 4)$ ضرب و تقسیم و $(3n - 3)$ جمع و تفریق نیاز دارد؛ در نتیجه، بخصوص برای مقادیر بزرگ n ، دارای مزیت فوق‌العاده محاسباتی نسبت به روشهایی که سه قطری بودن ماتریس را در نظر نمی‌گیرند ، دارد .

مثال ۰۳. برای بررسی روند موجود در الگوریتم ۷.۷ ، دستگاه معادلات سه‌قطری

$$\begin{aligned} 2x_1 - x_2 &= 1, \\ -x_1 + 2x_2 - x_3 &= 0, \\ -x_2 + 2x_3 - x_4 &= 0, \\ -x_3 + 2x_4 &= 1, \end{aligned}$$

را در نظر می‌گیریم ، که ماتریس افزوده آن عبارت است از

$$\left[\begin{array}{cccc|c} 2 & -1 & 0 & 0 & 1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & 1 \end{array} \right]$$

با دنبال کردن مراحل الگوریتم ۷.۷ ،

مرحله ۰۱. $l_{11} = 2, u_{11} = 1$

$$u_{12} = \frac{a_{12}}{l_{11}} = -\frac{1}{2}$$

مرحله ۰۲. $i = 2$

مرحله ۰۳. $l_{21} = a_{21} = -1,$

$$l_{22} = a_{22} - l_{21}u_{12} = 2 - (-1)\left(-\frac{1}{2}\right) = \frac{3}{2}$$

$$u_{23} = \frac{a_{23}}{l_{22}} = -\frac{1}{\left(\frac{3}{2}\right)} = -\frac{2}{3}$$

مرحله ۰۴. $i = 3$

مرحله ۰۵. $i = 3$ ؛ به مرحله ۳ می‌رویم .

مرحله ۳. $l_{32} = a_{32} = -1,$

$l_{33} = a_{33} - l_{32}u_{23} = 2 - (-1)(-\frac{2}{3}) = \frac{4}{3},$

$u_{34} = \frac{a_{34}}{l_{33}} = -\frac{3}{4}$

مرحله ۴. $i = 4$

مرحله ۵. $i = 4$ ؛ به مرحله ۶ می‌رویم.

مرحله ۶. $l_{43} = a_{43} = -1,$

$l_{44} = a_{44} - l_{43}u_{34} = 2 - (-1)(-\frac{3}{4}) = \frac{5}{4}.$

مرحله ۷. $z_1 = \frac{a_{15}}{l_{11}} = \frac{1}{2}.$

$z_2 = \frac{1}{l_{22}} [a_{25} - l_{21}z_1] = \frac{0 - (-1)(\frac{1}{2})}{(\frac{3}{2})} = \frac{1}{3}.$

$z_3 = \frac{1}{l_{33}} [a_{35} - l_{32}z_2] = \frac{0 - (-1)(\frac{1}{3})}{(\frac{4}{3})} = \frac{1}{4}.$

$z_4 = \frac{1}{l_{44}} [a_{45} - l_{43}z_3] = \frac{1 - (-1)(\frac{1}{4})}{(\frac{5}{4})} = 1.$

مرحله ۸. $x_4 = 1,$

$x_3 = z_3 - u_{34}x_4 = \frac{1}{4} - (-\frac{3}{4}) = 1,$

$x_2 = z_2 - u_{23}x_3 = \frac{1}{3} - (-\frac{2}{3}) = 1,$

$x_1 = z_1 - u_{12}x_2 = \frac{1}{2} - (-\frac{1}{2}) = 1.$

این الگوریتم ماتریس مورد نظر را به صورت

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & \frac{3}{2} & 0 & 0 \\ 0 & -1 & \frac{4}{3} & 0 \\ 0 & 0 & -1 & \frac{5}{4} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{2}{3} & 0 \\ 0 & 0 & 1 & -\frac{3}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

تجزیه می‌کند، و جواب درست $x_1 = x_2 = x_3 = x_4 = 1$ را ارائه می‌دهد.

الگوریتم ۷.۷ را وقتی به‌ازای هر $l_{ii} \neq 0, i = 1, 2, \dots, n$ می‌توان بکار برد. در شرطی که هر یک از آنها درستی این مطلب را تضمین می‌کنند عبارتند از اینکه، ماتریس

ضرایب دستگاه معین مثبت یا اکیدا "قطر غالب" باشد. یک شرط اضافی که اطمینان می‌دهد این الگوریتم را می‌توان اعمال کرد در قضیه بعد، که اثبات آن در تمرین ۸ مطرح می‌شود، داده شده است.

قضیه ۲۵.۷. هرگاه $|a_{11}| > |a_{12}|$ و به ازای هر $i = 2, 3, \dots, n-1$ ، $|a_{ii}| \geq |a_{i,i-1}| + |a_{i,i+1}|$ و $|a_{nn}| > |a_{n,n-1}|$ ، نگاه A نامنفرد است و مقادیر l_{ii} مذکور در الگوریتم ۷.۷ به ازای هر $i = 1, 2, \dots, n$ ناصفرند.

مجموعه تمرینات ۶.۷

۱. دستگاه خطی

$$\begin{aligned} 1.012x_1 - 2.132x_2 + 3.104x_3 &= 1.984, \\ -2.132x_1 + 4.096x_2 - 7.013x_3 &= -5.049, \\ 3.104x_1 - 7.013x_2 + .014x_3 &= -3.895. \end{aligned}$$

را با استفاده از الگوریتم ۵.۷، با $l_{ii} = 1$ به ازای هر $i = 1, 2, 3$ ، حل کنید.

۲. نشان دهید که ماتریس زیر معین مثبت است، و آن را با استفاده از روش چولسکی به صورت $A = L \cdot L'$ تجزیه کنید:

$$A = \begin{bmatrix} 6 & 2 & 1 & -1 \\ 2 & 4 & 1 & 0 \\ 1 & 1 & 4 & -1 \\ -1 & 0 & -1 & 3 \end{bmatrix}$$

۳. ماتریسهای سه‌قطری را، برای تاکید در اینکه لازم نیست این ماتریس تماما "درنظر گرفته شود، معمولاً" با نماد

$$A = \begin{bmatrix} a_1 & c_1 & 0 & \dots & 0 \\ b_2 & a_2 & c_2 & \dots & \vdots \\ 0 & b_3 & \dots & \dots & 0 \\ \vdots & \dots & \dots & \dots & c_{n-1} \\ 0 & \dots & 0 & b_n & a_n \end{bmatrix}$$

نشان می‌دهند. الگوریتم ۷.۷ را، با استفاده از این نمادگذاری و تغییر نماد l_{ii} و r_{ii} به روشی مشابه، مجدداً بنویسید.

۴. الگوریتمی برای تجزیه ماتریسهای سه‌قطری، مستقیماً از الگوریتم ۴.۷ با استفاده

از $u_{ii} = 1$ بهازای هر $i = 1, 2, \dots, n$ نتیجه بگیرید، و این روند را با الگوریتم ۷.۷ مقایسه کنید.

۵. الگوریتمی برای تجزیه ماتریسهای سه قطری با استفاده از الگوریتم ۴.۷ با $l_{ii} = 1$ بهازای هر $i = 1, 2, \dots, n$ نتیجه بگیرید.

۶. فرض کنید A ماتریس سه قطری داده شده با $a_{ii} = 2, a_{i,i+1} = a_{i,i-1} = -1$ بهازای هر $i = 2, \dots, 9$ و $a_{11} = a_{10,10} = 2, a_{12} = a_{10,9} = -1$ باشد. همچنین، برداری ستونی b با بعد ده داده شده با $b_1 = b_{10} = 1$ و $b_i = 0$ بهازای هر $i = 2, 3, \dots, 9$ باشد. با استفاده از الگوریتم ۷.۷، $Ax = b$ را حل کنید.

۷. تمرین ۶ را با استفاده از روشی مبتنی بر الگوریتم حاصل از تمرین ۵ تکرار کنید.

۸. قضیه ۲۵.۷ را ثابت کنید. [راهنمایی]. نشان دهید که بهازای هر $i = 1, 2, \dots, n-1$ ، $|u_{i,i+1}| < 1$ و بهازای هر $i = 1, 2, \dots, n$ ، $|l_{ii}| > 0$ ، و نتیجه بگیرید که $\det A = \det L \cdot \det U \neq 0$.

یک ماتریس سه قطری بلوکی (با افراز شده) به شکل

$$A = \begin{bmatrix} A_1 & C_1 & 0 & \cdots & 0 \\ B_2 & A_2 & C_2 & \ddots & \vdots \\ 0 & B_3 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & C_{N-1} \\ 0 & \cdots & 0 & B_N & A_N \end{bmatrix}$$

است، که در آن، بهازای گردآیمای از اعداد صحیح مثبت n_1, n_2, \dots, n_N ، هر A_i ماتریسی $n_i \times n_i$ و هر B_i ماتریسی $n_i \times n_{i-1}$ و هر C_i ماتریسی $n_i \times n_{i+1}$ است. (T) A را به LU تجزیه کنید، که در آن

$$L = \begin{bmatrix} L_1 & 0 & \cdots & 0 \\ B_2 & L_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & B_N & L_N \end{bmatrix}, \quad U = \begin{bmatrix} I_1 & \Gamma_1 & 0 & \cdots & 0 \\ 0 & I_2 & \Gamma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \Gamma_{N-1} \\ 0 & \cdots & 0 & 0 & I_N \end{bmatrix}$$

و هر L_i ماتریسی $n_i \times n_i$ ، هر Γ_i ماتریسی $n_i \times n_{i+1}$ است، و I_i ماتریس همانی $n_i \times n_i$ را مشخص می‌کند.

(ب) یک الگوریتم ماتریس سه قطری بلوکی، مشابه الگوریتم ۷.۷، برای حل دستگاه

خطی $Ax = y$ ، که در آن y نظیر A افزاز شده است ، ارائه دهید .

- ۱۰ . ماتریس سه قطری بلوکی تعریف شده در تمرین ۹ را در نظر بگیرید . نشان دهید هرگاه زیرماتریسهای اصلی پیشرو A به شکل

$$A^{(k)} = \begin{bmatrix} A_1 & C_1 & 0 & \cdots & 0 \\ B_2 & A_2 & C_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & C_{n-1} \\ 0 & \cdots & 0 & B_n & A_n \end{bmatrix}$$

بهزای هر $k = 1, 2, \dots, n$ ، همگی نامنفرد باشند ، آنگاه ماتریسهای L_1, L_2, \dots, L_n نامنفردند و الگوریتم حاصل از تمرین ۹ (ب) را می توان انجام داد .

- ۱۱ . فرض کنید A ماتریس سه قطری بلوکی تعریف شده با

$$A = \begin{bmatrix} A_1 & C_1 & 0 \\ B_2 & A_2 & C_2 \\ 0 & B_3 & A_3 \end{bmatrix}$$

باشد ، که در آن بهزای هر $i = 1, 2, 3$ ،

$$A_i = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{bmatrix}$$

و ، بهزای هر $i = 1, 2$ ،

$$B_{i+1} = C_i = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

با در نظر گرفتن بردار ستونی نه بعدی b ، با $b_1 = b_0 = 2$ و $b_i = 1$ بهزای هر $i = 2, 3, \dots, 8$ ، و با استفاده از الگوریتم حاصل از تمرین ۹ ، دستگاه $Ax = b$ را حل کنید .

- ۱۲ . نشان دهید که ضرب و تقسیم و جمع و تفریقهای لازم در الگوریتم ۵.۷ به همان تعداد موجود در حذف گاوسی است .

۱۳ . عمل شمارش را برای یک دستگاه خطی $n \times n$ ، با استفاده از الگوریتم ۶.۷ ، اجرا کنید .

۱۴. عمل شمارش را برای یک دستگاه خطی $n \times n$ ، با استفاده از الگوریتم ۷.۷، اجرا کنید.

۱۵. نشان دهید که الگوریتم ۵.۷ را با $l_{ii} > 0$ ، $u_{ii} > 0$ ، به ازای هر $i = 1, 2, \dots, n$ ، برای هر ماتریس معین مثبت A می توان اجرا کرد.

روشهای تکراری در جبر ماتریسی ۸

در آغاز فصل ۷ یک مسئله مدار الکتریکی شامل هفت مقاومت و یک ولتاژ موثر را بررسی کردیم که منجر به یک دستگاه خطی 5×5 شد که جواب آن پتانسیل را در هر اتصال مدار بدست می‌داد. در مداری شامل مقاومت‌های بسیار بیشتر وضعیت مشابه خواهد بود اما، البته، دستگاه خطی مربوطه مفصلتر است. معه‌ذا، حتی در مدارهای بزرگ، انتظار می‌رود که هر اتصال تعداد نسبتاً "کوچکی انشعاب داشته باشد. اگر ماتریس مربوط به دستگاه خطی 5×5 بحث شده در فصل ۷ را در نظر بگیریم،

$$\begin{matrix} & v_b & v_c & v_d & v_e & v_f \\ \begin{matrix} v_b \\ v_c \\ v_d \\ v_e \\ v_f \end{matrix} & \begin{bmatrix} -31 & 10 & 0 & 0 & 6 \\ 2 & -8 & 3 & 3 & 0 \\ 0 & 1 & -3 & 2 & 0 \\ 0 & 2 & 4 & -9 & 1 \\ 12 & 0 & 15 & 0 & -47 \end{bmatrix} \end{matrix}$$

ملاحظه می‌کنیم که در ماتریس، اگر دو اتصال مستقیماً "به هم مربوط نباشند، یک صفر ظاهر می‌شود. انتظار می‌رود که ماتریسهای متناظر با مدارهای الکتریکی بزرگ درصد بزرگی از مولفه‌های صفر داشته باشند. ماتریسهایی از این نوع را تنک نامند، و روشهای مناسب برای حل دستگاههای خطی شامل ماتریسهای تنک اغلب متضمن تکنیکهای تکراری بجای مستقیم هستند.

روشهای ارائه شده در فصل ۷ تکنیکهای مستقیم را برای حل یک دستگاه $n \times n$ از معادلات خطی به شکل $Ax = b$ مورد استفاده قرار دادند. در این فصل چند روش که می‌توانند برای حل یک دستگاه از این نوع به‌طور تکراری بکار روند ارائه خواهد شد.

۱۰.۸ نرمهای بردارها و ماتریسها

پیش از بررسی روشهای تکراری برای حل دستگاههای خطی، لازم است روشی برای اندازه‌گیری کمی فاصله بین بردارهای موجود در R^n ، یعنی مجموعه تمام بردارهای ستونی با مولفه‌های حقیقی، ارائه شود تا اینکه معین کنیم آیا یک دنباله از بردارها که از بکار بردن یک تکنیک تکراری نتیجه شده‌اند به‌جوابی از دستگاه همگرا هستند. فی‌الواقع، این اندازه زمانی که جواب با روشهای مستقیم ارائه شده در فصل قبل بدست آمده باشد نیز مورد نیاز است، زیرا این روشها محتاج انجام تعداد زیادی عملیات حسابی است، و استفاده از حساب ارقام منتهای فقط منجر به جوابی تقریبی از جواب واقعی دستگاه خواهد شد. برای تعریف یک فاصله در R^n ، ایده نرم یک بردار را بکار خواهیم برد که مشابه نرم یک تابع پیوسته است که در بخش ۱۰.۴ ارائه شد.

تعریف ۱۰.۸. یک نرم برداری روی R^n ، یعنی مجموعه تمام بردارهای n بعدی با مولفه‌های حقیقی، تابعی است مانند $\| \cdot \|$ از R^n بتوی R که در خواص زیر صدق می‌کند:

(A) به‌ازای هر $x \in R^n$ ، $\|x\| \geq 0$ ؛

(B) $\|x\| = 0$ اگر و فقط اگر $x = (0, 0, \dots, 0)$ ؛

- (پ) بهازای هر $\alpha \in R$ و $\mathbf{x} \in R^n$ ، $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$ ،
 (ت) بهازای هر $\mathbf{x}, \mathbf{y} \in R^n$ ، $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

برای مقاصد ما فقط دو نرم مشخص روی R^n احتیاج است، گرچه نرم سومی روی R^n در تمرین ۲ ارائه شده است .

چون بردارهای موجود در R^n بردارهای ستونی هستند، اگر یک بردار برحسب مولفه‌هایش نمایش داده شده باشد، مناسب است که از نماد ترانهاده، ارائه شده دربخش ۵.۷، استفاده کنیم. برای مثال، بردار

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

عموماً $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ نوشته خواهد شد.

تعریف ۲.۰۸. نرمهای l_2 و l_∞ برای بردار $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ به صورت زیر تعریف می‌شوند:

$$(۱.۰۸) \quad \|\mathbf{x}\|_2 = \left\{ \sum_{i=1}^n x_i^2 \right\}^{1/2}$$

$$(۲.۰۸) \quad \|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

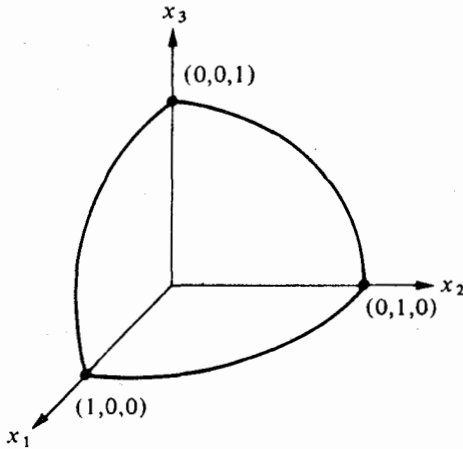
نرم l_2 اغلب نرم اقلیدسی بردار \mathbf{x} نامیده می‌شود، زیرا مفهوم عادی فاصله تا مبدا را درحالتی که \mathbf{x} در R^1 ، R^2 ، یا R^3 است نمایش می‌دهد.

مثال ۰.۱. بردار $\mathbf{x} = (-1, 1, -2)'$ در R^3 دارای نرمهای زیر است:

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

$$\|\mathbf{x}\|_\infty = \max\{|-1|, |1|, |-2|\} = 2.$$

تمام بردارها در R^3 که نرم l_2 کمتر از ۱ دارند داخل یک کره به مرکز $(0, 0, 0)'$ و شعاع ۱ قرار می‌گیرند (ر.ک. شکل ۱.۰۸)، در حالی که بردارهای با نرم l_∞



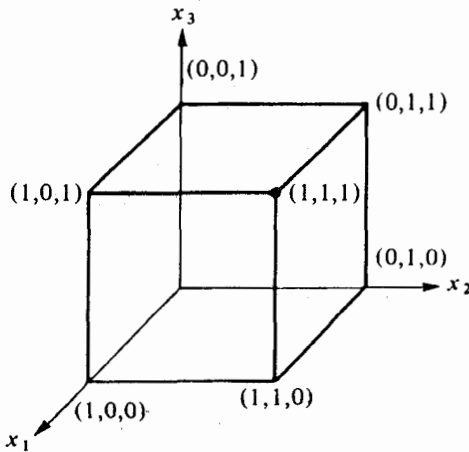
شکل ۱۰۸ بردارهای با نرم l_2 کمتر از یک و واقع در یک‌هشتم اول داخل این شکل هستند.

کمتر از ۱ داخل مکعبی که رئوس آن نقاط

$$(1, 1, 1), (1, 1, -1), (1, -1, -1), (1, -1, 1), (-1, 1, 1),$$

$$(-1, 1, -1), (-1, -1, -1), (-1, -1, 1).$$

هستند قرار دارند (ر.ک. شکل ۲۰۸).



شکل ۲۰۸ بردارهای با نرم l_∞ کمتر از یک و واقع در یک‌هشتم اول داخل این شکل

هستند.

توجیه اینکه مفهوم تعریف شده با $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$ یک نرم بر R^n است در قضیه زیر ارائه شده است.

قضیه ۳.۸ . بهازای هر $x, y \in R^n$ و $\alpha \in R$ ،

$$: \|x\|_\infty \geq 0 \quad (T)$$

$$: \|x\|_\infty = 0 \text{ اگر و فقط اگر } x = 0 \quad (پ)$$

$$: \|\alpha x\|_\infty = |\alpha| \|x\|_\infty \quad (پ)$$

$$\cdot \|x + y\|_\infty \leq \|x\|_\infty + \|y\|_\infty \quad (ت)$$

در نتیجه ، $\|\cdot\|_\infty$ یک نرم بر R^n است.

برهان . خواص فوق به سادگی از احکام مشابه در مورد تابع قدر مطلق نتیجه می شوند.

به عنوان مثال ، حکم (ت) را بررسی می کنیم . هرگاه $x = (x_1, x_2, \dots, x_n)$ و $y = (y_1, y_2, \dots, y_n)$ آنگاه

$$\begin{aligned} \|x + y\|_\infty &= \max_{1 \leq i \leq n} |x_i + y_i| \\ &\leq \max_{1 \leq i \leq n} \{|x_i| + |y_i|\} \\ &\leq \max_{1 \leq i \leq n} |x_i| + \max_{1 \leq i \leq n} |y_i| = \|x\|_\infty + \|y\|_\infty. \end{aligned}$$

خواص دیگر به روشی مشابه نشان داده می شوند .

نشان دادیم ، اینکه

$$\|x\|_2 = \left\{ \sum_{i=1}^n x_i^2 \right\}^{1/2}$$

یک نرم بر R^n را مشخص می کند مشکلتر است . مشکل در اثبات اینکه بهازای هر $x, y \in R^n$

$$\|x + y\|_2 \leq \|x\|_2 + \|y\|_2$$

ظاهر می شود .

قضیه ۴.۸ . بهازای هر $x, y \in R^n$ و $\alpha \in R$ ،

$$: \|x\|_2 \geq 0 \quad (T)$$

$$: \|x\|_2 = 0 \text{ اگر و فقط اگر } x = 0 \quad (پ)$$

$$: \|\alpha \mathbf{x}\|_2 = |\alpha| \|\mathbf{x}\|_2 \quad (\text{پ})$$

$$\cdot \|\mathbf{x} + \mathbf{y}\|_2 \leq \|\mathbf{x}\|_2 + \|\mathbf{y}\|_2 \quad (\text{ت})$$

برهان. بهازای هر $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ در R^n ،

$$\|\mathbf{x}\|_2 = \left\{ \sum_{i=1}^n x_i^2 \right\}^{1/2} \geq 0$$

و $\|\mathbf{x}\|_2 = 0$ اگر و فقط اگر بهازای هر $i = 1, 2, \dots, n$ ، $x_i^2 = 0$ ، بنابراین $\|\mathbf{x}\|_2 = 0$ اگر و فقط اگر $\mathbf{x} = 0$ ، و خواص (پ) و (ت) ثابت شده‌اند.

برای نشان دادن اینکه (پ) برقرار است، فقط توجه کنید که

$$\begin{aligned} \|\alpha \mathbf{x}\|_2 &= \left\{ \sum_{i=1}^n (\alpha x_i)^2 \right\}^{1/2} = \left\{ \sum_{i=1}^n \alpha^2 x_i^2 \right\}^{1/2} \\ &= (\alpha^2)^{1/2} \left\{ \sum_{i=1}^n x_i^2 \right\}^{1/2} = |\alpha| \|\mathbf{x}\|_2. \end{aligned}$$

برای اینکه خاصیت (ت) را نشان دهیم، ابتدا لازم است لم زیر را ثابت کنیم.

لم ۵.۸ (نامساوی کشی - بونیاکوفسکی - شوارتز برای مجموعها).

بهازای هر $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ و $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ در R^n ،

$$(3.8) \quad \sum_{i=1}^n |x_i y_i| \leq \left\{ \sum_{i=1}^n x_i^2 \right\}^{1/2} \left\{ \sum_{i=1}^n y_i^2 \right\}^{1/2}$$

برهان لم ۵.۸. اگر $\mathbf{y} = 0$ یا $\mathbf{x} = 0$ نتیجه فوری است، زیرا هر دو طرف معادله (۳.۸) صفر می‌شوند.

فرض کنیم $\mathbf{y} \neq 0$ و $\mathbf{x} \neq 0$. بهازای هر $\lambda \in R$ ،

$$0 \leq \|\mathbf{x} - \lambda \mathbf{y}\|_2^2 = \sum_{i=1}^n (x_i - \lambda y_i)^2 = \sum_{i=1}^n x_i^2 - 2\lambda \sum_{i=1}^n x_i y_i + \lambda^2 \sum_{i=1}^n y_i^2.$$

بنابراین، $2\lambda \sum_{i=1}^n x_i y_i \leq \sum_{i=1}^n x_i^2 + \lambda^2 \sum_{i=1}^n y_i^2 = \|\mathbf{x}\|_2^2 + \lambda^2 \|\mathbf{y}\|_2^2$ ، چون قضیه ۴.۸ نتیجه

می‌دهد که $\|\mathbf{x}\|_2 > 0$ و $\|\mathbf{y}\|_2 > 0$ ، با انتخاب $\lambda = \|\mathbf{x}\|_2 / \|\mathbf{y}\|_2$ داریم

$$\left(2 \frac{\|x\|_2}{\|y\|_2}\right) \left(\sum_{i=1}^n x_i y_i\right) \leq \|x\|_2^2 + \frac{\|x\|_2^2}{\|y\|_2^2} \|y\|_2^2 = 2\|x\|_2^2$$

یا

$$2 \sum_{i=1}^n x_i y_i \leq 2 \frac{\|y\|_2}{\|x\|_2} \|x\|_2^2 = 2\|x\|_2 \|y\|_2$$

بنابراین،

$$\sum_{i=1}^n x_i y_i \leq \|x\|_2 \|y\|_2$$

با جایگزین کردن x_i بوسیله $-x_i$ ، هر وقت $x_i y_i < 0$ ، و نامیدن آن بردار با \bar{x} ،

$$\sum_{i=1}^n |x_i y_i| \leq \|\bar{x}\|_2 \|y\|_2 = \|x\|_2 \|y\|_2$$

بطوری که $\sum_{i=1}^n |x_i y_i| \leq \|x\|_2 \|y\|_2$

برهان قضیه ۴.۸ (ادامه). به ازای هر $x, y \in \mathbb{R}^n$

$$\|x + y\|_2^2 = \sum_{i=1}^n (x_i + y_i)^2$$

$$= \sum_{i=1}^n x_i^2 + 2 \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2$$

$$\leq \|x\|_2^2 + 2\|x\|_2 \|y\|_2 + \|y\|_2^2 = (\|x\|_2 + \|y\|_2)^2 \quad (۵.۸)$$

بنابراین،

$$\|x + y\|_2 \leq \|x\|_2 + \|y\|_2$$

چون نرم یک بردار اندازه‌های برای "فاصله" بین آن بردار و مبدأ می‌دهد،

"فاصله" بین دو بردار را می‌توان به صورت نرم تفاضل بردارها تعریف کرد.

تعریف ۶.۸. هرگاه $x = (x_1, x_2, \dots, x_n)'$ و $y = (y_1, y_2, \dots, y_n)'$ بردارهایی در \mathbb{R}^n

باشند، فاصله‌های l_2 و l_∞ بین x و y چنین تعریف می‌شوند:

$$(۴.۸) \quad \|x - y\|_2 = \left\{ \sum_{i=1}^n |x_i - y_i|^2 \right\}^{1/2}$$

$$(۵.۸) \quad \|x - y\|_\infty = \max_{1 \leq i \leq n} |x_i - y_i|$$

و

مثال ۲. دستگاه خطی

$$3.3330x_1 + 15920x_2 - 10.333x_3 = 15913,$$

$$2.2220x_1 + 16.710x_2 + 9.6120x_3 = 28.544,$$

$$1.5611x_1 + 5.1791x_2 + 1.6852x_3 = 8.4254,$$

دارای جواب $y = (1.0000, 1.0000, 1.0000)$ است. اگر حذف گاوسی با حساب پنج رقمی انجام شود، با استفاده از محورگیری ماکزیم ستونی (الگوریتم ۲.۷ صفحه ۴۶۴)، جواب حاصل عبارت است از

$$\bar{x} = (\bar{x}_1, \bar{x}_2, \bar{x}_3)' = (1.2001, .99991, .92538)'$$

اندازه‌گیریهای $x - \bar{x}$ چنین داده می‌شود:

$$\begin{aligned} \|x - \bar{x}\|_{\infty} &= \max\{|1.0000 - 1.2001|, |1.0000 - .99991|, |1.0000 - .92538|\} \\ &= \max\{.2001, .00009, .07462\} \\ &= .2001, \end{aligned}$$

و

$$\begin{aligned} \|x - \bar{x}\|_2 &= (|1.0000 - 1.2001|^2 + |1.0000 - .99991|^2 + |1.0000 - .92538|^2)^{1/2} \\ &= ((.2001)^2 + (.00009)^2 + (.07462)^2)^{1/2} \\ &= .21356. \end{aligned}$$

با وجود اینکه مولفه‌های \bar{x}_2 و \bar{x}_3 تقریبهای نسبتاً خوبی به x_2 و x_3 هستند، توجه کنید که مولفه \bar{x}_1 تقریب بدی به x_1 است و $|x_1 - \bar{x}_1|$ بر نرمها غالب می‌شود. مفهوم فاصله در R^n را نیز می‌توان برای تعریف حد یک دنباله از بردارها در این فضا بکار برد.

تعریف ۷.۸. دنباله $\{x^{(k)}\}_{k=1}^{\infty}$ از بردارها در R^n را، نسبت به نرم $\|\cdot\|$ ، همگرا به x گویند اگر به‌ازای هر $\varepsilon > 0$ عدد صحیحی مانند $N(\varepsilon)$ یافت شود بطوری که به‌ازای هر $k \geq N(\varepsilon)$ ، $\|x^{(k)} - x\| < \varepsilon$.

می‌توان نشان داد که تمام نرمهای روی R^n نسبت به همگرایی معادلند؛ یعنی، هرگاه $\|\cdot\|$ و $\|\cdot\|'$ دو نرم دلخواه روی R^n باشند و $\{x^{(k)}\}_{k=1}^{\infty}$ نسبت به $\|\cdot\|$ دارای حد x باشد، آنگاه $\{x^{(k)}\}_{k=1}^{\infty}$ نسبت به $\|\cdot\|'$ دارای حد x است. اثبات این امر در حالت کلی

را می‌توان در اورتگا [۴۶]، صفحه ۴۸، یافت. این مطلب برای نرمهای $\| \cdot \|_2$ و $\| \cdot \|_\infty$ در تمرین ۳ داده شده است.

مثال ۳. فرض کنیم $x^{(k)} \in R^4$ با

$$\begin{aligned} x^{(k)} &= (x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, x_4^{(k)})' \\ &= \left(1, 2 + \frac{1}{k}, \frac{3}{k^2}, e^{-k} \sin k \right) \end{aligned}$$

تعریف شده باشد. چند نمونه از اعضای این دنباله عبارتند از

$$x^{(1)} = \begin{bmatrix} 1 \\ 3 \\ 3 \\ \frac{\sin 1}{e^2} \end{bmatrix}, \quad x^{(2)} = \begin{bmatrix} 1 \\ 2.5 \\ .75 \\ \frac{\sin 2}{e^2} \end{bmatrix} \approx \begin{bmatrix} 1 \\ 2.5 \\ .75 \\ .1231 \end{bmatrix}$$

$$x^{(10)} = \begin{bmatrix} 1 \\ 2.1 \\ .03 \\ \frac{\sin 10}{e^{10}} \end{bmatrix} \approx \begin{bmatrix} 1 \\ 2.1 \\ .03 \\ -2.5 \times 10^{-5} \end{bmatrix}$$

و

$$x^{(100)} = \begin{bmatrix} 1 \\ 2.01 \\ 3 \times 10^{-4} \\ \frac{\sin 100}{e^{100}} \end{bmatrix} \approx \begin{bmatrix} 1 \\ 2.01 \\ 3 \times 10^{-4} \\ -1.88 \times 10^{-44} \end{bmatrix}$$

مشهود است که بردار

$$x = (1, 2, 0, 0)'$$

حد دنباله $\{x^{(k)}\}_{k=1}^\infty$ است، زیرا تمام دنباله‌های مولفه‌ها به مولفه‌های x همگرا هستند؛ یعنی،

$$\lim_{k \rightarrow \infty} x_1^{(k)} = \lim_{k \rightarrow \infty} 1 = 1 = x_1.$$

$$\lim_{k \rightarrow \infty} x_2^{(k)} = \lim_{k \rightarrow \infty} \left(2 + \frac{1}{k} \right) = 2 = x_2.$$

$$\lim_{k \rightarrow \infty} x_3^{(k)} = \lim_{k \rightarrow \infty} \frac{3}{k^2} = 0 = x_3.$$

$$\lim_{k \rightarrow \infty} x_4^{(k)} = \lim_{k \rightarrow \infty} \frac{\sin k}{e^k} = 0 = x_4.$$

برای اثبات این مطلب توجه می‌کنیم که

$$\mathbf{x}^{(k)} - \mathbf{x} = \left(0, \frac{1}{k}, \frac{3}{k^2}, e^{-k} \sin k \right)^t$$

و، با بسط‌دادن e^k به سری تیلور در مجاورت صفر،

$$e^k \geq 1 + k + \frac{1}{2}k^2 \geq \frac{1}{2}k^2,$$

که به‌ازای $k \geq 3$ نتیجه می‌دهد

$$0 \leq |e^{-k} \sin k| \leq \frac{2}{k^2} |\sin k| \leq \frac{2}{k^2} \leq \frac{1}{k}$$

بنابراین، به‌ازای $k \geq 3$

$$\|\mathbf{x}^{(k)} - \mathbf{x}\|_\infty = \max \left\{ |0|, \left| \frac{1}{k} \right|, \left| \frac{3}{k^2} \right|, |e^{-k} \sin k| \right\} = \frac{1}{k}.$$

با $\varepsilon > 0$ مفروض، فرض کنیم N عدد صحیحی بزرگتر از $1/\varepsilon$ باشد. هرگاه $k \geq N$ ، آنگاه

$$\|\mathbf{x}^{(k)} - \mathbf{x}\|_\infty = \frac{1}{k} \leq \frac{1}{N} < \varepsilon.$$

و $\{\mathbf{x}^{(k)}\}_{k=1}^\infty$ به \mathbf{x} همگراست.

نشان دادن مستقیم اینکه این دنباله نسبت به نرم l_2 به $(1, 2, 0, 0)^t$ همگراست

کاملاً "پیچیده و بسیار ساده‌تر است که قضیهٔ بعدی را اثبات و آن را در این حالت خاص

اعمال کنیم.

قضیهٔ ۸.۸. به‌ازای هر $\mathbf{x} \in R^n$

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

برهان. فرض کنیم x_j مولفه‌های x باشد بطوری که $|x_j| = |x_i|$ $1 \leq i \leq n$ $\|x\|_\infty = \max$ در این صورت ،

$$\|x\|_\infty^2 = |x_j|^2 = x_j^2 \leq \sum_{i=1}^n x_i^2 \leq \sum_{i=1}^n x_j^2 = nx_j^2 = n\|x\|_\infty^2.$$

بنابراین ،

$$\|x\|_\infty \leq \left\{ \sum_{i=1}^n x_i^2 \right\}^{1/2} = \|x\|_2 \leq \sqrt{n} \|x\|_\infty.$$

برای استفاده از این نتیجه جهت نشان دادن اینکه دنباله $\{x^{(k)}\}_{k=1}^\infty$ با

$$x^{(k)} = \left(1, 2 + \frac{1}{k}, \frac{3}{k^2}, e^{-k} \sin k \right)'$$

به $(1, 2, 0, 0)$ همگراست ، فرض کنیم $\varepsilon > 0$ ، و $N(\varepsilon)$ را عدد صحیحی بزرگتر از $2/\varepsilon$ اختیار می‌کنیم . چون قبلاً" نشان داده‌ایم که $\|x^{(k)} - x\|_\infty < 1/k$ ، اگر $k \geq N(\varepsilon)$ ، قضیه ۸.۸ نتیجه می‌دهد

$$\|x^{(k)} - x\|_2 \leq \sqrt{4} \|x^{(k)} - x\|_\infty < \frac{2}{k} \leq \frac{2}{N(\varepsilon)} < \varepsilon$$

همچنین ، لازم است روشی برای اندازه‌گیری فاصله بین دو ماتریس $n \times n$ داشته باشیم ، که مجدداً" محتاج استفاده از مفهوم نرم است .

تعریف ۹.۰۸ . یک نرم ماتریسی بر مجموعه تمام ماتریسهای $n \times n$ حقیقی یک تابع حقیقی مانند $\|\cdot\|$ است که براین مجموعه تعریف شده است و به‌ازای تمام ماتریسهای $n \times n$ ، A و B و تمام اعداد حقیقی α در شرایط زیر صدق می‌کند :

$$\|A\| \geq 0 \quad (T)$$

$$\|A\| = 0 \quad \text{اگر و فقط اگر } A \text{ صفر باشد؛} \quad (ب)$$

$$\|\alpha A\| = |\alpha| \|A\| \quad ؛ \quad (پ)$$

$$\|A + B\| \leq \|A\| + \|B\| \quad ؛ \quad (ت)$$

$$\|AB\| \leq \|A\| \|B\| \quad . \quad (ث)$$

فاصله بین ماتریسهای $n \times n$ ، A و B می‌تواند به طریق معمولی $\|A - B\|$ تعریف

شود .

با وجود اینکه نرمهای ماتریسی می‌توانند به طرق گوناگون بدست آیند، تنها نرمهایی که بررسی خواهیم کرد آنهايي هستند که نتیجهٔ طبیعی نرمهای برداری l_2 و l_∞ می‌باشند.

قضیهٔ زیر به آسانی ثابت می‌شود و اثبات آن به تمرین ۸ موکول شده است.

قضیهٔ ۱۰۰۸. هرگاه $\|\cdot\|$ یک نرم برداری روی R^n باشد، آنگاه

$$\|A\| = \max_{\|x\|=1} \|Ax\|$$

یک نرم ماتریسی بر مجموعهٔ تمام ماتریسهای $n \times n$ حقیقی تعریف می‌کند، که یک نرم طبیعی نامیده می‌شود.

نرمهای ماتریسی که بررسی خواهیم کرد نتیجتاً به اشکال زیر هستند:

$$\|A\|_\infty = \max_{\|x\|_\infty=1} \|Ax\|_\infty, \quad l_\infty \text{ نرم}$$

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2, \quad l_2 \text{ نرم}$$

نرم l_∞ یک ماتریس نمایش جالبی نسبت به درایه‌های آن ماتریس دارد.

قضیهٔ ۱۱۰۸. هرگاه $A = (a_{ij})$ یک ماتریس $n \times n$ باشد، آنگاه

$$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|.$$

برهان. فرض کنیم x یک بردار ستونی n بعدی با $\|x\|_\infty = 1$ باشد. چون Ax نیز یک بردار ستونی n بعدی است،

$$\begin{aligned} \|Ax\|_\infty &= \max_{1 \leq i \leq n} |(Ax)_i| = \max_{1 \leq i \leq n} \left| \sum_{j=1}^n a_{ij} x_j \right| \\ &\leq \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| (\max_{1 \leq j \leq n} |x_j|) = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| \|x\|_\infty \end{aligned}$$

$$= \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|;$$

بنابراین، به ازای هر x که $\|x\|_\infty = 1$ ،

$$\|Ax\|_\infty \leq \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$$

و در نتیجه

$$(۶.۸) \quad \|A\|_\infty = \max_{\|x\|_\infty=1} \|Ax\|_\infty \leq \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|.$$

اما، هرگاه p آن عدد صحیح با $1 \leq p \leq n$ باشد که

$$\sum_{j=1}^n |a_{pj}| = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|,$$

و x با شرایط

$$x_j = \begin{cases} 1 & , a_{pj} \geq 0 \\ -1 & , a_{pj} < 0 \end{cases}$$

اختیار شده باشد، آنگاه $\|x\|_\infty = 1$ ، و به ازای هر $j = 1, 2, \dots, n$ ، $a_{p,j}x_j = |a_{pj}|$ ،
بعلاوه،

$$\begin{aligned} \|Ax\|_\infty &= \max_{1 \leq i \leq n} \left| \sum_{j=1}^n a_{ij}x_j \right| \\ &\geq \left| \sum_{j=1}^n a_{pj}x_j \right| = \sum_{j=1}^n |a_{pj}| = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|. \end{aligned}$$

این نتیجه می دهد که

$$\|A\|_\infty = \max_{\|x\|_\infty=1} \|Ax\|_\infty \geq \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|,$$

که، همراه با معادله (۶.۸)، می دهد

$$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|.$$

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & -1 \\ 5 & -1 & 1 \end{bmatrix}$$

آنگاه

$$\sum_{j=1}^3 |a_{1j}| = |1| + |2| + |-1| = 4,$$

$$\sum_{j=1}^3 |a_{2j}| = |0| + |3| + |-1| = 4,$$

$$\sum_{j=1}^3 |a_{3j}| = |5| + |-1| + |1| = 7,$$

$$\|A\|_{\infty} = \max\{4, 4, 7\} = 7.$$

لذا،

برای تحقیق در l_2 - نرم لازم است مفاهیم بیشتری از جبر خطی را مورد بحث قرار دهیم .

تعریف ۱۲.۰۸ . هرگاه A یک ماتریس $n \times n$ حقیقی باشد، چند جمله‌ای تعریف شده‌با

$$p(\lambda) = \det(A - \lambda I)$$

چند جمله‌ای مشخص A نامیده می‌شود .

به سادگی نشان داده می‌شود (ر.ک. تمرین ۱۱) که $p(\lambda)$ یک چند جمله‌ای درجه n با ضرایب حقیقی است و، در نتیجه، حداکثر n ریشه متمایز دارد، که بعضی از آنها ممکن است مختلط باشند. هرگاه λ یکی از این ریشه‌ها باشد، آنگاه، چون $\det(A - \lambda I) = 0$ ، قضیه ۱۶.۰۷ (ص ۴۵۲) نتیجه می‌دهد که دستگاه خطی تعریف شده با $(A - \lambda I)x = 0$ جوابی غیر از صفر دارد. مایلیم صفرهای p و جوابهای نابدییهی متناظر این دستگاهها را مطالعه کنیم .

تعریف ۱۳.۰۸ . هرگاه $p(\lambda)$ چند جمله‌ای مشخص ماتریس A باشد، صفرهای p مقادیر ویژه یا مقادیر مشخص ماتریس A نامیده می‌شوند. هرگاه λ یک مقدار ویژه از A بوده و $x \neq 0$ دارای این خاصیت باشد که $(A - \lambda I)x = 0$ ، آنگاه x یک بردار ویژه یا بردار

مشخص A ، نظیر مقدار ویژه λ ، نامیده می‌شود .

مثال ۵ . فرض کنیم

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

برای محاسبه مقادیر ویژه A ، در نظر می‌گیریم

$$\begin{aligned} p(\lambda) &= \det(A - \lambda I) = \det \begin{bmatrix} 1 - \lambda & 0 & 1 \\ 2 & 2 - \lambda & 1 \\ -1 & 0 & -\lambda \end{bmatrix} \\ &= (1 - \lambda) \begin{vmatrix} 2 - \lambda & 1 \\ 0 & -\lambda \end{vmatrix} + 1 \begin{vmatrix} 2 & 2 - \lambda \\ -1 & 0 \end{vmatrix} \\ &= (1 - \lambda)(2 - \lambda)(-\lambda) + (2 - \lambda) \\ &= (2 - \lambda)(-\lambda(1 - \lambda) + 1) \\ &= (2 - \lambda)(\lambda^2 - \lambda + 1). \end{aligned}$$

مقادیر ویژه A جوابهای $p(\lambda) = 0$ هستند ، که به سادگی حساب می‌شوند $\lambda_1 = 2$ ،

$$\lambda_2 = \frac{1}{2} + (\sqrt{3}/2)i \text{ و } \lambda_3 = \frac{1}{2} - (\sqrt{3}/2)i$$

یک بردار ویژه x از A مربوط به λ_1 جوابی از دستگاه $(A - \lambda_1 I)x = 0$ با

$$\begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & 1 \\ -1 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

است . نوشتن این دستگاه می‌دهد

$$-x_1 + x_3 = 0,$$

$$2x_1 + x_3 = 0,$$

$$-x_1 - 2x_3 = 0,$$

که جوابی با $x_1 = x_3 = 0$ و x_2 دلخواه دارد . بنابراین ، بخصوص ، $x = (0, 1, 0)$ یک بردار ویژه A نظیر مقدار ویژه $\lambda_1 = 2$ است .

بدست آوردن بردار ویژه‌ای برای $\lambda_2 = \frac{1}{2} + (\sqrt{3}/2)i$ محتاج حل دستگاه زیر است :

$$\begin{bmatrix} 1 - \left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) & 0 & 1 \\ 2 & 2 - \left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) & 1 \\ -1 & 0 & -\left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

با استفاده از حساب مختلط، می‌توان نشان داد (ر.ک. تمرین ۲۵) که یک جواب عبارت است از

$$x_1 = -\frac{1}{2} - \frac{\sqrt{3}}{2}i, \quad x_2 = -\frac{1}{2} + \frac{\sqrt{3}}{2}i, \quad x_3 = 1;$$

بنابراین، یک بردار ویژه نظیر به مقدار ویژه $\lambda_2 = \frac{1}{2} + (\sqrt{3}/2)i$ عبارت است از

$$\mathbf{x} = \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i, -\frac{1}{2} + \frac{\sqrt{3}}{2}i, 1\right)^t$$

به طریقی مشابه، می‌توان نشان داد که

$$\mathbf{x} = \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i, -\frac{1}{2} - \frac{\sqrt{3}}{2}i, 1\right)^t$$

یک بردار ویژه نظیر به مقدار ویژه $\lambda_3 = \frac{1}{2} - \frac{\sqrt{3}}{2}i$ است.

محاسبه مقادیر ویژه و بردارهای ویژه ماتریسها بخش مهمی از جبر خطی عددی است، و در بخشهای ۴۰۸ و ۵۰۸ بیشتر بررسی می‌شود.

تعریف ۱۴۰۸. شعاع طیفی $\rho(A)$ ماتریس A به صورت

$$\rho(A) = \max |\lambda|$$

تعریف می‌شود، که در آن λ یک مقدار ویژه A است. (بخاطر آوریده که، به‌ازای

$$(\cdot) \quad |\lambda| = \{\alpha^2 + \beta^2\}^{1/2}, \quad \text{مختلط، } \lambda = \alpha + \beta i$$

برای ماتریس مثال ۵،

$$\rho(A) = \max\left\{2, \left|\frac{1}{2} + \frac{\sqrt{3}}{2}i\right|, \left|\frac{1}{2} - \frac{\sqrt{3}}{2}i\right|\right\} = \max\{2, 1, 1\} = 2$$

شعاع طیفی با نرم ماتریس ارتباط نزدیکی دارد. همچنانکه در قضیه زیر نشان داده شده است.

قضیه ۱۵.۸. هرگاه A یک ماتریس $n \times n$ حقیقی باشد، آنگاه

$$[\rho(A^t A)]^{1/2} = \|A\|_2 \quad (T)$$

(ب) بهازای هر نرم طبیعی $\|\cdot\|$ ، $\rho(A) \leq \|A\|$.

برهان. اثبات قسمت (T) به اطلاع بیشتری از آنچه درباره مقادیر ویژه در دسترس است نیاز دارد. برای شرح موجود در اثبات، ر.ک. صفحه ۲۱ از اورتگا [۴۶].

برای اثبات قسمت (ب)، فرض کنیم λ یک مقدار ویژه A با بردار ویژه x باشد. چون $Ax = \lambda x$ ، پس $(A - \lambda I)x = 0$ ، و بهازای هر نرم طبیعی،

$$|\lambda| \|x\| = \|\lambda x\| = \|Ax\| \leq \|A\| \|x\|$$

(آخرین نامساوی نتیجه‌ای از تمرین ۱ است)، و

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

بنابراین، $\rho(A) = \max|\lambda| \leq \|A\|$.

یک نتیجه جالب و مفید که مشابه قسمت (ب) از قضیه ۱۵.۸ می‌باشد آن است که، بهازای هر A و هر $\varepsilon > 0$ نرمی وجود دارد مانند $\|\cdot\|$ با این خاصیت که $\|A\| < \rho(A) + \varepsilon$. در نتیجه، $\rho(A)$ بزرگترین کران پایینی برای نرمهای روی A است. اثبات این نتیجه را می‌توان در صفحه ۲۳ اورتگا [۴۶] یافت.

مثال ۶. هرگاه

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

آنگاه، چون $A^t = A$ ، متقارن است، و

$$A'A = A^2 = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 3 & 1 \\ 3 & 3 & 3 \\ 1 & 3 & 5 \end{bmatrix}.$$

برای محاسبه $\rho(A'A)$ به مقادیر ویژه $A'A$ نیاز داریم. هرگاه

$$\begin{aligned} 0 = \det(A'A - \lambda I) &= \det \begin{bmatrix} 5 - \lambda & 3 & 1 \\ 3 & 3 - \lambda & 3 \\ 1 & 3 & 5 - \lambda \end{bmatrix} \\ &= (5 - \lambda)^2(3 - \lambda) + 9 + 9 - (3 - \lambda) - 9(5 - \lambda) - 9(5 - \lambda) \\ &= -\lambda^3 + 13\lambda^2 - 36\lambda \\ &= -\lambda(\lambda - 4)(\lambda - 9), \end{aligned}$$

آنگاه λ عبارت است از 0 ، 4 ، یا 9 . بنابراین،

$$\|A\|_2 = \sqrt{\rho(A'A)} = \sqrt{\max\{0, 4, 9\}} = 3.$$

در مطالعه تکنیکهای تکراری، دانستن اینکه چه وقت توانهای یک ماتریس کوچک می‌شوند، یعنی چه وقت تمام درایه‌ها به صفر میل می‌کنند، اهمیت خاصی دارد. ماتریسهایی از این نوع همگرا نامیده می‌شوند.

تعریف ۱۶.۸. یک ماتریس $n \times n$ مانند A را همگرا نامیم اگر به ازای هر $i = 1, 2, \dots, n$

$$\lim_{k \rightarrow \infty} (A^k)_{ij} = 0, \quad j = 1, 2, \dots, n$$

مثال ۷. فرض کنیم

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

با محاسبه توانهای A بدست می‌آوریم

$$A^2 = \begin{bmatrix} \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}, \quad A^3 = \begin{bmatrix} \frac{1}{8} & 0 \\ \frac{3}{16} & \frac{1}{8} \end{bmatrix}, \quad A^4 = \begin{bmatrix} \frac{1}{16} & 0 \\ \frac{1}{8} & \frac{1}{16} \end{bmatrix},$$

و، بطور کلی،

$$A^k = \begin{bmatrix} (\frac{1}{2})^k & 0 \\ k/2^{k+1} & (\frac{1}{2})^k \end{bmatrix}.$$

چون

$$\lim_{k \rightarrow \infty} \left(\frac{1}{2}\right)^k = \lim_{k \rightarrow \infty} \frac{k}{2^{k+1}} = 0,$$

A یک ماتریس همگراست. توجه کنید که $\rho(A) = \frac{1}{2}$ ، زیرا $\frac{1}{2}$ تنها مقدار ویژه A است.

همچنانکه ممکن است به نظر تان آمده باشد، رابطه‌ای بین شعاع طیفی یک ماتریس و همگرایی آن وجود دارد.

قضیه ۱۷.۰۸. احکام زیر معادلند:

(آ) A یک ماتریس همگراست؛

(ب) $\lim_{n \rightarrow \infty} \|A^n\| = 0$ ، چون $\|\cdot\|$ یک نرمی است؛

(پ) $\rho(A) < 1$

برهان این قضیه را می‌توان در صفحه ۱۴ از ایزاکسون و کلر [۳۹] یافت.

مجموعه تمرینات ۱.۸

۱. نشان دهید که برای هر نرم طبیعی $\|Ax\| \leq \|A\| \|x\|$

۲. تحقیق کنید که تابع $\|\cdot\|$ ، تعریف شده روی R^n با

$$\|x\|_1 = \sum_{i=1}^n |x_i|,$$

یک نرم روی R^n است.

۳. نشان دهید هرگاه دنباله $\{x^{(k)}\}_{k=1}^{\infty}$ نسبت به نرم l_2 به $x \in R^n$ همگرا باشد، آنگاه

نسبت به نرم l_∞ به x همگراست و بالعکس.

۴. برای بردارهای زیر $\|x\|_1$ ، $\|x\|_\infty$ و $\|x\|_2$ را پیدا کنید:

$$(A) \quad x = (3, -4, 0, \frac{3}{2})'$$

(ب) $x = (4/(k+1), 2/k^2, k^2 e^{-k})'$ ، که در آن k یک عدد صحیح مثبت ثابت است.

۵. ثابت کنید دنباله‌های زیر همگرا هستند و حد آنها را پیدا کنید:

$$(A) \quad x^{(k)} = (1/k, e^{1-k}, -2/k^2)'$$

؛ $x^{(k)} = (ke^{-k^2}, (\cos k)/k, \sqrt{k^2 + k - k})'$ (ب)

$x^{(k)} = (e^{1/k}, (k^2 + 1)/(1 - k^2), (1/k^2)(1 + 3 + 5 + \dots + (2k - 1)))'$ (پ)

۶. نشان دهید که $\|\cdot\|_{\textcircled{2}}$ و $\|\cdot\|_{\textcircled{1}}$ نرمهای ماتریسی هستند، که بهازای هر ماتریس $A, n \times n$

$$\|A\|_{\textcircled{1}} = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| \quad \text{و} \quad \|A\|_{\textcircled{2}} = \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2}$$

۷. نشان دهید که $\|\cdot\|_{\textcircled{2}}$ ، تعریف شده بوسیله $\|A\|_{\textcircled{2}} = \max_{1 \leq i, j \leq n} |a_{ij}|$ ، یک نرم ماتریسی را تعریف نمی‌کند.

۸. قضیه ۱۰.۸ را ثابت کنید.

۹. نرم ماتریسی $\|\cdot\|_1$ را با

$$\|A\|_1 = \max_{\|x\|_1 = 1} \|Ax\|_1$$

تعریف کنید و نشان دهید که

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|.$$

۱۰. برای ماتریسهای زیر $\|\cdot\|_1$ و $\|\cdot\|_{\infty}$ را پیدا کنید:

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \quad (\text{ت})$$

$$B = \begin{bmatrix} -3 & 2 \\ \frac{3}{2} & \frac{7}{2} \end{bmatrix} \quad (\text{ب})$$

$$C = \begin{bmatrix} 10 & 15 \\ 0 & 1 \end{bmatrix} \quad (\text{پ})$$

۱۱. نشان دهید که چندجمله‌ای مشخص $P(\lambda) = \det(A - \lambda I)$ ، برای ماتریس $n \times n$ ، A ، یک چندجمله‌ای درجه n است. [راهنمایی. $\det(A - \lambda I)$ را نسبت به سطر اول بسط دهید و از استقرای روی n استفاده کنید.]

۱۲. نشان دهید هرگاه A یک ماتریس $n \times n$ باشد، آنگاه

$$\det A = \prod_{i=1}^n \lambda_i.$$

که در آن $\lambda_1, \dots, \lambda_n$ مقادیر ویژه A هستند. [راهنمایی. $\det A = p(0)$ و

$$[\cdot p(\lambda) = (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n)$$

۱۳. نشان دهید A منفرد است اگر و فقط اگر $\lambda = 0$ یک مقدار ویژه A باشد.

۱۴. مقادیر ویژه و بردارهای ویژه مربوطه ماتریسهای زیر را حساب کنید:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad (\text{ب}) \qquad \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \quad (\text{ت})$$

$$\begin{bmatrix} -1 & 2 & 0 \\ 0 & 3 & 4 \\ 0 & 0 & 7 \end{bmatrix} \quad (\text{پ})$$

$$\begin{bmatrix} 2 & 1 & 1 \\ 2 & 3 & 2 \\ 1 & 1 & 2 \end{bmatrix} \quad (\text{ث})$$

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad (\text{ج})$$

$$\begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \quad (\text{ح})$$

۱۵. شعاع طیفی هر یک از ماتریسهای تمرین ۱۴ را بیابید.

۱۶. $\|A\|_2$ را برای هر یک از ماتریسهای تمرین ۱۴ بیابید.

۱۷. نشان دهید اگر A متقارن باشد، $\rho(A) = \|A\|_2$.

۱۸. (آ) نشان دهید که مقادیر ویژه توانهای A^k ، $k = 1, 2, 3, \dots$ ، ماتریس A به

ترتیب توانهای مقادیر ویژه λ^k ، $k = 1, 2, 3, \dots$ ، با همان بردارهای ویژه A ، می‌باشند، که در آنها λ یک مقدار ویژه A است.

(ب) نشان دهید هرگاه A^{-1} موجود باشد، آنگاه مقادیر ویژه A^{-1} معکوسهای مقادیر ویژه A با همان بردارهای ویژه می‌باشند.

(پ) قسمتهای (آ) و (ب) را به $(A^{-1})^k$ ، که $k = 2, 3, 4, \dots$ ، تعمیم دهید.

(ت) فرض کنید $q(x) = q_0 + q_1x + \dots + q_kx^k$ یک چند جمله‌ای باشد، و

تعریف کنید $q(A) = q_0 I + q_1 A + \dots + q_k A^k$. نشان دهید که مقادیر ویژه $q(A)$ عبارتند از $q(\lambda)$ ، که در آن λ نمایش مقدار ویژه دلخواهی از A است .

۱۹ . ماتریسهای A و B را چنان پیدا کنید که به ازای آنها $\rho(A + B) > \rho(A) + \rho(B)$.
(این نشان می دهد که $\rho(A)$ نمی تواند نرم ماتریسی باشد .)

۲۰ . نشان دهید که

$$A_1 = \begin{bmatrix} 1 & 0 \\ \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

همگرا نیست ، اما

$$A_2 = \begin{bmatrix} \frac{1}{2} & 0 \\ 16 & \frac{1}{2} \end{bmatrix}$$

همگراست .

۲۱ . از ماتریسهای تمرین ۱۴ کدامها همگرايند ؟

۲۲ . نشان دهید که به ازای هر مقدار ویژه λ از ماتریس نامنفرد A ،
 $\|A\| \leq |\lambda| \leq \|A^{-1}\|^{-1}$ ، که در آن $\|\cdot\|$ نرم طبیعی دلخواهی است .

۲۳ . فرض کنید S یک ماتریس معین مثبت باشد .

(آ) نشان دهید که مقادیر ویژه S اعداد حقیقی مثبت هستند .

(ب) به ازای هر $x \in R^n$ تعریف کنید $\|x\| = \sqrt{x^T S x}$. نشان دهید که این یک نرم

روی R^n تعریف می کند .

۲۴ . فرض کنید S ماتریسی نامنفرد و حقیقی بوده و $\|\cdot\|$ نرم دلخواهی روی R^n باشد .

$\|\cdot\|'$ را با $\|Sx\| = \|x\|'$ تعریف کنید . نشان دهید که $\|\cdot\|'$ نیز یک نرم روی R^n است .

۲۵ . نشان دهید که

$$\begin{bmatrix} 1 - \left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) & 0 & 1 \\ 2 & 2 - \left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) & 1 \\ -1 & 0 & -\left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) \end{bmatrix} \begin{bmatrix} -\frac{1}{2} - \frac{\sqrt{3}}{2}i \\ -\frac{1}{2} + \frac{\sqrt{3}}{2}i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

۲۶ . در تمرین ۹ بخش ۵.۷ ماتریس مقارن

$$A = \begin{bmatrix} 1.59 & 1.69 & 2.13 \\ 1.69 & 1.31 & 1.72 \\ 2.13 & 1.72 & 1.85 \end{bmatrix}$$

برای بیان میانگین طول بالهای حشرات میوه‌جات که حاصل جفت‌گیری سه نوع از این حشرات بکار رفت. درایه a_{ij} میانگین طول بال یک حشره، که فرزند یک حشره و یک حشره ماده از نوع j است، را نمایش می‌دهد.

(آ) مقادیر ویژه و بردارهای ویژه مربوطه این ماتریس را پیدا کنید.

(ب) نتیجه تمرین ۲۳ را بکار ببرید تا به سوال طرح شده در قسمت (ب) تمرین ۹ بخش ۵.۷ پاسخ دهید؛ یعنی، آیا این ماتریس معین مثبت است؟

۲۷. در تمرین ۱۷ بخش ۳.۷ فرض کردیم سهمی که یک سوسک ماده از یک نوع معین در جمعیت سالهای بعد سوسکهای ماده دارد با ماتریس زیر بیان شود

$$A = \begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix}$$

که در آن درایه سطر i م و ستون j سهم احتمالی یک سوسک با سن i در جمعیت سوسکهای ماده سال بعد با سن j را نمایش می‌دهد.

(آ) آیا ماتریس A مقدار ویژه‌ای حقیقی دارد؟ اگر دارد، آنها و بردارهای ویژه مربوطه را تعیین کنید.

(ب) اگر یک نمونه از این انواع برای مقاصد آزمایشگاهی مورد نیاز باشد، چه معیاری باید روی جمعیت اولیه اعمال کرد تا اطمینان حاصل شود که این نمونه دارای نسبت ثابتی در هر گروه سنی از سالی به سالی است.

۲.۸ تخمینهای خطا و تصفیه تکراری

با معرفی نرم یک بردار و ماتریس، می‌توان تخمینهای خطای موجود در حل دستگاههای خطی را دقیقتر مورد بحث قرار داد. یک راه تخمین دقت یک جواب تقریبی به یک دستگاه خطی، تعیین آن است که تقریب به چه نزدیکی در دستگاه صدق می‌کند.

تعریف ۱۸.۸. هرگاه $\bar{x} \in R^n$ تقریبی به جواب دستگاه خطی تعریف شده با $Ax = b$

باشد، بردار مانده برای \bar{x} نسبت به این دستگاه با $r = b - A\bar{x}$ تعریف می شود.

به طور شهودی انتظار می رود که وقتی $\|r\|$ برای نرمی برداری کوچک باشد، $\|x - \bar{x}\|$ نیز به همان وضع کوچک باشد. اگرچه اغلب چنین است، دستگاههای خاص معینی که اغلب در عمل ظاهر می شوند این خاصیت را ندارند.

مثال ۱. دستگاه خطی $Ax = b$ داده شده با

$$\begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3.0001 \end{bmatrix}$$

دارای جواب منحصر بفرد $x = (1, 1)$ است. تقریب $\bar{x} = (3, 0)$ به این جواب دارای بردار مانده

$$r = b - A\bar{x} = \begin{bmatrix} 3 \\ 3.0001 \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ .0002 \end{bmatrix}$$

است. بنابراین، $\|r\|_\infty = .0002$. اگرچه نرم بردار مانده کوچک است، بوضوح تقریب $\bar{x} = (3, 0)$ "کاملاً" بد است؛ در واقع، $\|x - \bar{x}\|_\infty = 2$.

مشکلی که در این مثال بروز می کند، با توجه به اینکه جواب دستگاه محل برخورد

خطوط

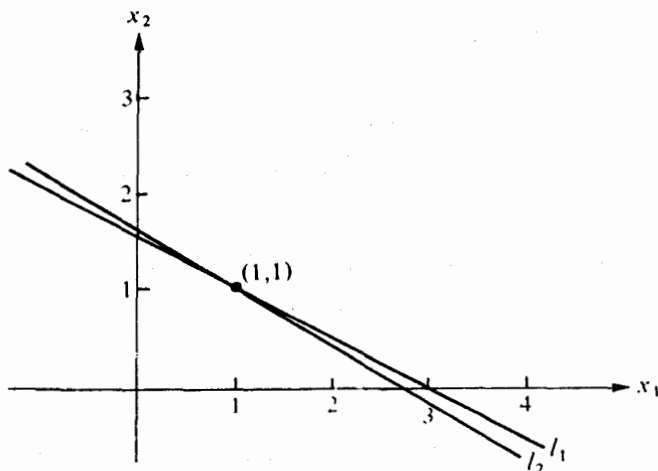
$$l_1: x_1 + 2x_2 = 3 \quad \text{و} \quad l_2: 1.0001x_1 + 2x_2 = 3.0001$$

را نمایش می دهد، قابل توضیح است.

نقطه $(3, 0)$ روی خط l_1 قرار دارد و خطوط فوق تقریباً موازیند. این نتیجه می دهد که $(3, 0)$ در نزدیکی l_2 نیز قرار دارد، هرچند به طور قابل ملاحظه ای با نقطه تلاقی $(1, 1)$ متفاوت است. (ر. ک. شکل ۳۰۸.)

این مثال آشکارا برای نشان دادن اینکه ممکن است، و در حقیقت حتماً، مشکلاتی رخ دهد ساخته شده بود. اگر دو خط تقریباً موازی نمی بودند، انتظار می رفت که یک بردار مانده کوچک یک تقریب دقیق را نتیجه دهد.

دروضیعت کلی، برای دادن نشانی از اینکه چه وقت ممکن است مشکلاتی رخ دهد، نمی توان به نمایش هندسی دستگاه تکیه کرد. اما، می توان این اطلاعات را بوسیله بررسی نرمهای ماتریس A و معکوس آن بدست آورد.



شکل ۳.۸

قضیه ۱۹.۸. هرگاه \bar{x} تقریبی به جواب $AX = b$ و A ماتریسی نامنفرد باشد، آنگاه، به ازای هر نرم طبیعی،

$$(۷.۸) \quad \|x - \bar{x}\| \leq \|r\| \|A^{-1}\|$$

و، به شرط آنکه $x \neq 0$ و $b \neq 0$ ،

$$(۸.۸) \quad \frac{\|x - \bar{x}\|}{\|x\|} \leq \|A\| \|A^{-1}\| \frac{\|r\|}{\|b\|}$$

که در آن r بردار مانده برای \bar{x} نسبت به دستگاه $AX = b$ است.

برهان. چون $x - \bar{x} = A^{-1}r$ ، $r = b - A\bar{x} = Ax - A\bar{x}$ ، تمرین بخش ۱.۸ نتیجه می‌دهد که

$$\|x - \bar{x}\| = \|A^{-1}r\| \leq \|A^{-1}\| \|r\|$$

بعلاوه، چون $b = Ax$ ، بنابراین، $\|b\| \leq \|A\| \|x\|$.

$$\frac{\|x - \bar{x}\|}{\|x\|} \leq \frac{\|A\| \|A^{-1}\|}{\|b\|} \|r\|.$$

معادله (۷.۸) و (۸.۸) نتیجه می‌دهند که از کمیت‌های $\|A^{-1}\|$ و $\|A\| \|A^{-1}\|$ می‌توان برای دادن نشانی از ارتباط بین بردار مانده و دقت تقریب استفاده کرد. در حالت

کلی، خطای نسبی ($\|x - \bar{x}\|/\|x\|$) از اهمیت بسزایی برخوردار است و، بنا بر (۸۰۸) این خطا بوسیله حاصل ضرب ($\|A\| \|A^{-1}\|$) در ماندهء نسبی برای این تقریب، یعنی ($\|r\|/\|b\|$)، محدود می‌شود. برای این تقریب هر نرم مناسبی را می‌توان بکار برد؛ تنها الزام آن است که آن را در همهجا به‌طور سازگار بکار برد.

تعریف ۲۰.۸. عدد شرطی $K(A)$ از ماتریس نامفرد A نسبت به نرم $\|\cdot\|$ با

$$K(A) = \|A\| \|A^{-1}\|.$$

تعریف می‌شود.

با این نمادگذاری، نامساویهای قضیه ۱۹.۸ چنین می‌شوند:

$$(۹۰۸) \quad \|x - \bar{x}\| \leq K(A) \frac{\|r\|}{\|A\|}.$$

$$(۱۰۰۸) \quad \frac{\|x - \bar{x}\|}{\|x\|} \leq K(A) \frac{\|r\|}{\|b\|}.$$

چون به‌ازای هر ماتریس نامفرد A

$$1 = \|I\| = \|A \cdot A^{-1}\| \leq \|A\| \|A^{-1}\| = K(A)$$

انتظار آن است که اگر $K(A)$ نزدیک ۱ باشد، ماتریس A خوشرفتار باشد (کفرسا) ماتریس خوش وضع نامیده می‌شود، و وقتی $K(A)$ به‌طور قابل ملاحظه‌ای بزرگتر از ۱ باشد، خوشرفتار نباشد (که بد وضع نامیده می‌شود). در این حالت رفتار به‌اطمینان نسبی به اینکه یک بردار ماندهء کوچک مستلزم جواب تقریبی دقیق متناظری است مربوط می‌شود.

مثال ۲. ماتریس دستگاهی که در مثال ۱ بررسی شد عبارت بود از

$$A = \begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix}$$

که دارای $\|A\|_\infty = 3.0001$ است. این نرم بزرگ اطلاق نمی‌شود؛ معیناً، چون

$$A^{-1} = \begin{bmatrix} -10000 & 10000 \\ 5000.5 & -5000 \end{bmatrix};$$

$$\|A^{-1}\|_\infty = 20,000,$$

و، برای نرم بی نهایت، $K(A) = (20,000)(3.0001) = 60,002$. اندازه عدد شرطی برای این مثال یقیناً " مانع می شود که براساس مانده یک تقریب در مورد دقت آن تصمیم بگیریم .

گرچه به طور نظری عدد شرطی یک ماتریس کاملاً" به نرمهای ماتریس و معکوس آن بستگی دارد ، عملاً" محاسبه معکوس مقید به خطای گرد کردن وابسته به دقتی است که محاسبات با آن انجام می شود . اگر عملیات شامل محاسبات با دقت t رقم با معنی باشد ، عدد شرطی تقریبی برای ماتریس در واقع نرم ماتریس ضریب نرم تقریب به معکوس A خواهد بود ، که با استفاده از حساب t رقمی بدست آمده است . در واقع ، عدد شرطی حتی به روشی که برای محاسبه معکوس A بکار رفته بستگی دارد . اگر فرض شود که جواب تقریبی به دستگاه خطی $AX = b$ با استفاده از حساب t رقمی و حذف گاوسی تعیین می شود ، نشان داده می شود (ر.ک. فورسایت^۱ و مولر^۲ [۲۸] ، صفحات ۵۱ - ۴۹) که بردار مانده r برای تقریب \bar{x} دارای خاصیت زیر است :

$$\|r\| \approx 10^{-t} \|A\| \|\bar{x}\|. \quad (11.08)$$

از این معادله تقریبی ، بدون الزام به معکوس کردن A ، می توان تخمینی برای عدد شرطی موثر در حساب t رقمی بدست آورد . در واقع ، تقریب در معادله (۱۱.۰۸) فرض می کند که تمام عملیات حسابی در تکنیک حذف گاوسی با استفاده از حساب t رقمی انجام شده است ، اما آن عملیاتی که برای تعیین مانده لازم است در دقت مضاعف (یعنی ، $2t$ رقمی) انجام گرفته است . این تکنیک به زحمات محاسباتی نمی افزاید ، و اکثر از دست دادن دقت (ر.ک. بخش ۲.۰۱) را که در تفریق اعداد تقریباً مساوی ، که در محاسبه مانده ظاهر می شوند ، حذف می کند .

تقریب برای عدد شرطی t رقمی $K(A)$ از بررسی دستگاه زیر می آید :

$$Ay = r.$$

جواب این دستگاه را می توان فوراً" تقریب زد ، زیرا مضارب برای روش حذف گاوسی قبلاً" محاسبه و خوشبختانه نگهداری شده اند .

در واقع ، \bar{y} ، یعنی جواب تقریبی $Ay = r$ ، در

$$\bar{y} \approx A^{-1}r = A^{-1}(b - A\bar{x}) = A^{-1}b - A^{-1}A\bar{x} = x - \bar{x}, \quad (12.08)$$

صدق می‌کند. بنابراین، \bar{y} تخمینی از خطا در تقریب جواب دستگاه اولیه است. لذا، می‌توان معادله (۱۱.۸) را بکار برد و نتیجه گرفت که

$$\|\bar{y}\| \approx \|x - \bar{x}\| = \|A^{-1}r\| \leq \|A^{-1}\| \|r\| \approx \|A^{-1}\| (10^{-4} \|A\| \|\bar{x}\|) \\ = 10^{-4} \|\bar{x}\| K(A).$$

این امر تقریبی برای عدد شرطی موجود در حل دستگاه $Ax = b$ با استفاده از حذف گاوسی، و آن نوع حساب l رقمی که در بالا ذکر شد، بدست می‌دهد:

$$K(A) \approx \frac{\|\bar{y}\|}{\|\bar{x}\|} 10^4. \quad (13.8)$$

مثال ۳. دستگاه خطی داده شده با

$$\begin{bmatrix} 3.3330 & 15920 & -10.333 \\ 2.2220 & 16.710 & 9.6120 \\ 1.5611 & 5.1791 & 1.6852 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 15913 \\ 28.544 \\ 8.4254 \end{bmatrix}$$

دارای جواب دقیق $x = (1, 1, 1)'$ است.

استفاده از حذف گاوسی و حساب 5 رقمی منجر به ماتریس افزوده زیر می‌شود:

$$\left[\begin{array}{ccc|c} 3.3330 & 15920 & -10.333 & 15913 \\ 0 & -10596 & 16.501 & -10580 \\ 0 & 0 & -5.0790 & -4.7000 \end{array} \right].$$

و جواب تقریبی این دستگاه عبارت است از

$$\bar{x} = (1.2001, .99991, .92538)'$$

بردار مانده نظیر به \bar{x} که در دقت مضاعف حساب شده عبارت است از

$$r = b - A\bar{x}$$

$$= \begin{bmatrix} 15913 \\ 28.544 \\ 8.4254 \end{bmatrix} - \begin{bmatrix} 3.3330 & 15920 & -10.333 \\ 2.2220 & 16.710 & 9.6120 \\ 1.5611 & 5.1791 & 1.6852 \end{bmatrix} \begin{bmatrix} 1.2001 \\ .99991 \\ .92538 \end{bmatrix} \\ = \begin{bmatrix} -.00518 \\ .27413 \\ -.18616 \end{bmatrix}.$$

بنابراین، $\|r\|_{\infty} = 0.27413$.

استفاده از حساب 5 رقمی برای محاسبات، تقریب زیر را برای معکوس A می دهد:

$$A^{-1} = \begin{bmatrix} -1.1701 \times 10^{-4} & -1.4983 \times 10^{-1} & 8.5416 \times 10^{-1} \\ 6.2782 \times 10^{-5} & 1.2124 \times 10^{-4} & -3.0662 \times 10^{-4} \\ -8.6631 \times 10^{-5} & 1.3846 \times 10^{-1} & -1.9689 \times 10^{-1} \end{bmatrix}$$

و می توان قضیه (۱۱.۰۸) را برای نشان دادن اینکه $\|A^{-1}\|_{\infty} = 1.0041$ و $\|A\|_{\infty} = 15934$ بکار برد. عدد شرطی در این حالت فوق العاده بزرگ است:

$$(14.08) \quad K(A) \approx (1.0041)(15934) = 15999.$$

تخمین عدد شرطی که در بحث قبل داده شد ابتدا با حل دستگاه

$$\begin{bmatrix} 3.3330 & 15920 & -10.333 \\ 2.2220 & 16.710 & 9.6120 \\ 1.5611 & 5.1791 & 1.6852 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -0.00518 \\ 0.27413 \\ -1.18616 \end{bmatrix}$$

که $\tilde{y} = (-0.20008, 8.9987 \times 10^{-5}, 0.074607)^T$ را نتیجه می دهد، و بعد، با استفاده از تخمین داده شده در (۱۳.۰۸) بدست می آید

$$(15.08) \quad K(A) \approx 10^5 \frac{\|\tilde{y}\|_{\infty}}{\|\tilde{x}\|_{\infty}} = \frac{10^5(0.20008)}{1.2001} = 16672.$$

تخمین کاملاً "نزدیک به مقدار واقعی است و زحمت محاسباتی فوق العاده کمتری لازم دارد.

چون جواب واقعی $x = (1, 1, 1)$ برای این دستگاه معلوم است، می توان حساب کرد

$$\frac{\|x - \tilde{x}\|}{\|x\|} = \frac{0.2001}{1} = 0.2001. \quad \text{و} \quad \|x - \tilde{x}\|_x = 0.2001$$

کرانه های خطای داده شده در قضیه ۱۹.۰۸ برای این مقادیر عبارتند از

$$\|x - \tilde{x}\|_{\infty} \leq K(A) \frac{\|r\|_{\infty}}{\|A\|_{\infty}} = \frac{(15999)(0.27413)}{15934} = 0.27525$$

$$\frac{\|x - \tilde{x}\|_{\infty}}{\|x\|_{\infty}} \leq K(A) \frac{\|r\|_{\infty}}{\|b\|_{\infty}} = \frac{(15999)(0.27413)}{15913} = 0.27561.$$

در معادله (۱۲.۰۸) تخمین $x - \tilde{x}$ را بکار بردیم، که در آن \tilde{y} جواب تقریبی

دستگاه $Av = r$ است. معقول است که از این نتیجه‌گمان بریم که $\bar{x} + \bar{r}$ یک تقریب دقیقتری از تقریب اولیه \bar{x} به جواب دستگاه خطی $Ax = b$ خواهد بود. روشی که طرف راست آن، برای تقریبهای متوالی، بردار ماندهء مربوطه است تشکیل می‌شود تا اینکه نتایج دقیق‌رضایت‌بخشی بدست آید. این روند، عموماً، فقط روی دستگاههایی که برای آنها گمان بدو وضعی ماتریس مربوطه می‌رود بکار خواهد رفت، زیرا تقریب برای یک دستگاه خوش وضع با این تکنیک به‌طور قابل ملاحظه‌ای بهتر نمی‌شود.

الگوریتم تصفیهء تکراری ۱۰.۸

برای تقریب جواب دستگاه $Ax = b$ ، وقتی گمان رود A بد وضع باشد، تقریب اولیهء $\bar{x}^{(1)}$ را بوسیلهء حذف گاوسی (مضارب m_{ji} ، $i = 1, 2, \dots, n-1$ ، $j = i+1, i+2, \dots, n$) را ذخیره و به تعویضهای سطری توجه کنید) بدست می‌آوریم.

مرحلهء ۱. قرار می‌دهیم $k = 1$.

مرحلهء ۲. مولفه‌های ماندهء $\bar{x}^{(k)}$ را به‌ازای هر $i = 1, 2, \dots, n$ حساب می‌کنیم،

$$r_i^{(k)} = b_i - \sum_{j=1}^n a_{ij} \bar{x}_j^{(k)}$$

(محاسبات را در حساب دقت مضاعف انجام می‌دهیم.)

مرحلهء ۳. دستگاه خطی

$$Ay^{(k)} = r^{(k)}$$

را با استفاده از حذف گاوسی، به همان ترتیبی که برای بدست آوردن $\bar{x}^{(1)}$ بکار رفته بود، حل می‌کنیم.

مرحلهء ۴. مختصات $\bar{x}^{(k+1)}$ ، یعنی $\bar{x}_i^{(k+1)} = \bar{x}_i^{(k)} + r_i^{(k)}$ را به‌ازای هر $i = 1, 2, \dots, n$ حساب می‌کنیم.

مرحلهء ۵. اگر $\bar{x}^{(k+1)} - \bar{x}^{(k)}$ کوچک نباشد، k را به $k+1$ اضافه کرده به مرحلهء ۲ می‌رویم.

مرحلهء ۶. روند کامل است؛ $\bar{x}^{(k+1)}$ تقریب پذیرفته شده به جواب $Ax = b$ است.

یک روند متوقف توصیه شده در مرحلهء ۵ انجام تکرار تا اینکه $|\bar{r}_i^{(k)}| \leq 10^{-4}$ ،

به‌ازای هر $i = 1, 2, \dots, n$ ، خواهد بود، اگر حساب k رقی انجام می‌شود. باید تأکید شود که تکنیک تصفیهء تکراری برای تمام دستگاههای شامل ماتریسهای بدوضع جواب

رضایت بخش نمی‌دهد. بخصوص، اگر $K(A) \geq 10^4$ ، روند احتمالا "به نتیجه نمی‌رسد و تنها راه دیگر، در صورت امکان، استفاده از دقت اضافه شده برای محاسبات است.

مثال ۴. در مثال ۳ تقریب به مسئله‌ای که تحت بررسی بود را با استفاده از حساب پنج رقمی و حذف گاوسی بدست آوردیم، که بود

$$\bar{x}^{(1)} = (1.2001, .99991, .92538)^T$$

و جواب $Ay = r^{(1)}$ عبارت بود از

$$y^{(1)} = (-.20008, 8.9987 \times 10^{-5}, .074607)^T.$$

بنابراین مرحله ۴ الگوریتم فوق، این امر نتیجه می‌دهد که

$$\bar{x}^{(2)} = \bar{x}^{(1)} + y^{(1)} = (1.0000, 1.0000, .99999)^T$$

و خطای واقعی در این تقریب عبارت است از

$$\|x - \bar{x}^{(2)}\|_{\infty} = 1 \times 10^{-5}.$$

اگر تکنیک توقف توصیه شده برای الگوریتم بکار رود، باید $r^{(2)} = b - A\bar{x}^{(2)}$ حساب شود و دستگاه $Ay^{(2)} = r^{(2)}$ حل گردد، که خواهد داد

$$y^{(2)} = (1.5002 \times 10^{-9}, 2.0951 \times 10^{-10}, 1.0000 \times 10^{-5})^T.$$

چون $\|y^{(2)}\|_{\infty} \leq 10^{-5}$ ، نتیجه می‌شود که

$$\bar{x}^{(3)} = \bar{x}^{(2)} + \bar{y}^{(2)} = (1.0000, 1.0000, 1.0000)^T$$

به حد کافی دقیق است، که یقیناً درست می‌باشد.

در سراسر این بخش فرض شده است که در دستگاه $Ax = b$ ، A و b به طور دقیق

نمایش داده می‌شوند. فی الواقع، درایه‌های a_{ij} و b_i ممکن است به مقدار δa_{ij} و δb_i تغییر یافته یا پریشیده باشند، و سبب شوند که دستگاه خطی

$$(A + \delta A)x = b + \delta b$$

به جای $Ax = b$ حل شود. به طور نرمال، اگر $\|\delta A\|$ و $\|\delta b\|$ کوچک باشند، از مرتبه 10^{-4} ، حساب r رقمی باید یک جواب \bar{x} حاصل کند که برای آن $\|x - \bar{x}\|$ متناظراً "کوچک است"، اما، در حالت دستگاههای بدوضع، دیده شده که حتی اگر A و b دقیقاً نمایش داده شده باشند، خطاهای گرد کردن ممکن است سبب شوند که $\|x - \bar{x}\|$ بزرگ باشد. قضیه زیر پریشیدگی دستگاههای خطی را به عدد شرطی یک ماتریس ارتباط می‌دهد (برهان این نتیجه را می‌توان در صفحه ۳۳ اورنگا [۴۶] یافت).

قضیه ۲۱.۸. فرض کنیم A نامنفرد باشد و

$$\|\delta A\| < \frac{1}{\|A^{-1}\|}.$$

جواب \bar{x} به $(A + \delta A)\bar{x} = b + \delta b$ از $Ax = b$ را با تخمین خطای زیر تقریب می‌کند:

$$(16.8) \quad \frac{\|x - \bar{x}\|}{\|x\|} \leq \frac{K(A)}{1 - K(A)(\|\delta A\|/\|A\|)} \left(\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|} \right).$$

تخمین مذکور در معادله (۱۶.۸) بیان می‌کند که هرگاه ماتریس A خوش وضع باشد، یعنی $K(A)$ خیلی بزرگ نباشد، آنگاه تغییرات کوچک در A و b تغییرات کوچک نظیر در جواب x حاصل می‌کنند. از سوی دیگر، هرگاه A بدوضع باشد، آنگاه تغییرات کوچک در A و b ممکن است تغییرات فاحش در x ایجاد کنند.

این قضیه از روند عددی خاص بکار رفته برای حل $Ax = b$ مستقل است. می‌توان نشان داد (مثلاً، ر.ک. ویلکسون^۱ [۶۸] یا [۶۹]) که اگر حذف گاوسی با محورگیری برای حل $Ax = b$ در حساب t رقمی بکار رود، جواب عددی \bar{x} جواب واقعی یک دستگاه خطی

$$(17.8) \quad (A + \delta A)\bar{x} = b$$

است، که در آن

$$(18.8) \quad \|\delta A\|_{\infty} \leq f(n)10^{t-1} \max_{i,j,k} |a_{ij}^{(k)}|.$$

ویلکسون عملاً^۱ دریافته است که $f(n) \approx n$ ، در بدترین وضع $f(n) \leq 1.01(n^3 + 3n^2)$.

مجموعه تمرینات ۲۰.۸

۱. دستگاههای خطی زیر را با استفاده از حذف گاوسی و بهتر کردن تکراری حل کنید. $K(A)$ را نیز تخمین بزنید.

$$x_1 + \frac{1}{2}x_2 + \frac{1}{3}x_3 = \frac{11}{6}, \quad (\bar{A})$$

$$5x_1 + \frac{10}{3}x_2 + \frac{5}{2}x_3 = \frac{65}{6},$$

$$(\text{حساب سه رقمی}) \quad \frac{100}{3}x_1 + 25x_2 + 20x_3 = \frac{235}{3}.$$

$$1.003x_1 + 58.09x_2 = 68.12, \quad (\varphi)$$

$$5.550x_1 + 321.8x_2 = 377.8 \quad (\text{حساب چهاررقمی})$$

$$3.9x_1 + 1.6x_2 = 5.5, \quad (\varphi)$$

$$6.8x_1 + 2.9x_2 = 9.7. \quad (\text{حساب دورقمی})$$

$$\frac{1}{4}x_1 + \frac{1}{5}x_2 + \frac{1}{6}x_3 = 9, \quad (\text{ت})$$

$$\frac{1}{3}x_1 + \frac{1}{4}x_2 + \frac{1}{5}x_3 = 8,$$

$$\frac{1}{2}x_1 + x_2 + 2x_3 = 8. \quad (\text{حساب کامپیوتری تک دقتی})$$

$$4.56x_1 + 2.18x_2 = 6.74, \quad (\text{ث})$$

$$2.79x_1 + 1.38x_2 = 4.13. \quad (\text{حساب سهرقمی})$$

۲. اعداد شرطی ماتریسهای زیر را نسبت به $\|\cdot\|_\infty$ حساب کنید:

$$T = \begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix} \quad (T)$$

$$P = \begin{bmatrix} 3.9 & 1.6 \\ 6.8 & 2.9 \end{bmatrix} \quad (P)$$

با (پ) مقایسه کنید.

$$Q = \begin{bmatrix} 1.003 & 58.09 \\ 5.550 & 321.8 \end{bmatrix} \quad (Q)$$

با (ب) مقایسه کنید.

۳. نشان دهید که اگر B منفرد باشد،

$$\frac{1}{K(A)} \leq \frac{\|A - B\|}{\|A\|}.$$

[راهنمایی. برداری مانند $x \neq 0$ ، با $\|x\| = 1$ ، هست بطوری که $Bx = 0$. تخمین

$$\|Ax\| \geq \frac{\|x\|}{\|A^{-1}\|} \quad \text{بدست آورید.}]$$

۴. با استفاده از تمرین ۳، اعداد شرطی را برای ماتریسهای زیر تخمین بزنید:

$$T = \begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix} \quad (T)$$

با (ت) مقایسه کنید.

$$\cdot \begin{bmatrix} 3.9 & 1.6 \\ 6.8 & 2.9 \end{bmatrix} \quad (\text{ب})$$

با ۱ (پ) مقایسه کنید.

۵. دستگاه خطی

$$\begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3.0001 \end{bmatrix}$$

دارای جواب $(1, 1)$ است. A را به طور جزئی به

$$\begin{bmatrix} 1 & 2 \\ .9999 & 2 \end{bmatrix}$$

تغییر دهید و دستگاه خطی زیر را بررسی کنید.

$$\begin{bmatrix} 1 & 2 \\ .9999 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3.00001 \end{bmatrix}$$

جواب جدید \bar{x} را با حساب هفت رقمی محاسبه و خطای واقعی را با تخمین (۱۶۰۸)

مقایسه کنید. آیا A بدوضع است؟

۶. دستگاه خطی $Ax = b$ داده شده با

$$\begin{bmatrix} 1 & 2 \\ 1.00001 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3.0001 \end{bmatrix}$$

دارای جواب $(1, 1)$ است. جواب دستگاه پیریشیده^۲ زیر را، با استفاده از حساب

هفت رقمی، پیدا کنید:

$$\begin{bmatrix} 1 & 2 \\ 1.000011 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3.00001 \\ 3.00003 \end{bmatrix}$$

و خطای واقعی را با تخمین (۱۶۰۸) مقایسه نمایید. آیا A بدوضع است؟

۷. با استفاده از کامپیوتر، دستگاه خطی ۱ (ت) را نسبت به تخمین (۱۸۰۸) تحلیل

کنید. آیا $f(n)$ به n یا به $1.01(n^3 + 3n^2)$ نزدیکتر است؟

۸. ماتریس $n \times n$ ، $H^{(n)}$ تعریف شده با

$$H_{ij}^{(n)} = \frac{1}{i+j-1}, \quad 1 \leq i, j \leq n$$

ماتریسی بدوضع است که در حل معادلات نرمال برای ضرایب چندجمله‌ای کمترین

مربعات (ر.ک. مثال ۱ بخش ۳.۴) ظاهر می شود.

(ت) نشان دهید که

$$[H^{(4)}]^{-1} = \begin{bmatrix} 16 & -120 & 240 & -140 \\ -120 & 1200 & -2700 & 1680 \\ 240 & -2700 & 6480 & -4200 \\ -140 & 1680 & -4200 & 2800 \end{bmatrix}$$

و $K(H^{(4)})$ را نسبت به $\|\cdot\|_{\infty}$ حساب کنید.

(ب) نشان دهید

$$[H^{(5)}]^{-1} = \begin{bmatrix} 52 & -300 & 1050 & -1400 & 630 \\ -300 & 4800 & -18900 & 26880 & -12600 \\ 1050 & -18900 & 79380 & -117600 & 56700 \\ -1400 & 26880 & -117600 & 179200 & -88200 \\ 630 & -12600 & 56700 & -88200 & 44100 \end{bmatrix}$$

و $K(H^{(5)})$ را نسبت به $\|\cdot\|_1$ حساب کنید.

(پ) دستگاه خطی

$$H^{(4)} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

را با استفاده از حساب سمرقمی حل کنید، و خطای واقعی را با خطای حساب شده در (۱۸.۸) مقایسه کنید.

۳.۸ روشهای تکراری برای حل دستگاههای خطی

یک روش تکراری برای حل دستگاه خطی $n \times n$ ، $Ax = b$ ، با یک تقریب اولیه $x^{(0)}$ برای جواب x شروع می کند، و دنباله ای از بردارهای $\{x^{(k)}\}_{k=0}^{\infty}$ تولید می کند که به x همگراست. اکثر این روشهای تکراری شامل فرایندی است که دستگاه $Ax = b$ را به دستگاهی معادل آن به شکل $x = Tx + c$ تبدیل می کند. با انتخاب بردار اولیه $x^{(0)}$ ، دنباله بردارهای تقریب به جواب با محاسبه

$$x^{(k)} = Tx^{(k-1)} + c \quad (19.8)$$

بازای $k = 1, 2, 3, \dots$ تولید می‌شود. چنین روندی باید یادآور تکرار نقطه ثابت، که در فصل ۲ مطالعه شد، باشد.

در عمل، روشهای تکراری بندرت برای حل دستگاههای خطی با بعد کوچک بکار می‌روند، زیرا زمان لازم برای دقت کافی بیش از زمان لازم برای روشهای مستقیم نظیر روش حذف گاوسی است. این روشها می‌توانند برای دستگاههای بزرگ معینی بخصوص آنهایی که نوع ویژه‌ای دارند که در حل عددی مسائل مقدار مرزی و معادلات دیفرانسیل جزئی رخ می‌دهند، از نظر نیاز به حافظه و زمان کامپیوتر موثر باشند.

دستگاه خطی

$$10x_1 - x_2 + 2x_3 = 6,$$

$$+ 11x_2 - x_3 + 3x_4 = 25,$$

$$x_2 + 10x_3 - x_4 = -11,$$

$$3x_2 - x_3 + 8x_4 = 15,$$

دارای جواب $x = (1, 2, -1, 1)'$ است. برای تبدیل $Ax = b$ به شکل $x = Tx + c$ ، معادله (E_i) را نسبت به x_i ، بازای $i = 1, 2, 3, 4$ ، حل و بدست می‌آوریم

$$x_1 = \frac{1}{10}x_2 - \frac{1}{5}x_3 + \frac{3}{5},$$

$$x_2 = \frac{1}{11}x_1 + \frac{1}{11}x_3 - \frac{3}{11}x_4 + \frac{25}{11},$$

$$x_3 = -\frac{1}{5}x_1 + \frac{1}{10}x_2 + \frac{1}{10}x_4 - \frac{11}{10},$$

$$x_4 = -\frac{3}{8}x_2 + \frac{1}{8}x_3 + \frac{15}{8}$$

دراین مثال،

$$T = \begin{bmatrix} 0 & \frac{1}{10} & -\frac{1}{5} & 0 \\ \frac{1}{11} & 0 & \frac{1}{11} & -\frac{3}{11} \\ -\frac{1}{5} & \frac{1}{10} & 0 & \frac{1}{10} \\ 0 & -\frac{3}{8} & \frac{1}{8} & 0 \end{bmatrix} \quad \text{و} \quad c = \begin{bmatrix} \frac{3}{5} \\ \frac{25}{11} \\ -\frac{11}{10} \\ \frac{15}{8} \end{bmatrix}$$

برای یک تقریب اولیه قرار می دهیم $x^{(0)} = (0, 0, 0, 0)$ و $x^{(1)}$ را تولید می کنیم:

$$x_1^{(1)} = \frac{1}{10}x_2^{(0)} - \frac{1}{3}x_3^{(0)} + \frac{3}{5} = 0.6000,$$

$$x_2^{(1)} = \frac{1}{11}x_1^{(0)} + \frac{1}{11}x_3^{(0)} - \frac{3}{11}x_4^{(0)} + \frac{25}{11} = 2.2727,$$

$$x_3^{(1)} = -\frac{1}{5}x_1^{(0)} + \frac{1}{10}x_2^{(0)} + \frac{1}{10}x_4^{(0)} - \frac{11}{10} = -1.1000,$$

$$x_4^{(1)} = -\frac{3}{8}x_2^{(0)} + \frac{1}{8}x_3^{(0)} + \frac{15}{8} = 1.8750.$$

تکرارهای بعدی $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, x_4^{(k)})$ به روش مشابهی تولید می شوند و در جدول (۱۰۸) ارائه شده اند.

جدول ۱۰۸

k	0	1	2	3	4	5	6	7	8	9	10
$x_1^{(k)}$	0.0000	.6000	1.0473	.9326	1.0152	.9890	1.0032	.9981	1.0006	.9997	1.0001
$x_2^{(k)}$	0.0000	2.2727	1.7159	2.0533	1.9537	2.0114	1.9922	2.0023	1.9987	2.0004	1.9998
$x_3^{(k)}$	0.0000	-1.1000	-.8052	-1.0493	-.9681	-1.0103	-.9945	-1.0020	-.9990	-1.0004	-.9998
$x_4^{(k)}$	0.0000	1.8750	.8852	1.1309	.9739	1.0214	.9944	1.0036	.9989	1.0006	.9998

تصمیم توقف بعد از 10 تکرار بر این واقعیت استوار است که

$$\frac{\|x^{(10)} - x^{(9)}\|_{\infty}}{\|x^{(10)}\|_{\infty}} = \frac{8.0 \times 10^{-4}}{1.9998} < 10^{-3}$$

که نشان می دهد $x^{(10)}$ احتمالاً "بحد کافی دقیق است". در واقع،

$$\|x^{(10)} - x\|_{\infty} = .0002,$$

لذا، $x^{(10)}$ تقریب دقیقی است.

روش مثال ۱ روش تکراری ژاکوبی^۱ نامیده می شود. این روش شامل حل معادله

$Ax = b$ م نسبت به x_i ، یعنی

$$(۲۰۰۸) \quad x_i = \sum_{\substack{j=1 \\ j \neq i}}^n \left(-\frac{a_{ij}x_j}{a_{ii}} \right) + \frac{b_i}{a_{ii}}, \quad i = 1, 2, \dots, n \text{ برای}$$

و تولید هر $x_i^{(k)}$ از $x_i^{(k-1)}$ ، به‌مازای $k \geq 1$ ، بوسیله

$$(21.8) \quad x_i^{(k)} = \frac{\sum_{j \neq i}^n (-a_{ij} x_j^{(k-1)}) + b_i}{a_{ii}}, \quad i = 1, 2, \dots, n \text{ به‌مازای}$$

می‌باشد.

این روش را می‌توان با تجزیه A به اعضای قطری و غیرقطری آن به شکل $x^{(k)} = Tx^{(k-1)} + c$ نوشت. برای ملاحظه این مطلب، فرض کنیم D ماتریسی قطری باشد که قطر آن همانند قطر A باشد، L - قسمت پایین مثلثی A و U - قسمت بالامثلثی A باشد. با این نمادگذاری، A به صورت زیر تجزیه می‌شود:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix} - \begin{bmatrix} 0 & \cdots & 0 \\ -a_{21} & \ddots & \vdots \\ \vdots & \ddots & \ddots \\ -a_{n1} & \cdots & -a_{n,n-1} & 0 \end{bmatrix} - \begin{bmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & & -a_{n-1,n} \\ 0 & & & 0 \end{bmatrix}$$

$$= D - L - U.$$

لذا، معادله $Ax = b$ یا $(D - L - U)x = b$ به

$$Dx = (L + U)x + b,$$

تبدیل می‌شود و سرانجام

$$(22.8) \quad x = D^{-1}(L + U)x + D^{-1}b.$$

در روش تکراری ژاکوبی، این مطلب نتیجه می‌دهد

$$(23.8) \quad x^{(k)} = D^{-1}(L + U)x^{(k-1)} + D^{-1}b, \quad k = 1, 2, \dots$$

در عمل، معادله (۲۱.۸) در محاسبات بکار می‌رود، و (۲۳.۸) برای مقاصد نظری در نظر گرفته می‌شود.

الگوریتم تکراری ژاکوبی ۲.۸

برای حل $Ax = b$ یک تقریب اولیه مانند $x^{(0)} = (x_1^{(0)}, \dots, x_n^{(0)})'$ انتخاب می‌کنیم.

مرحله ۱. قرار می دهیم $k = 1$.

مرحله ۲. به ازای $i = 1, 2, \dots, n$ ، محاسبه می کنیم

$$x_i^{(k)} = \frac{-\sum_{j \neq i}^n (a_{ij} x_j^{(k-1)}) + b_i}{a_{ii}}.$$

مرحله ۳. اگر $x^{(k)}$ بحد کافی دقیق باشد، به مرحله ۴ می رویم.

اگر $x^{(k)}$ بحد کافی دقیق نباشد، 1 را به k افزوده به مرحله ۲ می رویم.

مرحله ۴. روند تمام است.

این الگوریتم مستلزم آن است که به ازای $i = 1, 2, \dots, n$ ، $a_{ii} \neq 0$ ، چه در غیر این صورت مرحله ۲ انجام نخواهد شد. اگر چنین نباشد، می توان، به شرط نامنفرد بودن دستگاه، تغییری در ترتیب معادلات داد بطوری که هیچ a_{ii} صفر نباشد. توصیه می شود که معادلات طوری مرتب شوند که a_{ii} ها بزرگترین مقدار ممکن را داشته باشند تا همگرایی را سرعت بخشند. (ر.ک. تمرین ۱ (T) و ۱ (پ).) این موضوع بعداً با تفصیل بیشتر بررسی می شود.

یک معیار توقف ممکن برای مرحله ۳ آن است که تکرار کنیم تا اینکه

$$\frac{\|x^{(k)} - x^{(k-1)}\|}{\|x^{(k)}\|} \text{ یا } \|x^{(k)} - x^{(k-1)}\|$$

کوچکتر از ϵ معینی شود. برای این منظور هر نرم مناسبی را می توان بکار برد؛ نرم l_x معمولترین است. چون روش فوق ممکن است همیشه همگرا نباشد، عاقلانه است که کران بالایی نیز برای تعداد تکرارهایی که باید انجام شود در نظر گرفت.

با تحلیل مرحله ۲ پیشرفتی احتمالی در الگوریتم ۲۰.۸ توصیه می شود. برای محاسبه $x_i^{(k)}$ ، به ازای هر $i > 1$ ، مولفه های $x^{(k-1)}$ بکار می روند. چون $x_1^{(k)}, \dots, x_{i-1}^{(k)}$ قبلاً محاسبه شده اند و فرضاً تقریبهای بهتری برای جوابهای واقعی x_1, \dots, x_{i-1} از $x_1^{(k-1)}, \dots, x_{i-1}^{(k-1)}$ هستند، معقول بنظر می رسد که $x_i^{(k)}$ را به جای (۲۱.۸)، با بکار بردن آخرین مقادیر محاسبه شده حساب کنیم؛ یعنی،

$$(24.8) x_i^{(k)} = \frac{-\sum_{j=1}^{i-1} (a_{ij} x_j^{(k)}) - \sum_{j=i+1}^n (a_{ij} x_j^{(k-1)}) + b_i}{a_{ii}}, i = 1, 2, \dots, n$$

مثالی که این روند را روشن می کند به قرار ذیل است.

مثال ۲. دستگاه خطی

$$\begin{aligned} 10x_1 - x_2 + 2x_3 &= 6, \\ -x_1 + 11x_2 - x_3 + 3x_4 &= 25, \\ 2x_1 - x_2 + 10x_3 - x_4 &= -11, \\ 3x_2 - x_3 + 8x_4 &= 15, \end{aligned}$$

در مثال ۱ به روش تکراری ژاکوبی حل شد. با استفاده از (۲۴.۰۸) در الگوریتم ۲.۰۸ معادلات زیر بدست می‌آیند که به ازای هر $k = 1, 2, \dots$ بکار روند:

$$\begin{aligned} x_1^{(k)} &= \frac{1}{10}x_2^{(k-1)} - \frac{1}{5}x_3^{(k-1)} + \frac{3}{5}, \\ x_2^{(k)} &= \frac{1}{11}x_1^{(k)} + \frac{1}{11}x_3^{(k-1)} - \frac{3}{11}x_4^{(k-1)} + \frac{25}{11}, \\ x_3^{(k)} &= -\frac{1}{5}x_1^{(k)} + \frac{1}{10}x_2^{(k)} + \frac{1}{10}x_4^{(k-1)} - \frac{11}{10}, \\ x_4^{(k)} &= -\frac{3}{8}x_2^{(k)} + \frac{1}{8}x_3^{(k)} + \frac{15}{8}. \end{aligned}$$

با فرض $x^{(0)} = (0, 0, 0, 0)$ ، بردارهای تکراری در جدول ۲.۰۸ تولید می‌شوند.

جدول ۲.۰۸

k	0	1	2	3	4	5
$x_1^{(k)}$	0.0000	.6000	1.030	1.0065	1.0009	1.0001
$x_2^{(k)}$	0.0000	2.3272	2.037	2.0036	2.0003	2.0000
$x_3^{(k)}$	0.0000	-.9873	-1.014	-1.0025	-1.0003	-1.0000
$x_4^{(k)}$	0.0000	.8789	.9844	.9983	.9999	1.0000

چون

$$\frac{\|x^{(5)} - x^{(4)}\|_{\infty}}{\|x^{(5)}\|_{\infty}} = \frac{.0008}{2.000} = 4 \times 10^{-4}$$

کوچک است، $x^{(5)}$ به عنوان تقریب معقولی به جواب قابل قبول است. جالب است توجه شود که روش ژاکوبی برای همین تقریب در این مثال به دو برابر تکرار نیاز دارد.

تکنیک ارائه شده در مثال ۲ روش تکراری گاوس-سیدل^۱ نامیده می‌شود. برای

نوشتن این روش به شکل ماتریسی (۱۹.۸)، طرفین معادله (۲۴.۸) را در a_{ii} ضرب کرده و همه جملات تکراری k ام را دسته بندی می‌کنیم تا به‌ازای هر $i = 1, 2, \dots, n$ بدست آید

$$a_{i1}x_1^{(k)} + a_{i2}x_2^{(k)} + \dots + a_{in}x_n^{(k)} \\ = -a_{i,i+1}x_{i+1}^{(k-1)} - \dots - a_{in}x_n^{(k-1)} + b_i$$

با نوشتن تمام n معادله خواهیم داشت

$$\begin{aligned} a_{11}x_1^{(k)} &= -a_{12}x_2^{(k-1)} - a_{13}x_3^{(k-1)} - \dots - a_{1n}x_n^{(k-1)} + b_1, \\ a_{21}x_1^{(k)} + a_{22}x_2^{(k)} &= -a_{23}x_3^{(k-1)} - \dots - a_{2n}x_n^{(k-1)} + b_2, \\ \vdots & \\ a_{n1}x_1^{(k)} + a_{n2}x_2^{(k)} + \dots + a_{nn}x_n^{(k)} &= b_n \end{aligned}$$

و نتیجه می‌دهد که، به شکل ماتریسی، روش گاوس - سیدل را می‌توان با

$$(25.8) \quad (D - L)x^{(k)} = Ux^{(k-1)} + b$$

یا

$$(26.8) \quad x^{(k)} = (D - L)^{-1}Ux^{(k-1)} + (D - L)^{-1}b, \quad k = 1, 2, \dots$$

نمایش داد. برای آنکه ماتریس پایین مثلثی $D - L$ نامفرد باشد لازم و کافی است که به‌ازای هر $i = 1, 2, \dots, n$ ، $a_{ii} \neq 0$. حال روش گاوس - سیدل به شکل الگوریتمی ارائه می‌شود.

الگوریتم تکراری گاوس - سیدل ۳.۸

برای حل $Ax = b$ ، تقریب اولیه $x^{(0)} = (x_1^{(0)}, \dots, x_n^{(0)})'$ را انتخاب می‌کنیم.

مرحله ۱. قرار می‌دهیم $k = 1$.

مرحله ۲. به‌ازای هر $i = 1, 2, \dots, n$ ،

$$x_i^{(k)} = \frac{-\sum_{j=1}^{i-1} (a_{ij}x_j^{(k)}) - \sum_{j=i+1}^n (a_{ij}x_j^{(k-1)}) + b_i}{a_{ii}}$$

را محاسبه می‌کنیم.

مرحله ۳. اگر $x^{(k)}$ به حد کافی دقیق باشد، به مرحله ۴ می‌رویم.

اگر $x^{(k)}$ به حد کافی دقیق نباشد، به مرحله ۲ می‌رویم.

مرحله ۴. روند تمام است.

توضیحات مربوط به الگوریتم ۲۰.۸ در اینجا نیز اعمال می شود. بنظر می رسد که نتایج مثال ۲ نشانگر آن است که روش گاوس - سیدل بهتر از روش ژاکوبی است. این حکم عموماً درست نیست. در واقع، دستگاههایی خطی وجود دارند که برای آنها روش ژاکوبی همگراست ولی روش گاوس - سیدل همگرا نیست و بعکس (ز. ک. و ارگا. [۶۵]، صفحه ۷۴). برای مطالعه همگرایی روشهای تکراری، فرمول کلی

$$x^{(k)} = Tx^{(k-1)} + c, \quad k = 1, 2, \dots$$

که در آن $x^{(0)}$ دلخواه است، را در نظر می گیریم. ابتدا لازم است لم زیر بررسی شود.

لم ۲۲.۸. هرگاه شعاع طیفی $\rho(T) < 1$ در $\rho(T) < 1$ صدق کند. آنگاه $(I - T)^{-1}$ وجود دارد و

$$(I - T)^{-1} = I + T + T^2 + \dots \quad (27.8)$$

برهان. به ازای هر مقدار ویژه λ از T ، $1 - \lambda$ یک مقدار ویژه $I - T$ است. چون $|\lambda| \leq \rho(T) < 1$ ، نتیجه می گیریم که هیچ مقدار ویژه $I - T$ نمی تواند صفر باشد و، در نتیجه، $I - T$ نامنفرد است.

فرض کنیم $S_m = I + T + T^2 + \dots + T^m$ در این صورت،

$$(I - T)S_m = I - T^{m+1}$$

و، چون T همگراست، بنابر قضیه ۱۷.۸ (صفحه ۵۱۳)،

$$\lim_{m \rightarrow \infty} (I - T)S_m = \lim_{m \rightarrow \infty} (I - T^{m+1}) = I.$$

بنابراین، $\lim_{m \rightarrow \infty} S_m = (I - T)^{-1}$.

قضیه ۲۳.۸. دنباله $\{x^{(k)}\}_{k=0}^{\infty}$ ، که به ازای هر $k \geq 1$ با

(۲۸.۸)

$$x^{(k)} = Tx^{(k-1)} + c$$

تعریف شده است و $c \neq 0$ ، به ازای هر $x^{(0)} \in R^n$ به بردار x همگراست اگر و فقط اگر

$$\rho(T) < 1$$

برهان. بنابر (۲۷.۸)،

$$\begin{aligned} \mathbf{x}^{(k)} &= T\mathbf{x}^{(k-1)} + \mathbf{c} \\ &= T(T\mathbf{x}^{(k-2)} + \mathbf{c}) + \mathbf{c} \\ &= T^2\mathbf{x}^{(k-2)} + (T + I)\mathbf{c} \\ &\vdots \\ &= T^k\mathbf{x}^{(0)} + (T^{k-1} + \dots + T + I)\mathbf{c} \end{aligned}$$

(۲۹۰.۸)

با فرض $\rho(T) < 1$ ، می‌توان از قضیه^{۱۷.۸} (۱۷.۸) و لم^{۲۲.۸} استفاده کرده بدست آورد

$$\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \lim_{k \rightarrow \infty} T^k\mathbf{x}^{(0)} + \lim_{k \rightarrow \infty} \left(\sum_{j=0}^{k-1} T^j \right) \mathbf{c}$$

(۳۰۰.۸)

$$\begin{aligned} &= 0 \cdot \mathbf{x}^{(0)} + (I - T)^{-1}\mathbf{c} \\ &= (I - T)^{-1}\mathbf{c}. \end{aligned}$$

برای اثبات عکس، فرض کنیم $\{\mathbf{x}^{(k)}\}$ به‌ازای هر $\mathbf{x}^{(0)}$ به \mathbf{x} همگرا باشد. از معادله^{۲۸.۸}

$$\mathbf{x} = T\mathbf{x} + \mathbf{c} \quad ; \quad \text{لذا، به‌ازای هر } k,$$

$$\mathbf{x} - \mathbf{x}^{(k)} = T(\mathbf{x} - \mathbf{x}^{(k-1)}) = \dots = T^k(\mathbf{x} - \mathbf{x}^{(0)})$$

چون $\mathbf{x}^{(0)}$ دلخواه است، T باید همگرا باشد، و، بنا بر قضیه^{۱۷.۸}، این نتیجه می‌دهد

$$\rho(T) < 1.$$

نتیجه^{۲۴.۸}. هرگاه به‌ازای نرمی ماتریسی $\|T\| < 1$ ، آنگاه دنباله^{۲۴.۸} $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ در

(۲۸.۸) به‌ازای هر $\mathbf{x}^{(0)} \in R^n$ به یک بردار $\mathbf{x} \in R^n$ همگراست و گرانه‌های خطای زیربرقرارند:

(۳۱.۸)

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \|T\|^k \|\mathbf{x}^{(0)} - \mathbf{x}\|;$$

(۳۲.۸)

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \frac{\|T\|^k}{1 - \|T\|} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|.$$

برهان این نتیجه در تمرین ۲ بررسی می‌شود.

برای اعمال نتایج فوق روی روشهای ژاکوبی و گاوس-سیدل، مجدداً "خاطر نشان

می‌شود که ماتریسهای تکرار برای روش ژاکوبی، یعنی T_j ، و برای روش گاوس-سیدل،

یعنی T_g ، با

$$T_j = D^{-1}(L + U),$$

$$T_g = (D - L)^{-1}U.$$

داده می‌شوند. در صورتی که $\rho(T_j)$ یا $\rho(T_g)$ کوچکتر از واحد باشد، واضح است که دنباله^{۲۴.۸}

$\{x^{(k)}\}_{k=0}^{\infty}$ به جواب x از $Ax = b$ همگرا خواهد بود. مثلاً، در طرح ژاکوبی،

$$x^{(k)} = D^{-1}(L + U)x^{(k-1)} + D^{-1}b,$$

و، اگر $\{x^{(k)}\}_{k=0}^{\infty}$ همگرا به x باشد،

$$x = D^{-1}(L + U)x + D^{-1}b,$$

که نتیجه می‌دهد

$$Dx = (L + U)x + b$$

و

$$(D - L - U)x = b.$$

چون $D - L - U = A$ ، در $Ax = b$ صدق می‌کند. حال شرایط کافی، که آزمایش آنها ساده است، ارائه می‌شود.

قضیه ۲۵۰۸. هرگاه A اکیدا "قطر غالب" باشد، آنگاه روشهای ژاکوبی و گاوس-سیدل هر دو دنباله $\{x^{(k)}\}_{k=0}^{\infty}$ را می‌دهند که به‌ازای هر $x^{(0)}$ به جواب $Ax = b$ همگرا می‌باشد.

برای اثبات همگرایی روش ژاکوبی، ر.ک. تمرین ۳؛ و برای طرح گاوس-سیدل، ر.ک. اورتگا [۴۶]، صفحه ۱۲۵.

رابطه سرعت همگرایی با شعاع طیفی ماتریس تکرار T را می‌توان از معادله (۳۱۰۸) ملاحظه کرد. چون (۳۱۰۸) به‌ازای هر نرم ماتریسی طبیعی برقرار است، از حکم بعد از قضیه (۱۵۰۸) نتیجه می‌گیریم که

$$\|x^{(k)} - x\| \approx \rho(T)^k \|x^{(0)} - x\|. \quad (۳۳۰۸)$$

فرض کنیم $\rho(T) < 1$ و اینکه قرار است $x^{(0)} = 0$ در یک روش تکراری جهت تقریب x با حداکثر خطای نسبی 10^{-l} بکار رود. بنا بر تخمین (۳۳۰۸)، بعد از k تکرار خطای نسبی

تقریباً " $\rho(T)^k$ است؛ لذا، اگر $\rho(T)^k \leq 10^{-l}$ یا $k \geq \frac{l}{-\log_{10} \rho(T)}$ ، دقت 10^{-l}

انتظار می‌رود. بنابراین، برای یک دستگاه خاص $Ax = b$ مطلوب آن است که تکنیک تکراری با حداقل $\rho(T) < 1$ انتخاب شود. بطور کلی، معلوم نیست کدامیک از دو تکنیک، ژاکوبی یا گاوس-سیدل، باید بکار رود. معیناً، در حالت خاص زیر جواب معلوم است.

قضیه ۲۶۰۸ (اشتاین^۱ - روزنبرگ^۲). هرگاه به ازای هر z ، $a_{ij} \leq 0$ ، $i \neq j$ و ، به ازای هر $i = 1, 2, \dots, n$ ، $a_{ii} > 0$ ، نگاه یک و فقط یکی از احکام زیر برقرارند :

$$(A) \quad 0 < \rho(T_0) < \rho(T_j) < 1$$

$$(B) \quad 1 < \rho(T_j) < \rho(T_0)$$

$$(C) \quad \rho(T_j) = \rho(T_0) = 0$$

$$(D) \quad \rho(T_j) = \rho(T_0) = 1$$

اثبات این قضیه در اینجا ارائه نمی شود ، اما می توان آن را در یانگ^۳ [۷۰] ، صفحات ۱۲۷ - ۱۲۰ ، یافت .

برای حالت خاص تشریح شده در قضیه ۲۶۰۸ ، ملاحظه می کنیم که وقتی یک روش همگرا باشد ، هر دو همگرایند ، و روش گاوس - سیدل سریعتر از روش ژاکوبی است . چون سرعت همگرایی یک رونده شعاع طیفی ماتریس مربوط به آن روش بستگی دارد ، یک راه انتخاب روند که منجر به همگرایی سریع شود آن است که روشی انتخاب گردد که ماتریس مربوط به آن کوچکترین شعاع طیفی را داشته باشد .

در روندهایی چون روشهای ژاکوبی و گاوس - سیدل یک بردار مانده به محاسبه هر مولفه تقریبی بردار جواب مربوط می شود . اگر فرض کنیم

$$r_i^{(k)} = (r_{1i}^{(k)}, r_{2i}^{(k)}, \dots, r_{ni}^{(k)})^t$$

نمایانگر بردار مانده برای روش گاوس - سیدل متناظر با بردار جواب تقریبی

$$(x_1^{(k)}, x_2^{(k)}, \dots, x_{i-1}^{(k)}, x_i^{(k-1)}, \dots, x_n^{(k-1)})^t$$

باشد ، مولفه m $r_i^{(k)}$ عبارت خواهد بود از

$$r_{mi}^{(k)} = b_m - \sum_{j=1}^{i-1} a_{mj} x_j^{(k)} - \sum_{j=i}^n a_{mj} x_j^{(k-1)} \quad m = 1, 2, \dots, n$$

$$= b_m - \sum_{j=1}^{i-1} a_{mj} x_j^{(k)} - \sum_{j=i+1}^n a_{mj} x_j^{(k-1)} - a_{mi} x_i^{(k-1)}$$

بالاخص ، مولفه i $r_i^{(k)}$ عبارت است از

$$r_{ii}^{(k)} = b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k)} - \sum_{j=i+1}^n a_{ij} x_j^{(k-1)} - a_{ii} x_i^{(k-1)}$$

در نتیجه ،

$$(۳۴.۸) \quad a_{ii}x_i^{(k-1)} + r_{ii}^{(k)} = b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^n a_{ij}x_j^{(k-1)}.$$

بخاطر آورید که در روش گاوس - سیدل

$$x_i^{(k)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^n a_{ij}x_j^{(k-1)} \right]$$

لذا ، (۳۴.۸) را می توان نوشت

$$a_{ii}x_i^{(k-1)} + r_{ii}^{(k)} = a_{ii}x_i^{(k)}$$

یا

$$(۳۵.۸) \quad x_i^{(k)} = x_i^{(k-1)} + \frac{r_{ii}^{(k)}}{a_{ii}}.$$

روش گاوس - سیدل برحسب بردارهای مانده با الزام بر اینکه ، در هر مرحله ،

مختص i م بردار مانده $r_{i+1}^{(k)}$ صفر باشد بدست می آید . (ر. ک. تمرین ۰۷)

معهدا ، تقلیل یک مختص بردار مانده به صفر الزاما " کارآترین راه تقلیل نرم بردار

نیست . در واقع ، با تغییر روند گاوس - سیدل ، که با معادله (۳۵.۸) داده شد ، به

$$(۳۶.۸) \quad x_i^{(k)} = x_i^{(k-1)} + \omega \frac{r_{ii}^{(k)}}{a_{ii}}$$

برای انتخابهای ویژه ای از ω ی مثبت منجر به همگرایی سریعتری خواهد شد .

روشهای مشتمل بر معادله (۳۶.۸) روشهای تخفیف نامیده می شوند . بـبرای

انتخابهای $0 < \omega < 1$ روند را روشهای تحت - تخفیف نامند و می توان آنها را جهت

بدست آوردن همگرایی برای دستگاههایی که به روش گاوس - سیدل همگرا نیستند بکار

برد . برای انتخابهای $\omega > 1$ روندها را روشهای فوق تخفیف نامند ، و این روندها که

می توان برای سرعت بخشیدن به همگرایی دستگاههایی که به روش گاوس - سیدل همگرا

هستند بکار برد را بررسی خواهیم کرد. این روشها اغلب برای فوق - تخفیف متوالی به **SOR**

مختصر شده اند ، و برای حل دستگاههای خطی که در حل عددی معادلات دیفرانسیل

جزئی معینی ظاهر می شوند مفیدند .

قبل از تشریح مزیتهای روش **SOR** متذکر شویم که معادله (۳۶.۸) را می توان

با استفاده از (۳۴.۸) ، برای مقاصد محاسبه ای به صورت زیر نوشت :

$$(۳۷.۸) \quad x_i^{(k)} = (1 - \omega)x_i^{(k-1)} + \frac{\omega}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^n a_{ij}x_j^{(k-1)} \right].$$

مثال ۳. دستگاه خطی $Ax = b$ که به صورت زیر داده شده است

$$4x_1 + 3x_2 = 24,$$

$$3x_1 + 4x_2 - x_3 = 30,$$

$$-x_2 + 4x_3 = -24,$$

دارای جواب $(3, 4, -5)$ است. روش گاوس-سیدل و روش SOR با $\omega = 1.25$ ، با

استفاده از $x^{(0)} = (1, 1, 1)^T$ برای هر دو روش، برای حل این دستگاه بکار خواهد رفت.

برای روش گاوس-سیدل معادلات عبارتند از

$$x_1^{(k)} = -.75x_2^{(k-1)} + 6,$$

$$x_2^{(k)} = -.75x_1^{(k)} + .25x_3^{(k-1)} + 7.5,$$

$$x_3^{(k)} = .25x_2^{(k)} - 6, \quad \text{for each } k = 1, 2, \dots,$$

و برای روش SOR با $\omega = 1.25$ معادلات عبارتند از

$$x_1^{(k)} = -.25x_1^{(k-1)} - .9375x_2^{(k-1)} + 7.5,$$

$$x_2^{(k)} = -.9375x_1^{(k)} - .25x_2^{(k-1)} + .3125x_3^{(k-1)} + 9.375,$$

$$x_3^{(k)} = .3125x_2^{(k)} - .25x_3^{(k-1)} - 7.5.$$

هفت تکرار اول برای هر روش در جدول ۳.۸ و ۴.۸ لیست شده است.

جدول ۳.۸ گاوس-سیدل

k	0	1	2	3	4	5	6	7
$x_1^{(k)}$	1	5.250000	3.1406250	3.0878906	3.0549316	3.0343323	3.0214577	3.01341
$x_2^{(k)}$	1	3.812500	3.8828125	3.9267578	3.9542236	3.9713898	3.9821186	3.98882
$x_3^{(k)}$	1	-5.046875	-5.0292969	-5.0183105	-5.0114441	-5.0071526	-5.0044703	-5.00279

جدول ۴.۸ SOR با $\omega = 1.25$

k	0	1	2	3	4	5	6	7
$x_1^{(k)}$	1	6.312500	2.6223145	3.1333027	2.9570512	3.0037211	2.9963276	3.00004
$x_2^{(k)}$	1	3.5195313	3.9585266	4.0102646	4.0074838	4.0029250	4.0009262	4.00025
$x_3^{(k)}$	1	-6.6501465	-4.6004238	-5.0966863	-4.9734897	-5.0057135	-4.9982822	-5.00034

برای آنکه تکرارها تا هفت رقم اعشار دقیق باشند روش گاوس - سیدل ۳۴ تکرار لازم دارد، در مقایسه با ۱۴ تکرار برای روش فوق - تخفیف با $\omega = 1.25$ ، که پیشرفتی قابل توجه است.

سوال واضحی که مطرح می شود این است که مقدار مناسب ω چگونه انتخاب می شود. با آنکه برای یک دستگاه خطی $n \times n$ کلی جواب کامل این سوال معلوم نیست، نتایج زیر را می توان در حالت معینی بکار برد.

قضیه ۲۷۰۸ (کاهان^۱). هرگاه به ازای هر $i = 1, 2, \dots, n$ ، $a_{ii} \neq 0$ ، آنگاه $\rho(T_\omega) \geq |\omega - 1|$ ؛ بنابراین، $\rho(T_\omega) < 1$ فقط اگر $0 < \omega < 2$ ، که در آن $T_\omega = (D - \omega L)^{-1}\{(1 - \omega)D + \omega U\}$ ماتریس تکرار برای روش SOR است.

برهان این قضیه در تمرینهای ۴ و ۵ بررسی شده است. برهان دو نتیجه بعدی را می توان در اورنگا [۴۶]، ص ۱۳۳ - ۱۲۳، یافت.

قضیه ۲۸۰۸ (استروسکی^۲ - ریچ^۳). هرگاه A یک ماتریس معین مثبت بوده و $0 < \omega < 2$ ، آنگاه روش SOR برای هر انتخاب بردار جواب تقریبی اولیه $x^{(0)}$ همگرا می باشد.

قضیه ۲۹۰۸. هرگاه A معین مثبت و سه قطری باشد، آنگاه $\rho(T_\omega) = [\rho(T_j)]^2 < 1$ بهترین انتخاب ω برای روش SOR عبارت است از

$$\omega = \frac{2}{1 + \sqrt{1 - \rho(T_j)^2}},$$

و با این انتخاب ω ، $\rho(T_\omega) = \omega - 1$.

مثال ۴. در مثال ۳ ماتریس A بوسیله

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

داده شده بود. این ماتریس معین مثبت و سه قطری است، لذا، قضیه ۲۹.۸ اعمال می شود. چون

$$\begin{aligned} T_j &= D^{-1}(L + U) \\ &= \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} -3 & 0 \\ -3 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -0.75 & 0 \\ -0.75 & 0 & 0.25 \\ 0 & 0.25 & 0 \end{bmatrix}, \end{aligned}$$

داریم

$$T_j - \lambda I = \begin{bmatrix} -\lambda & -0.75 & 0 \\ -0.75 & -\lambda & 0.25 \\ 0 & 0.25 & -\lambda \end{bmatrix}$$

در نتیجه، $\det(T_j - \lambda I) = -\lambda(\lambda^2 - 0.625)$ ، لذا، $\rho(T_j) = \sqrt{0.625}$ و

$$\begin{aligned} \omega &= \frac{2}{1 + \sqrt{1 - \rho(T_j)}} = \frac{2}{1 + \sqrt{1 - 0.625}} = \frac{2}{1 + \sqrt{0.375}} \\ &\approx 1.24. \end{aligned}$$

این همگرایی نسبتاً "سریع بدست آمده در مثال ۳ را، با کاربرد $\omega = 1.25$ ، توجیه می کند.

این بخش را با یک الگوریتم برای روش SOR خاتمه می دهیم.

الگوریتم SOR ۴.۸

برای حل دستگاه خطی $Ax = b$ ، که در آن به ازای هر $i = 1, 2, \dots, n$ ، $a_{ii} \neq 0$ ، پارامتر

ω و بردار جواب اولیه $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$ را انتخاب می کنیم.

مرحله ۱. قرار می دهیم $k = 1$.

مرحله ۲. به ازای هر $i = 1, 2, \dots, n$ فرض می کنیم

$$x_i^{(k)} = (1 - \omega)x_i^{(k-1)} + \frac{\omega}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^n a_{ij}x_j^{(k-1)} \right).$$

مرحله ۳. اگر $x^{(k)}$ به قدر کافی دقیق باشد، به مرحله ۴ می رویم.

اگر $\mathbf{x}^{(k)}$ به قدر کافی دقیق نباشد، 1 را به k افزوده به مرحله ۲ می‌رویم.
مرحله ۴. روند تمام است.

معیار توقف برای روش SOR همانند روشهای ژاکوبی یا گاوس - سیدل است؛ و مجدداً توصیه می‌شود که روش توقف با کران بالایی برای k همراه باشد، زیرا این روند ممکن است واگرا باشد.

مجموعه تمرینات ۳۰۸

۱. دستگاههای خطی زیر را، در صورت امکان، به روش ژاکوبی و روش گاوس - سیدل حل کنید و سرعتهای همگرایی را مقایسه نمایید. به عنوان آزمایشی از همگرایی، تکرارها را انجام دهید تا اینکه $N > 25$ (واگرایی) یا $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2 \leq 0.1$.

$$\begin{aligned} 2x_1 - x_2 + 10x_3 &= -11, & (\bar{1}) \\ 3x_2 - x_3 + 8x_4 &= -11, \\ 10x_1 - x_2 + 2x_3 &= 6, \\ -x_1 + 11x_2 - x_3 + 3x_4 &= 25. \\ 10x_1 - x_2 &= 9, & (\bar{2}) \\ -x_1 + 10x_2 - 2x_3 &= 7, \\ -4x_2 + 10x_3 &= 6. \end{aligned}$$

$$(\bar{3}) \quad \begin{bmatrix} 4 & -1 & 0 & -1 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 \\ -1 & 0 & 0 & 4 & -1 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 \\ 0 & 0 & -1 & 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \\ 0 \\ 6 \\ -2 \\ 6 \end{bmatrix}$$

۲. قسمت (ت) $(\bar{1})$ را با آرایش مجدد معادلات جهت بهتر شدن نتایج تکرار کنید. ثابت کنید.

$$\|\mathbf{x}^{(k)} - \mathbf{x}\| \leq \|T\|^k \|\mathbf{x}^{(0)} - \mathbf{x}\|$$

$$\|\mathbf{x}^{(k)} - \mathbf{x}\| \leq \frac{\|T\|^k}{1 - \|T\|} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|,$$

که در آن T یک ماتریس $n \times n$ با $\|T\| < 1$ است و

$$x^{(k)} = Tx^{(k-1)} + c, k = 1, 2, \dots$$

با $x^{(0)}$ دلخواه، $c \in R^n$ ، و $x = Tx + c$ ، آیا این کرانه‌های خطا کافی یا مفیدند؟

این کرانه‌ها را در مورد تمرین ۱ اعمال کنید.

۳. نشان دهید هرگاه A اکیدا "قطر غالب" باشد، آنگاه $\|T_J\|_x < 1$.

۴. نشان دهید که برای روش SOR ماتریس تکرار عبارت است از

$$T_\omega = (D - \omega L)^{-1}\{(1 - \omega)D + \omega U\} \quad \text{و} \quad c = \omega(D - \omega L)^{-1}b$$

۵. قضیه ۲۷.۸ را ثابت کنید [راهنمایی]. هرگاه $\lambda_1, \dots, \lambda_n$ مقادیر ویژه T_ω باشند،

آنگاه $\det T_\omega = \prod_{i=1}^n \lambda_i$. چون $\det D^{-1} = \det(D - \omega L)^{-1}$ و دترمینان حاصل ضرب

ماتریسهای مساوی حاصل ضرب دترمینانهای عوامل آن است، نتیجه با بکار بردن

تمرین ۴ به سادگی حاصل می‌شود.]

۶. دستگاه خطی زیر را با استفاده از روش SOR با مقادیر ω ی داده شده، و همان

معیار همگرایی مذکور در تمرین ۱، حل کنید.

(آ) تمرین ۱ (ب)؛ $\omega = 1.1$ و $\omega = 0.5$.

(ب) تمرین ۱ (پ) با $\omega = 1.334$ ، $\omega = 1.95$ ، و $\omega = 0.95$.

۷. نشان دهید هرگاه $x_i^{(k)}$ چنان انتخاب شود که

$$x_i^{(k)} = x_i^{(k-1)} + \frac{r_{ii}^{(k)}}{a_{ii}},$$

آنگاه $r_{i,i+1}^{(k)} = 0$.

۴.۸ مقادیر ویژه و بردارهای ویژه

حل بسیاری از مسائل فیزیک مستلزم محاسبه، یا حداقل تخمین، مقادیر ویژه و بردارهای

ویژه متناظر با یک ماتریس مربوط به یک دستگاه معادلات خطی است. دیده‌ایم (ر.ک.

تعاریف ۱۲.۸ و ۱۳.۸ از بخش ۱.۸) که یک ماتریس $n \times n$ مانند A دقیقاً n مقدار

ویژه، نه الزاماً متمایز، دارد که صفرهای چندجمله‌ای $p(\lambda) = \det(A - \lambda I)$ می‌باشند.

به‌طور نظری، مقادیر ویژه A با پیدا کردن n ریشه $p(\lambda)$ بدست می‌آیند و بعد دستگاه

خطی مربوطه برای تعیین بردارهای ویژه نظیر حل می‌شود. در عمل، بدست آوردن $p(\lambda)$

مشکل است، و در هر حالت دیده‌ایم که تعیین ریشه‌های یک معادله n درجه n ، جز برای

مقادیر کوچک n ، مشکل است، لذا، لازم است تکنیکهای تقریبی برای پیدا کردن مقادیر ویژه یک ماتریس ساخته شود. محاسبه مقدار ویژه یک ماتریس که بزرگترین اندازه را دارد از اهمیت خاصی برخوردار است. نتیجه مهم زیر کرانی برای این مقدار ویژه می دهد.

قضیه ۳۰.۸ (قضیه دایره گرشگورین^۱). فرض کنیم A یک ماتریس $n \times n$ بوده و R_i نمایشگر دایره ای در صفحه مختلط به مرکز a_{ii} و شعاع $\sum_{j \neq i}^n |a_{ij}|$ باشد؛ یعنی،

$$R_i = \left\{ z \in \mathbb{C} \mid |z - a_{ii}| \leq \sum_{j \neq i}^n |a_{ij}| \right\}$$

که در آن \mathcal{C} برای نمایش صفحه مختلط بکار رفته است. مقادیر ویژه A در $R = \bigcup_{i=1}^n R_i$ قرار دارند و اجتماع هر k تا از این دایره که $(n - k)$ تای دیگر را قطع نکند باید دقیقاً شامل k (با شمارش مرتبه تکرار) مقدار ویژه باشد.

برهان. هرگاه λ مقدار ویژه ای از A و x یک بردار ویژه مربوطه باشد، آنگاه هر ضرب x ، یعنی $x x$ ، نیز یک بردار ویژه مربوطه λ است، زیرا $(A - \lambda I)x = 0$ ایجاب می کند که $(A - \lambda I)(x x) = x(A - \lambda I)x = 0$. در نتیجه، می توان یک بردار ویژه x مربوط به λ پیدا کرد که $\|x\|_\infty = 1$.

چون $Ax - \lambda x = 0$ ، نمایش مولفه ای معادل عبارت است از

$$(۳۸.۸) \quad \sum_{j=1}^n a_{ij} x_j = \lambda x_i, \quad i = 1, 2, \dots, n$$

و اگر k عدد صحیحی با $\|x\|_\infty = |x_k| = 1$ باشد، معادله (۳۸.۸)، با $i = k$ ، نتیجه می دهد که

$$\sum_{j=1}^n a_{kj} x_j = \lambda x_k$$

بنابراین،

$$\sum_{\substack{j=1, \\ j \neq k}}^n a_{kj} x_j = \lambda x_k - a_{kk} x_k$$

$$|(a_{kk} - \lambda)x_k| = \left| \sum_{\substack{j=1, \\ j \neq k}}^n -a_{kj}x_j \right| \leq \sum_{\substack{j=1, \\ j \neq k}}^n |a_{kj}| |x_j|.$$

چون به‌ازای هر $j = 1, 2, \dots, n$ ، $|x_k| \geq |x_j|$ ،

$$|a_{kk} - \lambda| \leq \sum_{\substack{j=1, \\ j \neq k}}^n |a_{kj}| \frac{|x_j|}{|x_k|} \leq \sum_{\substack{j=1, \\ j \neq k}}^n |a_{kj}|.$$

لذا ، $\lambda \in R_k$ ، که اولین حکم قضیه را ثابت می‌کند . قسمت دوم این قضیه به استدلال ماهرانه‌ای از پیوستگی نیاز دارد . برهانی کاملاً " خواندنی در صفحه ۴۸ اورنگا [۴۶] آمده است .

مثال ۱ . برای ماتریس

$$A = \begin{bmatrix} 4 & 1 & 1 \\ 0 & 2 & 1 \\ -2 & 0 & 9 \end{bmatrix}$$

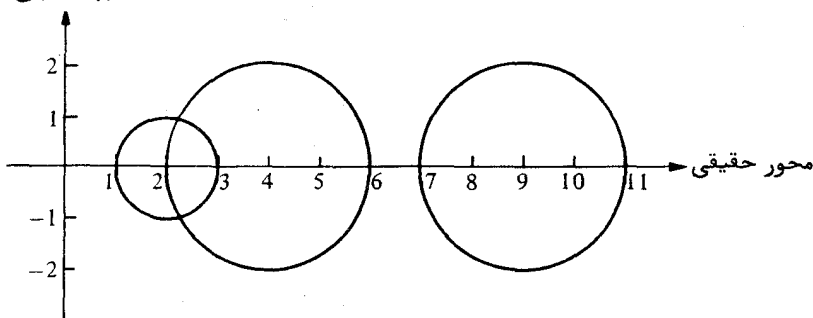
دوایر قضیه‌ء گرگورین عبارتند از (ر.ک. شکل ۴.۸)

$$R_1 = \{z \in \mathbb{C} \mid |z - 4| \leq 2\},$$

$$R_2 = \{z \in \mathbb{C} \mid |z - 2| \leq 1\},$$

$$R_3 = \{z \in \mathbb{C} \mid |z - 9| \leq 2\}.$$

محور موهومی



شکل ۴.۸

چون R_1 و R_2 از R_3 جدایند، باید دقیقاً " دو مقدار ویژه در R_2 و R_1 موجود باشد و یکی در R_3 . بعلاوه، چون $\rho(A) = \max_{1 \leq i \leq 3} |\lambda_i|$ ، $7 \leq \rho(A) \leq 11$.

پیش از بررسی نتایج بیشتری در مورد مقادیر ویژه و بردارهای ویژه، به چند تعریف و نتیجه از جبر خطی نیاز داریم . تمام نتایج کلی که تا انتهای این فصل مورد نیاز است، بخاطر راحتی مراجعه به آنها، در اینجا فهرست می شوند. برهان این نتایج داده نمی شود، لیکن آنها در اغلب کتب درسی متعارف جبر خطی موجودند (برای مثال، ر.ک. نوبل [۴۵] .)

تعریف ۳۱.۸ . فرض کنیم $\{v^{(1)}, v^{(2)}, v^{(3)}, \dots, v^{(k)}\}$ مجموعه ای از بردارها در R^n باشد. این مجموعه وابسته خطی نامیده می شود اگر اعداد $\alpha_1, \alpha_2, \dots, \alpha_k$ ، نه جملگی صفر، یافت شوند بطوری که

$$0 = \alpha_1 v^{(1)} + \alpha_2 v^{(2)} + \alpha_3 v^{(3)} + \dots + \alpha_k v^{(k)}.$$

مجموعه ای از بردارها که وابسته خطی نباشند مستقل خطی نامیده می شود .

تعریف ۳۲.۸ . هرگاه $\{v^{(1)}, v^{(2)}, v^{(3)}, \dots, v^{(n)}\}$ مجموعه ای از n بردار مستقل خطی در R^n باشد، آنگاه هر بردار $x \in R^n$ را می توان به طور منحصر بفرد، به ازای بردآیه ای از ثابتهای $\beta_1, \beta_2, \dots, \beta_n$ ، به صورت

$$x = \beta_1 v^{(1)} + \beta_2 v^{(2)} + \beta_3 v^{(3)} + \dots + \beta_n v^{(n)}$$

نوشت .

هر بردآیه از n بردار مستقل خطی در R^n یک پایه برای R^n نامیده می شود .

مثال ۲ . فرض کنیم $v^{(1)} = (1, 0, 0)^t$ ، $v^{(2)} = (-1, 1, 1)^t$ ، و $v^{(3)} = (0, 4, 2)^t$. هرگاه α_1 ، α_2 ، و α_3 اعدادی باشند که

$$0 = \alpha_1 v^{(1)} + \alpha_2 v^{(2)} + \alpha_3 v^{(3)}.$$

آنگاه

$$\begin{aligned} (0, 0, 0)^t &= \alpha_1 (1, 0, 0)^t + \alpha_2 (-1, 1, 1)^t + \alpha_3 (0, 4, 2)^t \\ &= (\alpha_1 - \alpha_2, \alpha_2 + 4\alpha_3, \alpha_2 + 2\alpha_3)^t, \end{aligned}$$

در نتیجه، $\alpha_1 - \alpha_2 = 0$ ، $\alpha_2 + 4\alpha_3 = 0$ ، و $\alpha_2 + 2\alpha_3 = 0$.

چون تنها جواب این دستگاه $\alpha_1 = \alpha_2 = \alpha_3 = 0$ است، مجموعه $\{v^{(1)}, v^{(2)}, v^{(3)}\}$ در R^3 مستقل خطی بوده و یک پایه برای R^3 است.

هر بردار $x = (x_1, x_2, x_3)$ در R^3 را می‌توان با انتخاب

$$\beta_3 = \frac{1}{2}(x_2 - x_3) \text{ و } \beta_2 = 2x_3 - x_2, \beta_1 = x_1 - x_2 + 2x_3$$

به صورت

$$x = \beta_1 v^{(1)} + \beta_2 v^{(2)} + \beta_3 v^{(3)}$$

نوشت.

تعریف ۳۳.۸. دو ماتریس $n \times n$ ، A و B را متشابه گویند اگر یک ماتریس نامنفرد S وجود داشته باشد که $A = S^{-1}BS$.

قضیه ۳۴.۸. فرض کنیم A و B ماتریسهای $n \times n$ متشابه بوده و λ یک مقدار ویژه A با بردار ویژه x مربوطه باشد. در این صورت، λ یک مقدار ویژه B نیز هست و هرگاه $A = S^{-1}BS$ ، آنگاه Sx یک بردار ویژه B مربوط به λ و ماتریس B است.

تعریف ۳۵.۸. یک مجموعه از بردارهای $\{v^{(1)}, v^{(2)}, \dots, v^{(n)}\}$ متعامد نامیده می‌شود اگر به ازای هر $i \neq j$ ، $(v^{(i)})^t v^{(j)} = 0$ ، بعلاوه، هرگاه به ازای هر $i = 1, 2, \dots, n$ ، $(v^{(i)})^t v^{(i)} = 1$ ، آنگاه مجموعه متعامد نرمال نامیده می‌شود.

قضیه ۳۶.۸. هرگاه A یک ماتریس و $\lambda_1, \dots, \lambda_k$ مقادیر ویژه متمایز A با بردارهای ویژه $x^{(1)}, x^{(2)}, \dots, x^{(k)}$ باشند، آنگاه $\{x^{(1)}, x^{(2)}, \dots, x^{(k)}\}$ مستقل خطی است.

قضیه ۳۷.۸. یک مجموعه متعامد از بردارها که شامل بردار صفر نباشد مستقل خطی است.

تعریف ۳۸.۸. یک ماتریس $n \times n$ مانند P ماتریس متعامد نامیده می‌شود اگر $P^{-1} = P^t$.

قضیه ۳۹۰۸. هرگاه A یک ماتریس متقارن $n \times n$ باشد، آنگاه یک ماتریس متعامد P وجود دارد بطوری که $D = P^{-1}AP$ ، که در آن D یک ماتریس قطری است که درایه‌های قطری آن مقادیر ویژه A هستند.

نتیجه ۴۰۰۸. هرگاه A یک ماتریس $n \times n$ متقارن باشد، آنگاه n بردار ویژه از A وجود دارد که یک مجموعه متعامد نرمال تشکیل می‌دهند.

قضیه ۴۱۰۸. مقادیر ویژه یک ماتریس متقارن جملگی اعداد حقیقی هستند.

این مطالب مورد نیاز از جبر خطی را کامل می‌کند. حال آنچه را که روش توانی برای پیدا کردن مقادیر ویژه و بردارهای ویژه یک ماتریس نامیده می‌شود بررسی می‌کنیم. برای اینکه روش توانی را اعمال کنیم باید فرض کنیم ماتریس $n \times n$ ، A دارای n مقدار ویژه $\lambda_1, \lambda_2, \dots, \lambda_n$ با یک‌گرددآیه از بردارهای ویژه مربوطه $\{v^{(1)}, v^{(2)}, v^{(3)}, \dots, v^{(n)}\}$ باشد، که مستقل خطی است، و نتیجتاً یک پایه برای R^n تشکیل می‌دهد. بعلاوه، فرض کنیم A دقیقاً یک مقدار ویژه دارد که از نظر اندازه بزرگترین است، و به این دلیل، فرض می‌کنیم $\lambda_1, \lambda_2, \dots, \lambda_n$ نمایشگر مقادیر ویژه A با $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n| \geq 0$ باشد.

هرگاه $x^{(0)}$ بردار دلخواهی در R^n باشد، این امر که $\{v^{(1)}, v^{(2)}, \dots, v^{(n)}\}$ یک پایه برای R^n تشکیل می‌دهد ایجاب می‌کند که ثابتایی، احتمالاً "مختلط، مانند $\alpha_1, \alpha_2, \dots, \alpha_n$ وجود داشته باشند که

$$(۳۹۰۸) \quad x^{(0)} = \sum_{j=1}^n \alpha_j v^{(j)}.$$

با ضرب طرفین معادله (۳۹۰۸) در A, A^2, \dots, A^k بدست می‌آوریم

$$Ax^{(0)} = \sum_{j=1}^n \alpha_j Av^{(j)} = \sum_{j=1}^n \alpha_j \lambda_j v^{(j)},$$

$$(۴۰۰۸) \quad A^2 x^{(0)} = \sum_{j=1}^n \alpha_j \lambda_j^2 Av^{(j)} = \sum_{j=1}^n \alpha_j \lambda_j^2 v^{(j)},$$

$$\vdots$$

$$A^k x^{(0)} = \sum_{j=1}^n \alpha_j \lambda_j^k v^{(j)}.$$

اگر λ_1^k از هر جمله سمت راست معادله (۴۰.۸) فاکتور گرفته شود،

$$A^k \mathbf{x}^{(0)} = \lambda_1^k \sum_{j=1}^n \alpha_j \left(\frac{\lambda_j}{\lambda_1} \right)^k \mathbf{v}^{(j)},$$

این امر که به ازای هر $j = 2, 3, \dots, n$ ، $|\lambda_1| > |\lambda_j|$ نتیجه می‌دهد که
 و $\lim_{k \rightarrow \infty} (\lambda_j/\lambda_1)^k = 0$

$$(41.8) \quad \lim_{k \rightarrow \infty} A^k \mathbf{x}^{(0)} = \lim_{k \rightarrow \infty} \lambda_1^k \alpha_1 \mathbf{v}^{(1)}.$$

اگر $|\lambda_1| < 1$ ، حد فوق صفر است و اگر $|\lambda_1| > 1$ ، این حد موجود نیست، البته بشرط آنکه $\alpha_1 \neq 0$.

با استفاده از چند نماد که به طور موثری توانهای $A^k \mathbf{x}^{(0)}$ را به روش مناسبی می‌سنجد، می‌توان از رابطه بیان شده در معادله (۴۱.۸) سود جست.

فرض کنیم $\mathbf{y}^{(1)} = A \mathbf{x}^{(0)}$ و p_0 عدد صحیحی باشد، $1 \leq p_0 \leq n$ ، که برای آن اندازه مولفه p_0 م $\mathbf{x}^{(0)}$ ، یعنی $x_{p_0}^{(0)}$ ، نرم l_∞ بردار $\mathbf{x}^{(0)}$ باشد؛ یعنی،

$$|x_{p_0}^{(0)}| = \max_{1 \leq i \leq n} |x_i^{(0)}| = \|\mathbf{x}^{(0)}\|_\infty.$$

$\mu^{(1)}$ را چنین تعریف می‌نیم:

$$\mu^{(1)} = \frac{y_{p_0}^{(1)}}{x_{p_0}^{(0)}}.$$

فرض کنیم p_1 کوچکترین عدد صحیحی باشد که $1 \leq p_1 \leq n$ و

$$|y_{p_1}^{(1)}| = \|\mathbf{y}^{(1)}\|_\infty$$

و $\mathbf{x}^{(1)} \in R^n$ را چنین تعریف می‌کنیم:

$$\mathbf{x}^{(1)} = \frac{1}{y_{p_1}^{(1)}} \mathbf{y}^{(1)}.$$

با این نمادگذاری، داریم $\|\mathbf{x}^{(1)}\|_\infty = 1$ ، و چون

$$\mathbf{x}^{(0)} = \sum_{j=1}^n \alpha_j \mathbf{v}^{(j)}$$

و

$$\mathbf{y}^{(1)} = A \mathbf{x}^{(0)} = \sum_{j=1}^n \alpha_j \lambda_j \mathbf{v}^{(j)},$$

داریم

$$\begin{aligned} \mu^{(1)} &= \frac{\sum_{j=1}^n \alpha_j \lambda_j v_{p_0}^{(j)}}{\sum_{j=1}^n \alpha_j v_{p_0}^{(j)}} \\ &= \frac{\alpha_1 \lambda_1 v_{p_0}^{(1)} + \sum_{j=2}^n \alpha_j \lambda_j v_{p_0}^{(j)}}{\alpha_1 v_{p_0}^{(1)} + \sum_{j=2}^n \alpha_j v_{p_0}^{(j)}} \\ &= \lambda_1 \left[\frac{\alpha_1 v_{p_0}^{(1)} + \sum_{j=2}^n \alpha_j (\lambda_j / \lambda_1) v_{p_0}^{(j)}}{\alpha_1 v_{p_0}^{(1)} + \sum_{j=2}^n \alpha_j v_{p_0}^{(j)}} \right]. \end{aligned}$$

به طریق مشابه، می‌توان دنباله‌های $\{x^{(m)}\}_{m=0}^\infty$ و $\{y^{(m)}\}_{m=1}^\infty$ از بردارها و دنباله‌ها $\{\mu^{(m)}\}_{m=1}^\infty$ از اسکالرها را به استقرا چنین تعریف کرد:

$$(۴۲.۸) \quad y^{(m)} = Ax^{(m-1)}$$

$$(۴۳.۸) \quad \mu^{(m)} = \frac{y_{p_m}^{(m)}}{x_{p_m}^{(m-1)}} = \lambda_1 \left[\frac{\alpha_1 v_{p_m-1}^{(1)} + \sum_{j=2}^n (\lambda_j / \lambda_1)^m \alpha_j v_{p_m-1}^{(j)}}{\alpha_1 v_{p_m-1}^{(1)} + \sum_{j=2}^n (\lambda_j / \lambda_1)^{m-1} \alpha_j v_{p_m-1}^{(j)}} \right],$$

و

$$(۴۴.۸) \quad x^{(m)} = \frac{y^{(m)}}{y_{p_m}^{(m)}} = (y_{p_1}^{(1)} \cdot y_{p_2}^{(2)} \cdots y_{p_m}^{(m)})^{-1} \sum_{j=1}^n \lambda_j^m \alpha_j v^{(j)},$$

که در آن p_m نمایش عدد صحیحی است که در هر مرحله برای آن

$$|y_{p_m}^{(m)}| = \|y^{(m)}\|_\infty.$$

با معاینه (۴۳.۸) ملاحظه می‌شود که، چون به‌ازای هر $j = 2, 3, \dots, n$ ، $|\lambda_j / \lambda_1| < 1$ ، وقتی $x^{(0)}$ طوری انتخاب شده باشد که $\alpha_1 \neq 0$ ، آنگاه $\lim_{m \rightarrow \infty} \mu^{(m)} = \lambda_1$.
 بعلاوه، می‌توان نشان داد (ر. ک. ویلکسون [۶۹]، صفحات ۵۷۲-۵۷۰) که دنباله $\{x^{(m)}\}_{m=0}^\infty$ از بردارها به بردار ویژه منحصراً با نرم یک مربوط به λ_1 میل خواهد کرد.

معیناً، باید متذکر شد که روش توانی در حالت کلی دارای این اشکال است که در ابتدا معلوم نیست ماتریس دارای یک مقدار ویژه غالب منحصراً با نرم یک هست یا نه، و معلوم نیست $x^{(0)}$ چگونه باید انتخاب شود تا اطمینان حاصل گردد که نمایش آن بر حسب بردارهای ویژه ماتریس سهم ناصفری از بردار ویژه مربوط به آن مقدار ویژه غالب، که باید وجود داشته باشد، را شامل شود.

الگوریتم روش توانی ۵.۸

برای تعیین مقدار ویژه غالب (مقدار ویژه با بزرگترین اندازه) از یک ماتریس A و یک بردار ویژه مربوطه با نرم یک، بردار ناصفر $x^{(0)}$ را انتخاب می‌کنیم.

مرحله ۱. قرار می‌دهیم $m = 1$ و کوچکترین عدد صحیح p_0 را پیدا می‌کنیم بطوری که

$$|x_{p_0}^{(0)}| = \|x^{(0)}\|_\infty.$$

مرحله ۲. $y^{(m)} = Ax^{(m-1)}$ را حساب می‌کنیم.

مرحله ۳. هرگاه $y^{(m)} = (0, 0, \dots, 0)^T$ ، آنگاه 0 یک مقدار ویژه A با بردار ویژه

$x^{(m-1)}$ است. بردار دیگری انتخاب می‌کنیم و به مرحله ۱ می‌رویم.

مرحله ۴. قرار می‌دهیم $\mu^{(m)} = \frac{y_{p_m}^{(m)}}{x_{p_m}^{(m-1)}}$.

مرحله ۵. کوچکترین عدد صحیح p_m را پیدا می‌کنیم که به‌ارای آن

$$|y_{p_m}^{(m)}| = \|y^{(m)}\|_\infty.$$

مرحله ۶. قرار می‌دهیم $x^{(m)} = \frac{y^{(m)}}{y_{p_m}^{(m)}}$.

مرحله ۷. اگر $\mu^{(m)}$ تخمین به‌حد کافی دقیقی از λ باشد، به مرحله ۱۰ می‌رویم.

مرحله ۸. 1 را به m می‌افزاییم.

مرحله ۹. به مرحله ۲ می‌رویم.

مرحله ۱۰. روند تمام است؛ $\mu^{(m)}$ تقریبی به مقدار ویژه غالب A است، و $x^{(m)}$ یک

برداریکه است که بردار ویژه مربوطه را تقریب می‌کند.

در مرحله ۵، انتخاب کوچکترین عدد صحیح p_m که برای آن $|y_{p_m}^{(m)}| = \|y^{(m)}\|_\infty$

عموماً "اطمینان می‌دهد که این اندیس سرانجام پایا خواهد شد. سرعتی که با آن

$\{\mu^{(m)}\}_{m=1}^\infty$ به λ_1 همگراست بوسیله نسبت‌های $|\lambda_j/\lambda_1|^m$ ، به‌ارای $j = 2, 3, \dots, n$ ، بویژه

با $|\lambda_2/\lambda_1|^m$ ، تعیین می‌شود، یعنی، همگرایی از مرتبه $O(\lambda_2/\lambda_1)$ است. زمانی که λ_2 و

λ_1 از حیث اندازه "نزدیک" هستند، کاربرد روند Δ^2 ایتنن، که در قضیه ۱۳.۲

(ص ۷۸) بحث شد، برای سرعت بخشیدن به همگرایی دنباله $\{\mu^{(m)}\}_{m=1}^\infty$ به مقدار ویژه

λ_1 مفید است.

در عمل، لازم نیست ماتریس n مقدار ویژه متمایز داشته باشد تا روش توانی همگرا

باشد. در واقع، اگر مقدار ویژه غالب منحصر بفرد λ_1 دارای مرتبه تکرار ν بزرگتر از یک بوده و بردارهای ویژه $v^{(1)}, v^{(2)}, \dots, v^{(\nu)}$ مربوط به λ_1 مستقل خطی باشند، روند باز هم به λ_1 همگراست. دنباله $\{x^{(m)}\}_{m=0}^{\infty}$ از بردارها در این حالت به یک بردار ویژه λ_1 با نرم یک میل خواهد کرد که ترکیبی خطی از $v^{(1)}, v^{(2)}, \dots, v^{(\nu)}$ است و به انتخاب بردار اولیه $x^{(0)}$ وابسته است.

مثال ۳. ماتریس

$$A = \begin{bmatrix} -4 & 14 & 0 \\ -5 & 13 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

دارای مقادیر ویژه $\lambda_1 = 6$ ، $\lambda_2 = 3$ ، و $\lambda_3 = 2$ است. در نتیجه، روش توانی تشریح شده در الگوریتم ۵.۸ همگرا خواهد بود.

فرض کنیم $x^{(0)} = (1, 1, 1)^t$. در این صورت،

$$y^{(1)} = Ax^{(0)} = (10, 8, 1)^t;$$

در نتیجه

$$\|y^{(1)}\|_{\infty} = 10, \mu^{(1)} = \frac{y_1^{(1)}}{x_1^{(0)}} = 10$$

و

$$x^{(1)} = \frac{y^{(1)}}{\|y^{(1)}\|_{\infty}} = (1, .8, .1)^t.$$

ادامه این روش منجر به مقادیر مندرج در جدول ۵.۸ می شود.

در این مرحله یک تقریب به مقدار ویژه غالب، یعنی 6، برابر $\mu^{(12)} = 6.000837$

خواهد بود، با بردار ویژه $(1, -.249895, -.714316)$.

وقتی A یک ماتریس متقارن باشد، می توان تغییری در انتخاب بردارهای $x^{(m)}, y^{(m)}$

و اسکالرهای $\mu^{(m)}$ داد تا سرعت همگرایی دنباله $\{\mu^{(m)}\}_{m=1}^{\infty}$ به مقدار ویژه غالب λ_1

به طور قابل ملاحظه ای بهتر شود. در واقع، در حالی که سرعت همگرایی روش توانی کلی

است، $O(|\lambda_2/\lambda_1|)$ سرعت همگرایی روند پیراسته ای که برای ماتریسهای متقارن در ذیل

داده شده $O(|\lambda_2/\lambda_1|^2)$ است.

جدول ۵-۸

m	$(y^{(m)})^t$	$\mu^{(m)}$	$(x^{(m)})^t$
0	—	—	(1, 1, 1)
1	(10., 8., 1.)	10	(1, .8, .1)
2	(7.2, 5.4, -.8)	7.2	(1, .75, -.111)
3	(6.5, 4.75, -1.222)	6.5	(1, .730769, -.188034)
4	(6.230769, 4.5, -1.376068)	6.230769	(1, .7222, -.220850)
5	(6.111, 4.388, -1.441701)	6.111	(1, .718182, -.235915)
6	(6.054546, 4.3364, -1.471829)	6.054546	(1, .716216, -.243095)
7	(6.027027, 4.310811, -1.486190)	6.027027	(1, .715247, -.246588)
8	(6.013453, 4.298206, -1.493175)	6.013453	(1, .714765, -.248306)
9	(6.006711, 4.291946, -1.496612)	6.006711	(1, .714525, -.249157)
10	(6.003352, 4.288827, -1.498313)	6.003352	(1, .714405, -.249579)
11	(6.001675, 4.287270, -1.499159)	6.001675	(1, .714346, -.249790)
12	(6.000837, 4.286492, -1.499580)	6.000837	(1, .714316, -.249895)

الگوریتم روش توانی برای ماتریسهای متقارن ۶-۸

برای تعیین مقدار ویژهء غالب از یک ماتریس متقارن A (مقدار ویژه با بزرگترین اندازه) و یک بردار ویژهء مربوطه با نرم یک، یک بردار ناصفر مانند $y^{(0)}$ انتخاب می‌کنیم .

مرحله ۱ . قرار می‌دهیم $x^{(0)} = \frac{y^{(0)}}{\|y^{(0)}\|_2}$.

مرحله ۲ . قرار می‌دهیم $m = 1$.

مرحله ۳ . $y^{(m)} = Ax^{(m-1)}$ را حساب می‌کنیم .

مرحله ۴ . هرگاه $y^{(m)} = (0, 0, 0, \dots, 0)^t$ ، آنگاه 0 یک مقدار ویژهء A با بردار ویژهء

$x^{(m-1)}$ است ، بردار دیگر $y^{(0)}$ را انتخاب کرده به مرحله ۱ می‌رویم .

مرحله ۵ . قرار می‌دهیم $\mu^{(m)} = (x^{(m-1)})^t y^{(m)}$.

مرحله ۶ . قرار می‌دهیم $x^{(m)} = \frac{y^{(m)}}{\|y^{(m)}\|_2}$.

مرحله ۷ . اگر $\mu^{(m)}$ تخمین به حد کافی دقیقی از λ_1 باشد، به مرحله ۱۰ می‌رویم .

مرحله ۸ . 1 را به m می‌افزاییم .

مرحله ۹ . به مرحله ۳ می‌رویم .

مرحله ۱۰. روند تمام است؛ $\mu^{(m)}$ تقریبی از مقدار ویژه غالب A و $x^{(m)}$ برداری با $\|x^{(m)}\|_2 = 1$ است که بردار ویژه مربوطه را تقریب می‌کند.

مثال ۰۴. ماتریس

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

مقارن است با مقادیر ویژه $\lambda_1 = 6$ ، $\lambda_2 = 3$ ، و $\lambda_3 = 1$. فهرست کردن نتایج ده تکرار اول آروشهای توانی ارائه شده در الگوریتمهای ۵۰۸ و ۶۰۸، با $y^{(0)} = x^{(0)} = (1, 0, 0)^t$ ، پیشرفت قابل توجه بدست آمده از الگوریتم دوم را نشان می‌دهد. با بکار بردن الگوریتم ۵۰۸، نتایج فهرست شده در جدول ۰۸ را داریم. نتایج فهرست شده در جدول ۰۸ با استفاده از الگوریتم ۶۰۸ بدست آمده‌اند.

جدول ۰۸

m	$(y^{(m)})^t$	$\mu^{(m)}$	$(x^{(m)})^t$
0	—	—	(1, 0, 0)
1	(4, -1, 1)	4	(1, -.25, .25)
2	(4.5, -2.25, 2.25)	4.5	(1, -.5, .5)
3	(5, -3.5, 3.5)	5	(1, -.7, .7)
4	(5.4, -4.5, 4.5)	5.4	(1, -.8333, .8333)
5	(5.666, -5.1666, 5.1666)	5.666	(1, -.911765, .911765)
6	(5.823529, -5.558824, 5.558824)	5.823529	(1, -.954545, .954545)
7	(5.909091, -5.772727, 5.772727)	5.909091	(1, -.976923, .976923)
8	(5.953846, -5.884615, 5.884615)	5.953846	(1, -.988372, .988372)
9	(5.976744, -5.941861, 5.941861)	5.976744	(1, -.994163, .994163)
10	(5.988327, -5.970817, 5.970817)	5.988327	(1, -.997076, .997076)

کران خطای ذیل برای تقریب مقادیر ویژه یک ماتریس مقارن را می‌توان به‌عنوان معیار توقف در مرحله ۷ الگوریتم ۶۰۸ بکار برد. برهان این نتیجه را می‌توان در صفحات ۱۴۱-۱۴۰ ایزاکسون و کلر [۳۹] یافت.

قضیه ۰۴۲۰۸. هرگاه A یک ماتریس مقارن $n \times n$ با مقادیر ویژه $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ بوده

و به ازای برداری چون $\|x\|_2 = 1$ و عددی حقیقی مانند λ ، $\|Ax - \lambda x\|_2 \leq \varepsilon$ ،
 آنگاه

$$(۴۵.۸) \quad \min_{1 \leq i \leq n} |\lambda_i - \lambda| \leq \varepsilon$$

جدول ۷.۸

m	$(y^{(m)})^t$	$\mu^{(m)}$	$(x^{(m)})^t$
0	(1, 0, 0)	—	(1, 0, 0)
1	(4, -1, 1)	4	(.942809, -.235702, .235702)
2	(4.242641, -2.121320, 2.121320)	5	(.816497, -.408248, .408248)
3	(4.082483, -2.857738, 2.857738)	5.666667	(.710669, -.497468, .497468)
4	(3.837613, -3.198011, 3.198011)	5.909091	(.646997, -.539164, .539164)
5	(3.666314, -3.342816, 3.342816)	5.976744	(.612836, -.558763, .558763)
6	(3.568871, -3.406650, 3.406650)	5.994152	(.595247, -.568190, .568190)
7	(3.517370, -3.436200, 3.436200)	5.998536	(.586336, -.572805, .572805)
8	(3.4909518, -3.450359, 3.450359)	5.999634	(.581852, -.575086, .575086)
9	(3.4775802, -3.457283, 3.457283)	5.999908	(.579603, -.576220, .576220)
10	(3.470854, -3.460706, 3.460706)	5.999977	(.578477, -.576786, .576786)

زمانی که تقریبی به مقدار ویژه غالب حساب شده باشد ، تکنیکهای متعددی برای بدست آوردن تقریبهایی از دیگر مقادیر ویژه یک ماتریس در دسترس است . بعضی از این تکنیکها در تمرینات ۱۲ و ۱۳ بررسی می شوند ، و آنچه خود را در اینجا به ارائه اش محدود می کنیم به تکنیکهای تقلیل معروف است . آن تکنیکهایی که شامل تشکیل یک ماتریس جدید B از ماتریس اصلی A است که مقادیر ویژه اش همان مقادیر ویژه A است با این استثناء که مقدار ویژه غالب A با مقدار ویژه صفر در B جایگزین شده است .

قضیه ۴۳.۸ . فرض کنیم $\lambda_1, \lambda_2, \dots, \lambda_n$ مقادیر ویژه A با بردارهای ویژه مربوطه $v^{(1)}, v^{(2)}, \dots, v^{(n)}$ بوده و λ_1 بستایی یک داشته باشد . هرگاه x بردار دلخواهی با خاصیت $1 = x^t v^{(1)}$ باشد ، آنگاه

$$(۴۶.۸) \quad B = A - \lambda_1 v^{(1)} x^t$$

دارای مقادیر ویژه $\lambda_2, \lambda_3, \dots, \lambda_n$ و 0 با بردارهای ویژه $w^{(1)}, w^{(2)}, w^{(3)}, \dots, w^{(n)}$ است ، که در آنها $v^{(i)}$ و $w^{(i)}$ ، به ازای هر $i = 2, 3, \dots, n$ ، با معادلات زیر مربوط شده اند :

$$(۴۷.۸) \quad v^{(i)} = (\lambda_i - \lambda_1) w^{(i)} + \lambda_1 (x^t w^{(i)}) v^{(1)}$$

برهان این نتیجه را می‌توان در صفحات ۴۸۲ - ۴۸۵ رالستون [۴۹] یافت. یک تکنیک تقلیل مفید خاص، به نام روند تقلید ویلاند^۱، با تعریف

$$\mathbf{x} = \frac{1}{\lambda_1 v_i^{(1)}} \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix},$$

نتیجه می‌شود، که در آن $v_i^{(1)}$ مختصی از $\mathbf{v}^{(1)}$ است که صفر نیست و مقادیر $a_{i1}, a_{i2}, \dots, a_{in}$ درایه‌های سطر i م A هستند. با این تعریف، داریم

$$\begin{aligned} \mathbf{x}'\mathbf{v}^{(1)} &= \frac{1}{\lambda_1 v_i^{(1)}} [a_{i1}, a_{i2}, \dots, a_{in}] (v_1^{(1)}, v_2^{(1)}, \dots, v_n^{(1)})' \\ &= \frac{1}{\lambda_1 v_i^{(1)}} \sum_{j=1}^n a_{ij} v_j^{(1)}, \end{aligned}$$

که در آن مجموع مولفه^۲ i م حاصل ضرب $A\mathbf{v}^{(1)}$ است. چون $A\mathbf{v}^{(1)} = \lambda_1 \mathbf{v}^{(1)}$ ، این ایجاب می‌کند که

$$\mathbf{x}'\mathbf{v}^{(1)} = \frac{1}{\lambda_1 v_i^{(1)}} (\lambda_1 v_i^{(1)}) = 1,$$

در نتیجه، \mathbf{x} در فرض قضیه^۳ ۴۳.۸ صدق می‌کند. بعلاوه، همانطور که در تمرین ۱۴ دیده شد، سطر i م $B = A - \lambda_1 \mathbf{v}^{(1)}\mathbf{x}'$ تماماً از درایه‌های صفر تشکیل شده است، و اگر $\lambda \neq 0$ مقدار ویژه دلخواهی با بردار ویژه^۴ مربوطه^۵ \mathbf{w} باشد، رابطه^۶ $B\mathbf{w} = \lambda\mathbf{w}$ نتیجه می‌دهد که مولفه^۷ i م \mathbf{w} نیز باید صفر باشد. در نتیجه، ستون i م ماتریس B سهمی در حاصل ضرب $B\mathbf{w} = \lambda\mathbf{w}$ ندارد. بنابراین، ماتریس B را می‌توان با یک ماتریس $(n-1) \times (n-1)$ چون B' ، که از حذف سطر و ستون i م B بدست می‌آید، جایگزین کرد. B' دارای مقادیر ویژه^۸ $\lambda_2, \lambda_3, \dots, \lambda_n$ خواهد بود، و اگر $|\lambda_2| > |\lambda_3|$ ، می‌توان روش توانی را جهت تعیین این مقدار ویژه^۹ غالب جدید و یک بردار ویژه^{۱۰} $\mathbf{w}^{(2)}$ ، مربوط به λ_2 ، نسبت به ماتریس B' ، مجدداً روی B' اعمال کرد. بدست آوردن بردار ویژه^{۱۱} مربوطه برای ماتریس اصلی A فقط محتاج ساختن $\mathbf{w}^{(2)}$ بوسیله^{۱۲} درج یک مولفه^{۱۳} صفر بین مولفه‌های $w_{i-1}^{(2)}$

1. Wielandt

و $w_1^{(2)}$ از بردار $(n-1)$ بعدی $w^{(2)}$ و سپس محاسبه $v^{(2)}$ با استفاده از معادله (۴۷.۸) است.

مثال ۵. از مثال ۴ می‌دانیم که ماتریس

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

دارای مقادیر ویژه $\lambda_1 = 6$ ، $\lambda_2 = 3$ ، و $\lambda_3 = 1$ است. با فرض آنکه مقدار ویژه غالب $\lambda_1 = 6$ و بردار ویژه $v^{(1)}$ یک‌ه مربوطه $v^{(1)} = (1, -1, 1)^t$ محاسبه شده است، روند بالا جهت تعیین λ_2 به طریق زیر پیش می‌رود:

$$x = \frac{1}{6} \begin{bmatrix} 4 \\ -1 \\ 1 \end{bmatrix} = \left(\frac{2}{3}, -\frac{1}{6}, \frac{1}{6}\right)^t,$$

$$v^{(1)}x^t = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \left[\frac{2}{3}, -\frac{1}{6}, \frac{1}{6}\right] = \begin{bmatrix} \frac{2}{3} & -\frac{1}{6} & \frac{1}{6} \\ -\frac{2}{3} & \frac{1}{6} & -\frac{1}{6} \\ \frac{2}{3} & -\frac{1}{6} & \frac{1}{6} \end{bmatrix},$$

۹

$$\begin{aligned} B = A - \lambda_1 v^{(1)}x^t &= \begin{bmatrix} 4 & -1 & 1 \\ -1 & 3 & -2 \\ 1 & -2 & 3 \end{bmatrix} - 6 \begin{bmatrix} \frac{2}{3} & -\frac{1}{6} & \frac{1}{6} \\ -\frac{2}{3} & \frac{1}{6} & -\frac{1}{6} \\ \frac{2}{3} & -\frac{1}{6} & \frac{1}{6} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 3 & 2 & -1 \\ -3 & -1 & 2 \end{bmatrix} \end{aligned}$$

حذف سطر و ستون اول نتیجه می‌دهد

$$B' = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix},$$

که دارای مقادیر ویژه $\lambda_2 = 3$ و $\lambda_3 = 1$ است. برای $\lambda_2 = 3$ بردار ویژه $w^{(2)}$ را می‌توان با حل دستگاه خطی مرتبه دوم

$$(B' - 3I)w^{(2)} = 0,$$

بدست آورد، که در نتیجه

$$w^{(2)'} = (1, -1)'$$

لذا، $w^{(2)} = (0, 1, -1)'$ و، از معادله (47.8) ، داریم

$$\begin{aligned} v^{(2)} &= (3 - 6)(0, 1, -1)' + 6\left(\frac{2}{3}, -\frac{1}{6}, \frac{1}{6}\right)(0, 1, -1)')(1, -1, 1)' \\ &= (-2, -1, 1)' \end{aligned}$$

توجه کنید که گرچه این فرایند تقلیل را می توان در بعضی حالات برای یافتن تقریبهایی به تمام مقادیر ویژه و بردارهای ویژه یک ماتریس بکار برد، این فرایند فوق العاده به خطای گرد کردن حساس است. تکنیکهای متکی به تبدیلات متشابه در بخش بعد ارائه خواهند شد، و این روشها وقتی تقریبهایی به تمام مقادیر ویژه و بردارهای ویژه نیاز باشد عموماً "اولویت دارند".

این بخش را با یک الگوریتم برای محاسبه مقدار ویژه غالب دوم و بردار ویژه مربوطه برای یک ماتریس، وقتی مقدار ویژه غالب و بردار ویژه مربوطه تعیین شده باشند، خاتمه می دهیم.

الگوریتم تقلیل ویلاند ۷.۸

با داشتن λ_1 به عنوان تقریبی به مقدار ویژه غالب λ_1 از A ، v به عنوان تقریبی به بردار ویژه متناظر $v^{(1)}$ از A ، و یک تقریب اولیه $x^{(0)} \in R^{n-1}$ شروع می کنیم.

مرحله ۱. فرض کنیم i کوچکترین عدد صحیحی باشد که $1 \leq i \leq n$ و به ازای آن $v_i \neq 0$.

مرحله ۲. اگر $i = 1$ ، به مرحله ۶ می رویم.

مرحله ۳. به ازای $k = 1, 2, \dots, i - 1$ و $j = 1, 2, \dots, i - 1$ قرار می دهیم

$$b_{kj} = a_{kj} - \frac{v_k}{v_i} a_{ij}.$$

مرحله ۴. به ازای هر $k = i, \dots, n - 1$ و $j = 1, 2, \dots, i - 1$ قرار می دهیم

$$b_{kj} = a_{k+1,j} - \frac{v_{k+1}}{v_i} a_{ij}.$$

مرحله ۵. به ازای هر $k = 1, 2, \dots, i - 1$ و $j = i, \dots, n - 1$ قرار می دهیم.

$$b_{kj} = a_{k,j+1} - \frac{v_k}{v_i} a_{i,j+1}$$

مرحله ۶. به ازای هر $k = i, \dots, n-1$ و $j = i, \dots, n-1$ قرار می دهیم

$$b_{kj} = a_{k+1,j+1} - \frac{v_{k+1}}{v_i} a_{i,j+1}$$

مرحله ۷. روش توانی را روی ماتریس $(n-1) \times (n-1)$ ، $B' = (b_{kj})$ ، با $x^{(0)}$ به عنوان تقریب اولیه انجام می دهیم و فرض می کنیم $\bar{\lambda}_2$ و $w^{(2)'} = (w_1^{(2)'}, w_2^{(2)'}, \dots, w_{n-1}^{(2)'})'$ نمایانگر تقریبهای بدست آمده به این طریق باشند.

مرحله ۸. اگر $i = 1$ ، قرار می دهیم $w_1^{(2)} = 0$ و به ازای $k = 2, \dots, n$ ، $w_k^{(2)} = w_{k-1}^{(2)}$.

اگر $i > 1$ ، به ازای $k = 1, 2, \dots, i-1$ قرار می دهیم $w_k^{(2)} = w_k^{(2)'}$ ،

$$w_i^{(2)} = 0,$$

و به ازای $k = i+1, i+2, \dots, n$ ، $w_k^{(2)} = w_{k-1}^{(2)'}$.

مرحله ۹. به ازای هر $k = 1, \dots, n$ قرار می دهیم

$$v_k^{(2)} = (\bar{\lambda}_2 - \bar{\lambda}_1)w_k^{(2)} + \left(\frac{\sum_{j=1}^n a_{ij}w_j^{(2)}}{v_i^{(1)}} \right) v_k^{(1)}$$

مرحله ۱۰. روند تمام است؛ $\bar{\lambda}_2$ تقریب به دومین مقدار ویژه غالب A و $v^{(2)}$ تقریب به بردار ویژه یکه مربوطه است.

مجموعه تمرینات ۴.۸

۱. نشان دهید که ماتریس

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

کمتر از سه بردار ویژه مستقل خطی دارد.

۲. (آ) فرض کنید B ماتریسی $n \times n$ باشد. نشان دهید که $\det B = \det B'$.

(ب) نشان دهید که A و A' مقادیر ویژه یکسان دارند.

۳. با به کار بردن قضیه ۳.۰۸، کرانهایی برای مقادیر ویژه ماتریسهای زیر تعیین کنید.

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \\ -1 & -1 & 2 \end{bmatrix} \quad (\text{آ})$$

$$\begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ -1 & -1 & 4 \end{bmatrix} \quad (\text{ب})$$

$$\begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 0 \\ 1 & 0 & 3 \end{bmatrix} \quad (\text{پ})$$

۴. نشان دهید که هر چهار بردار در R^3 وابسته خطی اند.
۵. نشان دهید هرگاه A ماتریسی $n \times n$ با n مقدار ویژه متمایز باشد، آنگاه A دارای n بردار ویژه مستقل خطی است.
۶. روش توانی را برای محاسبه بزرگترین مقدار ویژه (از نظر قدر مطلق) ماتریسهای زیر با دقت ۰.۰۰۰۱، یا تا وقتی $N > 25$ ، با استفاده از $v^{(0)} = (1, 1, 1)^T$ ، بکار برید.

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ -2 & -1 & -3 \end{bmatrix} \quad (\text{آ})$$

$$\begin{bmatrix} 1 & 1 & 3 \\ 1 & -2 & 1 \\ 3 & 1 & 3 \end{bmatrix} \quad (\text{ب})$$

$$\begin{bmatrix} 1 & 1 & .5 \\ 1 & 1 & .25 \\ .5 & .25 & 2 \end{bmatrix} \quad (\text{پ})$$

۷. تمرین ۶ (ب) و ۶ (پ) را، با بکار بردن الگوریتم ۶.۸، تکرار کنید. سرعتهای همگرایی را مقایسه نمایید.
۸. فرض کنید v_1, \dots, v_k بردار متعام باشند. نشان دهید که $\{v_1, \dots, v_k\}$ یک مجموعه مستقل خطی است.
۹. فرض کنید P یک ماتریس متعام باشد. نشان دهید که ستونهای P یک مجموعه متعام نرمال از بردارها تشکیل می‌دهند. همچنین، نشان دهید که $\|P\|_\infty \leq 1$ و $\|P^2\|_\infty \leq 1$.

۱۰. فرض کنید v_1, \dots, v_n بردار متعامد نرمال در R^n باشند. همچنین، $x \in R^n$. نشان دهید که $x = \sum_{k=1}^n c_k v_k$ ، که در آن $c_k = v_k^T x$.

۱۱. فرض کنید مقادیر ویژه A ، $\lambda_1 > \lambda_2 > \dots > \lambda_{n-1} < \lambda_n$ یا $\lambda_1 < \lambda_2 < \dots < \lambda_{n-1} < \lambda_n$ و غالب باشد.

(A) نشان دهید که مقدار ویژه λ_n را می‌توان با اعمال روش توانی روی $A - cI$ ، که در آن c عددی است بطوری که $|c| > |\lambda_1|$ و c و λ_1 دارای یک علامت هستند، تقریب زد.

(ب) این تکنیک را روی ماتریسهای تمرین ۶ اعمال کنید.

۱۲. تکنیک صفرسازی. فرض کنید ماتریس $n \times n$ ، A دارای مقادیر ویژه $\lambda_1, \dots, \lambda_n$ مرتب شده بوسیله

$$|\lambda_1| > |\lambda_2| > |\lambda_3| \geq \dots \geq |\lambda_n|$$

با بردارهای ویژه مستقل خطی $v^{(1)}, v^{(2)}, \dots, v^{(n)}$ باشد.

(A) نشان دهید هرگاه روش توانی با بردار اولیه

$$x^{(0)} = x_2 v^{(2)} + x_3 v^{(3)} + \dots + x_n v^{(n)}$$

اعمال شود، آنگاه دنباله $\{\mu^{(m)}\}$ به λ_2 همگرا خواهد بود.

(ب) نشان دهید که، به ازای هر بردار $x = \sum_{i=1}^n \alpha_i v^{(i)}$ ، بردار $x^{(0)} = (A - \lambda_1 I)x$ در شرط داده شده در قسمت (A) صدق می‌کند.

(پ) برای ماتریسهای تمرین ۶ تقریبی به λ_2 بدست آورید.

(ت) نشان دهید که این روش ممکن است برای پیدا کردن λ_3 ، با استفاده از

$$x^{(0)} = (A - \lambda_2 I)(A - \lambda_1 I)x$$

۱۳. تقلیل هتلینگ^۱. فرض کنید بزرگترین مقدار ویژه λ_1 از حیث اندازه و یک بردار ویژه مربوطه $v^{(1)}$ برای ماتریس متقارن $n \times n$ ، A بدست آمده باشد. نشان دهید که ماتریس

$$B = A - \frac{\lambda_1}{(v^{(1)})^T v^{(1)}} v^{(1)}(v^{(1)})^T$$

همان مقادیر ویژه $\lambda_2, \dots, \lambda_n$ از A را داراست، جز آنکه B دارای مقدار ویژه

۰ ، با بردار ویژه^(۱) $v^{(1)}$ ، به جای مقدار ویژه^۱ λ_1 است . این روش تقلیل را جهت پیدا کردن λ_2 ، برای هر ماتریس تمرین ۶ ، بکار برید . به طور نظری ، این روش ممکن است برای پیدا کردن مقادیر ویژه^۲ بیشتری ادامه یابد ، اما خطای گرد کردن بزودی ادامه کار را بی ارزش می کند .

۱۴ . نشان دهید که سطر i م $B = A - \lambda_1 v^{(1)} x^t$ ، که در آن λ_1 بزرگترین مقدار ویژه^۳ A از نظر قدر مطلق ، $v^{(1)}$ بردار ویژه^۴ مربوطه^۵ A برای λ_1 ، و x برداری است که در (۴۸.۸) تعریف شد ، صفر است .

۱۵ . تقلیل ویلاند را روی ماتریسهای تمرین ۶ اعمال کنید .

۱۶ . با دنبال کردن تمرین ۱۷ در بخش ۳۰.۷ و تمرین ۲۷ در بخش ۱۰.۸ ، فرض کنید نوعی از سوسکها دوران زندگی چهار ساله دارند ؛ یعنی ، یک ماده در سال اول نرخ حیات $\frac{1}{4}$ دارد ، در سال دوم نرخ حیات $\frac{1}{4}$ ، و در سال سوم نرخ حیات $\frac{1}{8}$. بعلاوه ، فرض کنید یک ماده ، به طور متوسط ، دو ماده^۶ جدید در سال سوم و چهار ماده^۷ جدید در سال چهارم بدنیای می آورد . ماتریسی که سهم یکساله^۸ یک ماده را در جمعیت ماده های سال بعد نشان می دهد

$$A = \begin{bmatrix} 0 & 0 & 2 & 4 \\ \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 \end{bmatrix}$$

می باشد ، که در آن درایسه^۹ واقع در سطر i م و ستون j م مجدداً "نمایانگر سهم احتمالی یک ماده به سن j به جمعیت ماده به سن i در سال بعد است .

(آ) قضیه^{۱۰} دایره^{۱۱} گرشگورین ۳۰.۸ را جهت تعیین ناحیه ای در صفحه^{۱۲} مختلط شامل تمام مقادیر ویژه^{۱۳} A بکار برید .

(ب) روش توانی را جهت تعیین مقدار ویژه^{۱۴} غالب این ماتریس و بردار ویژه^{۱۵} مربوطه بکار برید .

(پ) الگوریتم ۷.۸ را جهت تعیین هر مقدار ویژه^{۱۶} و بردار ویژه^{۱۷} دیگر A بکار برید .

(ت) مقادیر ویژه^{۱۸} A را با استفاده از چند جمله ای مشخص A و روش نیوتن - رفسون پیدا کنید .

(ث) پیش بینی دراز مدت شما برای جمعیت این سوسکها ، البته با فرض یک جمعیت اولیه^{۱۹} متناهی ، چیست ؟

۱۷. یک ماتریس تمام مقارن ماتریسی است که در اطراف هر دو قطر متقارن باشد؛ یعنی، یک ماتریس $N \times N$ ، $A = (a_{ij})$ تمام مقارن است اگر به ازای هر $i, j = 1, 2, \dots, N$ و $a_{ij} = a_{ji} = a_{N+1-i, N+1-j}$ ، تعدادی از مسائل در نظریه ارتباطات دارای جوابهایی هستند که شامل مقادیر ویژه و بردارهای ویژه ماتریسهایی است که به شکل تمام مقارن هستند. مثلاً، بردار ویژه متناظر با کمترین مقدار ویژه ماتریس تمام مقارن 4×4

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

عکس العمل تونل انرژی واحد را برای یک دنباله خطای مفروض به طول 2، و در نتیجه وزن مینیمم هر دنباله خطای ممکن را، می دهد.

(آ) قضیه دایره گرسگورین (قضیه ۳۰.۸) را بکار برده نشان دهید هرگاه A ماتریس داده شده در بالا و λ مقدار ویژه مینیمال آن باشد، آنگاه $\rho(A - 4I) = |\lambda - 4|$ ، که در آن ρ شعاع طیفی است.

(ب) مقدار ویژه مینیمال ماتریس A را با یافتن تمام مقادیر ویژه $A - 4I$ و محاسبه شعاع طیفی آن پیدا کنید. سپس بردار ویژه متناظر را پیدا نمایید.

(پ) مقدار ویژه مینیمال ماتریس A را با انجام روش توانی روی $A - 4I$ پیدا کنید.

(ت) قضیه دایره گرسگورین را بکار برده نشان دهید که اگر λ مقدار ویژه مینیمال ماتریس

$$B = \begin{bmatrix} 3 & -1 & -1 & 1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ 1 & -1 & -1 & 3 \end{bmatrix}$$

باشد، آنگاه $\rho(B - 6I) = |\lambda - 6|$.

(ث) سعی کنید قسمتهای (ب) و (پ) این تمرین را با بکار بردن ماتریس B و نتیجه قسمت (ت) تکرار نمایید.

۱۸. یک دستگاه دینامیک خطی را می توان با معادلات زیر نمایش داد:

$$\frac{dx}{dt} = A(t)x(t) + B(t)u(t),$$

$$y(t) = C(t)x(t) + D(t)u(t),$$

که در آنها A یک ماتریس متغیر $n \times n$ ، B یک ماتریس متغیر $n \times r$ ، C یک ماتریس متغیر $m \times n$ ، D یک ماتریس متغیر $m \times r$ ، x یک متغیر برداری n بعدی، y یک متغیر برداری m بعدی، و u یک متغیر برداری r بعدی است. برای آنکه این دستگاه پایدار باشد باید تمام مقادیر ویژه A ، به ازای تمام مقادیر r ، دارای قسمت حقیقی نامثبت باشند.

(ا) اگر

$$A(t) = \begin{bmatrix} -1 & 2 & 0 \\ -2.5 & -7 & 4 \\ 0 & 0 & -5 \end{bmatrix},$$

آیا دستگاه پایدار است؟

(ب) اگر

$$A(t) = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & -5 & 1 \\ -1 & -1 & -2 & -3 \end{bmatrix},$$

آیا دستگاه پایدار است؟

۵.۸ الگوریتم QR

روش تقلیل مطرح شده در بخش قبل عموماً "برای محاسبه تمام مقادیر ویژه یک ماتریس، بخاطر رشد خطای گرد کردن در هر مرحله، نامناسب است. در این بخش روندی ارائه خواهد شد که برای زمانی طرح ریزی شده است که یافتن همه مقادیر ویژه یک ماتریس ضرورت دارد. با آنکه تعدادی از این نوع روندها را می‌شناسیم، بررسی خود را به روندی به نام الگوریتم QR محدود کرده و فرض می‌کنیم ماتریس اصلی متقارن باشد. الگوریتم QR برای این نوع ماتریسها پایدار است و بیش از هر روش دیگر بکار می‌رود. در صورت نیاز به بررسی مشروح این روش، می‌توان ارائه اساسی مطلب را در ویلکسون [۶۹] یافت. الگوریتم QR برای یافتن مقادیر ویژه یک ماتریس متقارن نیاز به این دارد که ماتریس به شکل سه قطری باشد؛ یعنی، تنها درایه‌های صفر در ماتریس روی قطر یا روی قطرهای بلافاصله بالا یا پایین قطر باشند. چون از یک ماتریس متقارن این شکل انتظار نمی‌رود، ابتدا از یک روند، به نام روش هاوس هولدر^۱، برای یافتن یک ماتریس سه قطری

1. Householder

که همان مقادیر ویژه^۶ ماتریس اصلی را دارد استفاده می شود. قضیه^۶ زیر امکان این امر را از حیث نظری تضمین می کند.

قضیه^۶ ۴۴.۰۸ (شور^۱). فرض کنیم A یک ماتریس $n \times n$ دلخواه باشد. ماتریس نامنفردی چون U هست بطوری که

$$UAU^{-1} = T,$$

که در آن T یک ماتریس بالا مثلثی است که درایه های قطریش مقادیر ویژه^۶ A می باشند.

برهان این قضیه را می توان در رالستون [۴۹]، صفحه ۴۷۱، یافت. با آنکه روشهایی برای ساختن ماتریس U که UAU^{-1} بالا مثلثی شد وجود دارند، متداولترین روش بکاررفته برای ماتریس حقیقی A ساختن ماتریس متعامد U است بطوری که $T = UAU^{-1}$ به شکل بالا هسنبرگی^۲ باشد، یعنی، همه درایه های پایین زیر قطر صفر باشند (به ازای $i, j = 0, i \geq j + 2$). روش هاوس هولدر را می توان در این ساختن بکار برد. سپس الگوریتم QR را می توان برای تقریب مقادیر ویژه^۶ ماتریس بالا هسنبرگی T ، که همان مقادیر ویژه^۶ A اند، بکار گرفت.

برای ماتریس متقارن A ، قضیه^۶ ۳۹.۰۸، ص ۵۵۵، وجود ماتریس متعامدی چون U که $T = UAU^{-1}$ یک ماتریس قطری باشد را تضمین می کند.

متأسفانه، برهان قضیه^۶ ۴۴.۰۸ ساختنی نیست و در حالت کلی بدست آوردن ماتریس U و در نتیجه T ، در صورت امکان، بسیار مشکل است. به جای ساختن ماتریس U ، روش هاوس هولدر دنباله ای از ماتریسهای متعامد متقارن P_1, P_2, \dots, P_{n-1} را می سازد بطوری که حاصل ضرب

$$B = P_{n-1} \cdots P_2 P_1 A P_1 P_2 \cdots P_{n-1}$$

یک ماتریس متقارن سه قطری با همان مقادیر ویژه^۶ A بدست می دهد.

گرچه روش هاوس هولدر یا الگوریتم QR را به تفصیل ارائه نمی کنیم، اما طرز انجام این روشها را به طور کلی توصیف می نمایم.

با مفروض بودن ماتریس $n \times n$ متقارن A ، یک ماتریس متعامد و متقارن

$$P_1 = I - 2w^{(1)}[w^{(1)}]^t$$

که در آن

$$w_1^{(1)} = 0 \text{ و } [w^{(1)}]^t w^{(1)} = 1$$

ساخته می شود یا

$$w^{(1)} = \frac{1}{2r} (0, a_{21} + s, a_{31}, \dots, a_{n1})^t \in R^n,$$

که در آن

$$s = \text{sign}(a_{21}) \left(\sum_{i=2}^n a_{i1}^2 \right)^{1/2}$$

و

$$r = \frac{1}{2}(2s^2 + 2sa_{21})^{1/2}.$$

بنابراین،

$$(۴۹.۸) \quad A_2 = P_1 A P_1 = P_1 A P_1^t = P_1 A P_1^{-1} = \begin{bmatrix} a_{11}^{(2)} & a_{12}^{(2)} & 0 & \dots & 0 \\ a_{21}^{(2)} & a_{22}^{(2)} & a_{23}^{(2)} & \dots & a_{2n}^{(2)} \\ 0 & a_{32}^{(2)} & a_{33}^{(2)} & \dots & a_{3n}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(2)} & a_{n3}^{(2)} & \dots & a_{nn}^{(2)} \end{bmatrix}$$

همان مقادیر ویژه $A_1 \equiv A$ را داراست. (ر.ک. تمرین ۰۶.)

با پیدا کردن P_1 و محاسبه A_2 ، روند برای یافتن ماتریس متعامد و متقارن

A_2 ، که در آن $[w^{(2)}]^t w^{(2)} = 1$ و $w_1^{(2)} = w_2^{(2)} = 0$ با این خاصیت

که A_3 دارای همان مقادیر ویژه A_2 و A است تکرار می شود و

$$A_3 = P_2 A_2 P_2 = \begin{bmatrix} a_{11}^{(3)} & a_{12}^{(3)} & 0 & \dots & 0 \\ a_{21}^{(3)} & a_{22}^{(3)} & a_{23}^{(3)} & 0 & \dots & 0 \\ 0 & a_{32}^{(3)} & a_{33}^{(3)} & a_{34}^{(3)} & \dots & a_{3n}^{(3)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & a_{n3}^{(3)} & a_{n4}^{(3)} & \dots & a_{nn}^{(3)} \end{bmatrix}$$

با ادامه به این طریق، ماتریس سه قطری و متقارن A_{n-1} با همان مقادیر ویژه A تشکیل

خواهد شد. الگوریتم زیر روش هاوس هولدر را آنچنان که تشریح شد انجام می دهد، اما

ضربهای ماتریسی کسل کننده اند .

الگوریتم هاوس هولدر ۸.۸

برای بدست آوردن یک ماتریس سه قطری متقارن A_{n-1} متشابه با ماتریس متقارن $A \equiv A_1$ ، ماتریسهای A_2, A_3, \dots, A_{n-1} را می سازیم که در آنها ، به ازای هر $k = 1, 2, \dots, n-1$ ، $A_k = (a_{ij}^{(k)})$.

مرحله ۱. قرار می دهیم $k = 1$.

مرحله ۲. قرار می دهیم $q^2 = \sum_{i=k+1}^n (a_{ik}^{(k)})^2$.

مرحله ۳. اگر $q^2 = 0$ ، قرار می دهیم $A_{k+1} = A_k$ ، ۱ رابه k می افزاییم ، و به مرحله ۲ می رویم .

مرحله ۴. اگر $q^2 \neq 0$ ، قرار می دهیم $s = \frac{a_{k+1,k}^{(k)}}{|a_{k+1,k}^{(k)}|} \sqrt{q^2}$ ،

مرحله ۵. قرار می دهیم

$$r^2 = \frac{1}{2}(s^2 + sa_{k+1,k}^{(k)}) .$$

مرحله ۶. $w = (w_1, w_2, \dots, w_n)^t$ را حساب می کنیم ، که در آن

$$w_i = \begin{cases} 0 & , i = 1, 2, \dots, k \text{ اگر} \\ a_{k+1,k}^{(k)} + s & , i = k + 1 \text{ اگر} \\ a_{ik}^{(k)} & , i = k + 2, \dots, n \text{ اگر} \end{cases}$$

مرحله ۷. $v = (v_1, v_2, \dots, v_n)^t$ را ، که در آن

$$v_i = \frac{1}{2r^2} \sum_{j=k+1}^n a_{ij}^{(k)} w_j , \quad i = 1, 2, \dots, n$$

حساب می کنیم .

مرحله ۸. $z = (z_1, z_2, \dots, z_n)^t$ را ، که در آن

$$z_i = v_i - \frac{1}{4r^2} \left(\sum_{j=k+1}^n w_j v_j \right) w_i , \quad i = 1, 2, \dots, n$$

حساب می کنیم

مرحله ۹. $A_{k+1} = (a_{ij}^{(k+1)})$ را ، که در آن به ازای هر $i = 1, 2, \dots, n$ و $j = 1, 2, \dots, n$.

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - w_i z_j - z_i w_j .$$

می سازیم .

- مرحله ۱۰. ۱ را به k می‌افزاییم .
 مرحله ۱۱. اگر $k < n - 1$ ، به مرحله ۲ می‌رویم .
 اگر $k = n - 1$ ، به مرحله ۱۲ می‌رویم .
 مرحله ۱۲. روند تمام است

مثال ۱. ماتریس 4×4

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

مقارن است. برای استفاده از الگوریتم ۸.۸ جهت تبدیل این ماتریس به یک ماتریس که مقارن سه قطری بوده و همان مقادیر ویژه A را دارا باشد، ابتدا حساب می‌کنیم

$$q^2 = \sum_{i=2}^4 (a_{i1})^2 = 1 + (-2)^2 + (2)^2 = 9.$$

$$\text{پس } r^2 = \frac{1}{|1|} \sqrt{q^2} = 3 \text{ ، } s = \frac{1}{|1|} \sqrt{q^2} = 3 \text{ ، } r^2 = \frac{1}{2}(s^2 + a_{21} \cdot s) = 6$$

$$w = (w_1, w_2, w_3, w_4)^t = (0, 4, -2, 2)^t.$$

از مرحله ۷ الگوریتم درمی‌یابیم که

$$v_1 = \frac{1}{\sqrt{2}}[(4)(1) + (-2)(-2) + (2)(2)] = 1,$$

$$v_2 = \frac{1}{\sqrt{2}}[(4)(2) + (-2)(0) + (2)(1)] = \frac{5}{\sqrt{2}},$$

$$v_3 = \frac{1}{\sqrt{2}}[(4)(0) + (-2)(3) + (2)(-2)] = -\frac{5}{\sqrt{2}},$$

$$v_4 = \frac{1}{\sqrt{2}}[(4)(1) + (-2)(-2) + (2)(-1)] = \frac{1}{\sqrt{2}};$$

در نتیجه ،

$$v = (1, \frac{5}{\sqrt{2}}, -\frac{5}{\sqrt{2}}, \frac{1}{\sqrt{2}}).$$

از مرحله ۸ ملاحظه می‌شود که

$$\begin{aligned} z &= v - \frac{1}{24}[(\frac{5}{\sqrt{2}})(4) + (-\frac{5}{\sqrt{2}})(-2) + (\frac{1}{\sqrt{2}})(2)]w \\ &= (1, -\frac{1}{6}, -\frac{1}{3}, 0)^t, \end{aligned}$$

و درایه‌های A_2 را می‌توان از مرحله ۹ حساب کرد که عبارتند از

$$A_2 = \begin{bmatrix} 4 & -3 & 0 & 0 \\ -3 & \frac{10}{3} & 1 & \frac{4}{3} \\ 0 & 1 & \frac{5}{3} & -\frac{4}{3} \\ 0 & \frac{4}{3} & -\frac{4}{3} & -1 \end{bmatrix}$$

در محاسبه A_3 در می‌یابیم که

$$s = \frac{5}{3}, \quad w = (0, 0, \frac{8}{3}, \frac{4}{3})'$$

$$z = (0, 1, \frac{14}{25}, -\frac{28}{25})', \quad \text{و} \quad v = (0, 1, \frac{3}{4}, -\frac{11}{10})'$$

بنابراین،

$$A_3 = \begin{bmatrix} 4 & -3 & 0 & 0 \\ -3 & \frac{10}{3} & -\frac{5}{3} & 0 \\ 0 & -\frac{5}{3} & -\frac{99}{75} & \frac{68}{75} \\ 0 & 0 & \frac{68}{75} & \frac{149}{75} \end{bmatrix}$$

ماتریس سه قطری حاصل است.

تا پایان این بخش فرض می‌شود که ماتریسی که مقادیر ویژه‌اش باید حساب شود متقارن و سه قطری باشد، بطوری که هر وقت چنین نباشد، ابتدا روش هاوس هولدر بکار گرفته می‌شود. اگر A را نمایانگر ماتریسی از این نوع بگیریم، می‌توانیم با شماره‌گذاری درایه‌های A به قرار زیر، نمادگذاری را تا حدودی ساده کنیم.

$$A = \begin{bmatrix} a_1 & b_2 & 0 & \dots & 0 \\ b_2 & a_2 & b_3 & & \vdots \\ 0 & b_3 & a_3 & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \dots & 0 & b_n & a_n \end{bmatrix}$$

اولین بررسی آن است که، وقتی به ازای مقداری از z که $2 \leq z \leq n$ ، $b_j = 0$ ، مسئله را می‌توان به بررسی ماتریسهای کوچکتر ذیل، به جای A ، تقلیل داد:

(۵۰۰.۸)

$$\begin{bmatrix} a_1 & b_2 & 0 & \dots & 0 \\ b_2 & a_2 & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & \ddots & b_{j-1} \\ 0 & \dots & 0 & b_{j-1} & a_{j-1} \end{bmatrix}$$

$$\begin{bmatrix} a_j & b_{j+1} & 0 & \dots & 0 \\ b_{j+1} & a_{j+1} & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & \ddots & b_n \\ 0 & \dots & 0 & b_n & a_n \end{bmatrix}$$

و

اگر چنین حالتی نباشد، الگوریتم QR با محاسبه مقادیر ویژه μ_1 و μ_2 از ماتریس

$$E = \begin{bmatrix} a_{n-1} & b_n \\ b_n & a_n \end{bmatrix}$$

شروع می شود.

فرض کنیم λ نزدیکترین مقدار ویژه E به a_n باشد؛ یعنی، طوری که

$$|\lambda - a_n| = \min\{|\mu_1 - a_n|, |\mu_2 - a_n|\}.$$

در این صورت، مقادیر ویژه ماتریس

$$A_1 = A - \lambda I = \begin{bmatrix} x_1 & b_2 & 0 & \dots & 0 \\ b_2 & d_2 & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & \ddots & b_n \\ 0 & \dots & 0 & b_n & d_n \end{bmatrix}$$

که در آن $x_1 = a_1 - \lambda$, $d_2 = a_2 - \lambda$, ..., $d_n = a_n - \lambda$ هستند که به اندازه λ واحد انتقال یافته اند. مطلوبیت این روند انتقال در شتاب بخشیدن به همگرایی الگوریتم QR نهفته است.

ایده‌ای که پشتوانه الگوریتم QR است تبدیل ماتریس $n \times n$ ، A به یک جفت ماتریس $(j-1) \times (j-1)$ و $(n-j+1) \times (n-j+1)$ ، به‌ازای مقداری از j که $2 \leq j \leq n$ ، است که چنین نمایش داده می‌شوند.

$$\hat{A}^{(1)} = \begin{bmatrix} \hat{a}_1^{(1)} & \hat{b}_2^{(1)} & 0 & \dots & 0 \\ \hat{b}_2^{(1)} & \hat{a}_2^{(1)} & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \hat{b}_{j-1}^{(1)} & \hat{a}_{j-1}^{(1)} \end{bmatrix}$$

$$\hat{A}^{(2)} = \begin{bmatrix} \hat{a}_j^{(2)} & \hat{b}_{j+1}^{(2)} & 0 & \dots & 0 \\ \hat{b}_{j+1}^{(2)} & \hat{a}_{j+1}^{(2)} & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \hat{b}_n^{(2)} & \hat{a}_n^{(2)} \end{bmatrix}$$

با تبدیل درایه z م زیر قطر به صفر، به طریقی که مقادیر ویژه $\hat{A}^{(1)}$ و $\hat{A}^{(2)}$ همانند مقادیر ویژه A باشند. سپس این روند روی ماتریسهای $\hat{A}^{(1)}$ و $\hat{A}^{(2)}$ اعمال می‌شود تا چهار ماتریس که سه قطری هستند بدست آید و به همین ترتیب تا سرانجام ماتریسهایی با درجه به حد کافی تقلیل یافته که مقادیر ویژه آنها، و در نتیجه مقادیر ویژه ماتریس اصلی A ، را بتوان مستقیماً پیدا کرد بدست آیند.

برای اجرای روند QR، دنباله‌ای از ماتریسهای متعامد P_2, P_3, \dots, P_n با خواص

زیر می‌سازیم:

$$(1) \quad R = P_n P_{n-1} \dots P_2 (A - \lambda I) \text{ بالا مثلثی باشد؛}$$

$$(2) \quad Q = P_n P_{n-1} \dots P_2 \text{ به‌ازای}$$

$$\hat{A} = RQ' + \lambda I = Q(A - \lambda I)Q' + \lambda I = QAQ'$$

با A متشابه و سه قطری باشد؛

(3) \hat{A} تقریباً "قطری یا حداقل دارای یک عنصر زیر قطر که تقریباً صفر است باشد.

ماتریس P_2 به شکل

$$P_2 = \begin{bmatrix} c_2 & s_2 & 0 & \cdots & 0 \\ -s_2 & c_2 & 0 & & \vdots \\ 0 & 0 & 1 & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix}$$

برای قراردادن یک صف در محل (2, 1) از $A_1 = A - \lambda I$ بکار می‌رود. ضرب $P_2 A_1 = A_2$ را می‌توان چنین بیان کرد:

$$(10.8) \quad \begin{bmatrix} z_1 & q_1 & r_1 & 0 & \cdots & 0 \\ 0 & x_2 & y_2 & 0 & \ddots & \vdots \\ 0 & b_3 & d_3 & b_4 & \ddots & 0 \\ \vdots & & & & \ddots & 0 \\ 0 & \cdots & 0 & b_n & d_n \end{bmatrix}$$

$$= \begin{bmatrix} c_2 & s_2 & 0 & \cdots & 0 \\ -s_2 & c_2 & 0 & & \vdots \\ 0 & 0 & 1 & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_1 & b_2 & 0 & \cdots & 0 \\ b_2 & d_2 & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & \ddots & b_n \\ 0 & \cdots & 0 & b_n & d_n \end{bmatrix}$$

تا معادلات زیر حاصل شود

$$z_1 = c_2 x_1 + s_2 b_2,$$

$$q_1 = c_2 b_2 + s_2 d_2,$$

$$r_1 = s_2 b_3,$$

$$0 = -s_2 x_1 + c_2 b_2,$$

$$x_2 = -s_2 b_2 + c_2 d_2,$$

$$y_2 = c_2 b_3.$$

چون c_2 و s_2 مجهولند، می‌توانیم از معادله

$$(۵۲.۸) \quad 0 = -s_2 x_1 + c_2 b_2$$

و خواسته اضافه

$$(۵۳.۸) \quad c_2^2 + s_2^2 = 1$$

جهت تعیین c_2 و s_2 ، بطوری که (۵۱.۸) بتواند برقرار باشد، استفاده کنیم. انتخابهای

$c_2 = x_1(x_1^2 + b_2^2)^{-1/2}$ و $s_2 = b_2(x_1^2 + b_2^2)^{-1/2}$ اطمینان می‌دهند که (۵۲.۸) و

(۵۳.۸) برقرارند و نیز مقادیر z_1, q_1, r_1, x_2, y_2 را مشخص می‌کنند. با ادامه

این روند حاصل می‌شود

$$A_{k-1} = \begin{bmatrix} z_1 & & q_1 & r_1 & 0 & \cdots & 0 \\ & \ddots & & & & & \\ 0 & z_{k-2} & & q_{k-2} & r_{k-2} & \cdots & \vdots \\ & & \ddots & & & & \\ 0 & 0 & & x_{k-1} & y_{k-1} & 0 & 0 \\ & & & b_k & d_k & b_{k+1} & 0 \\ \vdots & & & & & \ddots & b_n \\ & & & & & & d_n \\ 0 & \cdots & & 0 & & b_n & d_n \end{bmatrix}$$

و اسخاب می‌کنیم

$$(۵۴.۸)$$

$$P_k = \left[\begin{array}{cccc|cccc} 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \ddots & 0 & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & c_k & s_k & 0 & \cdots & 0 & \\ 0 & \cdots & 0 & -s_k & c_k & 0 & \cdots & 0 & \\ \hline 0 & \cdots & 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & \vdots & \vdots & 0 & \ddots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{array} \right] \xrightarrow[\mu^k]{\text{سطر}} \left[\begin{array}{ccc|ccc} I_{k-2} & O & O & & & \\ \hline O & c_k & s_k & & & O \\ O & -s_k & c_k & & & \\ \hline O & O & I_{n-k} & & & \end{array} \right]$$

بطوری که درایه $(k, k-1)$ از A_{k-1} صفر می شود؛ یعنی،

$$A_k = P_k A_{k-1} = \begin{bmatrix} z_1 & q_1 & r_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & z_{k-1} & q_{k-1} & r_{k-1} & \cdots & \vdots \\ 0 & \vdots & 0 & x_k & y_k & 0 & 0 \\ \vdots & \vdots & \vdots & b_{k+1} & d_{k+1} & b_{k+2} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & b_n \\ 0 & \cdots & 0 & 0 & b_n & d_n \end{bmatrix}$$

ضرب ماتریسی نموده شده، با استفاده از مقادیر

$$s_k = b_k(x_{k-1}^2 + b_k^2)^{-1/2} \quad \text{و} \quad c_k = x_{k-1}(x_{k-1}^2 + b_k^2)^{-1/2}$$

نتیجه می دهد

$$z_{k-1} = (x_{k-1}^2 + b_k^2)^{1/2},$$

$$q_{k-1} = \frac{x_{k-1}y_{k-1} + d_k b_k}{z_{k-1}},$$

$$r_{k-1} = \frac{b_k b_{k+1}}{z_{k-1}},$$

$$x_k = \frac{-y_{k-1}b_k + d_k x_{k-1}}{z_{k-1}}$$

$$y_k = \frac{x_{k-1}b_{k+1}}{z_{k-1}}.$$

سرانجام،

$$P_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & 0 \\ \vdots & & 0 & c_n & s_n \\ 0 & \cdots & 0 & -s_n & c_n \end{bmatrix}$$

ساخته می شود بطوری که

$$R = A_n = P_n A_{n-1} = \begin{bmatrix} z_1 & q_1 & r_1 & 0 & \dots & 0 \\ 0 & & & & & \vdots \\ \vdots & & & & & 0 \\ \vdots & & & & & r_{n-2} \\ 0 & \dots & & z_{n-1} & q_{n-1} \\ 0 & & & 0 & x_n \end{bmatrix}$$

تشکیل شود .

هرگاه روند خلاصه شده در بالا نتیجه رضایت بخشی در قسمت (۳) ندهد ، آنگاه الگوریتم مجدداً روی ماتریس \hat{A} ، به جای A ، اعمال می شود تا اینکه درایه زیر قطر بحد کافی کوچک شود .

هنگامی که یک عنصر رضایت بخش در زیر قطر بدست آمد ، روند را می توان مجدداً روی دو ماتریس سه قطری ، که از تجزیه ماتریس \hat{A} در سطر و ستون شامل این درایه نتیجه می شوند ، اعمال کرد .

الگوریتم زیرروش بکاررفته برای تبدیل یک ماتریس سه قطری به دو ماتریس سه قطری با مرتبه کمتر و با همان مقادیر ویژه ماتریس اولیه را شرح می دهد . این الگوریتم حالتی که A را می توان به دو دستگاه از نوع فهرست شده در معادله (۵۰.۸) تجزیه کرد را نیز در نظر می گیرد ، اما اگر A را بتوان به این طریق تجزیه کرد ، الگوریتم ۷۰.۸ لازم نیست .

۹۰.۸ الگوریتم QR

برای تحویل ماتریس متقارن سه قطری

$$A \equiv \hat{A}_1 \equiv \begin{bmatrix} a_1^{(1)} & b_2^{(1)} & 0 & \dots & 0 \\ b_2^{(1)} & a_2^{(1)} & & & \vdots \\ 0 & & & & 0 \\ \vdots & & & & b_n^{(1)} \\ 0 & \dots & 0 & b_n^{(1)} & a_n^{(1)} \end{bmatrix}$$

به دو ماتریس متقارن سه قطری با بعد کمتر از n ، که دارای همان مقادیر ویژه A باشند .

مرحله ۱. قرار می دهیم $k = 1$.

مرحله ۲. هرگاه $b_2^{(k)} \approx 0$ ، آنگاه $a_1^{(k)}$ تقریبی به مقدار ویژه A است و می توانیم

ماتریس \hat{A}_k را با ماتریسی که از حذف سطر و ستون اول \hat{A}_k حاصل می شود جایگزین کنیم.

همچنین، عدد صحیح n را با $n - 1$ جایگزین می کنیم. هرگاه $b_n^{(k)} \approx 0$ ، آنگاه $a_n^{(k)}$

تقریبی به مقدار ویژه A است و می توانیم ماتریس \hat{A}_k را با ماتریس حاصل از حذف سطر

و ستون آخر \hat{A}_k جایگزین کنیم. همچنین، عدد صحیح n را با $n - 1$ جایگزین می کنیم.

هرگاه به ازای عدد صحیحی چون $j = 3, 4, \dots, n - 1$ ، $b_j^{(k)} \approx 0$ ، به مرحله ۱۶ می رویم.

مرحله ۳. مقادیر ویژه μ_1 و μ_2 از ماتریس $\begin{bmatrix} a_{n-1}^{(k)} & b_n^{(k)} \\ b_n^{(k)} & a_n^{(k)} \end{bmatrix}$ را حساب کرده و λ را طوری

انتخاب می کنیم که

$$|\lambda - a_n^{(k)}| = \text{minimum}\{|\mu_1 - a_n^{(k)}|, |\mu_2 - a_n^{(k)}|\}.$$

مرحله ۴. قرار می دهیم

$$A_1 = \hat{A}_k - \lambda I = \begin{bmatrix} x_1 & y_1 & 0 & \dots & 0 \\ b_2 & d_2 & b_3 & & \vdots \\ 0 & & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & b_n \\ 0 & \dots & 0 & b_n & d_n \end{bmatrix},$$

که در آن $x_1 = a_1 - \lambda$ ، $y_1 = b_2$ ، و به ازای هر $i = 2, 3, \dots, n$ ، $d_i = a_i - \lambda$.

مرحله ۵. قرار می دهیم $j = 2$.

مرحله ۶. محاسبه می کنیم

$$z_{j-1} = \sqrt{x_{j-1}^2 + b_j^2},$$

$$c_j = \frac{x_{j-1}}{z_{j-1}},$$

$$s_j = \frac{b_j}{z_{j-1}},$$

$$q_{j-1} = c_j y_{j-1} + s_j d_j,$$

$$x_j = -s_j y_{j-1} + c_j d_j.$$

مرحله ۷. اگر $z = n$ ، به مرحله ۹ می‌رویم.

مرحله ۸. قرار می‌دهیم $r_{j-1} = s_j b_{j+1}$ ،

$$y_j = c_j b_{j+1}.$$

مرحله ۹. P_j را تعریف می‌کنیم:

$$P_j = \begin{bmatrix} 1 & 0 & & \dots & & & 0 \\ 0 & \ddots & & & & & \vdots \\ & & 1 & 0 & & \dots & 0 \\ & & & & & & \vdots \\ & & & & & & \vdots \\ & & 0 & c_j & s_j & 0 & \dots & 0 \\ \vdots & & & & & & & \vdots \\ & & & -s_j & c_j & 0 & 0 & \leftarrow \text{سطر } m \\ & & & & & & & \vdots \\ & & & & 0 & 0 & 1 & \ddots \\ & & & & \vdots & \vdots & & \ddots \\ 0 & \dots & 0 & 0 & 0 & & 0 & 1 \end{bmatrix}$$

\uparrow
 سطر j

مرحله ۱۰. قرار می‌دهیم $A_j = P_j A_{j-1}$

$$= \begin{bmatrix} z_1 & q_1 & r_1 & 0 & & & \dots & & 0 \\ 0 & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ & & & & & & & & \vdots \\ & & & & z_{j-1} & q_{j-1} & r_{j-1} & & \\ 0 & \dots & 0 & x_j & y_j & 0 & & & \\ 0 & \dots & 0 & b_{j+1} & d_{j+1} & b_{j+2} & 0 & \dots & 0 \\ \vdots & & & & & & & & 0 \\ & & & & & & & & b_n \\ 0 & \dots & & & & & 0 & b_n & d_n \end{bmatrix}$$

مرحله ۱۱. ۱ را به z می‌افزاییم.

مرحله ۱۲. اگر $z \leq n$ ، به مرحله ۶ می‌رویم.

اگر $z > n$ ، به مرحله ۱۳ می‌رویم .

مرحله ۱۳ . حساب می‌کنیم $Q_k = P_n P_{n-1} \dots P_2$.

مرحله ۱۴ . حساب می‌کنیم $\tilde{A}_{k+1} = A_n Q_k^t + \lambda I$.

$$\equiv \begin{bmatrix} a_1^{(k+1)} & b_2^{(k+1)} & 0 & \dots & 0 \\ b_2^{(k+1)} & a_2^{(k+1)} & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & \ddots & b_n^{(k+1)} \\ 0 & \dots & 0 & b_n^{(k+1)} & a_n^{(k+1)} \end{bmatrix}$$

مرحله ۱۵ . به مرحله ۲ می‌رویم .

مرحله ۱۶ . روند تمام است . ماتریسهای تحویل یافته عبارتند از

$$\begin{bmatrix} a_1^{(k+1)} & b_2^{(k+1)} & 0 & \dots & 0 \\ b_2^{(k+1)} & a_2^{(k+1)} & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & \ddots & b_{j-1}^{(k+1)} \\ 0 & \dots & 0 & b_{j-1}^{(k+1)} & a_{j-1}^{(k+1)} \end{bmatrix}$$

و

$$\begin{bmatrix} a_j^{(k+1)} & b_{j+1}^{(k+1)} & 0 & \dots & 0 \\ b_{j+1}^{(k+1)} & a_{j+1}^{(k+1)} & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & \ddots & b_n^{(k+1)} \\ 0 & \dots & 0 & b_n^{(k+1)} & a_n^{(k+1)} \end{bmatrix}$$

مثال ۲ . فرض کنیم

$$A = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 3 \end{bmatrix} = \begin{bmatrix} a_1 & b_2 & 0 \\ b_2 & a_2 & b_3 \\ 0 & b_3 & a_3 \end{bmatrix}$$

برای یافتن پارامتر شتاب λ بدست آوردن مقادیر ویژه

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

لازم است، که می‌توان نشان داد $\mu_1 = 2$ و $\mu_2 = 4$. انتخاب نزدیکترین مقدار ویژه به $a_3 = 3$ در این حالت اختیاری است، و $\mu_1 = 2$ را انتخاب می‌کنیم. با نمادگذاری الگوریتم ۹.۸،

$$A - \lambda I = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 0 \\ b_2 & d_2 & b_3 \\ 0 & b_3 & d_3 \end{bmatrix}$$

با انجام مرحله‌های ۱۰-۶ از آن الگوریتم بدست می‌آوریم

$$z_1 = \sqrt{x_1^2 + b_2^2} = \sqrt{2}, \quad c_2 = \frac{x_1}{z_1} = \frac{\sqrt{2}}{2},$$

$$s_2 = \frac{b_2}{z_1} = \frac{\sqrt{2}}{2}, \quad q_1 = c_2 y_1 + s_2 d_2 = \sqrt{2}.$$

$$r_1 = s_2 b_3 = \frac{\sqrt{2}}{2}, \quad x_2 = -s_2 y_1 + c_2 d_2 = 0,$$

$$y_2 = c_2 b_3 = \frac{\sqrt{2}}{2} \quad \text{و}$$

بنابراین،

$$A_2 = \begin{bmatrix} \sqrt{2} & \sqrt{2} & \frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 1 \end{bmatrix} \quad \text{و} \quad P_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

با انجام مجدد مرحله‌های ۱۰-۶ با $j = 3$ بدست می‌آوریم

$$z_2 = \sqrt{x_2^2 + b_3^2} = 1, \quad c_3 = \frac{x_2}{z_2} = 0,$$

$$s_3 = \frac{b_3}{z_2} = 1, \quad q_2 = c_3 y_2 + s_3 d_3 = 1,$$

$$x_3 = -s_3 y_2 + c_3 d_3 = -\frac{\sqrt{2}}{2},$$

$$A_3 = \begin{bmatrix} \sqrt{2} & \sqrt{2} & \frac{\sqrt{2}}{2} \\ 0 & 1 & 1 \\ 0 & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix} \text{ و } P_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

از مرحله‌های ۱۳ و ۱۴ الگوریتم داریم

$$Q_1 = P_3 P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

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

$$\hat{A}_2 = A_3 Q_1' + 2I = \begin{bmatrix} \sqrt{2} & \sqrt{2} & \frac{\sqrt{2}}{2} \\ 0 & 1 & 1 \\ 0 & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \end{bmatrix}$$

$$+ \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & \frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & 2 & -\frac{\sqrt{2}}{2} \\ 0 & -\frac{\sqrt{2}}{2} & 0 \end{bmatrix}.$$

و بایک تکرار الگوریتم QR روند تمام است. چون هیچیک از $b_2^{(2)} = \frac{\sqrt{2}}{2}$ و $b_3^{(2)} = -\frac{\sqrt{2}}{2}$ را نمی توان نزدیک به صفر تلفی کرد، تکرار دیگری از الگوریتم QR، با بکار بردن \hat{A}_2 به جای $A \equiv \hat{A}_1$ ، انجام می شود.

انجام این تکرار نتیجه می دهد

$$\hat{A}_3 = \begin{bmatrix} 4.30602 & .37598 & 0 \\ .37598 & 3.10759 & .030405 \\ 0 & .030405 & 1.58641 \end{bmatrix},$$

و $b_3^{(3)} = .030405 \approx 0$. نتیجتاً، ماتریسهای تحویل یافته‌ای که باید در نظر گرفت عبارتند از ماتریس 2×2

$$\begin{bmatrix} 4.30602 & .37598 \\ .37598 & 3.10759 \end{bmatrix}$$

و ماتریس 1×1

$$[1.58641].$$

ماتریس آخری بوضوح دارای یک مقدار ویژه 1.58641 است، و می توان نشان داد که مقادیر ویژه 2×2 عبارتند از 4.41420 و 2.99941 . چون مقادیر ویژه واقعی A عبارتند از 1.58579 ، 4.41420 ، و 3.00000 ، الگوریتم QR در این مثال دقت چهار رقم با معنا را بدست می دهد.

این بخش را با یک قضیه که مجوز این فرض است که می توان انتظار همگرایی برای الگوریتم QR را داشت می بندیم. این قضیه را می توان در ویلکسون [۶۹]، صفحات ۵۱۹-۵۱۶، یافت.

قضیه ۴۵۰۸. هرگاه A یک ماتریس $n \times n$ متقارن سه قطری بوده و $\hat{A}_m = Q_m Q_{m-1} \cdots Q_1 A Q_1' \cdots Q_{m-1}' Q_m'$ تکرار m مد الگوریتم QR باشد:

$$\hat{A}_m = \begin{bmatrix} a_1^{(m)} & b_2^{(m)} & 0 & \cdots & 0 \\ b_2^{(m)} & a_2^{(m)} & & & \vdots \\ 0 & & \ddots & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & 0 & b_n^{(m)} & a_n^{(m)} \end{bmatrix};$$

نگاه ۶: $\lim_{m \rightarrow +\infty} b_n^{(m)} = 0$ و $a_n^{(m)}$ به یک مقدار ویژه A همگراست.

مجموعه تمرینات ۵۰۸

۱. روش هاوس هولدر را برای پیدا کردن یک ماتریس سه قطری متقارن B متشابه با

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

بکار برید.

۲. با استفاده از روش هاوس هولدر، ماتریسهای زیر را به شکل سه قطری تحویل کنید.

$$A = \begin{bmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{bmatrix} \quad (\text{ت})$$

$$B = \begin{bmatrix} 6 & 1 & 0 & -1 \\ 1 & 5 & 1 & 1 \\ 0 & 1 & -4 & 2 \\ -1 & 1 & 2 & 5 \end{bmatrix} \quad (\text{ب})$$

$$C = \begin{bmatrix} 8 & .25 & .5 & 2 & -1 \\ .25 & -4 & 0 & 1 & 2 \\ .5 & 0 & 5 & .75 & -1 \\ 2 & 1 & .75 & 5 & -.5 \\ -1 & 2 & -1 & -.5 & 6 \end{bmatrix} \quad (\text{پ})$$

۳. ثابت کنید هرگاه P_1, \dots, P_n ماتریسهای متعامد باشند، آنگاه $Q = P_1 : P_2 \cdots P_n$ نیز متعامد است.

۴. اگر $b_n = 0$ ، نشان دهید که a_n یک مقدار ویژه

$$A = \begin{bmatrix} a_1 & b_2 & 0 & \cdots & 0 \\ b_2 & a_2 & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & \ddots & b_n \\ 0 & \cdots & 0 & b_n & a_n \end{bmatrix}$$

است و نیز $(n - 1)$ مقدار ویژه^۶ دیگر A مقادیر ویژه^۶ ماتریس

$$\begin{bmatrix} a_1 & b_2 & 0 & \dots & 0 \\ b_2 & a_2 & & & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & b_{n-1} & a_{n-1} \end{bmatrix}$$

هستند.

۵. تمام مقادیر ویژه^۶ ماتریسهای زیر را با استفاده از الگوریتم QR تعیین کنید.

$$\begin{bmatrix} 4 & 2 & 0 & 0 & 0 \\ 2 & 4 & 2 & 0 & 0 \\ 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 2 & 4 & 2 \\ 0 & 0 & 0 & 2 & 4 \end{bmatrix} \quad (\text{ا})$$

$$\begin{bmatrix} 5 & -1 & 0 & 0 & 0 \\ -1 & 4.5 & .2 & 0 & 0 \\ 0 & .2 & 1 & -4 & 0 \\ 0 & 0 & -4 & 3 & 1 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix} \quad (\text{ب})$$

$$\begin{bmatrix} 4 & -3 & 0 & 0 \\ -3 & \frac{10}{3} & -\frac{5}{3} & 0 \\ 0 & -\frac{5}{3} & -\frac{99}{75} & \frac{68}{75} \\ 0 & 0 & \frac{68}{75} & \frac{149}{75} \end{bmatrix} \quad (\text{ج})$$

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & -1 & -2 \\ 0 & -2 & 3 \end{bmatrix} \quad (\text{د})$$

۶. نشان دهید که ماتریس $A_2 = P_1 A P_1$ ، که در توضیح روش هاوس هولدر ساخته شد،

متقارن و به شکل داده شده در معادله (۴۹.۸) است.

۷. یک ماتریس دوران P ماتریسی متعامد است با $\det P = 1$. نشان دهید که ماتریس

P_4 تعریف شده در (۵۴.۸) یک ماتریس دوران است بشرطی که $c_4^2 + s_4^2 = 1$.

حل عددی دستگاههای معادلات غیرخطی^۹

مقدار فشار لازم برای پایین راندن یک شی سنگین بزرگ در یک خاک نرم و همگن ، که روی پایه^{۱۰} خاکی محکمی قرار دارد ، را می توان با مقدار فشار لازم برای پایین راندن اشیاء کوچکتر در همان خاک پیشبینی کرد . بالاخص ، مقدار فشار p برای پایین راندن یک صفحه^{۱۱} مستدیر به شعاع r و فاصله^{۱۲} d در خاکی نرم ، که قسمت خاکی سخت در فاصله^{۱۳} $D > d$ از سطح خاک قرار دارد ، را می توان با معادلهای به شکل

$$p = k_1 e^{k_2 r} + k_3 r,$$

تقریب زد ، که در آن k_1 ، k_2 ، و k_3 ثابتهایی وابسته به d و سازگاری خاک هستند اما به شعاع صفحه وابسته نیستند .

جهت تعیین اندازه^{۱۴} مینیمال صفحه که تحمل بار بزرگی را داشته باشد ، سه صفحه^{۱۵} کوچک با شعاعهای متفاوت به فاصله^{۱۶} یکسان فرو برده می شوند و وزنهای لازم برای این فرو رفتنها اندازه گیری می شوند . این محاسبات منجر به سه معادله از نوع بالا نسبت به سه مجهول k_1 ، k_2 ، و k_3 می شود . اما ، بخاطر طبیعت این معادلات ، نمی توان مجهولها را به سادگی بدست آورد . روشهای عددی لازم است تا بتوان آنها را در حل دستگاههای معادلات ، وقتی خطی نیستند ، بکار گرفت .

در فصل ۲ روشهای تکرار نقطه ثابت در مطالعه حل عددی معادلات به شکل $f(x) = 0$ غالب بودند. روشهایی که برای حل دستگاههای معادلات از این نوع بکار خواهیم برد اساساً "تعمیم روش نیوتن خواهد بود. با وجودی که تکنیکهای دیگر در دسترس هستند، آنها عموماً "برحسب انواع مسائلی که حل می کنند کاملاً" اختصاصی هستند، و احتمالاً "برای ورود در یک درس مقدماتی آنالیز عددی نامناسب می باشند. اما توضیح کاملاً "مبسوطی از این روشها، که در یک سطح نسبتاً "مقدماتی ارائه شده است، را می توان در صفحات ۴۵۹ - ۳۶۱ اکتون^۱ [۲۱] یافت.

اغلب برهانهای نتایج نظری این فصل حذف خواهند شد، زیرا آنها عموماً "شامل روشهایی هستند که معمولاً" در یک درس حساب دیفرانسیل و انتگرال پیشرفته مطالعه می شوند. یک مرجع عمومی خوب برای این مواد کتاب اورتگا با عنوان دومین درس در آنالیز عددی [۴۶] می باشد، و یک مرجع کامل اورتگا و رین بولت^۲ [۴۷] است.

۱۰۹ نقاط ثابت برای توابع چند متغیره

شکل کلی یک دستگاه از معادلات غیر خطی عبارت است از

$$(109) \quad \begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0, \\ f_2(x_1, x_2, \dots, x_n) &= 0, \\ &\vdots \\ f_n(x_1, x_2, \dots, x_n) &= 0, \end{aligned}$$

1. Acton

2. Rheinboldt

که در آن هر تابع f_i فضای n بعدی R^n را بتوی خط حقیقی R می نگارد. اغلب مطلوب است که دستگاه به گونه ای دیگر با تعریف یک تابع F نمایش داد که R^n را بتوی R^n با $F(x_1, x_2, \dots, x_n) = (f_1(x_1, x_2, \dots, x_n), f_2(x_1, x_2, \dots, x_n), \dots, f_n(x_1, x_2, \dots, x_n))$ می نگارد. با استفاده از نماد بردار جهت نمایش متغیرهای x_1, x_2, \dots, x_n ، می نویسیم $x = (x_1, x_2, \dots, x_n)^t$ ، یا برای سادگی، (x_1, x_2, \dots, x_n) . لذا، معادله (۱.۹) شکل زیر را پیدا می کند:

$$(۲.۹) \quad F(x) = 0.$$

مثال ۱. دستگاه غیرخطی سه در سه

$$\begin{aligned} 3x_1 - \cos(x_2 x_3) - \frac{1}{2} &= 0, \\ x_1^2 - 81(x_2 + .1)^2 + \sin x_3 + 1.06 &= 0, \\ e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0, \end{aligned}$$

را می توان با تعریف سه تابع f_1 ، f_2 ، و f_3 از R^3 به R به صورت

$$\begin{aligned} f_1(x_1, x_2, x_3) &= 3x_1 - \cos(x_2 x_3) - \frac{1}{2}, \\ f_2(x_1, x_2, x_3) &= x_1^2 - 81(x_2 + .1)^2 + \sin x_3 + 1.06, \\ f_3(x_1, x_2, x_3) &= e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3}, \end{aligned}$$

و F از R^3 به R^3 با

$$\begin{aligned} F(x) &= F(x_1, x_2, x_3) \\ &= (f_1(x_1, x_2, x_3), f_2(x_1, x_2, x_3), f_3(x_1, x_2, x_3)) \\ &= \left(3x_1 - \cos(x_2 x_3) - \frac{1}{2}, x_1^2 - 81(x_2 + .1)^2 \right. \\ &\quad \left. + \sin x_3 + 1.06, e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} \right) \end{aligned}$$

به شکل (۲.۹) درآورد.

قبل از بحث در حل یک دستگاه که به شکل (۱.۹) یا (۲.۹) ارائه شده، لازم است نتایجی را در مورد پیوستگی و مشتق پذیری توابع از R^n بتوی R^n در نظر بگیریم. گرچه این

بررسی باید مستقیماً "ارائه شود (ر.ک. تمرین ۴) ، در عوض ترجیح می‌دهیم روش دیگری بکار ببریم که به ما اجازه می‌دهد مفاهیم نظراً "مشکلاتر حدود و پیوستگی را برحسب توابع از R^n بتوی R ارائه دهیم .

تعریف ۱۰۹ . فرض کنیم f یک تابع تعریف شده بر مجموعه $D \subset R^n$ بتوی R باشد . گوئیم تابع f در نقطه x_0 دارای حد L است ، و می‌نویسیم

$$\lim_{x \rightarrow x_0} f(x) = L,$$

اگر به‌ازای هر $\varepsilon > 0$ عددی مانند $\delta > 0$ با این خاصیت وجود داشته باشد که هر وقت $x \in D$ و

$$0 < \|x - x_0\| < \delta,$$

داشته باشیم

$$|f(x) - L| < \varepsilon.$$

باید توجه شود که وجود حد در یک نقطه ، بخاطر معادل بودن نرم‌های برداری در R^n (ر.ک. بخش ۱۰۸) ، مستقل از نرم برداری بکار رفته است .

تعریف ۲۰۹ . فرض کنیم f تابعی از مجموعه $D \subset R^n$ بتوی R باشد . تابع f را در $x_0 \in D$ پیوسته نامیم اگر $\lim_{x \rightarrow x_0} f(x)$ موجود بوده و

$$\lim_{x \rightarrow x_0} f(x) = f(x_0).$$

بعلاوه ، گوئیم f بر مجموعه D پیوسته است اگر f در هر نقطه D پیوسته باشد .

حال می‌توانیم مفاهیم حد و پیوستگی برای توابع از R^n بتوی R^n را با بررسی توابع مختصات از R^n بتوی R تعریف کنیم .

تعریف ۳۰۹ . فرض کنیم F یک تابع از $D \subset R^n$ بتوی R^n بوده و دارای نمایش

$$F(x) = (f_1(x), f_2(x), \dots, f_n(x)),$$

باشد ، که در آن f_i یک نگاشت از R^n بتوی R ، به‌ازای $i = 1, 2, \dots, n$ ، است . تعریف

می‌کنیم

$$\lim_{x \rightarrow x_0} F(x) = L = (L_1, L_2, \dots, L_n)$$

اگر و فقط اگر به‌ازای هر $i = 1, 2, \dots, n$ $\lim_{x \rightarrow x_0} f_i(x) = L_i$

تعریف ۴.۹. فرض کنیم F یک تابع از $D \subset R^n$ بتوی R^n با نمایش $F(x) = (f_1(x), f_2(x), \dots, f_n(x))$ تابع F را در $x_0 \in D$ پیوسته نامیم اگر $\lim_{x \rightarrow x_0} F(x) = F(x_0)$ موجود بوده و $\lim_{x \rightarrow x_0} F(x) = F(x_0)$ را بر D پیوسته گوئیم اگر F در هر نقطه از D پیوسته باشد.

برای توابع از R بتوی R ، پیوستگی در یک نقطه را می‌توان با اثبات اینکه تابع در آن نقطه مشتقپذیر است نشان داد (ر.ک. قضیه ۶.۱، ص ۵). این قضیه به توابع چندمتغیره تعمیم پیدا می‌کند، اما کاملاً "درگیر مشتق یا مشتق کامل یک تابع چندمتغیره است و در اینجا ارائه نخواهد شد. در عوض، قضیه زیر را که پیوستگی یک تابع n متغیره در یک نقطه را با مشتقات جزئی تابع در آن نقطه مربوط می‌کند عنوان خواهیم کرد.

قضیه ۵.۹. فرض کنیم f یک تابع از $D \subset R^n$ بتوی R بوده و $x_0 \in D$. هرگاه ثابتهای $\delta > 0$ و $K > 0$ موجود باشند بطوری که هر وقت $\|x - x_0\| < \delta$ و $x \in D$

$$\left| \frac{\partial f(x)}{\partial x_j} \right| \leq K, \quad j = 1, 2, \dots, n$$

به‌ازای هر x_0 پیوسته است.

مثال ۲. فرض کنیم $F: R^3 \rightarrow R^3$ به‌ازای $x = (x_1, x_2, x_3) \in R^3$

$$F(x) = \left(\frac{1}{3} \cos(x_2 x_3) + \frac{1}{6}, \frac{1}{9} \sqrt{x_1^2 + \sin x_3} + 1.06, -1, -\frac{1}{20} e^{-x_1 x_2} - \frac{10\pi - 3}{60} \right)$$

تعریف شده باشد. برای اثبات اینکه F بر مجموعه

$$D = \{(x_1, x_2, x_3) \mid -1 \leq x_i \leq 1, i = 1, 2, 3\}$$

پیوسته است، کافی است نشان دهیم که توابع

$$f_1(x_1, x_2, x_3) = \frac{1}{3} \cos(x_2 x_3) + \frac{1}{6},$$

$$f_2(x_1, x_2, x_3) = \frac{1}{9} \sqrt{x_1^2 + \sin x_3} + 1.06 - 1,$$

$$f_3(x_1, x_2, x_3) = -\frac{1}{20} e^{-x_1 x_2} - \frac{10\pi - 3}{60},$$

در هر نقطه از D پیوسته هستند. با گرفتن مشتقهای جزئی داریم

$$\begin{aligned} \frac{\partial f_1}{\partial x_1} &= 0, & \frac{\partial f_1}{\partial x_2} &= -\frac{1}{3} x_3 \sin(x_2 x_3), & \frac{\partial f_1}{\partial x_3} &= -\frac{1}{3} x_2 \sin(x_2 x_3), \\ \frac{\partial f_2}{\partial x_1} &= \frac{x_1}{9\sqrt{x_1^2 + \sin x_3} + 1.06}, & \frac{\partial f_2}{\partial x_2} &= 0, & \frac{\partial f_2}{\partial x_3} &= \frac{\cos x_3}{18\sqrt{x_1^2 + \sin x_3} + 1.06}, \\ \frac{\partial f_3}{\partial x_1} &= \frac{x_2}{20} e^{-x_1 x_2}, & \frac{\partial f_3}{\partial x_2} &= \frac{x_1}{20} e^{-x_1 x_2}, & \frac{\partial f_3}{\partial x_3} &= 0. \end{aligned}$$

فرض کنیم $\mathbf{x}_0 = (x_{01}, x_{02}, x_{03}) \in D$. برای اعمال قضیه ۵.۹ می‌توانیم δ

را ۱ قرار دهیم و l_r را بکار ببریم. با این نرم، $\|\mathbf{x} - \mathbf{x}_0\|_x < \delta = 1$ ، دقیقاً "وقتی

بمازای هر $i = 1, 2, 3$ ، $|x_i - x_{0i}| < 1$ ، (۳.۹)

در نتیجه، بمازای هر $\mathbf{x} = (x_1, x_2, x_3)$ در D با $\|\mathbf{x} - \mathbf{x}_0\|_\infty < 1$ ، کرانه‌های زیر برای

مشتقهای جزئی بدست می‌آیند:

$$\left| \frac{\partial f_1}{\partial x_1}(x_1, x_2, x_3) \right| = 0,$$

$$\left| \frac{\partial f_1}{\partial x_2}(x_1, x_2, x_3) \right| = \left| -\frac{1}{3} x_3 \sin(x_2 x_3) \right| \leq \frac{1}{3} |x_3| \leq \frac{1}{3} (|x_{03}| + 1) \leq \frac{2}{3},$$

$$\left| \frac{\partial f_1}{\partial x_3}(x_1, x_2, x_3) \right| = \left| -\frac{1}{3} x_2 \sin(x_2 x_3) \right| \leq \frac{1}{3} |x_2| \leq \frac{1}{3} (|x_{02}| + 1) \leq \frac{2}{3},$$

$$\left| \frac{\partial f_2}{\partial x_1}(x_1, x_2, x_3) \right| = \frac{|x_1|}{9\sqrt{x_1^2 + \sin x_3} + 1.06} \leq \frac{|x_{01}| + 1}{9\sqrt{.218}} < .48,$$

$$\left| \frac{\partial f_2}{\partial x_2}(x_1, x_2, x_3) \right| = 0,$$

$$\left| \frac{\partial f_2}{\partial x_3}(x_1, x_2, x_3) \right| = \frac{|\cos x_3|}{18\sqrt{x_1^2 + \sin x_3} + 1.06} \leq \frac{1}{18\sqrt{.218}} < .12,$$

$$\left| \frac{\partial f_3}{\partial x_1}(x_1, x_2, x_3) \right| = \left| \frac{x_2}{20} e^{-x_1 x_2} \right| \leq \frac{|x_{02}| + 1}{20} e^{|x_{01}| |x_{02}|} \leq \frac{1}{10} e^{(|x_{01}| + 1)(|x_{02}| + 1)}$$

$$= \frac{1}{10}e^4 < 5.5,$$

$$\left| \frac{\partial f_3}{\partial x_2}(x_1, x_2, x_3) \right| = \left| \frac{x_1}{20} e^{-x_1 x_2} \right| \leq \frac{|x_{01}| + 1}{20} e^{(|x_{01}| + 1)(|x_{02}| + 1)}$$

$$\leq \frac{1}{10}e^4 < 5.5,$$

$$\cdot \left| \frac{\partial f_3}{\partial x_3} \right| = 0$$

با $K = 5.5$ و $\delta = 1$ ، مفروضات قضیه ۵.۹، هر وقت $x_0 \in D$ ، توسط $f_i(x_1, x_2, x_3)$ ، بازای هر $i = 1, 2, 3$ ، برقرار است. در نتیجه، هر یک از توابع f_i ، بازای $i = 1, 2, 3$ ، بر D پیوسته است و F بر D پیوسته است.

در فصل ۲، یک روند تکراری برای حل یک معادله $f(x) = 0$ به وسیله تبدیل این معادله به معادله‌ای به شکل $x = g(x)$ ، که دقیقاً در آن نقاطی که جوابهای معادله اصلی بودند دارای نقاط ثابت بود، بدست آمد. در اینجا روندی مشابه برای توابع از R^n بتوی R^n تحقیق می‌شود.

تعریف ۶.۹. گوئیم تابع G از $R^n \supset D$ بتوی R^n دارای یک نقطه ثابت در $p \in D$ است اگر $G(p) = p$.

قضیه زیر قضیه نقطه ثابت (قضیه ۴.۲، ص ۳۹) را به حالت n بعدی تعمیم می‌دهد. این قضیه حالت خاصی از قضیه معروف نگاشت انقباض است، و برهان آن را می‌توان در کتاب اورتگا [۴۶]، صفحه ۱۵۳، که قبلاً ذکر شد، یافت.

قضیه ۷.۹. فرض کنیم بازای گردآیه‌ای از ثابتهای a_1, a_2, \dots, a_n و b_1, b_2, \dots, b_n $\{ \text{بازای هر } i = 1, 2, \dots, n, a_i \leq x_i \leq b_i \}$ و نیز G تابع پیوسته‌ای با مشتقهای جزئی اول پیوسته از $R^n \supset D$ بتوی R^n باشد با این خاصیت که هر وقت $x \in D$ ، $G(x) \in D$. در این صورت، G دارای یک نقطه ثابت در D است.

بعلاوه، هرگاه یک ثابت $K < 1$ موجود باشد که هر وقت $x \in D$

$$\left| \frac{\partial g_i(x)}{\partial x_j} \right| \leq \frac{K}{n}, \quad g_i \text{ هر تابع مولفه‌ای } z = 1, 2, \dots, n$$

نگاه نقطه ثابت منحصر بفرد است.

مثال ۳. دستگاه غیرخطی مثال ۱ که با

$$3x_1 - \cos(x_2 x_3) - \frac{1}{2} = 0,$$

$$(۴۰۹) \quad x_1^2 - 81(x_2 + .1)^2 + \sin x_3 + 1.06 = 0,$$

$$e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0.$$

داده شده بود را در نظر می‌گیریم. اگر معادله i م نسبت به x_i حل شود، دستگاه را می‌توان به مسئله نقطه ثابت زیر تغییر داد:

$$x_1 = \frac{1}{3} \cos(x_2 x_3) + \frac{1}{6},$$

$$x_2 = \frac{1}{9} \sqrt{x_1^2 + \sin x_3 + 1.06} - .1,$$

$$x_3 = -\frac{1}{20} e^{-x_1 x_2} - \frac{10\pi - 3}{60}.$$

فرض کنیم $G: R^3 \rightarrow R^3$ با $G(x) = (g_1(x), g_2(x), g_3(x))$ تعریف شده باشد، که در آن

$$g_1(x_1, x_2, x_3) = \frac{1}{3} \cos(x_2 x_3) + \frac{1}{6},$$

$$g_2(x_1, x_2, x_3) = \frac{1}{9} \sqrt{x_1^2 + \sin x_3 + 1.06} - .1,$$

$$g_3(x_1, x_2, x_3) = -\frac{1}{20} e^{-x_1 x_2} - \frac{10\pi - 3}{60}.$$

قضیه ۷.۹ برای نشان دادن اینکه G دارای یک نقطه ثابت منحصر بفرد در

$$D = \{(x_1, x_2, x_3) \mid -1 \leq x_i \leq 1, \quad i = 1, 2, 3\}$$

است بکار خواهد رفت.

بمازای $x = (x_1, x_2, x_3)$ در D

$$|g_1(x_1, x_2, x_3)| \leq \frac{1}{3} |\cos(x_2 x_3)| + \frac{1}{6} \leq \frac{1}{2},$$

$$|g_2(x_1, x_2, x_3)| = \left| \frac{1}{9} \sqrt{x_1^2 + \sin x_3 + 1.06} - .1 \right|$$

$$\leq \frac{1}{9} \sqrt{1 + \sin 1 + 1.06} - .1$$

$$< .090,$$

$$|g_3(x_1, x_2, x_3)| = \frac{1}{20}e^{-x_1x_2} + \frac{10\pi - 3}{60}$$

$$\leq \frac{1}{20}e + \frac{10\pi - 3}{60}$$

$$< .61;$$

در نتیجه، به ازای هر $i = 1, 2, 3$ ، $-1 \leq g_i(x_1, x_2, x_3) \leq 1$ ، پس هر وقت $x \in D$

$$G(x) \in D$$

در مثال ۲ نشان داده شد که G بر D پیوسته است؛ لذا اولین قسمت قضیه ۷.۹

نتیجه می‌دهد که G حداقل یک نقطه ثابت در D دارد.

پیدا کردن کرانه‌ها برای مشتق‌های جزئی بر D نتایج زیر را می‌دهد:

$$\left| \frac{\partial g_1}{\partial x_1} \right| = 0,$$

$$\left| \frac{\partial g_1}{\partial x_2} \right| \leq \frac{1}{3}|x_3| |\sin x_2 x_3| \leq \frac{1}{3} \sin 1 < .281,$$

$$\left| \frac{\partial g_1}{\partial x_3} \right| = \frac{1}{3}|x_2| |\sin x_2 x_3| \leq \frac{1}{3} \sin 1 < .281,$$

$$\left| \frac{\partial g_2}{\partial x_1} \right| = \frac{|x_1|}{9\sqrt{x_1^2 + \sin x_3 + 1.06}} < \frac{1}{9\sqrt{.218}} < .238,$$

$$\left| \frac{\partial g_2}{\partial x_2} \right| = 0,$$

$$\left| \frac{\partial g_2}{\partial x_3} \right| = \frac{|\cos x_3|}{18\sqrt{x_1^2 + \sin x_3 + 1.06}} < \frac{1}{18\sqrt{.218}} < .119,$$

$$\left| \frac{\partial g_3}{\partial x_1} \right| = \frac{|x_2|}{20} e^{-x_1x_2} \leq \frac{1}{20} e < .14,$$

$$\left| \frac{\partial g_3}{\partial x_2} \right| = \frac{|x_1|}{20} e^{-x_1x_2} \leq \frac{1}{20} e < .14,$$

$$\left| \frac{\partial g_3}{\partial x_3} \right| = 0.$$

بنابراین، برای هر $x \in D$ ،

$$\left| \frac{\partial g_i(x)}{\partial x_j} \right| \leq 0.281, \quad j = 1, 2, 3 \text{ و } i = 1, 2, 3$$

و شرط قسمت دوم قضیه ۷.۹ با $K = 0.843$ برقرار است. با ادامه روش مثال ۲، همچنین می‌توان نشان داد که، به‌ازای هر $i = 1, 2, 3$ و $j = 1, 2, 3$ بر D پیوسته‌است. (این مطلب در تمرین ۳ بررسی خواهد شد.) در نتیجه، G دارای نقطه ثابت منحصر بفرد در D است و دستگاه غیرخطی (۴.۹) جواب منحصر بفرد در D دارد.

بعد از اینکه یک دستگاه غیرخطی $F(x) = 0$ به یک مسئله نقطه ثابت معادل $G(x) = x$ تبدیل شد، طبیعی است که یک تابعی یا روند تکراری نقطه ثابت روی G اعمال شود. برای انجام این کار، $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$ را انتخاب می‌کنیم و دنباله بردارهای $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})$ را به وسیله

$$(5.9) \quad x^{(k)} = G(x^{(k-1)}), \quad k = 1, 2, \dots, n$$

یا، مولفه وار، به صورت

$$(6.9) \quad \begin{aligned} x_1^{(k)} &= g_1(x_1^{(k-1)}, x_2^{(k-1)}, \dots, x_n^{(k-1)}), \\ x_2^{(k)} &= g_2(x_1^{(k-1)}, x_2^{(k-1)}, \dots, x_n^{(k-1)}), \\ &\vdots \\ x_n^{(k)} &= g_n(x_1^{(k-1)}, x_2^{(k-1)}, \dots, x_n^{(k-1)}) \end{aligned}$$

تولید می‌کنیم.

قضیه‌ای شبیه قضیه ۵.۲، ص ۴۴، در حالت یک بعدی، که شرایطی را برای همگرایی تکرارهای D ، ذیلاً "ارائه می‌شود. برهان این قضیه را می‌توان در اورنگا [۴۶]، صفحه ۱۵۳، یافت.

قضیه ۸.۹. فرض کنیم به‌ازای بردارهای از ثابتهای a_1, a_2, \dots, a_n و b_1, b_2, \dots, b_n ، $\{ \text{به‌ازای هر } i = 1, 2, \dots, n \mid a_i \leq x_i \leq b_i \}$ ، و نیز G تابع پیوسته‌ای با مشتقهای جزئی اول پیوسته از $D \subset R^n$ بتوی R^n باشد با این خاصیت که هر وقت $x \in D$ ، $G(x) \in D$ ، هرگاه یک ثابت $K < 1$ موجود باشد که، به‌ازای هر $j = 1, 2, \dots, n$ و هر تابع مولفه‌ای g_i ،

$$\left| \frac{\partial g_i(x)}{\partial x_j} \right| \leq \frac{K}{n}, \quad x \in D$$

آنگاه دنباله $\{x^{(k)}\}_{k=0}^{\infty}$ تعریف شده در (۵.۹) به ازای هر $x^{(0)}$ در D به نقطه ثابت منحصر بفرد $p \in D$ همگراست و

$$(۷.۹) \quad \|x^{(j)} - p\|_{\infty} \leq \frac{K^j}{1-K} \|x^{(1)} - x^{(0)}\|_x.$$

مثال ۴. در مثال ۳، با بکار بردن قضیه ۷.۹ روی مسئله نقطه ثابت

$$x_1 = \frac{1}{3} \cos x_2 x_3 + \frac{1}{6},$$

$$(۸.۹) \quad x_2 = \frac{1}{9} \sqrt{x_1^2 + \sin x_3 + 1.06} - .1,$$

$$x_3 = -\frac{1}{20} e^{-x_1 x_2} - \frac{10\pi - 3}{60}.$$

نشان داده شد که دستگاه غیر خطی

$$3x_1 - \cos(x_2 x_3) - \frac{1}{6} = 0,$$

$$x_1^2 - 81(x_2 + .1)^2 + \sin x_3 + 1.06 = 0,$$

$$e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0,$$

دارای جواب منحصر بفرد در

$$D = \{(x_1, x_2, x_3) \mid -1 \leq x_i \leq 1, i = 1, 2, 3\}$$

می باشد. چون قضیه (۸.۹) با $K = .843$ برای (۸.۹) برقرار است، روش تکرار تابعی را می توان اعمال کرد. انتخاب $x^{(0)}$ در D اختیاری است و ما فرض می کنیم $x^{(0)} = (.1, .1, -.1)$ دنباله بردارهای تولید شده به وسیله

$$x_1^{(k)} = \frac{1}{3} \cos x_2^{(k-1)} x_3^{(k-1)} + \frac{1}{6},$$

$$x_2^{(k)} = \frac{1}{9} \sqrt{(x_1^{(k-1)})^2 + \sin x_3^{(k-1)} + 1.06} - .1,$$

$$x_3^{(k)} = -\frac{1}{20} e^{-x_1^{(k-1)} x_2^{(k-1)}} - \frac{10\pi - 3}{60}.$$

به جواب منحصر بفرد (۸.۹) همگرا خواهد بود. در این مثال دنباله تا وقتی k چنان یافت شود که

$$\| \mathbf{x}^{(k)} - \mathbf{x}^{(k-1)} \|_x < 10^{-5}$$

تولید می‌شود. نتایج در جدول ۱.۹ داده شده است.

جدول ۱.۹

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\ \mathbf{x}^{(k)} - \mathbf{x}^{(k-1)} \ _x$
0	.10000000	.10000000	-.10000000	—
1	.49998333	.00944115	-.52310127	.423
2	.49999593	.00002557	-.52336331	9.4×10^{-3}
3	.50000000	.00001234	-.52359814	2.3×10^{-4}
4	.50000000	.00000003	-.52359847	1.2×10^{-5}
5	.50000000	.00000002	-.52359877	3.1×10^{-7}

بکاربردن کران خطای (۷.۹) با $K = .843$ نتیجه می‌دهد

$$\| \mathbf{x}^{(5)} - \mathbf{p} \|_x \leq \frac{(.843)^5}{1 - .843} (.423) < 1.15,$$

که دقت واقعی $\mathbf{x}^{(5)}$ را نشان نمی‌دهد. هرگاه

$$\mathbf{x}^{(3)} = (.50000000, 1.234 \times 10^{-5}, -.52359814)$$

به‌عنوان تقریب اولیه انتخاب شده بود، آنگاه $\mathbf{x}^{(5)}$ تکرار دوم می‌بود و نامساوی (۹.۷) نتیجه می‌داد که

$$\| \mathbf{x}^{(5)} - \mathbf{p} \|_x \leq \frac{(.843)^2}{1 - .843} (1.50 \times 10^{-5}) < 6.8 \times 10^{-5},$$

که یک تخمین مناسبتر است.

چون

$$\mathbf{p} = \left(.5, 0, -\frac{\pi}{6} \right) \approx (.5, 0, -.5235987757),$$

جواب واقعی است، خطای واقعی عبارت است از

$$\| \mathbf{x}^{(5)} - \mathbf{p} \|_x \leq 2 \times 10^{-8}.$$

یکراه ممکن برای شتاب بخشیدن به همگرایی آن است که آخرین تخمینهای

$x_1^{(k)}, \dots, x_{i-1}^{(k)}$ را به جای $x_1^{(k-1)}, \dots, x_{i-1}^{(k-1)}$ برای محاسبه $x_i^{(k)}$ ، همانند روش گاوس

سیدل برای دستگاههای خطی (ر.ک. بخش ۳۰.۸)، بکار بریم. در این صورت، معادلات مولفهای خواهند شد

$$x_1^{(k)} = \frac{1}{3} \cos(x_2^{(k-1)} x_3^{(k-1)}) + \frac{1}{6}.$$

$$x_2^{(k)} = \frac{1}{9} \sqrt{(x_1^{(k)})^2 + \sin x_3^{(k-1)}} + 1.06 - 1.$$

$$x_3^{(k)} = -\frac{1}{20} e^{-x_1^{(k)} x_2^{(k)}} - \frac{10\pi - 3}{60};$$

و با $x^{(0)} = (.1, .1, -.1)$ نتایج محاسبات در جدول ۲۰.۹ فهرست شده است.

جدول ۲۰.۹

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\ x^{(k)} - x^{(k-1)}\ _r$
0	.10000000	.10000000	-.10000000	—
1	.49998333	.02222979	-.52304613	.423
2	.49997747	.00002815	-.52359807	2.2×10^{-2}
3	.50000000	.00000004	-.52359877	2.8×10^{-5}
4	.50000000	.00000000	-.52359877	3.8×10^{-8}

تکرار $x^{(4)}$ در نرم l_r در واقع تا حدود 10^{-7} دقیق است؛ لذا برای این مسئله با استفاده از روش سیدل واقعا "به همگرایی سرعت داده شده است. معبدا، باید متذکر شده، هرچند روش سیدل در این مثال همگرایی را بهتر کرد، اما همیشه به همگرایی شتاب نمی‌بخشد.

مجموعه تمرینات ۱۰.۹

۱. نشان دهید که تابع $F: R^3 \rightarrow R^3$ تعریف شده با

$$F(x_1, x_2, x_3) = (x_1 + 2x_3, x_1 \cos x_2, x_2^2 + x_3)$$

در هر نقطه R_3 پیوسته است.

۲. مثالی از یک تابع $F: R^2 \rightarrow R^2$ بزنید که در هر نقطه از R^2 جز نقطه $(1, 0)$ پیوسته باشد.

۳. نشان دهید که مشتقهای جزئی اول در مثال ۳ پیوسته هستند.

۴. نشان دهید که تعریف (۴.۹) معادل عبارت زیر است: تابع $F: D \subset R^n \rightarrow R^n$ در $x_0 \in D$ پیوسته نامیده می‌شود اگر به ازای هر $\varepsilon > 0$ عددی مانند $\delta > 0$ بتوان یافت با این

خاصیت که هر وقت $x \in D$ و $\|x - x_0\| < \delta$ ،

$$\|F(x) - F(x_0)\| < \varepsilon .$$

۵ . نشان دهید که دستگاه غیرخطی

$$x_1^2 + x_2^2 - x_1 = 0 ,$$

$$x_1^2 - x_2^2 - x_2 = 0 ,$$

جواب منحصر بفرد دارد . جواب را به طور نموداری تقریب کنید . جواب نموداری را به عنوان تقریب اولیه برای یک تکرار تابعی مناسب بکار ببرید .

۶ . دستگاه غیرخطی

$$x_1^2 - 10x_1 + x_2^2 + 8 = 0 ,$$

$$x_1x_2^2 + x_1 - 10x_2 + 8 = 0 ,$$

را می توان به مسئله نقطه ثابت زیر تبدیل کرد .

$$x_1 = g_1(x_1, x_2) = \frac{x_1^2 + x_2^2 + 8}{10} ,$$

$$x_2 = g_2(x_1, x_2) = \frac{x_1x_2^2 + x_1 + 8}{10} .$$

(آ) نشان دهید که $G = (g_1, g_2): D \subset R^2 \rightarrow R^2$ دارای نقطه ثابت منحصر بفردی

در $D = \{(x_1, x_2) | 0 \leq x_1, x_2 \leq 1.5\}$ است .

(ب) تکرار تابعی را جهت تقریب جواب اعمال کنید .

(پ) آیا روش سیدل به همگرایی شتاب می دهد؟

۷ . نشان دهید که توابع $G: D \subset R^3 \rightarrow R^3$ زیر دارای نقاط ثابت منحصر بفرد در D

هستند :

$$G(x_1, x_2, x_3) = \left(\frac{\cos(x_2, x_3) + .5}{3} , \frac{1}{25} \sqrt{x_1^2 + .3125} - .03 , -\frac{1}{20} e^{-x_1x_2} - \frac{10\pi - 3}{60} \right) \quad (\text{آ})$$

$$D = \{(x_1, x_2, x_3) | -1 \leq x_i \leq 1, i = 1, 2, 3\} .$$

$$G(x_1, x_2, x_3) = \left(\frac{7.17 + 3x_2^2 + 4x_3}{12} , \frac{11.54 + x_3 - x_1^2}{10} , \frac{7.631 - x_3^2}{7} \right) \quad (\text{ب})$$

$$D = \{(x_1, x_2, x_3) | 0 \leq x_i \leq 1.5, i = 1, 2, 3\} .$$

$$G(x_1, x_2, x_3) \quad (\text{پ})$$

$$= (1 - \cos(x_1, x_2, x_3), 1 - (1 - x_1)^4 - .05x_2^2 + .15x_3, x_1^2 + .1x_2^2 - .01x_2 + 1)$$

$$D = \{(x_1, x_2, x_3) | -.1 \leq x_1 \leq .1, -.1 \leq x_2 \leq .3, .5 \leq x_3 \leq 1.1\} .$$

۸. تکرار تابعی و روش سیدل را برای تقریب نقاط ثابت تمرین ۷ بکار ببرید .
 ۹. در تمرین ۶، بخش ۸.۰۶، مسئله پیش بینی جمعیت دو نمونه که برای یک منبع تامین غذایی رقابت می کنند را بررسی کردیم. در آن مسئله فرض کردیم که پیش بینی لازم در مورد جمعیتها می تواند با حل دستگاه معادلات

$$\frac{dx_1(t)}{dt} = x_1(t)(4 - .0003x_1(t) - .0004x_2(t))$$

و

$$\frac{dx_2(t)}{dt} = x_2(t)(2 - .0002x_1(t) - .0001x_2(t))$$

انجام شود. در این تمرین می خواهیم مسئله تعیین تعادل جمعیتهای دو نمونه را بررسی کنیم. محکهای ریاضی که باید برقرار باشند تا اینکه جمعیتها در حال تعادل باشند عبارتند از اینکه، به طور همزمان،

$$\frac{dx_1(t)}{dt} = 0 \quad \text{و} \quad \frac{dx_2(t)}{dt} = 0 .$$

این بوضوح وقتی اتفاق می افتد که نمونه اول از بین رفته باشد و نمونه دوم دارای جمعیت 20,000 باشد، یا وقتی نمونه دوم از بین رفته باشد و نمونه اول دارای جمعیت 13,333 باشد. آیا این تعادل تحت وضعیت دیگری اتفاق می افتد؟

۲۰۹ روش نیوتن

اگرچه مسئله ارائه شده در مثال ۳، بخش ۱۰.۹، را توانستیم به سادگی به شکل نقطه - ثابت همگرا تبدیل کنیم، اما وضعیت اغلب چنین نیست. به این دلیل لازم است روندی الگوریتمی طراحی کنیم که بتواند جهت انجام تبدیل برای یک مسئله کلی بکار رود. برای ساختن این الگوریتم که به روش نقطه - ثابت مناسبی در حالت یک بعدی منجر می شود، سعی می کنیم تابعی مانند ϕ پیدا کنیم، با این خاصیت که تابع g ، داده شده با

$$g(x) = x - \phi(x)f(x),$$

همگرایی درجه دوم به نقطه ثابت p از g دهد. روش نیوتن از این شرط، با انتخاب $\phi(x) = 1/f'(x)$ ، بشرط آنکه در نقطه ثابت p ، $f'(p) \neq 0$ ، ناشی شد.

با بکار بردن روشی مشابه در حالت n بعدی، یک ماتریس

$$(۹.۹) \quad A(x) = \begin{bmatrix} a_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x) \\ a_{21}(x) & a_{22}(x) & \cdots & a_{2n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(x) & a_{n2}(x) & \cdots & a_{nn}(x) \end{bmatrix}$$

وارد می‌شود که در آن هریک از درایه‌های $a_{ij}(x)$ یک تابع از R^n بتوی R است. در این روند لازم است $A(x)$ چنان پیدا شود که

$$(۱۰.۹) \quad G(x) = x - A(x)^{-1}F(x)$$

همگرایی درجه دوم به جواب $F(x) = 0$ دهد، البته بشرطی که $A(x)$ در نقطه ثابت نامنفرد باشد.

قضیه زیر نشان می‌دهد که این گرایش می‌تواند به عنوان انگیزهای برای انتخاب A بکار رود.

قضیه ۹.۹. فرض کنیم p جوابی از $G(x) = x$ باشد، به ازای تابعی چون $G = (g_1, g_2, \dots, g_n)$ که R^n را بتوی R^n می‌نگارد. اگر $\delta > 0$ موجود باشد با این خاصیت که

(آ) $N_\delta = \{x \mid \|x - p\| < \delta\}$ بر $z = 1, 2, \dots, n$ و $i = 1, 2, \dots, n$ به ازای $\partial g_i / \partial x_j$ پیوسته باشد،

(ب) $\partial^2 g_i(x) / (\partial x_j \partial x_k)$ پیوسته بوده و، به ازای ثابتی چون M ، هر وقت $x \in N_\delta$ ، به

ازای هر $i = 1, 2, \dots, n$ ، $k = 1, 2, \dots, n$ ، $z = 1, 2, \dots, n$ ، $|\partial^2 g_i(x) / (\partial x_j \partial x_k)| \leq M$ ،

(پ) به ازای هر $i = 1, 2, \dots, n$ و $j = 1, 2, \dots, n$ ، $\partial g_i(p) / \partial x_j = 0$ ، آنگاه دنباله تولید شده به وسیله $x^{(k)} = G(x^{(k-1)})$ ، به ازای هر انتخاب $x^{(0)} \in N_\delta$ ، به طور درجه

دوم به p همگراست و به ازای هر $k \geq 1$

$$\|x^{(k)} - p\|_x \leq \frac{n^2 M}{2} \|x^{(k-1)} - p\|_x^2$$

این قضیه نظیر قضیه ۱۱.۲ در بخش ۵.۲ است و برهان آن احتیاج به بیان G

برحسب سری تیلور با n متغیر حول نقطه p دارد.

برای اینکه قضیه ۹.۹ را مورد استفاده قرار دهیم، فرض می‌کنیم $A(x)$ ماتریسی

$n \times n$ از توابع از R^n بتوی R به شکل (۹.۹) باشد که درایه‌های مشخصی از آن بعداً

انتخاب خواهند شد. همچنین، $A(\mathbf{x})$ در نزدیکی جواب \mathbf{p} از $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ نامنفرد بوده و $b_{ij}(\mathbf{x})$ نمایانگر درایه $A(\mathbf{x})^{-1}$ در سطر i و ستون j م باشد.

چون $\mathbf{G}(\mathbf{x}) = \mathbf{x} - A(\mathbf{x})^{-1}\mathbf{F}(\mathbf{x})$

$$g_i(\mathbf{x}) = x_i - \sum_{j=1}^n b_{ij}(\mathbf{x})f_j(\mathbf{x});$$

در نتیجه،

$$\frac{\partial g_i(\mathbf{x})}{\partial x_k} = \begin{cases} 1 - \sum_{j=1}^n \left(b_{ij}(\mathbf{x}) \frac{\partial f_j}{\partial x_k}(\mathbf{x}) + \frac{\partial b_{ij}}{\partial x_k}(\mathbf{x}) f_j(\mathbf{x}) \right), & \text{اگر } i = k \\ - \sum_{j=1}^n \left(b_{ij}(\mathbf{x}) \frac{\partial f_j}{\partial x_k}(\mathbf{x}) + \frac{\partial b_{ij}(\mathbf{x})}{\partial x_k} f_j(\mathbf{x}) \right), & \text{اگر } i \neq k \end{cases}$$

قضیه ۹.۹ نتیجه می‌دهد که مایلم به‌ازای هر $i = 1, 2, \dots, n$ و $k = 1, 2, \dots, n$ داشته باشیم $\partial g_i(\mathbf{p}) / \partial x_k = 0$ ؛ یعنی، به‌ازای $i = k$ ،

$$0 = 1 - \sum_{j=1}^n b_{ij}(\mathbf{p}) \frac{\partial f_j}{\partial x_i}(\mathbf{p}),$$

در نتیجه،

$$(11.9) \quad \sum_{j=1}^n b_{ij}(\mathbf{p}) \frac{\partial f_j}{\partial x_i}(\mathbf{p}) = 1,$$

و، وقتی $i \neq k$ ،

$$(12.9) \quad 0 = - \sum_{j=1}^n b_{ij}(\mathbf{p}) \frac{\partial f_j}{\partial x_k}(\mathbf{p}),$$

بنابراین،

$$\sum_{j=1}^n b_{ij}(\mathbf{p}) \frac{\partial f_j}{\partial x_k}(\mathbf{p}) = 0.$$

با تعریف ماتریس $J(\mathbf{x})$ با

$$(13.9) \quad J(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial f_2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n(\mathbf{x})}{\partial x_1} & \frac{\partial f_n(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial f_n(\mathbf{x})}{\partial x_n} \end{bmatrix},$$

ملاحظه می‌کنیم که شرایط (۱۱.۹) و (۱۲.۹) لازم دارند که

$$A(\mathbf{p})^{-1}J(\mathbf{p}) = I \text{ (ماتریس همانی).}$$

بنابراین، $A(\mathbf{p}) = J(\mathbf{p})$.

در نتیجه، یک انتخاب مناسب برای $A(\mathbf{x})$ ، $A(\mathbf{x}) = J(\mathbf{x})$ است، زیرا شرط (پ) از قضیه (۹.۹) با این انتخاب برقرار خواهد بود. لذا، تابع \mathbf{G} عبارت خواهد بود از

$$\mathbf{G}(\mathbf{x}) = \mathbf{x} - J(\mathbf{x})^{-1}\mathbf{F}(\mathbf{x}),$$

و روند تکرار تابعی از معادله زیر ناشی خواهد شد:

$$(14.9) \quad \mathbf{x}^{(k)} = \mathbf{G}(\mathbf{x}^{(k-1)}) = \mathbf{x}^{(k-1)} - J(\mathbf{x}^{(k-1)})^{-1}\mathbf{F}(\mathbf{x}^{(k-1)}).$$

این روش کاملاً "بحق"، روش نیوتن برای دستگاه‌های غیرخطی نامیده شده است و انتظار می‌رود که بطور کلی همگرایی درجه دوم بدهد، بشرط آنکه مقدار آغازین به حد کافی دقیقی معلوم باشد.

ماتریس $J(\mathbf{x})$ ، داده شده در (۱۲.۹)، ماتریس ژاکوبین نامیده می‌شود و کاربردی چند در آنالیز دارد. بخصوص، ممکن است به علت کاربرد در انتگرالگیری چندگانه از یک تابع چند متغیره روی یک ناحیه که برای انجام آن یک تغییر متغیر لازم است، برای خواننده آشنا باشد.

یک ضعف مشخص در روش نیوتن از لزوم معکوس کردن ماتریس $J(\mathbf{x})$ در هر مرحله نشأت می‌گیرد. در عمل، این روش عموماً "به یک طریق دومرحله‌ای انجام می‌شود. ابتدا، یک بردار \mathbf{y} پیدا می‌شود. که در $J(\mathbf{x}^{(k)})\mathbf{y} = -\mathbf{F}(\mathbf{x}^{(k)})$ صدق می‌کند. بعد از این کار، تقریب جدید، یعنی $\mathbf{x}^{(k+1)}$ ، را می‌توان با افزودن \mathbf{y} به $\mathbf{x}^{(k)}$ بدست آورد. الگوریتم زیر این روند دومرحله‌ای را بکار می‌برد.

الگوریتم روش نیوتن برای دستگاهها ۱.۹

برای بدست آوردن تقریبی به جواب دستگاه غیر خطی $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ ، یک تقریب اولیه $\mathbf{x}^{(0)}$ انتخاب می‌کنیم.

- مرحله ۱. قرار می‌دهیم $k = 0$.
- مرحله ۲. $\mathbf{F}(\mathbf{x}^{(k)})$ و

$$J(\mathbf{x}^{(k)}) = \begin{bmatrix} \frac{\partial f_1(\mathbf{x}^{(k)})}{\partial x_1} & \frac{\partial f_1(\mathbf{x}^{(k)})}{\partial x_2} & \dots & \frac{\partial f_1(\mathbf{x}^{(k)})}{\partial x_n} \\ \frac{\partial f_2(\mathbf{x}^{(k)})}{\partial x_1} & \frac{\partial f_2(\mathbf{x}^{(k)})}{\partial x_2} & \dots & \frac{\partial f_2(\mathbf{x}^{(k)})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(\mathbf{x}^{(k)})}{\partial x_1} & \frac{\partial f_n(\mathbf{x}^{(k)})}{\partial x_2} & \dots & \frac{\partial f_n(\mathbf{x}^{(k)})}{\partial x_n} \end{bmatrix}$$

را پیدا می‌کنیم .

مرحله ۳ . دستگاه خطی $n \times n$

$$J(\mathbf{x}^{(k)})\mathbf{y}^{(k)} = -\mathbf{F}(\mathbf{x}^{(k)})$$

را حل می‌کنیم .

مرحله ۴ . قرار می‌دهیم $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{y}^{(k)}$.

مرحله ۵ . اگر $\mathbf{x}^{(k+1)}$ به قدر کافی دقیق نباشد ، 1 را به k افزوده به مرحله ۲ می‌رویم .

مرحله ۶ . روند تمام است .

این الگوریتم دو ضعف دارد ، که باید به آنها اشاره شود . اولین ضعف آن است که در هر تکرار ، لازم است ماتریس $J(\mathbf{x}^{(k)})$ نامفرد باشد . برای مقابله با این ضعف توصیه می‌شود که آزمایشی برای منفرد بودن در مرحله ۳ از الگوریتم گنجانده شود . مشکل دیگر مهمتر است و ، برای اینکه همگرایی بدست آید ، متضمن لزوم نزدیکی تقریب اولیه $\mathbf{x}^{(0)}$ به جواب است . با وجودی که این در حالت یک بعدی نیز لازم بود ، برای دستگاههای معادلات به‌طور فزاینده‌ای مهم است و منجر به مشکلات قابل توجهی می‌شود . شرایطی مشابه شرایط فوریه برای حالت یک بعدی که در قضیه ۲ ، ۹۰ ، صفحه ۵۹ ، داده شده در دسترس است ، اما عموماً " برای اجرا مناسب نیست . برای بحث در مورد این شرایط و روندهای شروع بطور کلی ، ر. ک. قضیه نیوتن - کانتروویچ^۱ در اورنگا [۴۶] ، صفحه ۱۵۵ و مراجع آن کتاب فهرست شده در صفحات ۱۶۰ - ۱۵۹ .

مثال ۰۱ . در مثال ۳ ، بخش ۱۰۹ ، نشان داده شد که دستگاه غیرخطی

$$3x_1 - \cos(x_2 x_3) - \frac{1}{2} = 0,$$

$$x_1^2 - 81(x_2 + .1)^2 + \sin x_3 + 1.06 = 0,$$

$$e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0,$$

در

$$D = \{(x_1, x_2, x_3) \mid -1 \leq x_i \leq 1, i = 1, 2, 3\}$$

دارای یک جواب منحصر بفرد است. روش نیوتن برای تقریب جواب، وقتی تقریب اولیه $\mathbf{x}^{(0)} = (.1, .1, -.1)$ است، بکار خواهد رفت.

ماتریس ژاکوبین $J(\mathbf{x})$ برای این دستگاه یا

$$J((x_1, x_2, x_3)) = \begin{bmatrix} 3 & x_3 \sin x_2 x_3 & x_2 \sin x_2 x_3 \\ 2x_1 & -162(x_2 + .1) & \cos x_3 \\ -x_2 e^{-x_1 x_2} & -x_1 e^{-x_1 x_2} & 20 \end{bmatrix}$$

داده می شود و

$$\begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{bmatrix} = \begin{bmatrix} x_1^{(k-1)} \\ x_2^{(k-1)} \\ x_3^{(k-1)} \end{bmatrix} + \begin{bmatrix} y_1^{(k-1)} \\ y_2^{(k-1)} \\ y_3^{(k-1)} \end{bmatrix},$$

که در آن

$$\begin{bmatrix} y_1^{(k-1)} \\ y_2^{(k-1)} \\ y_3^{(k-1)} \end{bmatrix} = -J((x_1^{(k-1)}, x_2^{(k-1)}, x_3^{(k-1)}))^{-1}.$$

بنابراین، در مرحله k ام، دستگاه خطی

$$\begin{bmatrix} 3 & x_3^{(k-1)} \sin x_2^{(k-1)} x_3^{(k-1)} & x_2^{(k-1)} \sin x_2^{(k-1)} x_3^{(k-1)} \\ 2x_1^{(k-1)} & -162(x_2^{(k-1)} + .1) & \cos x_3^{(k-1)} \\ -x_2^{(k-1)} e^{-x_1^{(k-1)} x_2^{(k-1)}} & -x_1^{(k-1)} e^{-x_1^{(k-1)} x_2^{(k-1)}} & 20 \end{bmatrix}$$

$$\times \begin{bmatrix} x_1^{(k)} - x_1^{(k-1)} \\ x_2^{(k)} - x_2^{(k-1)} \\ x_3^{(k)} - x_3^{(k-1)} \end{bmatrix} = - \begin{bmatrix} 3x_1^{(k-1)} - \cos x_2^{(k-1)} x_3^{(k-1)} - \frac{1}{2} \\ (x_1^{(k-1)})^2 - 81(x_2^{(k-1)} + .1)^2 + \sin x_3^{(k-1)} + 1.06 \\ e^{-x_1^{(k-1)} x_2^{(k-1)}} + 20x_3^{(k-1)} + \frac{10\pi - 3}{3} \end{bmatrix}$$

باید حل شود، نتایج بدست آمده با استفاده از این روند تکراری در جدول ۳.۹ نشان داده شده است.

جدول ۳.۹

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\ x^{(k)} - x^{(k-1)}\ _x$
0	.10000000	.10000000	-.10000000	
1	.50003702	.01946686	-.52152047	.422
2	.50004593	.00158859	-.52355711	1.79×10^{-2}
3	.50000034	.00001244	-.52359845	1.58×10^{-3}
4	.50000000	.00000000	-.52359877	1.24×10^{-5}
5	.50000000	.00000000	-.52359877	0

جواب واقعی مثال عبارت است از

$$p = \left(.5, 0, -\frac{\pi}{6} \right) \approx (.5, 0, -.52359877)$$

توجه کنید که همگرایی روش نیوتن وقتی یک تکرار نزدیک p باشد خیلی سریع می شود. این امر همگرایی درجه دوم روش را نزدیک یک جواب توجیه می کند.

مجموعه تمرینات ۲.۹

۱. دستگاههای غیرخطی زیر را با استفاده از روش نیوتن حل کرده و کران خطای داده

شده در قضیه ۲.۹ را حساب کنید.

$$x_1^2 - 10x_1 + x_2^2 + 8 = 0, \quad (\text{آ})$$

$$x_1 x_2^2 + x_1 - 10x_2 + 8 = 0;$$

$$x_1^2 + x_2^2 - x_1 = 0, \quad (\text{ب})$$

$$x_1^2 - x_2^2 - x_2 = 0;$$

$$3x_1^2 - x_2^2 = 0, \quad (\text{پ})$$

$$3x_1 x_2^2 - x_1^3 - 1 = 0.$$

$$x_1^2 + x_2 - 37 = 0, \quad (\text{ت})$$

$$x_1 - x_2^2 - 5 = 0,$$

$$x_1 + x_2 + x_3 - 3 = 0.$$

$$x_1 + \cos(x_1, x_2, x_3) - 1 = 0, \quad (\text{ش})$$

$$(1 - x_1)^4 + x_2 + .05x_3^2 - .15x_3 - 1 = 0,$$

$$-x_1^2 - .1x_2^2 + .01x_2 + x_3 - 1 = 0;$$

$$12x_1 - 3x_2^2 - 4x_3 = 7.17, \quad (\text{ج})$$

$$x_1^2 + 10x_2 - x_3 = 11.54,$$

$$x_3^2 + 7x_3 = 7.631;$$

$$4x_1 - x_2 + x_3 - 2 = 0, \quad (\text{چ})$$

$$-x_1 + 6x_2 - 3x_3 - 11 = 0,$$

$$x_1 - 3x_2 + 5x_3 + 5 = 0.$$

۲. آیا روش نیوتن را می‌توان برای حل دستگاه غیرخطی

$$3x_1 - \cos(x_2, x_3) - .5 = 0,$$

$$x_1^2 - 625x_2^2 = 0,$$

$$e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0.$$

بکار برد؟

۳. ماتریس ژاکوبین دستگاه خطی زیر را پیدا کنید:

$$c_{11}x_1 + c_{12}x_2 + \dots + c_{1n}x_n - b_1 = 0,$$

$$c_{21}x_1 + c_{22}x_2 + \dots + c_{2n}x_n - b_2 = 0,$$

⋮

⋮

$$c_{n1}x_1 + c_{n2}x_2 + \dots + c_{nn}x_n - b_n = 0.$$

چه وقت این ماتریس معکوس خواهد داشت؟

۴. سی. چیارلا، دبلیو. چارلتون^۲، و ا. دبلیو. رابرتس^۳ [۱۹]، در محاسبه شکل

جریان جاذبه شیب تخلیه باری که زمان انتقال نقاط مادی دانه‌ای تخلیه شده را

مینیمم می‌کند، معادلات زیر را به روش نیوتن حل کردند:

(یک) به‌ازای هر $n = 1, 2, \dots, N - 1$

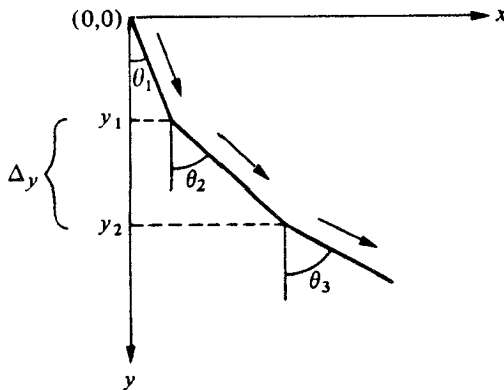
$$f_n(\theta_1, \dots, \theta_N) = \frac{\sin \theta_{n+1}}{v_{n+1}} (1 - \mu w_{n+1}) - \frac{\sin \theta_n}{v_n} (1 - \mu w_n) = 0$$

$$\text{(دو)} \quad f_N(\theta_1, \dots, \theta_N) = \Delta y \sum_{i=1}^N \tan \theta_i - X = 0, \quad \text{که در آن}$$

(سه) به ازای هر $n = 1, 2, \dots, N$ ، $v_n^2 = v_0^2 + 2gn\Delta y - 2\mu\Delta y \sum_{j=1}^n \frac{1}{\cos \theta_j}$ ، و

(چهار) به ازای هر $n = 1, 2, \dots, N$ ، $w_n = -\Delta y v_n \sum_{i=n}^N \frac{1}{v_i^3 \cos \theta_i}$ ،

ثابت v_0 سرعت اولیه ماده دانه‌ای است ، X عبارت است از مختص x انتهای شیب ، μ نیروی اصطکاک است ، N تعداد قطعات شیب‌دار است ، و g ثابت ثقل می‌باشد . متغیر θ_i زاویه قطعه شیب‌دار i م با خط قائم است و v_i سرعت ماده در قطعه شیب‌دار i م می‌باشد . (یک) و (دو) را برای $\theta_1, \dots, \theta_N$ ، با استفاده از روش نیوتن با $\mu = 0$ ، $X = 2$ ، $\Delta y = .2$ ، $N = 20$ ، و $(32.2 \text{ ثانیه} / \text{فوت})^2$ حل کنید ، که در آن v_n و w_n را می‌توان مستقیماً از (سه) و (چهار) بدست آورد .



۵ . مقدار فشار لازم برای فروکردن یک شی سنگین بزرگ در خاکی همگن و نرم که روی پایه خاکی سختی قرار دارد را می‌توان با مقدار فشار لازم برای پایین راندن اشیاء کوچکتر در همان خاک پیش‌بینی کرد . بخصوص ، مقدار فشار p لازم برای پایین راندن یک صفحه مستدیر به شعاع r به فاصله d در خاکی نرم ، که قسمت خاکی سخت در فاصله $D > d$ از سطح خاک قرار دارد ، را می‌توان با یک معادله به شکل

$$p = k_1 e^{k_2 r} + k_3 r$$

تقریب کرد ، که در آن k_1 ، k_2 ، و k_3 ثابتهایی هستند با $k_2 > 0$ ، که به d و سازگاری خاک بستگی دارند اما به شعاع صفحه وابسته نیستند (ر.ک. بکر^۱ [۹] ،

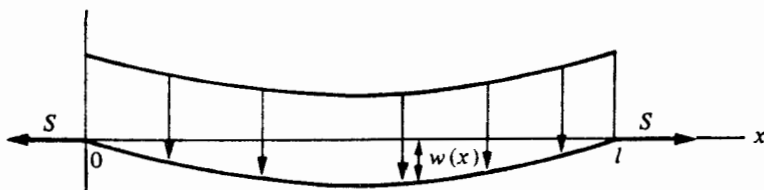
صفحات ۹۴-۰۸۹)

(T) مقادیر k_1 ، k_2 ، و k_3 را در صورتی بیابید که یک صفحه به شعاع 1 اینچ فشاری به مقدار 10^2 (اینچ)/پوند لازم دارد تا 1 پا در زمینی گلی فرورود، یک صفحه به شعاع 2 اینچ فشاری به مقدار 12^2 (اینچ)/پوند لازم دارد تا یک پا فرو رود، و یک صفحه به شعاع 3 اینچ فشاری به مقدار 15^2 (اینچ)/پوند نیاز دارد تا همین فاصله را فرورود، با فرض اینکه گل بیش از یک پا عمق دارد.

(ب) با استفاده از محاسبات قسمت (T)، اندازه مینیمال صفحه مستدیری را پیش‌بینی کنید که لازم است وزنی معادل 500 پوند را در این زمین با فرو رفتن کمتر از یک پا تحمل کند.

مسائل مقدارمرزی برای معادلات دیفرانسیل معمولی^{۱۰}

یک مسئله عادی در مهندسی راه و ساختمان به خمیدگی یک میله آهنی با مقطع عرضی مستطیلی مربوط می شود که مقید به تحمل باری یکنواخت است ، درحالی که دو انتهای میله آهنی چنان نگهداری می شوند که متحمل هیچ خمیدگی نمی شوند .



معادله دیفرانسیلی که این وضعیت فیزیکی را تقریب می کند به شکل

$$\frac{d^2w}{dx^2} - \frac{S}{EI} w = \frac{qx}{2EI} (x - l),$$

است ، که در آن $w = w(x)$ خمیدگی به فاصله x از انتهای چپ میله آهنی است و l ، q ، E ، S ، و I ، به ترتیب ، معرف طول میله ، فشار باریکنواخت ، مدول الاستیسیته ، قدرت در نقاط انتهایی ، و ممان اینرسی مرکزی است . به این معادله دیفرانسیل دو شرط مربوط می شود که با این فرض که هیچ خمیدگی در دو انتهای میله رخ نمی دهد ، یعنی $w(0) = w(l) = 0$ ، داده می شود .

وقتی کلفتی میله یکنواخت باشد ، حاصل ضرب EI ثابت خواهد بود و جواب دقیق را می توان به سادگی بدست آورد . اما ، در کاربردهای زیادی کلفتی

یکنواخت نیست؛ لذا، ممان اینرسی I تابعی از x است و تکنیکهای تقریبی لازم است.

گرچه روشهای تعیین تقریب جوابهای معادلات دیفرانسیل در فصل ۶ مطالعه شدند، لیکن در آنها لازم است که تمام شرایط وضع شده روی معادله دیفرانسیل در یک نقطه اولیه اتفاق افتد. برای یک معادله مرتبه دوم نیاز به دانستن هر دوی $w(0)$ و $w'(0)$ داشتیم که در این مسئله چنین حالتی لازم نیست. تکنیکهای تازه برای حل مسائلی که شرایط وضع شده روی آنها از نوع مقدار مرزی به جای مقدار اولیه هستند لازم است.

مسائل فیزیکی که به جای وابسته زمانی بودن وابسته مکانی هستند غالباً " بر حسب معادلات دیفرانسیل با شرایط اعمال شده در بیش از یک نقطه بیان می شوند. مسائل مقدار مرزی دو نقطه ای کلی که در این فصل مورد بحث قرار می دهیم شامل یک معادله دیفرانسیل مرتبه دوم به شکل

$$(1.10) \quad y'' = f(x, y, y'), \quad a \leq x \leq b,$$

است به همراه شرایط مرزی

$$(2.10) \quad y(b) = \beta \quad \text{و} \quad y(a) = \alpha$$

اکثر مطالب مربوط به مسائل مقدار مرزی مرتبه دوم را می توان به مسائل با شرایط

مرزی به شکل

$$(3.10) \quad \alpha_2 y(b) + \beta_2 y'(b) = \beta \quad \text{و} \quad \alpha_1 y(a) - \beta_1 y'(a) = \alpha,$$

که در آن $|\alpha_1| + |\beta_1| \neq 0$ و $|\alpha_2| + |\beta_2| \neq 0$ ، توسعه داد اما، بعضی از تکنیکها کاملاً پیچیده می شوند و خواننده ای که مشتاق این نوع مسائل است به بررسی یک کتاب تخصصی در مسائل مقدار مرزی، مانند کلر [۴۰]، راهنمایی می شود.

۱.۱۰ روش تیراندازی خطی

قضیه زیر شرایطی کلی که اطمینان می دهد جواب یک مسئله مقدار مرزی مرتبه دوم وجود دارد و منحصر بفرد است را بدست می دهد. برهان این قضیه در حالت کلی را می توان در کتاب کلر که در بالا ذکر شد یافت.

۱.۱۰ قضیه. فرض کنیم تابع f در مسئله مقدار مرزی

$$y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

بر مجموعه^۴

$$D = \{(x, y, y') \mid a \leq x \leq b, \quad -\infty < y < \infty, \quad -\infty < y' < \infty\},$$

پیوسته بوده و $\partial f/\partial y$ و $\partial f/\partial y'$ بر D پیوسته باشند. هرگاه

$$(یک) \quad \text{به‌زای هر } (x, y, y') \in D \text{ و } \frac{\partial f}{\partial y}(x, y, y') > 0$$

(دو) یک ثابت M وجود داشته باشد که

$$\left| \frac{\partial f}{\partial y'}(x, y, y') \right| \leq M, \quad (x, y, y') \in D \text{ به‌زای هر}$$

آنگاه مسئله^۴ مقدار مرزی فوق جواب منحصر بفرد دارد.

مثال ۱. مسئله^۴ مقدار مرزی

$$y'' + e^{-xy} + \sin y' = 0, \quad 1 \leq x \leq 2, \quad y(1) = y(2) = 0,$$

دارای

$$f(x, y, y') = -e^{-xy} - \sin y',$$

است، و چون

$$\frac{\partial f}{\partial y}(x, y, y') = xe^{-xy} > 0$$

و

$$\left| \frac{\partial f}{\partial y'}(x, y, y') \right| = |-\cos y'| \leq 1,$$

این مسئله دارای جواب منحصر بفرد است.

وقتی $f(x, y, y')$ را بتوان به شکل

$$f(x, y, y') = p(x)y' + q(x)y + r(x),$$

بیان کرد، معادله^۴ دیفرانسیل

$$y'' = f(x, y, y')$$

خطی نامیده می‌شود. مسائل از این نوع کرارا^۴ در عمل رخ می‌دهند و این نمایش باعث

سادگی قابل ملاحظه^۴ قضیه^۴ ۱۰۱۰ می‌شود.

نتیجه^۴ ۲۰۱۰. اگر مسئله^۴ مقدار مرزی

$$(۴۰۱۰) \quad y'' = p(x)y' + q(x)y + r(x), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

در شرایط زیر صدق کند :

(A) $p(x)$ ، $q(x)$ ، و $r(x)$ بر $[a, b]$ پیوسته باشند ؛

(ب) $q(x) > 0$ بر $[a, b]$ ؛

نگاه مسئله فوق دارای جواب منحصر بفرد است .

برای اینکه جواب منحصر بفردی که توسط برقراری مفروضات نتیجه ۲۰۱۰ تضمین

شده است را تقریب کنیم ابتدا مسائل مقدار اولیه

$$(۵۰۱۰) \quad y'' = p(x)y' + q(x)y + r(x), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y'(a) = 0,$$

و

$$(۶۰۱۰) \quad y'' = p(x)y' + q(x)y, \quad a \leq x \leq b, \quad y(a) = 0, \quad y'(a) = 1$$

را در نظر می گیریم .

قضیه ۱۷۰۶ (ص ۳۸۳) اطمینان می دهد که تحت مفروضات نتیجه ۲۰۱۰ هر دوی این

مسائل جواب منحصر بفرد دارند . اگر $y_1(x)$ جواب (۵۰۱۰) و $y_2(x)$ جواب (۶۰۱۰)

باشد ، تحقیق اینکه

$$(۷۰۱۰) \quad y(x) = y_1(x) + \frac{\beta - y_1(b)}{y_2(b)} y_2(x)$$

جواب منحصر بفرد مسئله مقدار مرزی ماست مشکل نیست ، البته بشرطی که $y_2(b) \neq 0$.

اینکه $y_2(b) \neq 0$ مغایر مفروضات نتیجه ۲۰۱۰ است در تمرین ۸ بررسی می شود .

روش تیراندازی برای معادلات خطی بر جایگزینی مسئله مقدار مرزی به وسیله دو

مسئله مقدار اولیه (۵۰۱۰) و (۶۰۱۰) استوار است .

روشهای زیادی برای تقریب جوابهای $y_1(x)$ و $y_2(x)$ در فصل ۶ موجودند ، و زمانی که

این تقریبها در دسترس باشند ، جواب مسئله مقدار مرزی را می توان با استفاده از (۷۰۱۰)

تقریب کرد . به طور نموداری ، روش دارای شمای نشان داده شده در شکل ۱۰۱۰ می باشد .

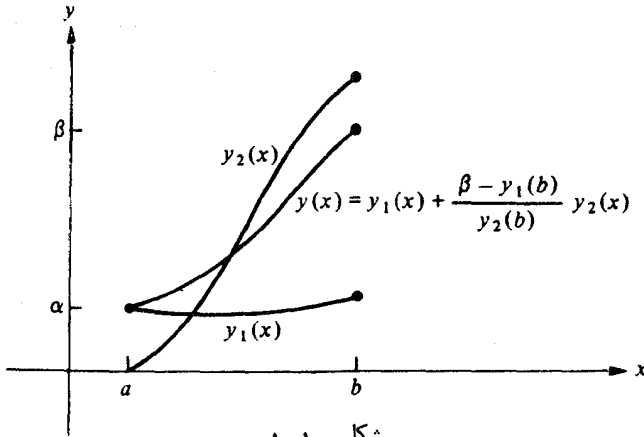
الگوریتم ارائه شده در زیر تکنیک چهار نقطه ای رونگ - کوتا را برای بدست آوردن

تقریبهایی به $y_1(x)$ و $y_2(x)$ بکار می برد ، اما هر تکنیک دیگری برای تقریب جوابهای

مسائل مقدار اولیه را می توان در مرحله ۴ جایگذاری کرد (ر.ک. تمرینهای ۴ و ۵) . این

الگوریتم دارای ویژگی اضافی تعیین تقریبهایی برای مشتق جواب مسئله مقدار مرزی ،

علاوه بر تقریب جواب مسئله ، می باشد .



استفاده از این الگوریتم به آن مسائلی که مفروضات نتیجه ۲.۱۰ برای آنها قابل تحقیق است محدود نمی‌شود؛ در واقع، نتایج رضایت بخشی برای مسائل زیادی که در این مفروضات صدق نمی‌کنند بدست می‌دهد.

الگوریتم تیراندازی خطی ۱.۱۰

برای بدست آوردن تقریبهایی به جواب مسئله مقدار مرزی

$$-y'' + p(x)y' + q(x)y + r(x) = 0, \quad a \leq x \leq b, \quad y(a) = \alpha, y(b) = \beta,$$

عدد صحیحی مانند $N > 0$ اختیار می‌کنیم.

مرحله ۱. قرار می‌دهیم $h = (b - a)/N$ و به ازای هر $i = 0, 1, \dots, N$ $x_i = a + ih$:

مرحله ۲. قرار می‌دهیم $i = 0$:

مرحله ۳. قرار می‌دهیم $u_{1,0} = \alpha, u_{2,0} = 0, v_{1,0} = 0, v_{2,0} = 1$

مرحله ۴. قرار می‌دهیم $k_{1,1} = hu_{2,i}$:

$$k_{1,2} = h[p(x_i)u_{2,i} + q(x_i)u_{1,i} + r(x_i)],$$

$$k_{2,1} = h[u_{2,i} + \frac{1}{2}k_{1,2}],$$

$$k_{2,2} = h \left[p \left(x_i + \frac{h}{2} \right) \left(u_{2,i} + \frac{1}{2}k_{1,2} \right) + q \left(x_i + \frac{h}{2} \right) \left(u_{1,i} + \frac{1}{2}k_{1,1} \right) + r \left(x_i + \frac{h}{2} \right) \right],$$

$$k_{3,1} = h[u_{2,i} + \frac{1}{2}k_{2,2}],$$

$$k_{3,2} = h \left[p \left(x_i + \frac{h}{2} \right) \left(u_{2,i} + \frac{1}{2}k_{2,2} \right) + q \left(x_i + \frac{h}{2} \right) \left(u_{1,i} + \frac{1}{2}k_{2,1} \right) + r \left(x_i + \frac{h}{2} \right) \right],$$

$$k_{4,1} = h[u_{2,i} + k_{3,2}],$$

$$k_{4,2} = h[p(x_i + h)(u_{2,i} + k_{3,2}) + q(x_i + h)(u_{1,i} + k_{3,1}) + r(x_i + h)],$$

$$u_{1,i+1} = u_{1,i} + \frac{1}{6}[k_{1,1} + 2k_{2,1} + 2k_{3,1} + k_{4,1}],$$

$$u_{2,i+1} = u_{2,i} + \frac{1}{6}[k_{1,2} + 2k_{2,2} + 2k_{3,2} + k_{4,2}],$$

$$k'_{1,1} = hv_{2,i},$$

$$k'_{1,2} = h[p(x_i)v_{2,i} + q(x_i)v_{1,i}],$$

$$k'_{2,1} = h[v_{2,i} + \frac{1}{2}k'_{1,2}],$$

$$k'_{2,2} = h \left[p \left(x_i + \frac{h}{2} \right) \left(v_{2,i} + \frac{1}{2}k'_{1,2} \right) + q \left(x_i + \frac{h}{2} \right) \left(v_{1,i} + \frac{1}{2}k'_{1,1} \right) \right],$$

$$k'_{3,1} = h[v_{2,i} + \frac{1}{2}k'_{2,2}],$$

$$k'_{3,2} = h \left[p \left(x_i + \frac{h}{2} \right) \left(v_{2,i} + \frac{1}{2}k'_{2,2} \right) + q \left(x_i + \frac{h}{2} \right) \left(v_{1,i} + \frac{1}{2}k'_{2,1} \right) \right],$$

$$k'_{4,1} = h[v_{2,i} + k'_{3,2}],$$

$$k'_{4,2} = h[p(x_i + h)(v_{2,i} + k'_{3,2}) + q(x_i + h)(v_{1,i} + k'_{3,1})],$$

$$v_{1,i+1} = v_{1,i} + \frac{1}{6}[k'_{1,1} + 2k'_{2,1} + 2k'_{3,1} + k'_{4,1}],$$

$$v_{2,i+1} = v_{2,i} + \frac{1}{6}[k'_{1,2} + 2k'_{2,2} + 2k'_{3,2} + k'_{4,2}].$$

مرحله ۱.۵ را به i می‌افزاییم.

مرحله ۱.۶. اگر $i < N$ ، به مرحله ۱ می‌رویم.

مرحله ۱.۷. به‌ازای هر $i = 0, 1, \dots, N$ ، قرار می‌دهیم

$$w_{1,i} = u_{1,i} + \left(\frac{\beta - u_{1,N}}{v_{1,N}} \right) v_{1,i},$$

$$w_{2,i} = u_{2,i} + \left(\frac{\beta - u_{1,N}}{v_{1,N}} \right) v_{2,i}.$$

مرحله ۸. روند تمام است، $w_{1,i}$ مقدار $y(x_i)$ و $w_{2,i}$ مقدار $y'(x_i)$ را به ازای هر $i = 0, 1, \dots, N$ تقریب می‌کند.

مثال ۲. مسئله مقدار مرزی

$$(۸.۱۰) \quad y'' = -\frac{2}{x}y' + \frac{2}{x^2}y + \frac{\sin(\ln x)}{x^2},$$

$$1 \leq x \leq 2, \quad y(1) = 1, \quad y(2) = 2,$$

دارای جواب دقیق

$$y = c_1x + \frac{c_2}{x^2} - \frac{3}{10} \sin(\ln x) - \frac{1}{10} \cos(\ln x),$$

است، که در آن

$$c_2 = \frac{1}{70}[8 - 12 \sin(\ln 2) - 4 \cos(\ln 2)] \approx -0.3920701320$$

و

$$c_1 = \frac{11}{10} - c_2 \approx 1.139207013.$$

اعمال الگوریتم ۱۰.۱۰ روی این مسئله به تقریب جوابهای مسائل مقدار اولیه زیر

نیاز دارد:

$$y_1'' = -\frac{2}{x}y_1' + \frac{2}{x^2}y_1 + \frac{\sin(\ln x)}{x^2},$$

$$1 \leq x \leq 2, \quad y_1(1) = 1, \quad y_1'(1) = 0,$$

و

$$y_2'' = -\frac{2}{x}y_2' + \frac{2}{x^2}y_2, \quad 1 \leq x \leq 2, \quad y_2(1) = 0, \quad y_2'(1) = 1.$$

نتایج محاسبات با بکار بردن $N = 10$ و $h = 0.1$ در جدول ۱۰.۱۰ داده شده

است.

x_i	$u_{1,i}$	$v_{1,i}$	w_i	$y(t_i)$	$ y(t_i) - w_i $
1.0	1.0000000	.0000000	1.0000000	1.0000000	—
1.1	1.00896058	.09117986	1.09262917	1.09262930	1.43×10^{-7}
1.2	1.03245472	.16851175	1.18708471	1.18708484	1.34×10^{-7}
1.3	1.06674375	.23608704	1.28338227	1.28338236	9.78×10^{-8}
1.4	1.10928795	.29659067	1.38144589	1.38144595	6.02×10^{-8}
1.5	1.15830000	.35184379	1.48115939	1.48115942	3.06×10^{-8}
1.6	1.21248372	.40311695	1.58239245	1.58239246	1.08×10^{-8}
1.7	1.27087454	.45131840	1.68501396	1.68501396	5.43×10^{-10}
1.8	1.33273851	.49711137	1.78889854	1.78889853	5.05×10^{-9}
1.9	1.39750618	.54098928	1.89392951	1.89392951	4.41×10^{-9}
2.0	1.46472815	.58332538	2.00000000	2.00000000	—

جدول ۱۰.۱۰

چون روش مرتبه چهار رونگ - کوتا دقت $O(h^4)$ به جوابهای مسائل مقدار اولیه می‌دهد، دقت نمایش داده شده در مثال ۲ انتظار می‌رفت. متأسفانه، به علت خطاهای گرد کردن، مسائل نهفته‌ای در این تکنیک وجود دارد. اگر β از نظر اندازه در مقایسه با $u_{1,N}$ کوچک باشد، $-u_{1,N}/v_{1,N}$ بر جمله $(\beta - u_{1,N})/v_{1,N}$ غالب می‌شود، و هرگاه $y_1(x)$ ، همچنانکه x از a به b می‌رود، سریعاً "افزایش یابد، آنگاه $u_{1,N}$ بزرگ خواهد بود. این امر می‌تواند سبب حذف ارقام با معنی در محاسبات مرحله γ شود. اما، چون $u_{1,i}$ تقریبی از $y_1(x_i)$ است، رفتار y_1 را به سادگی می‌توان کنترل کرد و هرگاه $u_{1,i}$ از a به b سریعاً "افزایش یابد، آنگاه می‌توان تکنیک تیراندازی را در جهت دیگر بکار برد؛ یعنی، در عوض مسائل مقدار مرزی

$$-y'' + p(x)y' + q(x)y + r(x) = 0,$$

$$a \leq x \leq b, \quad y(b) = \beta, \quad y'(b) = 0,$$

و

$$-y'' + p(x)y' + q(x)y = 0, \quad a \leq x \leq b, \quad y(b) = 0, \quad y'(b) = 1$$

را حل کرد.

اگر تکنیک تیراندازی معکوس باز هم حذف ارقام با معنی را دهد، و اگر نکته سنجی بیشتر دقت بیشتری را نتیجه ندهد، تکنیکهای دیگری، که در دو بخش آخر این فصل ارائه خواهند شد، باید بکار گرفته شوند. معیناً، در حالت کلی، هرگاه $u_{1,i}$ و $v_{1,i}$

بمازای هر $i = 0, 1, \dots, N$ ، بترتیب تقریبی به $y_1(x_i)$ و $y_2(x_i)$ از مرتبه $O(h^n)$ باشند، آنگاه می‌توان نشان داد که $w_{1,i}$ تقریبی از مرتبه $O(h^n)$ به $y_{1,i}$ خواهد بود. بالاخص، بمازای ثابت K ای،

$$|w_{1,i} - y(x_i)| \leq Kh^n \left| 1 + \frac{v_{1,i}}{v_{1,N}} \right|$$

(ر.ک. ایزاکسون و کلر [۳۹]، صفحه ۴۲۶).

مجموعه تمرینات ۱۰۱۰

۱. از مسائل مقدار مرزی زیر کدامها جواب منحصر بفرد دارند:

$$y'' = -\frac{4}{x}y' + \frac{2}{x^2}y - \frac{2 \ln x}{x^2}, \quad 1 < x < 2, \quad (T)$$

$$y(1) = -\frac{1}{2}, \quad y(2) = \ln 2;$$

$$y'' = -y - x, \quad 0 < x < \pi, \quad (A)$$

$$y(0) = 1, \quad y(\pi) = 2;$$

$$y'' = -\sin xy, \quad 0 < x < \pi, \quad (B)$$

$$y(0) = 0, \quad y(\pi) = 1;$$

$$y'' = \frac{1}{2}y^3, \quad 1 < x < 2, \quad (C)$$

$$y(1) = -\frac{2}{3}, \quad y(2) = -1;$$

$$y'' = y^3 - yy', \quad 1 < x < 2, \quad (D)$$

$$y(1) = \frac{1}{2}, \quad y(2) = \frac{1}{3};$$

$$y'' = -y, \quad 0 < x < \pi, \quad (E)$$

$$y(0) = 1, \quad y(\pi) = 1;$$

$$y'' = -y, \quad 0 < x < \pi, \quad (F)$$

$$y(0) = 1, \quad y(\pi) = -1;$$

$$y'' = -y, \quad 0 < x < \frac{\pi}{4}, \quad (G)$$

$$y(0) = 1, \quad y\left(\frac{\pi}{4}\right) = 1;$$

مسائل مقدار مرزی برای معادلات دیفرانسیل معمولی ۶۲۱

$$y'' = 2y^3 - 6y - 2x^3, \quad 1 < x < 2, \quad (\text{خ})$$

$$y(1) = 2, \quad y(2) = \frac{1}{2}.$$

۲. با استفاده از الگوریتم ۱۰۱۰، جواب مسائل مقدار مرزی زیر را تقریب کنید:

$$y'' = -\frac{4}{x}y' + \frac{2}{x^2}y - \frac{2 \ln x}{x^2}, \quad 1 < x < 2, \quad (\text{آ})$$

$$y(1) = -\frac{1}{2}, \quad y(2) = \ln 2;$$

$h = .05$ را بکار ببرید.

$$y'' = -4y' + 4y, \quad 0 < x < 5, \quad (\text{ب})$$

$$y(0) = 1, \quad y(5) = 0;$$

$h = .2$ را بکار ببرید.

$$y'' = -3y' + 2y + 2x + 3, \quad 0 < x < 1, \quad (\text{پ})$$

$$y(0) = 2, \quad y(1) = 1;$$

$h = .1$ را بکار ببرید.

۳. نشان دهید که نتیجه ۲۰۱۰ قابل اعمال روی مسائل مقدار مرزی زیر نیست، اما

الگوریتم ۱۰۱۰ را همچنان می‌توان بکار برد.

$$y'' = -\frac{4}{x}y' - \frac{2}{x^2}y + \frac{2}{x^2} \ln x, \quad 1 < x < 2, \quad (\text{آ})$$

$$y(1) = \frac{1}{2}, \quad y(2) = \ln 2;$$

$h = .05$ را بکار ببرید.

$$y'' = -y, \quad 0 < x < \frac{\pi}{4}, \quad (\text{ب})$$

$$y(0) = 1, \quad y\left(\frac{\pi}{4}\right) = 1;$$

$h = \frac{\pi}{40}$ را بکار ببرید.

۴. مسائل مقدار اولیه مرتبه دوم (۵۰۱۰) و (۶۰۱۰) را به صورت دستگاههای مرتبه اول بنویسید. و معادلات لازم برای حل دستگاهها را بدست آورید، برای دستگاهها

از روش مرتبه چهار رونگ - گوتا استفاده کنید.

۵. الگوریتمی شبیه الگوریتم ۱۰۱۰، با استفاده از مسائل مقدار اولیه ذیل، درست

کنید:

(آ) روش پیشگو - اصلاحگر آدامز - بشفورث و آدامز مولتون؛

(ب) برونیابی گراگ؛

(پ) روش رونگ - کوتا - فلبرگ.

۶. با استفاده از الگوریتمهای ساخته شده در تمرین ۵، تمرینهای ۲ و ۳ را تکرار کنید.

۷. با استفاده از الگوریتم ۱۰.۱۰ یا الگوریتمهای ساخته شده در تمرین ۵، جواب

$$y = e^{-10x}$$

برای مسئله مقدار مرزی

$$y'' = 100y, \quad 0 < x < 1,$$

$$y(0) = 1, \quad y(1) = e^{-10}$$

را تقریب کنید. $h = .1$ و $h = .05$ را بکار ببرید. آیا می‌توانید نتایج را توجیه کنید؟

۸. نشان دهید هرگاه y_2 جواب (۶.۱۰) بوده و $y_2(b) = 0$ ، آنگاه نتیجه ۲.۱۰

مستلزم $y_2 \equiv 0$ است.

۹. یک روش تیراندازی برای مسئله مقدار مرزی کلی زیر طراحی کنید:

$$y'' = p(x)y' + q(x)y + r(x), \quad a < x < b,$$

$$\alpha_1 y(a) - \beta_1 y'(a) = \alpha, \quad |\alpha_1| + |\beta_1| \neq 0,$$

$$\alpha_2 y(b) + \beta_2 y'(b) = \beta, \quad |\alpha_2| + |\beta_2| \neq 0.$$

۱۰. فرض کنید u نمایش الکترواستاتیک بین دو کره متحداالمرکز به شعاعهای R_1 و R_2

($R_1 < R_2$) باشد، بطوری که پتانسیل کره داخلی در V_1 ولت و پتانسیل کره

خارجی در 0 ولت ثابت نگهداشته شود. پتانسیل در ناحیه بین دو کره از معادله

لاپلاس^۱ تبعیت می‌کند (ر.ک. بخش ۲.۱۱) که در این کاربرد خاص چنین است:

$$\frac{d^2 u}{dr^2} + \frac{2}{r} \frac{du}{dr} = 0, \quad R_1 < r < R_2,$$

$$u(R_1) = V_1, \quad u(R_2) = 0.$$

فرض کنید 2 اینچ $R_1 =$ و 4 اینچ $R_2 =$ و 110 ولت $V_1 =$.

(آ) $u(3)$ را با استفاده از الگوریتم تیراندازی خطی ۱۰.۱۰، با $N = 20$ ، تقریب کنید.

(ب) $u(3)$ را با استفاده از الگوریتم تیراندازی خطی، با $N = 40$ ، تقریب کنید.

(پ) نتایج (T) و (ب) را با پتانسیل واقعی $u(3)$ که

$$u(r) = \frac{V_1 R_1}{r} \left(\frac{R_2 - r}{R_2 - R_1} \right).$$

مقایسه کنید.

۲۰۱۰ روش تیراندازی برای مسائل غیرخطی

ایده^۱ استفاده از تکنیک تیراندازی برای مسئله^۲ مقدار مرزی مرتبه^۳ دوم غیرخطی

$$(۹۰۱۰) \quad y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

مشابه حالت خطی است جز آنکه جواب یک مسئله^۴ غیرخطی را نمی‌توان به سادگی به صورت یک ترکیب خطی از جوابهای دو مسئله^۵ مقدار اولیه بیان کرد. در عوض، احتیاج داریم

که از جوابهای دنباله‌ای از مسائل مقدار اولیه به شکل

$$(۱۰۰۱۰) \quad y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y'(a) = t_k,$$

که شامل پارامتر t_k است استفاده کرده جواب مسئله^۶ مقدار مرزی را تقریب کنیم. این

عمل را با انتخاب پارامترهای t_k به گونه‌ای انجام می‌دهیم که اطمینان داشته باشیم

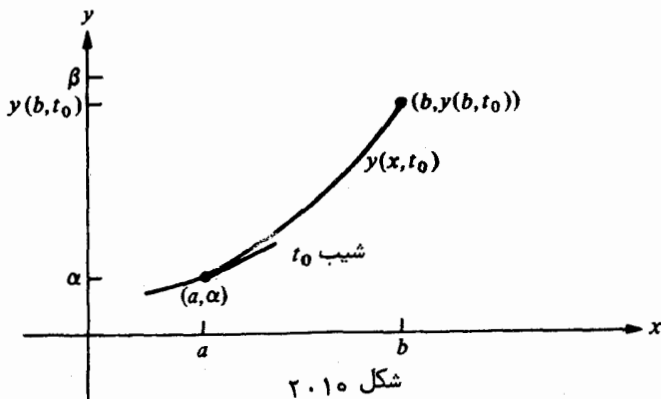
$$\lim_{k \rightarrow +\infty} y(b, t_k) = y(b) = \beta,$$

که در آن $y(x, t_k)$ نمایانگر جواب مسئله^۷ مقدار اولیه^۸ (۱۰۰۱۰) و $y(x)$ نمایانگر جواب

مسئله^۹ مقدار مرزی (۹۰۱۰) است.

این روش، بخاطر شباهت با تیراندازی به طرف اشیایی که هدف ثابتی هستند،

روش "تیراندازی" نامیده می‌شود (ر.ک. شکل ۲۰۱۰). با پارامتر t_0 ، که زاویه تیر-

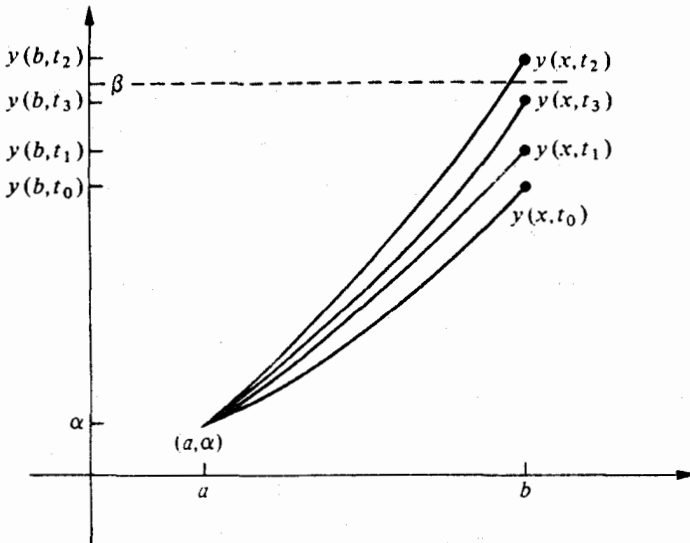


اندازی به طرف شیء در نقطه (a, α) و در امتداد منحنی مشخص شده با جواب مسئلهء مقدار اولیهء

$$y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y'(a) = t_0$$

را تعیین می‌کند، شروع می‌کنیم.

اگر $y(b, t_0)$ به قدر کافی به β نزدیک نباشد، سعی می‌کنیم تقریب خود را، با انتخاب زاویهء دیگر t_1 و به همین ترتیب، تا اینکه $y(b, t_k)$ به قدر کافی به β ی "هدف" نزدیک باشد تصحیح کنیم (ر.ک. شکل ۳۰۱۰).



شکل ۳۰۱۰

البته، در عمل نمی‌توان برحل دنبالهء مسائل مقدار اولیه به طور دقیق تکیه کرد؛ لذا، تقریب به مسئلهء مقدار مرزی خود بر تقریبیابی استوار است. برای اطمینان از اینکه این روند به نتایج رضایت‌بخش منجر خواهد شد لازم است شرایطی جهت اعمال بر $f(x, y, y')$ داشته باشیم تا حداقل مطمئن شویم که مسائل مقدار اولیه‌ای که تقریب می‌کنیم، و همچنین مسئلهء مقدار مرزی، در بازهء $[a, b]$ جوابهای منحصر بفرد دارند. قضیهء زیر حالت خاص قضیهء ۱۰۱۰ است که انجام روند را ضمانت خواهد کرد. خواننده را برای اثبات این قضیه مجدداً به کلر [۴۰]، صفحهء ۴۸، ارجاع می‌دهیم.

قضیه ۳.۱۰. مسئله مقدار مرزی

$$(11.10) \quad y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

و مسئله مقدار اولیه

$$(12.10) \quad y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y'(a) = t,$$

به ازای یک پارامتر دلخواه t ، بر $[a, b]$ جوابهای منحصر بفرد دارند اگر بر

$$D = \{(x, y, y') \mid a \leq x \leq b, \quad -\infty < y < \infty, \quad -\infty < y' < \infty\},$$

(A) f ، $\partial f / \partial y$ و $\partial f / \partial y'$ پیوسته باشند؛

(ب) ثابت M وجود داشته باشد بطوری که

$$\left| \frac{\partial f}{\partial y'} \right| \leq M;$$

و

(پ) ثابت L وجود داشته باشد بطوری که

$$0 < \partial f / \partial y < L.$$

برای تعیین چگونگی انتخاب پارامترهای t_k ، فرض کنیم یک مسئله مقدار مرزی به شکل (11.10) موجود باشد که در مفروضات قضیه ۳.۱۰ صدق می‌کنند. اگر $y(x, t)$ جهت نمایش جواب مسئله مقدار اولیه (12.10) بکار رود، سوال این است که t چگونه انتخاب شود که

$$(13.10) \quad y(b, t) - \beta = 0.$$

چون این یک معادله غیرخطی از نوع بررسی شده در فصل ۲ است، برای حل آن تعدادی روش در دسترس هستند. اگر بخواهیم روش خطاطع (الگوریتم ۴.۲ از بخش ۴.۲) را برای حل این مسئله بکار گیریم، احتیاج خواهیم داشت که تقریبهای اولیه t_0 و t_1 انتخاب و بعد بقیه دنباله را به وسیله

$$t_k = t_{k-1} - \frac{(y(b, t_{k-1}) - \beta)(t_{k-1} - t_{k-2})}{y(b, t_{k-1}) - y(b, t_{k-2})}, \quad k = 2, 3, \dots$$

تولید کنیم.

برای اینکه از روش تواناتر نباشد جهت تولید دنباله $\{t_k\}$ استفاده کنیم، تنها یک

اولیه t_0 لازم است. لیکن، تکرار به شکل زیر خواهد بود:

$$(14.10) \quad t_k = t_{k-1} - \frac{(y(b, t_{k-1}) - \beta)}{(dy/dt)(b, t_{k-1})}$$

که در آن

$$(dy/dt)(b, t_{k-1}) \equiv \frac{dy(b, t_{k-1})}{dt}$$

و دانستن $(dy/dt)(b, t_{k-1})$ لازم خواهد بود. این امر مشکل ایجاد می کند، زیرا $y(b, t)$ تنها برای $t = t_0, t_1, \dots, t_{k-1}$ معلوم است.

فرض کنیم مسئله مقدار اولیه (12.10) را، با تاکید روی اینکه جواب وابسته به هردوی x و t است، مجدداً "بنویسیم

$$(15.10) \quad y''(x, t) = f(x, y(x, t), y'(x, t)), \quad a \leq x \leq b, \quad y(a, t) = \alpha, \\ y'(a, t) = t,$$

با نگهداری علامت پریم جهت نشان دادن دیفرانسیل برحسب x . چون علاقه مند به تعیین $(dy/dt)(b, t)$ وقتی $t = t_{k-1}$ می باشیم، ابتدا از معادله (15.10) برحسب t مشتق جزئی می گیریم.

این نتیجه می دهد که به ازای $a \leq x \leq b$

$$(16.10) \quad \frac{\partial y''}{\partial t}(x, t) = \frac{\partial}{\partial t} f(x, y(x, t), y'(x, t)) \\ = \frac{\partial f}{\partial x}(x, y(x, t), y'(x, t)) \frac{\partial x}{\partial t} \\ + \frac{\partial f}{\partial y}(x, y(x, t), y'(x, t)) \frac{\partial y}{\partial t}(x, t) \\ + \frac{\partial f}{\partial y'}(x, y(x, t), y'(x, t)) \frac{\partial y'}{\partial t}(x, t) \\ = \frac{\partial f}{\partial y}(x, y(x, t), y'(x, t)) \frac{\partial y}{\partial t}(x, t) \\ + \frac{\partial f}{\partial y'}(x, y(x, t), y'(x, t)) \frac{\partial y'}{\partial t}(x, t)$$

و از شرایط اولیه نتیجه می شود

$$\frac{\partial y'}{\partial t}(a, t) = 1 \quad \text{و} \quad \frac{\partial y}{\partial t}(a, t) = 0$$

اگر نمادگذاری را با استفاده از $z(x, t)$ برای نمایاندن $(\partial y / \partial t)(x, t)$ ساده کرده و فرض کنیم ترتیب مشتقگیری x و t را بتوان عکس کرد، مسئله مقدار اولیه زیر را خواهیم داشت:

$$(17.10) \quad z'' = \frac{\partial f}{\partial y}(x, y, y')z + \frac{\partial f}{\partial y'}(x, y, y')z',$$

$$a \leq x \leq b, \quad z(a) = 0, \quad z'(a) = 1.$$

لذا، روش نیوتن برای هر تکرار به حل دو مسئله مقدار اولیه نیاز دارد، زیرا

(۱۴.۱۰) خواهد شد

$$(18.10) \quad t_k = t_{k-1} - \frac{y(b, t_{k-1}) - \beta}{z(b, t_{k-1})}.$$

در عمل، احتمالاً "هیچیک از این مسائل مقدار اولیه به طور دقیق حل نمی شوند؛ در عوض، جوابها به وسیله یکی از روشهای بحث شده در فصل ۶ تقریب می شوند. الگوریتمی که ذیلاً ارائه می شود از روش مرتبه چهار رونگ-کوتا، جهت تقریب هر دو جواب لازم به وسیله روش نیوتن، استفاده می کند. در تمرین ۲، روندی مشابه برای روش خط قاطع بررسی می شود.

الگوریتم تیراندازی غیرخطی با روش نیوتن ۲.۱۰

برای تقریب جواب مسئله مقدار مرزی غیرخطی

$$y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

عدد صحیح $N > 0$ را انتخاب می کنیم.

مرحله ۱. قرار می دهیم $h = (b - a)/N$ و به ازای هر $i = 0, 1, \dots, N$ $x_i = a + ih$.

مرحله ۲. قرار می دهیم $k = 0$ و $t_k = (\beta - \alpha)/(b - a)$.

مرحله ۳. قرار می دهیم $i = 1$.

مرحله ۴. قرار می دهیم $w_{1,0}(t_k) = \alpha$, $w_{2,0}(t_k) = t_k$.

$$u_{1,0}(t_k) = 0, \quad u_{2,0}(t_k) = 1.$$

مرحله ۵. قرار می دهیم

$$k_{1,1} = hw_{2,i-1},$$

$$k_{1,2} = hf(x_{i-1}, w_{1,i-1}, w_{2,i-1}),$$

$$\begin{aligned}
 k_{2,1} &= h(w_{2,i-1} + \frac{1}{2}k_{1,2}), \\
 k_{2,2} &= hf\left(x_{i-1} + \frac{h}{2}, w_{1,i-1} + \frac{1}{2}k_{1,1}, w_{2,i-1} + \frac{1}{2}k_{1,2}\right), \\
 k_{3,1} &= h(w_{2,i-1} + \frac{1}{2}k_{2,2}), \\
 k_{3,2} &= hf\left(x_{i-1} + \frac{h}{2}, w_{1,i-1} + \frac{1}{2}k_{2,1}, w_{2,i-1} + \frac{1}{2}k_{2,2}\right), \\
 k_{4,1} &= h(w_{2,i-1} + k_{3,2}), \\
 k_{4,2} &= hf(x_{i-1} + h, w_{1,i-1} + k_{3,1}, w_{2,i-1} + k_{3,2}), \\
 w_{1,i}(t_k) &= w_{1,i-1}(t_k) + \frac{1}{6}(k_{1,1} + 2k_{2,1} + 2k_{3,1} + k_{4,1}), \\
 w_{2,i}(t_k) &= w_{2,i-1}(t_k) + \frac{1}{6}(k_{2,1} + 2k_{2,2} + 2k_{3,2} + k_{4,2}), \\
 k'_{1,1} &= hu_{2,i-1}, \\
 k'_{1,2} &= h[f_y(x_{i-1}, w_{1,i-1}(t_k), w_{2,i-1}(t_k))u_{1,i-1} \\
 &\quad + f_y(x_{i-1}, w_{1,i-1}(t_k), w_{2,i-1}(t_k))u_{2,i-1}], \\
 k'_{2,1} &= h[u_{2,i-1} + \frac{1}{2}k'_{1,2}], \\
 k'_{2,2} &= h\left[f_y\left(x_{i-1} + \frac{h}{2}, w_{1,i-1}(t_k), w_{2,i-1}(t_k)\right)\left(u_{1,i-1} + \frac{1}{2}k'_{1,1}\right) \right. \\
 &\quad \left. + f_y\left(x_{i-1} + \frac{h}{2}, w_{1,i-1}(t_k), w_{2,i-1}(t_k)\right) \right. \\
 &\quad \left. \times \left(u_{2,i-1} + \frac{1}{2}k'_{2,1}\right)\right], \\
 k'_{3,1} &= h(u_{2,i-1} + \frac{1}{2}k'_{2,2}), \\
 k'_{3,2} &= h\left[f_y\left(x_{i-1} + \frac{h}{2}, w_{1,i-1}(t_k), w_{2,i-1}(t_k)\right)\left(u_{1,i-1} + \frac{1}{2}k'_{2,1}\right) \right. \\
 &\quad \left. + f_y\left(x_{i-1} + \frac{h}{2}, w_{1,i-1}(t_k), w_{2,i-1}(t_k)\right) \right. \\
 &\quad \left. \times \left(u_{2,i-1} + \frac{1}{2}k'_{2,2}\right)\right], \\
 k'_{4,1} &= h(u_{2,i-1} + k'_{3,2}),
 \end{aligned}$$

$$k'_{3,2} = h[f_y(x_{i-1} + h, w_{1,i-1}(t_k), w_{2,i-1}(t_k))(u_{1,i-1} + k'_{3,1}) + f_y(x_{i-1} + h, w_{1,i-1}(t_k), w_{2,i-1}(t_k))(u_{2,i-1} + k'_{3,2})]$$

$$u_{1,i} = u_{1,i-1} + \frac{1}{6}[k'_{1,1} + 2k'_{2,1} + 2k'_{3,1} + k'_{4,1}],$$

$$u_{2,i} = u_{2,i-1} + \frac{1}{6}[k'_{1,2} + 2k'_{2,2} + 2k'_{3,2} + k'_{4,2}].$$

مرحله ۶. ۱ را به i می‌افزاییم .

مرحله ۷. اگر $i \leq N$ ، به مرحله ۵ می‌رویم .

مرحله ۸. اگر $|w_{1,N}(t_k) - \beta|$ به قدر کافی کوچک باشد ، به مرحله ۱۱ می‌رویم .

مرحله ۹. قرار می‌دهیم

$$t_{k+1} = t_k - \left(\frac{w_{1,N}(t_k) - \beta}{u_{1,N}(t_k)} \right).$$

مرحله ۱۰. ۱ را به k افزوده به مرحله ۳ می‌رویم .

مرحله ۱۱. روند تمام است ؛ مقدار $w_{1,N}(t_k)$ و مقدار $w_{2,N}(t_k)$ مقدار $y'(x_i)$ را ، به‌ازای

$i = 0, 1, \dots, N$ ، تقریب می‌کند .

در مرحله ۸ ، بهترین تقریبی که می‌توان انتظار داشت $w_{1,N}(t_k)$ به β بدهد $O(h^n)$

است ، اگر روش تفاضل متناهی انتخاب شده در مرحله ۵ سرعت همگرایی $O(h^n)$ داشته

باشد . مقدار t_0 که در مرحله ۲ انتخاب شده ضریب زاویه خط مستقیم ماربر (a, α) و

(b, β) است . اگر مسئله در مفروضات قضیه ۳.۱۰ صدق کند ، هر انتخاب t_0 همگرایی

خواهد داد ؛ اما ، در حالت کلی ، روند برای بسیاری از مسائل که برای آنها این مفروضات

برقرار نیستند نیز کار می‌کند ، هرچند که انتخاب خوبی از t_0 لازم است .

مثال ۱. مسئله مقدار مرزی

$$(19.10) \quad y'' = \frac{1}{8}(32 + 2x^3 - yy'), \quad 1 \leq x \leq 3, \quad y(1) = 17, \quad y(3) = \frac{43}{3},$$

که دارای جواب دقیق $y(x) = x^2 + (16/x)$ می‌باشد ، را در نظر می‌گیریم .

اعمال روش تیراندازی ارائه شده در الگوریتم ۲.۱۰ روی این مسئله نیاز به تقریب

مسائل مقدار اولیه

$$y'' = \frac{1}{8}(32 + 2x^3 - yy'), \quad 1 \leq x \leq 3, \quad y(1) = 17, \quad y'(1) = t_k,$$

$$z'' = -\frac{1}{8}(yz' + y'z), \quad 1 \leq x \leq 3, \quad z(1) = 0, \quad z'(1) = 1,$$

در هر مرحله تکرار دارد. اگر تکنیک توقف

$$|w_{1,N}(t_k) - y(3)| \leq 10^{-5}$$

بکار رود، این مسئله ۴ تکرار لازم دارد و $t_4 = -14.000203$. نتایج بدست آمده برای این مقدار t در جدول ۲.۱۰ نشان داده شده است.

x_i	w_{1i}	$y(x_i)$	$ w_{1i} - y(x_i) $
1.0	17.000000	17.000000	—
1.1	15.755495	15.755455	4.06×10^{-5}
1.2	14.773389	14.773333	5.60×10^{-5}
1.3	13.997752	13.997692	5.94×10^{-5}
1.4	13.388629	13.388571	5.71×10^{-5}
1.5	12.916719	12.916667	5.23×10^{-5}
1.6	12.560046	12.560000	4.64×10^{-5}
1.7	12.301805	12.301765	4.02×10^{-5}
1.8	12.128923	12.128889	3.41×10^{-5}
1.9	12.031081	12.031053	2.84×10^{-5}
2.0	12.000023	12.000000	2.32×10^{-5}
2.1	12.029066	12.029048	1.84×10^{-5}
2.2	12.112741	12.112727	1.40×10^{-5}
2.3	12.246532	12.246522	1.01×10^{-5}
2.4	12.426673	12.426667	6.68×10^{-6}
2.5	12.650004	12.650000	3.61×10^{-6}
2.6	12.913847	12.913846	9.17×10^{-7}
2.7	13.215924	13.215926	1.43×10^{-6}
2.8	13.554282	13.554286	3.47×10^{-6}
2.9	13.927236	13.927241	5.21×10^{-6}
3.0	14.333327	14.333333	6.69×10^{-6}

جدول ۲.۱۰

مسئله موجود در این مثال به این خاطر انتخاب شده بود که یک مینیمم نسبی، برای جواب دقیق، در $x = 2$ اتفاق می افتد. با یکنوا بودن جواب دقیق، در بازه مورد نظر، انتظار نمی رفت که برای این درجه دقت ۴ تکرار لازم باشد.

باید متذکر شد که گرچه استفاده از روش نیوتن با تکنیک تیراندازی محتاج حل یک مسئله مقدار اولیه اضافی است، اما در حالت کلی سریعتر از روش خط قاطع است. هر دو روش فقط به طور موضعی همگرا هستند به این معنی که هر وقت مفروضات قضیه ۳.۱۰

برقرار نیستند، به تقریبهای اولیه خوب نیاز دارند.

برای یک بحث کلی از همگرایی تکنیکهای تیراندازی در مسائل غیر خطی، خواننده به کتاب درسی عالی توسط کالر [۴۰] ارجاع داده می‌شود. در این مرجع شرایط مرزی کلیتر نیز مورد بحث قرار گرفته است و نیز توجه داده شده است که تکنیک تیراندازی برای مسائل غیر خطی، در صورتی که $y(x)$ و $z(x)$ توابعی سریعاً "صعودی بر $[a, b]$ باشند، نسبت به خطاهای گرد کردن حساس است.

مجموعه تمرینات ۲۰۱۰

۱. الگوریتم تیراندازی غیرخطی با روش نیوتن (الگوریتم ۲۰۱۰) را برای تقریب جوابهای مسائل مقدار مرزی زیر بکار ببرید. جوابهای واقعی، جهت مقایسه با نتایج شما، داده شده است. آیا مفروضات قضیه ۳۰۱۰ برقرار است؟

$$y'' = \frac{1}{2}y^3, \quad 1 < x < 2, \quad (A)$$

$$y(1) = -\frac{2}{3}, \quad y(2) = -1;$$

۰.۵ h را بکار ببرید و با $y(x) = 2/(x-4)$ مقایسه کنید.

$$y'' = 2y^3, \quad 1 < x < 5, \quad (B)$$

$$y(1) = \frac{1}{4}, \quad y(5) = \frac{1}{8};$$

۰.۲ h را بکار ببرید و با $y(x) = 1/(x+3)$ مقایسه کنید.

$$y'' = y^3 - yy', \quad 1 < x < 2, \quad (C)$$

$$y(1) = \frac{1}{2}, \quad y(2) = \frac{1}{3};$$

۰.۱ h را بکار ببرید و با $y(x) = 1/(x+1)$ مقایسه کنید.

$$y'' = 2y^3 - 6y - 2x^3, \quad 1 < x < 2, \quad (D)$$

$$y(1) = 2, \quad y(2) = \frac{2}{3};$$

۰.۵ h را بکار ببرید و با $y(x) = x + (1/x)$ مقایسه کنید.

۲. الگوریتم ۲۰۱۰ را جهت آمیختن با روش خطاطع به جای روش نیوتن تغییر دهید. تمرین ۱ را با استفاده از الگوریتم جدید تکرار کنید و تعداد تکرارهای لازم را مقایسه نمایید.

۳. الگوریتم ۲۰۱۰ را جهت آمیختن با روشهای

(A) پیشگو - اصلاحگر، آدامز - بشفورث و آدامز - مولتون؛

(ب) برونیایی گراگ ؛

(پ) رونگ - کوتا - فلبرگ

به جای روش مرتبه چهار رونگ - کوتا تغییر دهید .

۴ . قسمتهای (آ) ، (ب) ، و (پ) از الگوریتم بدست آمده در تمرین ۳ را جهت تقریب جوابهای مسائل داده شده در تمرین ۱ بکار برید . برای هر مسئله معین کنید کدام یک دقیقترین است .

۳.۱۰ روشهای تفاضل متناهی برای مسائل مقدار مرزی

هرچند روشهای تیراندازی ، ارائه شده در قسمت اول این فصل ، را می توان برای هر دو مسئله مقدار مرزی خطی و غیرخطی بکار برد ، ولی غالباً " مشکلات ناپایداری در بردارند . روشهایی که اینجا ارائه خواهیم کرد خصوصیات پایداری بهتری دارند اما عموماً " جهت بدست آوردن دقت مشخص شده به کار بیشتری نیاز دارند .

روشهای مشتعل بر تفاضلات متناهی ، برای حل مسائل مقدار مرزی ، از جایگزینی هریک از مشتقات در معادله دیفرانسیل با یک تقریب تفاضل - خارج قسمت مناسب تشکیل می شود . خارج قسمت تفاضلی بطور کلی چنان انتخاب می شود که مرتبه معینی ر برشی تا مین شود .

مسئله مقدار مرزی مرتبه دوم خطی

$$(20.10) \quad y'' = p(x)y' + q(x)y + r(x), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

محتاج کاربرد تقریبهای تفاضل - خارج قسمت برای هر دوی y' و y'' است . برای انجام این کار ، عدد صحیح $N > 0$ را انتخاب و بازه $[a, b]$ را به $(N + 1)$ زیر بازه مساوی تقسیم می کنیم که نقاط انتهایی آنها نقاط شبکه های $x_i = a + ih$ ، به ازای $i = 0, 1, \dots, N + 1$ است که در آن $h = (b - a)/(N + 1)$. انتخاب h به این طریق ، کاربرد یک الگوریتم ماتریسی از فصل ۷ ، که در این شکل به حل یک دستگاه خطی شامل یک ماتریس $N \times N$ نیاز دارد ، سهولت می بخشد .

در نقاط شبکه های داخلی x_i ، $i = 1, 2, \dots, N$ ، معادله دیفرانسیلی که باید

تقریب کرد عبارت است از

$$(21.10) \quad y''(x_i) = p(x_i)y'(x_i) + q(x_i)y(x_i) + r(x_i).$$

با بسط y' حول نقطه x_i به صورت یک چند جمله ای تیلور درجه سوم که در x_{i-1}

و x_{i-1} حساب شده است و فرض اینکه $y \in C^4[x_{i-1}, x_{i+1}]$ ، به‌ازای نقطه‌ای چون ξ_i^+ داریم $x_i < \xi_i^+ < x_{i+1}$

$$(22.10) \quad \begin{aligned} y(x_{i+1}) = y(x_i + h) = & y(x_i) + hy'(x_i) + \frac{h^2}{2} y''(x_i) \\ & + \frac{h^3}{6} y'''(x_i) + \frac{h^4}{24} y^{(4)}(\xi_i^+), \end{aligned}$$

و ، به‌ازای نقطه‌ای چون ξ_i^- که $x_{i-1} < \xi_i^- < x_i$ ،

$$(23.10) \quad \begin{aligned} y(x_{i-1}) = y(x_i - h) = & y(x_i) - hy'(x_i) + \frac{h^2}{2} y''(x_i) - \frac{h^3}{6} y'''(x_i) \\ & + \frac{h^4}{24} y^{(4)}(\xi_i^-), \end{aligned}$$

اگر این معادلات با هم جمع شوند ، جملات شامل $y'(x_i)$ و $y'''(x_i)$ حذف می‌شوند و از عملیات جبری ساده‌ای نتیجه می‌شود

$$\begin{aligned} y''(x_i) = & \frac{1}{h^2} [y(x_{i+1}) - 2y(x_i) + y(x_{i-1})] \\ & - \frac{h^2}{24} [y^{(4)}(\xi_i^+) + y^{(4)}(\xi_i^-)]. \end{aligned}$$

قضیه مقدار میانی را می‌توان جهت ساده کردن حتی بیشتر این معادله بکار برد : به‌ازای نقطه‌ای چون ξ_i که $x_{i-1} < \xi_i < x_{i+1}$ ،

$$(24.10) \quad y''(x_i) = \frac{1}{h^2} [y(x_{i+1}) - 2y(x_i) + y(x_{i-1})] - \frac{h^2}{12} y^{(4)}(\xi_i),$$

معادله (24.10) فرمول تفاضل مرکزی برای $y''(x_i)$ نامیده می‌شود .

یک فرمول تفاضل مرکزی برای $y'(x_i)$ را می‌توان به‌طریقی مشابه بدست آورد (تفصیلات در تمرین ۴ بررسی می‌شوند) که نتیجه می‌دهد که به‌ازای η_i ای که $x_{i-1} < \eta_i < x_{i+1}$ ،

$$(25.10) \quad y'(x_i) = \frac{1}{2h} [y(x_{i+1}) - y(x_{i-1})] - \frac{h^2}{6} y'''(\eta_i)$$

کاربرد این فرمولهای تفاضل مرکزی در معادله (21.10) معادله زیر را نتیجه

می‌دهد :

$$\frac{y(x_{i+1}) - 2y(x_i) + y(x_{i-1}))}{h^2} = p(x_i) \left[\frac{y(x_{i+1}) - y(x_{i-1}))}{2h} \right] + q(x_i)y(x_i) + r(x_i) - \frac{h^2}{12} [2p(x_i)y'''(\eta_i) - y^{(4)}(\xi_i)].$$

با استفاده از این معادله به همراه شرایط مرزی $y(a) = \alpha$ و $y(b) = \beta$ ، یک روش تفاضل متناهی با خطای برشی از مرتبه $O(h^2)$ با تعریف

$$w_0 = \alpha, \quad w_{N+1} = \beta,$$

و ، به ازای هر $i = 1, 2, \dots, N$

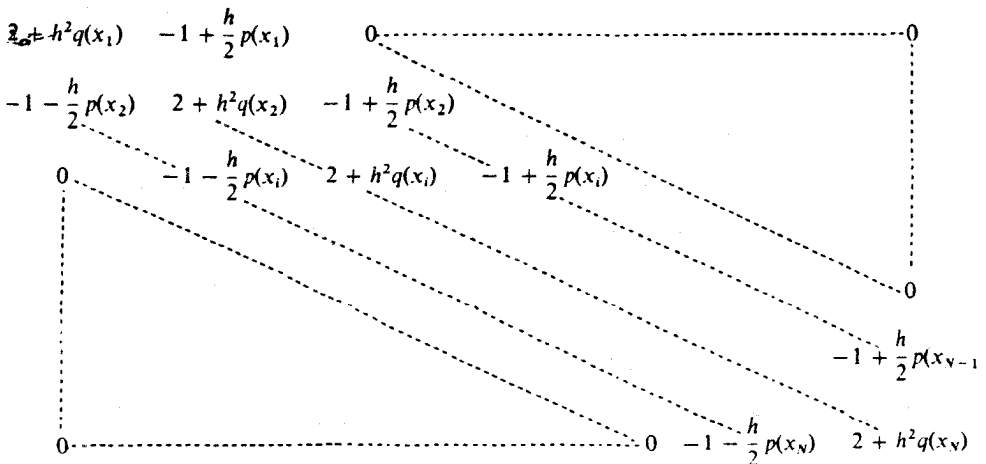
$$(26.10) \quad \left(\frac{2w_i - w_{i+1} - w_{i-1}}{h^2} \right) + p(x_i) \left(\frac{w_{i+1} - w_{i-1}}{2h} \right) + q(x_i)w_i = -r(x_i)$$

نتیجه می شود .

معادله (26.10) را مجدداً به شکلی که بررسی خواهیم کرد می نویسیم :

$$\begin{aligned} & - \left(1 + \frac{h}{2} p(x_i) \right) w_{i-1} + (2 + h^2 q(x_i)) w_i - \left(1 - \frac{h}{2} p(x_i) \right) w_{i+1} \\ & = -h^2 r(x_i), \end{aligned}$$

و دستگاه معادلات حاصل به شکل ماتریس سه قطری $N \times N$ بیان شده در معادله (27.10) نشان داده شده است .



$$(27.10) \quad \begin{matrix} w_1 \\ w_2 \\ \vdots \\ w_i \\ \vdots \\ w_{N-1} \\ w_N \end{matrix} \times = \begin{bmatrix} -h^2 r(x_1) + \left(1 + \frac{h}{2} p(x_1)\right) w_0 \\ -h^2 r(x_2) \\ \vdots \\ -h^2 r(x_i) \\ \vdots \\ -h^2 r(x_{N-1}) \\ -h^2 r(x_N) + \left(1 - \frac{h}{2} p(x_N)\right) w_{N+1} \end{bmatrix}$$

قضیه زیرشرایطی را می‌دهد که تحت آنها دستگاه خطی سه قطری (۲۶.۱۰) جواب منحصر بفرد دارد. برهان آن نتیجه‌ای از قضیه ۲۵.۷ (ص ۴۹۱) است و در تمرین ۵ بررسی می‌شود.

قضیه ۴.۱۰. فرض کنیم p ، q ، و r بر $[a, b]$ پیوسته باشند. هرگاه بر $[a, b]$ ، $q(x) \geq 0$ ، آنگاه دستگاه خطی سه قطری (۲۷.۱۰) دارای جواب منحصر بفرد است بشرط آنکه $h < 2/L$ که در آن $L = \max_{a \leq x \leq b} |p(x)|$.

باید توجه شود که مفروضات قضیه ۴.۱۰ جواب منحصر بفرد برای مسئله مقدار مرزی (۲۰.۱۰) را تضمین می‌کند، اما تضمین نمی‌کند که $y \in C^2[a, b]$ برای اطمینان از اینکه خطای برشی دارای مرتبه $O(h^2)$ است لازم است ثابت کنیم که $y^{(4)}$ بر $[a, b]$ پیوسته است.

الگوریتم تفاضل متناهی خطی ۳.۱۰

برای تقریب جواب مسئله مقدار مرزی

$$y'' = p(x)y' + q(x)y + r(x), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

عدد صحیح $N > 0$ را انتخاب می‌کنیم.

مرحله ۱. قرار می‌دهیم $h = (b - a)/(N + 1)$ و به ازای هر $i = 0, 1, \dots, N + 1$

$$x_i = a + ih$$

مرحله ۲. قرار می‌دهیم

بمازای هر $i = 1, 2, \dots, N$ ، $a_i = 2 + h^2 q(x_i)$ ،

بمازای هر $i = 2, 3, \dots, N - 1$ ، $b_i = -1 + \frac{h}{2} p(x_i)$ ،

بمازای هر $i = 2, 3, \dots, N$ ، $c_i = -1 - \frac{h}{2} p(x_i)$ ،

$$d_1 = -h^2 r(x_1) + \left(1 + \frac{h}{2} p(x_1)\right) \alpha,$$

$$d_N = -h^2 r(x_N) + \left(1 - \frac{h}{2} p(x_N)\right) \beta,$$

بمازای هر $i = 1, 2, \dots, N - 1$ ، $d_i = -h^2 r(x_i)$ ،

مرحله ۳ . قرار می دهیم

$$l_1 = a_1,$$

$$u_1 = b_1/a_1.$$

مرحله ۴ . قرار می دهیم $i = 2$.

مرحله ۵ . قرار می دهیم

$$l_i = a_i - c_i u_{i-1},$$

$$u_i = b_i/l_i.$$

مرحله ۶ . 1 را به i می افزاییم.

مرحله ۷ . اگر $i < N$ ، به مرحله ۵ می رویم .

مرحله ۸ . قرار می دهیم $l_N = a_N - c_N u_{N-1}$.

مرحله ۹ . قرار می دهیم $z_1 = d_1/l_1$ ،

بمازای هر $i = 2, 3, \dots, N$ ، $z_i = \frac{1}{l_i} [d_i - c_i z_{i-1}]$ ،

مرحله ۱۰ . قرار می دهیم $w_0 = \alpha, w_{N+1} = \beta, w_N = z_N$ ،

بمازای هر $i = N - 1, N - 2, \dots, 1$ ، $w_i = z_i - u_i w_{i+1}$ ،

مرحله ۱۱ . روند تمام است ؛ w_i ، بمازای هر $i = 0, 1, \dots, N + 1$ ، $y(x_i)$ را تقریب

می کند .

مثال ۱ . برای تقریب جواب مسئله مقدار مرزی خطی

$$y'' = -\frac{2}{x}y' + \frac{2}{x^2}y + \frac{\sin(\ln x)}{x^2}, \quad 1 \leq x \leq 2, \quad y(1) = 1, \quad y(2) = 2,$$

که در مثال ۲، بخش ۱۰۱۰، با روش تیراندازی نیز تقریب شده بود، از الگوریتم ۳۰۱۰ استفاده خواهد شد. برای این مثال از $N = 9$ استفاده می‌کنیم بطوری که $h = 0.1$ و فاصله‌های همانند مثال ۲، بخش ۱۰۱۰، داریم. نتایج در جدول ۳۰۱۰ فهرست شده‌اند.

x_i	w_i	$y(x_i)$	$ w_i - y(x_i) $
1.0	1.00000000	1.00000000	—
1.1	1.09260052	1.09262930	2.88×10^{-5}
1.2	1.18704313	1.18708484	4.17×10^{-5}
1.3	1.28333687	1.28338236	4.55×10^{-5}
1.4	1.38140205	1.38144595	4.39×10^{-5}
1.5	1.48112026	1.48115942	3.92×10^{-5}
1.6	1.58235990	1.58239246	3.26×10^{-5}
1.7	1.68498902	1.68501396	2.49×10^{-5}
1.8	1.78888175	1.78889853	1.68×10^{-5}
1.9	1.89392110	1.89392951	8.41×10^{-6}
2.0	2.00000000	2.00000000	—

جدول ۳۰۱۰

باید توجه شود که دقت این نتایج به‌طور قابل ملاحظه‌ای از نتایج بدست آمده در مثال ۲، بخش ۱۰۱۰، کمتر است. این مطلب تعجب‌انگیز نیست، زیرا روش بکار برده شده در آن مثال شامل تکنیک رونگ - کوتا با خطای برشی از مرتبه $O(h^4)$ بوده در حالی که روش تفاضلی بکار برده شده خطای برشی مرتبه $O(h^2)$ دارد.

برای بدست آوردن یک روش تفاضلی با دقت بیشتر می‌توان به چند طریق عمل کرد. استفاده از سری تیلور مرتبه پنجم جهت تقریب $y'(x_i)$ و $y''(x_i)$ یک جمله خطای برشی شامل h^4 نتیجه خواهد داد. اما این روش نه فقط به‌کاربرد مضاربی از $y(x_{i+1})$ و $y(x_{i-1})$ بلکه به $y(x_{i+2})$ و $y(x_{i-2})$ نیز، در فرمولهای تقریب $y'(x_i)$ و $y''(x_i)$ ، نیاز دارد که منجر به اشکالاتی در $i = 0$ و $i = N$ می‌شود. علاوه، دستگاه معادلات حاصل مشابه (۲۷.۱۰) به شکل سه قطری نخواهد بود و حل این دستگاه محاسبات بسیار بیشتری لازم خواهد داشت. همچنین، این روند لازم دارد که جواب مسئله مقدار مرزی دارای مشتق ششم کراندار باشد که تحقیق آن، حتی وقتی درست باشد، برای اکثر مسائل ناشی شده از

وضعیت‌های فیزیکی مشکل است .

به‌جای کوشش جهت بدست آوردن یک روش تفاضلی با خطای برشی مرتبه بالاتر به همین طریق ، کلا " رضایت‌بخش‌تر است که تقلیلی در اندازه گام در نظر گرفته شود . بعلاوه ، می‌توان نشان داد (مثلا ، " ر.ک. کلا [۴۰] ، صفحه ۸۱) که تکنیک برون‌یابی ریچاردسون را برای این روش می‌توان مورد استفاده قرار داد ، زیرا جمله خطا برحسب توانهای زوج h با ضرایب مستقل از h ، بشرطی که y به قدر کافی مشتق‌پذیر باشد ، بیان می‌شود .

مثال ۲ . اگر از روش برون‌یابی ریچاردسون ، که در بخشهای ۲.۵ ، ۶.۵ ، و ۷.۶ مورد بحث قرار گرفت ، جهت تقریب جواب مسئله مقدار مرزی

$$y'' = -\frac{2}{x}y' + \frac{2}{x^2}y + \frac{\sin(\ln x)}{x^2}, \quad 1 \leq x \leq 2, \quad y(1) = 1, \quad y(2) = 2,$$

با $h = .1$ ، $h = .05$ ، و $h = .025$ استفاده کنیم ، نتایج فهرست شده در جدول ۴.۱۰

x_i	$w_i(h = .1)$	$w_i(h = .05)$	$w_i(h = .025)$	Ext_{1i}	Ext_{2i}	Ext_{3i}
1.0	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000
1.1	1.09260052	1.09262207	1.09262749	1.09262925	1.09262930	1.09262930
1.2	1.18704313	1.18707436	1.18708222	1.18708477	1.18708484	1.18708484
1.3	1.28333687	1.28337094	1.28337950	1.28338230	1.28338236	1.28338236
1.4	1.38140205	1.38143493	1.38144319	1.38144589	1.38144595	1.38144595
1.5	1.48112026	1.48114959	1.48115696	1.48115937	1.48115941	1.48115942
1.6	1.58235990	1.58238429	1.58239042	1.58239242	1.58239246	1.58239246
1.7	1.68498902	1.68500770	1.68501240	1.68501393	1.68501396	1.68501396
1.8	1.78888175	1.78889432	1.78889748	1.78889852	1.78889853	1.78889853
1.9	1.89392110	1.89392740	1.89392898	1.89392950	1.89392951	1.89392951
2.0	2.00000000	2.00000000	2.00000000	2.00000000	2.00000000	2.00000000

جدول ۴.۱۰

را بدست می‌آوریم . اولین برون‌یابی عبارت است از

$$Ext_{1i} = \frac{4w_i(h = .05) - w_i(h = .1)}{3},$$

دوم - مشابهی عبارت است از

$$\text{Ext}_{2i} = \frac{4w_i(h = .025) - w_i(h = .05)}{3},$$

و برونمایی نهایی عبارت است از

$$\text{Ext}_{3i} = \frac{16\text{Ext}_{2i} - \text{Ext}_{1i}}{15}.$$

تمام نتایج Ext_{3i} تا ارقام اعشار فهرست شده درست هستند؛ در واقع، این تقریب با خطای ماکزیم 6.3×10^{-11} ، در نقاط شبکه‌ای با جواب دقیق منطبق است.

برای مسئله مقدار مرزی غیرخطی کلی

$$(28.10) \quad y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

روش تفاضلی مشابه است. اما در اینجا دستگاه معادلاتی که بدست می‌آید خطی نخواهد بود. لذا، یکفرایند تکراری برای حل آن لازم است.

برای بدست آوردن این روند، از این به بعد فرض می‌کنیم f در شرایط زیر صدق

می‌کند:

(A) f و مشتقات جزئی $f_{y'}$ و f_y بر

$$D = \{(x, y, y') | a \leq x \leq b, -\infty < y < \infty, -\infty < y' < \infty\};$$

پیوسته هستند؛

(ب) بهازای $\delta > 0$ ، $f_y \geq \delta > 0$ بر D ؛

(پ) ثابتهای k و L که

$$k = \max_{(x, y, y') \in D} \left| \frac{\partial f}{\partial y} (x, y, y') \right|, \quad L = \max_{(x, y, y') \in D} \left| \frac{\partial f}{\partial y'} (x, y, y') \right|.$$

وجود دارند.

این مطلب، طبق قضیه ۱۰.۱۰، اطمینان می‌دهد که یک جواب منحصر بفرد به (28.10) وجود دارد.

همانند حالت خطی، $[a, b]$ را به $(N + 1)$ زیربازه مساوی تقسیم می‌کنیم طوری

که نقاط انتهایی آنها در نقاط شبکه‌ای $x_i = a + ih$ ، بهازای $i = 0, 1, \dots, N + 1$ ، قرار داشته باشند. فرض اینکه جواب دقیق مشتق چهارم کراندار دارد به ما اجازه می‌دهد

که $y''(x_i)$ و $y'(x_i)$ را درهریک از معادلات

$$(29.10) \quad y''(x_i) = f(x_i, y(x_i), y'(x_i))$$

با فرمول تفاضل مرکزی مناسب، داده شده در (۲۴.۱۰) و (۲۵.۱۰)، جایگزین کنیم تا بهمازای هر $i = 1, 2, \dots, N$ و نقاطی چون ξ_i و η_i در بازه (x_{i-1}, x_{i+1}) بدست آوریم

$$(30.10) \quad \frac{y(x_{i+1}) - 2y(x_i) + y(x_{i-1}))}{h^2} = f\left(x_i, y(x_i), \frac{y(x_{i+1}) - y(x_{i-1}))}{2h} - \frac{h^2}{6} y'''(\eta_i)\right) + \frac{h^2}{12} y^{(4)}(\xi_i).$$

همانند مسألت خطی روش تفاضلی، وقتی جملات خطا حذف و شرایط مرزی بکار گرفته شوند، نتیجه می شود

$$w_0 = \alpha, \quad w_{N+1} = \beta,$$

و، بهمازای هر $i = 1, 2, \dots, N$

$$-\frac{w_{i+1} - 2w_i + w_{i-1}}{h^2} + f\left(x_i, w_i, \frac{w_{i+1} - w_{i-1}}{2h}\right) = 0$$

دستگاه غیرخطی $N \times N$ حاصل از این روش، یعنی

$$2w_1 - w_2 + h^2 f\left(x_1, w_1, \frac{w_2 - w_0}{2h}\right) - \alpha = 0,$$

$$-w_1 + 2w_2 - w_3 + h^2 f\left(x_2, w_2, \frac{w_3 - w_1}{2h}\right) = 0,$$

⋮

$$-w_{N-2} + 2w_{N-1} - w_N + h^2 f\left(x_{N-1}, w_{N-1}, \frac{w_N - w_{N-2}}{2h}\right) = 0,$$

$$-w_{N-1} + 2w_N + h^2 f\left(x_N, w_N, \frac{\beta - w_{N-1}}{2h}\right) - \beta = 0,$$

همچنانکه در کتاب ارجاعی قبلی توسط کلر [۴۰]، صفحه ۸۶، نشان داده شده است، بشرط آنکه $h \leq 2/L$ ، جواب منحصر بفردی خواهد داشت.

برای تقریب جواب این دستگاه، روش نیوتن برای دستگاههای غیر خطی را، که در

بخش ۲۰.۹ مطالعه شد، بکار خواهیم برد. دنباله‌ای از تکرارهای $\{(w_1^{(k)}, w_2^{(k)}, \dots, w_N^{(k)})\}$ تولید می شود که به جواب (۳۱.۱۰) همگرا خواهد بود، بشرط آنکه $(w_1^{(0)}, w_2^{(0)}, \dots, w_N^{(0)})$

مسائل مقدار مرزی برای معادلات دیفرانسیل معمولی ۶۴۱

به قدر کافی به جواب (w_1, w_2, \dots, w_N) نزدیک باشد و ماتریس ژاکوبی دستگاه، که توسط معادله (۱۳.۹) تعریف شد، نامنفرد باشد. اما، برای دستگاه (۳۱.۱۵)، ماتریس ژاکوبی سه قطری است و مفروضات ارائه شده در ابتدای این بحث اطمینان می دهد که J یک ماتریس نامنفرد است.

$$J(w_1, \dots, w_N) =$$

$$\begin{bmatrix} 2 + h^2 f_x(x_1, w_1, \frac{w_2 - \alpha}{2h}) & -1 + \frac{h}{2} f_x(x_1, w_1, \frac{w_2 - \alpha}{2h}) & 0 & \dots & 0 \\ -1 - \frac{h}{2} f_x(x_2, w_2, \frac{w_3 - w_1}{2h}) & 2 + h^2 f_x(x_2, w_2, \frac{w_3 - w_1}{2h}) & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & -1 + \frac{h}{2} f_x(x_{N-1}, w_{N-1}, \frac{w_N - w_{N-2}}{2h}) \\ 0 & \dots & -1 - \frac{h}{2} f_x(x_N, w_N, \frac{\beta - w_{N-1}}{2h}) & 2 + h^2 f_x(x_N, w_N, \frac{\beta - w_{N-1}}{2h}) & 0 \end{bmatrix}$$

روش نیوتن برای دستگاههای غیرخطی لازم دارد که در هر تکرار دستگاه خطی

$N \times N$ زیر

$$\begin{aligned} & J(w_1, w_2, \dots, w_N)(v_1, v_2, \dots, v_N)^t \\ &= - \left(2w_1 - w_2 - \alpha + h^2 f \left(x_1, w_1, \frac{w_2 - \alpha}{2h} \right), \right. \\ &\quad \left. -w_1 + 2w_2 - w_3 + h^2 f \left(x_2, w_2, \frac{w_3 - w_1}{2h} \right), \dots, \right. \\ &\quad \left. -w_{N-2} + 2w_{N-1} - w_N + h^2 f \left(x_{N-1}, w_{N-1}, \frac{w_N - w_{N-2}}{2h} \right), \right. \\ &\quad \left. -w_{N-1} + 2w_N + h^2 f \left(x_N, w_N, \frac{\beta - w_{N-1}}{2h} \right) - \beta \right)^t \end{aligned}$$

نسبت به v_1, v_2, \dots, v_N حل شود زیرا،

$$\cdot w_i^{(k)} = w_i^{(k-1)} + v_i, \quad i = 1, 2, \dots, N$$

چون J سه قطری است، آنطور که ممکن است در آغاز بنظر رسد، این امر مسئله دشواری نیست؛ الگوریتم تبدیل کروت برای دستگاههای سه قطری (الگوریتم ۷.۷) را می توان

به آسانی بکار برد. روند کامل در الگوریتم زیر شرح داده شده است.

الگوریتم تفاضل متناهی غیر خطی ۴.۱۰

برای تقریب جواب مسئله مقدار مرزی غیر خطی

$$y'' = f(x, y, y'), \quad a \leq x \leq b, \quad y(a) = \alpha, \quad y(b) = \beta,$$

عدد صحیح $N > 0$ را انتخاب می‌کنیم.

مرحله ۱. قرار می‌دهیم $h = (b - a)/(N + 1)$ و، به‌ازای هر $i = 0, 1, \dots, N + 1$

$$x_i = a + ih$$

مرحله ۲. قرار می‌دهیم $w_0 = \alpha, w_{N+1} = \beta$

مرحله ۳. به‌ازای هر $i = 1, 2, \dots, N$ قرار می‌دهیم $w_i^{(0)} = \alpha + i \left(\frac{\beta - \alpha}{b - a} \right) h$

مرحله ۴. قرار می‌دهیم $k = 0$

مرحله ۵. قرار می‌دهیم

$$a_i = 2 + h^2 f_y \left(x_i, w_i^{(k)}, \frac{w_{i+1}^{(k)} - w_{i-1}^{(k)}}{2h} \right), \quad i = 1, 2, \dots, N$$

$$b_i = -1 + \frac{h}{2} f_{y'} \left(x_i, w_i^{(k)}, \frac{w_{i+1}^{(k)} - w_{i-1}^{(k)}}{2h} \right), \quad i = 1, 2, \dots, N - 1$$

$$c_i = -1 - \frac{h}{2} f_{y''} \left(x_i, w_i^{(k)}, \frac{w_{i+1}^{(k)} - w_{i-1}^{(k)}}{2h} \right), \quad i = 2, 3, \dots, N$$

به‌ازای هر $i = 1, 2, \dots, N$

$$d_i = - \left[2w_i^{(k)} - w_{i+1}^{(k)} - w_{i-1}^{(k)} + h^2 f \left(x_i, w_i^{(k)}, \frac{w_{i+1}^{(k)} - w_{i-1}^{(k)}}{2h} \right) \right]$$

مرحله ۶. قرار می‌دهیم

$$l_1 = a_1,$$

$$u_1 = b_1/a_1.$$

مرحله ۷. قرار می‌دهیم $j = 2$

مرحله ۸. قرار می‌دهیم

$$l_j = a_j - c_j u_{j-1},$$

$$u_j = b_j / l_j.$$

مرحله ۹. ۱ را به z می‌افزاییم.

مرحله ۱۰. اگر $z < N$ ، به مرحله ۸ می‌رویم.

مرحله ۱۱. قرار می‌دهیم $l_N = a_N - c_N u_{N-1}$.

مرحله ۱۲. قرار می‌دهیم $z_1 = d_1 / l_1$ ،

$$z_j = \frac{1}{l_j} [d_j - c_j z_{j-1}], \quad j = 2, 3, \dots, N$$

مرحله ۱۳. قرار می‌دهیم $v_N = z_N$ ،

$$v_j = z_j - u_j v_{j+1}, \quad j = N-1, N-2, \dots, 1$$

مرحله ۱۴. قرار می‌دهیم $w_0^{(k+1)} = \alpha, w_{N+1}^{(k+1)} = \beta$

$$w_j^{(k+1)} = w_j^{(k)} + v_j, \quad j = 1, 2, \dots, N$$

مرحله ۱۵. اگر $|w_j^{(k+1)} - w_j^{(k)}| = |v_j|$ برای $j = 1, 2, \dots, N$ به قدر کافی کوچک

باشد، به مرحله ۱۷ می‌رویم.

مرحله ۱۶. ۱ را به k افزوده به مرحله ۵ می‌رویم.

مرحله ۱۷. روند تمام است؛ اعداد $w_0^{(k+1)}, w_1^{(k+1)}, \dots, w_N^{(k+1)}, w_{N+1}^{(k+1)}$ اعداد

$w_0, w_1, \dots, w_N, w_{N+1}$ را تقریب می‌کنند؛ و در نتیجه، $y(x_0), y(x_1), \dots, y(x_N), y(x_{N+1})$

را تقریب می‌کنند.

در مرحله ۳ تقریب‌های اولیه $w_i^{(0)}$ به w_i ، به‌ازای هر $i = 1, 2, \dots, N$ ، با گذراندن

یک خط مستقیم از (α, x) و (b, β) و محاسبه آن در x_i بدست می‌آید. می‌توان نشان داد

(ر.ک. صفحه ۴۳۳ از ایزاکسون و کلر [۳۹]) که این روش تفاضل متناهی غیرخطی از

مرتبه $O(h^2)$ است. لذا، معیار توقف در مرحله ۱۵ برای این شرط که به‌ازای هر

$j = 1, 2, \dots, N$ ، $|v_j| = O(h^2)$ استوار است. چون وقتی نمی‌توان برقراری شرایط (ت)،

(ب) و (پ)، که در آغاز این مبحث داده شد، را تحقیق کرد، یک تقریب اولیه خوب

لازم است، باید یک کران بالایی برای k مشخص شود و در صورت رد شدن از آن یک

تقریب اولیه جدید یا تقلیلی در اندازه گام در نظر گرفته شود.

مثال ۳. اعمال الگوریتم ۴.۱۰ با $h = 0.1$ روی مسئله مقدار مرزی غیر خطی

$$y'' = \frac{1}{8}(32 + 2x^3 - yy'), \quad 1 \leq x \leq 3, \quad y(1) = 17, \quad y(3) = \frac{43}{3}$$

نتایج جدول ۵.۱۰ را می‌دهد. فرایند توقف بکار رفته در این مثال انجام تکرار تا زمانی که تمام مقادیر تکرارهای متوالی اختلافی کمتر از 10^{-8} داشته باشند می‌باشد. این عمل برای چهار تکرار انجام شده است. توجه کنید که مسئله این مثال همان است که در مثال ۱، بخش ۲.۱۰، برای روش تیراندازی غیرخطی بررسی شد.

x_i	w_i	$y(x_i)$	$ w_i - y(x_i) $
1.0	17.000000	17.000000	—
1.1	15.754503	15.755455	9.520×10^{-4}
1.2	14.771740	14.773333	1.594×10^{-3}
1.3	13.995677	13.997692	2.015×10^{-3}
1.4	13.386297	13.388571	2.275×10^{-3}
1.5	12.914252	12.916667	2.414×10^{-3}
1.6	12.557538	12.560000	2.462×10^{-3}
1.7	12.299326	12.301765	2.438×10^{-3}
1.8	12.126529	12.128889	2.360×10^{-3}
1.9	12.028814	12.031053	2.239×10^{-3}
2.0	11.997915	12.000000	2.085×10^{-3}
2.1	12.027142	12.029048	1.905×10^{-3}
2.2	12.111020	12.112727	1.707×10^{-3}
2.3	12.245025	12.246522	1.497×10^{-3}
2.4	12.425388	12.426667	1.278×10^{-3}
2.5	12.648944	12.650000	1.056×10^{-3}
2.6	12.913013	12.913846	8.335×10^{-4}
2.7	13.215312	13.215926	6.142×10^{-4}
2.8	13.553885	13.554286	4.006×10^{-4}
2.9	13.927046	13.927241	1.953×10^{-4}
3.0	14.333333	14.333333	—

جدول ۵.۱۰

فرایند بیرونیابی ریچاردسون را نیز می‌توان برای روش تفاضلی غیرخطی بکار برد. اگر این روش با استفاده از $h = .1$ ، $h = .05$ ، و $h = .025$ ، و چهار تکرار در هر حالت، روی این مسئله اعمال شود، نتایج جدول ۶.۱۰ بدست می‌آید. نمادگذاری مانند آن است که در مثال ۲ داشتیم و مقادیر Ext_3 با ماکزیمم خطای واقعی 3.68×10^{-10} در نقاط شبکه‌ای تا تمام ارقام فهرست شده دقیق هستند. مقادیر w_i ($h = .1$) از جدول حذف شده‌اند، زیرا قبلاً "فهرست شده بودند".

x_i	$w_i(h = .05)$	$w_i(h = .025)$	Ext_{1i}	Ext_{2i}	Ext_{3i}
1.0	17.00000000	17.00000000	17.00000000	17.00000000	17.00000000
1.1	15.75521721	15.75539525	15.75545543	15.75545460	15.75545455
1.2	14.77293601	14.77323407	14.77333479	14.77333342	14.77333333
1.3	13.99718996	13.99756680	13.99769413	13.99769242	13.99769231
1.4	13.38800424	13.38842973	13.38857346	13.38857156	13.38857143
1.5	12.91606471	12.91651628	12.91666881	12.91666680	12.91666667
1.6	12.55938618	12.55984665	12.56000217	12.56000014	12.56000000
1.7	12.30115670	12.30161280	12.30176684	12.30176484	12.30176471
1.8	12.12830042	12.12874287	12.12889094	12.12888902	12.12888889
1.9	12.03049438	12.03091316	12.03105457	12.03105275	12.03105263
2.0	11.99948020	11.99987013	12.00000179	12.00000011	12.00000000
2.1	12.02857252	12.02892892	12.02904924	12.02904772	12.02904762
2.2	12.11230149	12.11262089	12.11272872	12.11272736	12.11272727
2.3	12.24614846	12.24642848	12.24652299	12.24652182	12.24652174
2.4	12.42634789	12.42658702	12.42666773	12.42666673	12.42666667
2.5	12.64973666	12.64993420	12.65000086	12.65000005	12.65000000
2.6	12.91363828	12.91379422	12.91384683	12.91384620	12.91384615
2.7	13.21577275	13.21588765	13.21592641	13.21592596	13.21592593
2.8	13.55418579	13.55426075	13.55428603	13.55428573	13.55428571
2.9	13.92719268	13.92722921	13.92724153	13.92724139	13.92724138
3.0	14.33333333	14.33333333	14.33333333	14.33333333	14.33333333

جدول ۶.۱۰

مجموعه تمرینات ۳.۱۰

۱. نشان دهید که مسائل مقدار مرزی خطی زیر در مفروضات قضیه ۴.۱۰ صدق می‌کنند و با استفاده از الگوریتم ۳.۱۰ جواب آنها را تقریب کنید.

$$y'' = -\frac{4}{x}y' + \frac{2}{x^2}y - \frac{2 \ln x}{x^2}, \quad 1 < x < 2, \quad (\Gamma)$$

$$y(1) = -\frac{1}{2}, \quad y(2) = \ln 2;$$

۰.۵ h را بکار برده و نتایج را با جواب واقعی مقایسه کنید.

$$y'' = -4y' + 4y, \quad 0 < x < 5, \quad (-)$$

$$y(0) = 1, \quad y(5) = 0;$$

۰.۲ h را بکار برده و نتایج را با جواب واقعی مقایسه کنید.

$$y'' = -3y' + 2y + 2x + 3, \quad 0 < x < 1, \quad (\Psi)$$

$$y(0) = 2, \quad y(1) = 1;$$

۱. h را بکار برده و نتایج را با جواب واقعی مقایسه کنید.

$$y'' = 100y, \quad 0 < x < 1, \quad (ت)$$

$$y(0) = 1, \quad y(1) = e^{-10};$$

۱. $h = .05$ و $h = .1$ را بکار برده و نتایج را با جواب واقعی $y = e^{-10x}$ و تمرین ۷، بخش ۱۰.۱، مقایسه کنید.

۲. تمرین ۱ را با استفاده از برونمایی تکرار کنید.

۳. نشان دهید که مسائل مقدار مرزی خطی زیر در مفروضات قضیه ۴.۱۰ صدق نمی کنند، اما باز هم کاربرد الگوریتم ۳.۱۰ را آزمایش کنید.

$$y'' = -\frac{4}{x}y' - \frac{2}{x^2}y + \frac{2}{x^2} \ln x, \quad 1 < x < 2, \quad (T)$$

$$y(1) = \frac{1}{2}, \quad y(2) = \ln 2;$$

۰.۵ h را بکار برید.

$$y'' = -y, \quad 0 < x < \frac{\pi}{4}, \quad (r)$$

$$y(0) = 1, \quad y\left(\frac{\pi}{4}\right) = 1;$$

$h = \frac{\pi}{40}$ را بکار برید.

۴. نشان دهید هرگاه $y \in C^3[x_{i-1}, x_{i+1}]$ ، آنگاه یک نقطه مانند η_i که $x_{i-1} < \eta_i < x_{i+1}$ با این خاصیت که

$$y'(x_i) = \frac{1}{2h} [y(x_{i+1}) - y(x_{i-1})] - \frac{h^2}{6} y'''(\eta_i).$$

وجود دارد.

۵. قضیه ۴.۱۰ را ثابت کنید.

۶. نشان دهید هرگاه $y \in C^6[a, b]$ و w_0, w_1, \dots, w_{N+1} در (۲۷.۱۰) صدق کند، آنگاه

$$w_i - y(x_i) = Ah^2 + O(h^4),$$

که در آن بشرط آنکه w از w ای بر $[a, b]$ ، $g(x) \geq w > 0$ ، A مستقل از h است.

۷. نشان دهید که مسائل مقدار مرزی غیر خطی ذیل در مفروضات زیر صدق می کنند:

(یک) $f, f_y, f_{y'}$ همه بر

$$D = \{(x, y, y') | a \leq x \leq b, -\infty < y, y' < \infty\}$$

پيوستہ اند؛

(دو) به ازای $\delta > 0$ ای، بر D ، $(\partial f / \partial y)(x, y, y') \geq \delta > 0$ ؛

(سه) ثابتهای k و L وجود دارند که

$$k = \max_{(x, y, y') \in D} \left| \frac{\partial f}{\partial y}(x, y, y') \right|, \quad L = \max_{(x, y, y') \in D} \left| \frac{\partial f}{\partial y'}(x, y, y') \right|;$$

و با استفاده از الگوریتم ۴.۱۰ جواب آنها را تقریب کنید.

$$y'' = \frac{1}{2}y^3, \quad 1 < x < 2, \quad (\bar{1})$$

$$y(1) = -\frac{2}{3}, \quad y(2) = -1;$$

۰.۵ h را بکار برده و نتایج را با تمرین ۱ ($\bar{1}$)، بخش ۲.۱۰، مقایسه کنید.

$$y'' = 2y^3, \quad 1 < x < 5, \quad (\bar{2})$$

$$y(1) = \frac{1}{4}, \quad y(2) = \frac{1}{8};$$

۰.۲ h را بکار برده و نتایج را با تمرین ۱ ($\bar{2}$)، بخش ۲.۱۰، مقایسه کنید.

$$y'' = y^3 - yy', \quad 1 < x < 2, \quad (\bar{3})$$

$$y(1) = \frac{1}{2}, \quad y(2) = \frac{1}{3};$$

۱. h را بکار برده و نتایج را با تمرین ۱ ($\bar{3}$)، بخش ۲.۱۰، مقایسه کنید.

$$y'' = 2y^3 - 6y - 2x^3, \quad 1 < x < 2, \quad (\bar{4})$$

$$y(1) = 2, \quad y(2) = \frac{2}{3};$$

۰.۵ h را بکار برده و نتایج را با تمرین ۱ ($\bar{4}$)، بخش ۲.۱۰، مقایسه کنید.

۸. تمرین ۷ را با استفاده از برونیایی تکرار کنید.

۹. نشان دهید که مفروضات فهرست شده در تمرین ۷ به ازای $h < 2/L$ نامنفرد بودن

ماتریس ژاکوبین J را اطمینان می‌دهد

۱۰. خمیدگی یک صفحه^۱ مستطیلی طویل به طور یکنواخت وزندار تحت یک نیروی کشش

محوری، به ازای خمیدگیهای کوچک، از یک معادله^۲ دیفرانسیل مرتبه دوم تبعیت می‌کند. فرض کنید S نمایش نیروی محوری و q فشار وزن یکنواخت باشد. خمیدگی

W در امتداد طول اصلی یا

$$W''(x) - \frac{S}{D} W(x) = \frac{-ql}{2D} x + \frac{q}{2D} x^2, \quad 0 < x < l,$$

$$W(0) = W(l) = 0,$$

داده می‌شود، که در آن l طول صفحه و D سختی صفحه است. فرض کنید 200 (ثانیه) / پوند = q ، 100 اینچ / پوند = S ، 8.8×10^7 اینچ / پوند = D ، و 50 اینچ = l . خمیدگی را در بازه‌های یک اینچی به وسیله الگوریتم تفاضل متناهی خطی 3.10 تقریب کنید.

۴۰۱۰ روش ریلی ۱ - ریتس ۲

یک مسئله مقدار مرزی دو نقطه‌ای مهم در تحلیل فشار - میله آهنی توسط معادله دیفرانسیل

$$(32.10) \quad -\frac{d}{dx} \left(p(x) \frac{dy}{dx} \right) + q(x)y = f(x), \quad 0 \leq x \leq 1$$

با شرایط مرزی

$$(33.10) \quad y(0) = y(1) = 0$$

ارائه می‌شود. این معادله دیفرانسیل خمیدگی $y(x)$ روی یک میله آهنی به طول یک با مقطع عرضی متغیرنموده شده با $q(x)$ ، که بخاطر فشارهای اضافه شده $p(x)$ و $f(x)$ می‌باشد ، را بیان می‌کند.

در بحثی که خواهیم داشت فرض می‌کنیم $p \in C^1[0, 1]$ و $q, f \in C[0, 1]$. بعلاوه ، لازم است یک ثابت $\delta > 0$ وجود داشته باشد بطوری که به ازای $0 \leq x \leq 1$ ،

$$p(x) \geq \delta > 0$$

و اینکه به ازای $0 \leq x \leq 1$ ،

$$q(x) \geq 0.$$

این مفروضات برای تضمین آنکه مسئله مقدار مرزی (۳۲.۱۰) و (۳۳.۱۰) دارای جواب یکتا باشد کافی است (ر.ک. بیلی ، شامپاین ۳ ، و والتس ۴ [۷]).

همچنانکه حالت بسیاری از مسائل مقدار مرزی بیان کننده پدیده‌های فیزیکی است ، جواب معادله میله آهنی در یک خاصیت تغییراتی صدق می‌کند. اصل تغییراتی برای معادله میله آهنی در بنای روش ریلی - ریتس اساسی است و جواب معادله میله آهنی

1. Rayleigh

2. Ritz

3. Shampine

4. Waltman

را به عنوان تابعی که انتگرال معینی را روی تمام توابع متعلق به $C_0^2[0, 1]$ ، یعنی مجموعه^۲ توابعی از $C^2[0, 1]$ با این خاصیت که $u(0) = u(1) = 0$ ، مینیمم می کند را مشخص می نماید. قضیه^۳ زیر این خصوصیت را می دهد. برهان این قضیه با آنکه مشکل نیست طولانی است؛ آن را می توان در شولتز [۵۶] یافت.

۵.۱۰ قضیه. فرض کنیم $p \in C^1[0, 1]$ ، $q, f \in C[0, 1]$ ، و

$$(۳۴.۱۰) \quad p(x) \geq \delta > 0, \quad q(x) \geq 0, \quad 0 \leq x \leq 1$$

تابع $y \in C_0^2[0, 1]$ جواب منحصر بفرد مسئله مقدار مرزی

$$(۳۵.۱۰) \quad -\frac{d}{dx} \left(p(x) \frac{dy}{dx} \right) + q(x)y = f(x), \quad 0 \leq x \leq 1,$$

$$y(0) = y(1) = 0,$$

است اگر و فقط اگر y تابع منحصر بفردی از $C_0^2[0, 1]$ باشد که انتگرال زیر را مینیمم می سازد:

$$(۳۶.۱۰) \quad I[u] = \int_0^1 \{p(x)[u'(x)]^2 + q(x)[u(x)]^2 - 2f(x)u(x)\} dx.$$

در فرایند ریلی - ریثس، انتگرال I روی تمام توابع متعلق به $C_0^2[0, 1]$ مینیمم نمی شود بلکه روی یک مجموعه کوچکتر از توابع متشکل از ترکیبهای خطی از توابع $\phi_1, \phi_2, \dots, \phi_n$ مینیمم می شود. توابع پایه^۴ $\{\phi_i\}_{i=1}^n$ باید مستقل خطی باشند و به ازای هر $i = 1, 2, \dots, n$

$$\phi_i(0) = \phi_i(1) = 0$$

صدق کنند. یک تقریب $\phi(x) = \sum_{i=1}^n c_i \phi_i(x)$ به جواب $y(x)$ از (۳۵.۱۰) با پیدا کردن c_1, c_2, \dots, c_n که $I[\sum_{i=1}^n c_i \phi_i]$ را مینیمم می کند، بدست می آید. از (۳۶.۱۰) داریم

$$(۳۷.۱۰) \quad I[\phi] = I \left[\sum_{i=1}^n c_i \phi_i \right] \\ = \int_0^1 \left\{ p(x) \left[\sum_{i=1}^n c_i \phi_i'(x) \right]^2 + q(x) \left[\sum_{i=1}^n c_i \phi_i(x) \right]^2 - 2f(x) \sum_{i=1}^n c_i \phi_i(x) \right\} dx.$$

و برای اینکه یک مینیمم اتفاق افتد لازم است که، وقتی I به عنوان تابعی از c_1, c_2, \dots, c_n در نظر گرفته شود، داشته باشیم

$$(38.10) \quad \frac{\partial I}{\partial c_j} = 0, \quad j = 1, 2, \dots, n$$

بمازای هر $j = 1, 2, \dots, n$

مشقگیری از (37.10) نتیجه می دهد

$$\frac{\partial I}{\partial c_j} = \int_0^1 \left\{ 2p(x) \sum_{i=1}^n c_i \phi'_i(x) \phi'_j(x) + 2q(x) \sum_{i=1}^n c_i \phi_i(x) \phi_j(x) - 2f(x) \phi_j(x) \right\} dx,$$

و جایگذاری در (38.10) نتیجه می دهد که

$$, \quad j = 1, 2, \dots, n$$

بمازای هر $j = 1, 2, \dots, n$

$$(39.10) \quad 0 = \sum_{i=1}^n \left[\int_0^1 \{ p(x) \phi'_i(x) \phi'_j(x) + q(x) \phi_i(x) \phi_j(x) \} dx \right] c_i - \int_0^1 f(x) \phi_j(x) dx,$$

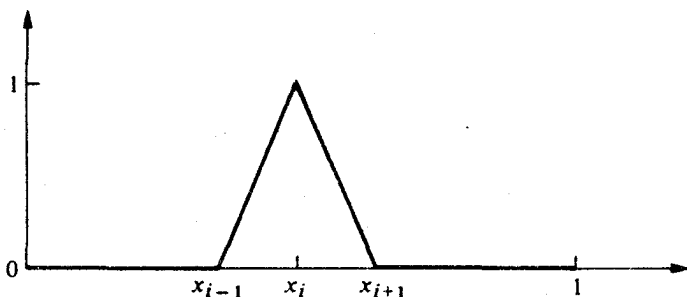
اولین انتخاب توابع پایه را که مورد بحث قرار می دهیم شامل چند جمله ایهای قطعه قطعه خطی است. اولین گام تشکیل دادن یک افراز $[0, 1]$ با انتخاب نقاط x_0, x_1, \dots, x_{n+1} است که

$$0 = x_0 < x_1 < \dots < x_n < x_{n+1} = 1$$

با فرض $h_i = x_{i+1} - x_i$ ، بمازای هر $i = 0, 1, \dots, n$ ، توابع پایه $\phi_1(x), \phi_2(x), \dots, \phi_n(x)$ را، بمازای هر $i = 1, 2, \dots, n$ ،

$$(40.10) \quad \phi_i(x) = \begin{cases} 0, & 0 \leq x \leq x_{i-1}, \\ \frac{(x - x_{i-1})}{h_{i-1}}, & x_{i-1} \leq x \leq x_i, \\ \frac{(x_{i+1} - x)}{h_i}, & x_i \leq x \leq x_{i+1}, \\ 0, & x_{i+1} \leq x \leq 1, \end{cases}$$

تعریف می کنیم. (ر.ک. شکل 40.10).



شکل ۴.۱۰

چون توابع ϕ_i قطعه قطعه خطی اند، مشتقات ϕ'_i ، درحالی که پیوسته نیستند، بر زیربازه باز (x_j, x_{j+1}) ، به ازای هر $j = 0, 1, \dots, n$ ، ثابت می‌باشند. پس، به ازای هر $i = 1, 2, \dots, n$ داریم

$$(۴۱.۱۰) \quad \phi'_i(x) = \begin{cases} 0, & 0 < x < x_{i-1}, \\ \frac{1}{h_{i-1}}, & x_{i-1} < x < x_i, \\ -\frac{1}{h_i}, & x_i < x < x_{i+1}, \\ 0, & x_{i+1} < x < 1. \end{cases}$$

چون ϕ_i و ϕ'_i فقط بر (x_{i-1}, x_{i+1}) غیر صفر هستند، دستگاه خطی (۳۹.۱۰) دستگاه خطی سه قطری $n \times n$ زیر تحویل می‌شود:

$$(۴۲.۱۰)$$

$$Ac = b.$$

درایه‌های غیر صفر A چنین داده می‌شوند:

$$i = 1, 2, \dots, n$$

$$\begin{aligned} a_{ii} &= \int_0^1 \{p(x)[\phi'_i(x)]^2 + q(x)[\phi_i(x)]^2\} dx \\ &= \int_{x_{i-1}}^{x_i} \left(\frac{1}{h_{i-1}}\right)^2 p(x) dx + \int_{x_i}^{x_{i+1}} \left(\frac{-1}{h_i}\right)^2 p(x) dx \\ &\quad + \int_{x_{i-1}}^{x_i} \left(\frac{1}{h_{i-1}}\right)^2 (x - x_{i-1})^2 q(x) dx \end{aligned}$$

$$+ \int_{x_i}^{x_{i+1}} \left(\frac{1}{h_i}\right)^2 (x_{i+1} - x)^2 q(x) dx ,$$

، $i = 1, 2, \dots, n - 1$ به‌ازای هر

$$\begin{aligned} a_{i,i+1} &= \int_0^1 \{p(x)\phi'_i(x)\phi'_{i+1}(x) dx + q(x)\phi_i(x)\phi_{i+1}(x)\} dx \\ &= \int_{x_i}^{x_{i+1}} - \left(\frac{1}{h_i}\right)^2 p(x) dx \\ &\quad + \int_{x_i}^{x_{i+1}} \left(\frac{1}{h_i}\right)^2 (x_{i+1} - x)(x - x_i)q(x) dx , \end{aligned}$$

و به‌ازای هر $i = 2, \dots, n$

$$\begin{aligned} a_{i,i-1} &= \int_0^1 \{p(x)\phi'_i(x)\phi'_{i-1}(x) dx + q(x)\phi_i(x)\phi_{i-1}(x)\} dx \\ &= \int_{x_{i-1}}^{x_i} - \left(\frac{1}{h_{i-1}}\right)^2 p(x) dx \\ &\quad + \int_{x_{i-1}}^{x_i} \left(\frac{1}{h_{i-1}}\right)^2 (x_i - x)(x - x_{i-1})q(x) dx . \end{aligned}$$

درایه‌های **b** عبارتند از

، $i = 1, 2, \dots, n$ به‌ازای هر

$$\begin{aligned} b_i &= \int_0^1 f(x)\phi_i(x) dx \\ &= \int_{x_{i-1}}^{x_i} \frac{1}{h_{i-1}} (x - x_{i-1})f(x) dx + \int_{x_i}^{x_{i+1}} \frac{1}{h_i} (x_{i+1} - x)f(x) dx . \end{aligned}$$

درایه‌های **c** ضرایب مجهول c_1, c_2, \dots, c_n هستند که از آن تقریب ریلی - ریتس ϕ که
 با

$$\phi(x) = \sum_{i=1}^n c_i \phi_i(x)$$

داده شده است ساخته می‌شود .

الگوریتم زیر دستگاه خطی سه قطری را بنا کرده و الگوریتم سه قطری ۷.۷ را جهت حل آن وارد می‌کند، در مرحله ۳ باید یک انتگرالگیری انجام شود، لذا باید یک فرمول

انتگرالگیری موجود باشد. توصیه می‌شود که انتگرالگیری به وسیله درونیابی p ، q ، و f با چند جمله‌ایهای قطعی ϕ_1, \dots, ϕ_n انجام شود مگر در مورد انتگرالهایی که به سادگی حساب شوند (ر.ک. بخش ۶.۳).

الگوریتم قطعه قطعه خطی ریلی - ریتس ۵.۱۰
برای تقریب جواب مسئله مقدار مرزی

$$-\frac{d}{dx} \left(p(x) \frac{dy}{dx} \right) + q(x)y = f(x), \quad 0 \leq x \leq 1.$$

$$y(0) = y(1) = 0,$$

با تابع قطعه قطعه خطی

$$\phi(x) = \sum_{i=1}^n c_i \phi_i(x)$$

یک عدد صحیح n و نقاط $\{x_i\}_{i=1}^n$ را، که در آن $0 = x_0 < x_1 < \dots < x_n < x_{n+1} = 1$ انتخاب می‌کنیم.

مرحله ۱. به ازای هر $i = 0, 1, \dots, n$ ، قرار می‌دهیم $h_i = x_{i+1} - x_i$.

مرحله ۲. پایه قطعه قطعه خطی $\{\phi_{ij}\}_{i=1}^n$ را به ازای هر $i = 1, 2, \dots, n$ تعریف می‌کنیم:

$$\phi_{ij}(x) = \begin{cases} 0, & 0 \leq x \leq x_{i-1}, \\ \frac{(x - x_{i-1})}{h_{i-1}}, & x_{i-1} \leq x \leq x_i, \\ \frac{(x_{i+1} - x)}{h_i}, & x_i \leq x \leq x_{i+1}, \\ 0, & x_{i+1} \leq x \leq 1, \end{cases}$$

مرحله ۳. انتگرالهای زیر را حساب می‌کنیم:

$$I_1^{(i)} = \left(\frac{1}{h_i}\right)^2 \int_{x_i}^{x_{i+1}} p(x) dx, \quad i = 0, 1, \dots, n$$

$$I_2^{(i)} = \left(\frac{1}{h_i}\right)^2 \int_{x_i}^{x_{i+1}} (x - x_i)^2 q(x) dx, \quad i = 0, 1, \dots, n-1$$

$$I_3^{(i)} = \left(\frac{1}{h_i}\right)^2 \int_{x_i}^{x_{i+1}} (x_{i+1} - x)^2 q(x) dx, \quad i = 1, 2, \dots, n$$

$$I_4^{(i)} = \left(\frac{1}{h_i}\right)^2 \int_{x_i}^{x_{i+1}} (x_{i+1} - x)(x - x_i)q(x) dx, \quad i = 1, 2, \dots, n-1 \text{ به ازای هر}$$

$$I_5^{(i)} = \frac{1}{h_{i-1}} \int_{x_{i-1}}^{x_i} (x - x_{i-1})f(x) dx, \quad i = 1, 2, \dots, n \text{ به ازای هر}$$

$$I_6^{(i)} = \frac{1}{h_i} \int_{x_i}^{x_{i+1}} (x_{i+1} - x)f(x) dx, \quad i = 1, 2, \dots, n \text{ به ازای هر}$$

مرحله ۴. قرار می دهیم

$$\alpha_i = I_1^{(i-1)} + I_1^{(i)} + I_2^{(i-1)} + I_3^{(i)}, \quad i = 1, 2, \dots, n \text{ به ازای هر}$$

$$\beta_i = -I_1^{(i)} + I_4^{(i)}, \quad i = 1, 2, \dots, n-1 \text{ به ازای هر}$$

$$b_i = I_5^{(i)} + I_6^{(i)}, \quad i = 1, 2, \dots, n \text{ به ازای هر}$$

$$\zeta_1 = \beta_1/\alpha_1 \text{ و } a_1 = x_1 \text{ قرار می دهیم}$$

$$\text{مرحله ۵. قرار می دهیم } i = 2$$

$$\text{مرحله ۶. قرار می دهیم } a_i = x_i - \beta_{i-1}\zeta_{i-1}$$

$$\zeta_i = \beta_i a_i.$$

$$\text{مرحله ۷. } 1 \text{ را به } i \text{ می افزاییم. اگر } i \leq n-1, \text{ به مرحله ۷ می رویم.}$$

$$\text{مرحله ۸. قرار می دهیم } a_n = x_n - \beta_n \zeta_n$$

$$\text{مرحله ۹. قرار می دهیم } z_1 = b_1/a_1$$

$$\text{مرحله ۱۰. قرار می دهیم } i = 2$$

$$\text{مرحله ۱۱. قرار می دهیم } z_i = (b_i - \beta_{i-1}z_{i-1})/a_i$$

$$\text{مرحله ۱۲. } 1 \text{ را به } i \text{ می افزاییم. اگر } i \leq n, \text{ به مرحله ۱۲ می رویم.}$$

$$\text{مرحله ۱۳. قرار می دهیم } c_n = z_n$$

$$\text{مرحله ۱۴. قرار می دهیم } i = n-1$$

$$\text{مرحله ۱۵. قرار می دهیم } c_i = z_i - \xi_i c_{i+1}$$

$$\text{مرحله ۱۶. } 1 \text{ را از } i \text{ کم می کنیم. اگر } i \geq 1, \text{ به مرحله ۱۶ می رویم.}$$

$$\text{مرحله ۱۷. روند تمام است؛ تابع قطعه قطعه خطی}$$

$$\phi(x) = \sum_{i=1}^n c_i \phi_i(x)$$

تقریبی به $y(x)$ می باشد.

مثال زیر از الگوریتم ۵.۱۰ استفاده می‌کند، اما، بخاطر طبیعت این مثال، انتگرالهای مرحله ۳ مستقیماً بدست آمده‌اند.

مثال ۱. مسئله مقدار مرزی

$$-y'' + \pi^2 y = 2\pi^2 \sin(\pi x), \quad 0 \leq x \leq 1$$

$$y(0) = y(1) = 0.$$

را در نظر می‌گیریم. فرض کنیم $h = h_i = .1$ بطوری که به ازای هر $i = 0, 1, \dots, 9$ ، $x_i = .1i$. انجام مراحل مندرج در الگوریتم ۵.۱۰ می‌دهد.

مرحله ۳.

$$I_1^{(i)} = 100 \int_{.1i}^{.1i+.1} dx = 10,$$

$$I_2^{(i)} = 100 \int_{.1i}^{.1i+.1} (x - .1i)^2 \pi^2 dx = \frac{\pi^2}{30},$$

$$I_3^{(i)} = 100 \int_{.1i}^{.1i+.1} (.1i + .1 - x)^2 \pi^2 dx = \frac{\pi^2}{30},$$

$$I_4^{(i)} = 100 \int_{.1i}^{.1i+.1} (.1i + .1 - x)(x - .1i) \pi^2 dx = \frac{\pi^2}{60},$$

$$I_5^{(i)} = 10 \int_{.1i-.1}^{.1i} (x - .1i + .1) 2\pi^2 \sin(\pi x) dx$$

$$= -2\pi \cos .1\pi i + 20[\sin(.1\pi i) - \sin((.1i - .1)\pi)],$$

$$I_6^{(i)} = 2\pi \cos .1\pi i - 20[\sin((.1i + .1)\pi) - \sin(.1\pi i)].$$

مرحله ۴.

$$\alpha_i = 20 + \frac{\pi^4}{15}, \quad i = 1, 2, \dots, 9$$

بمازای هر

$$\beta_i = -10 + \frac{\pi^4}{60}, \quad i = 1, 2, \dots, 8$$

بمازای هر

$$b_i = 40 \sin(.1\pi i)(1 - \cos(.1\pi)), \quad i = 1, 2, \dots, 9$$

بمازای هر

مراحل ۵، ۶، ۷، ۸، و ۹.

اینها مقادیر زیر را برای a_i و ζ_i می‌دهند:

$$\begin{aligned} a_1 &= 20.65797363, & \zeta_1 &= -.4761118767 \\ a_2 &= 15.97517213, & \zeta_2 &= -.6156745300 \\ a_3 &= 14.60250273, & \zeta_3 &= -.6735493754 \\ a_4 &= 14.03327431, & \zeta_4 &= -.7008704010 \\ a_5 &= 13.76455818, & \zeta_5 &= -.7145530183 \\ a_6 &= 13.6299871, & \zeta_6 &= -.7216081489, \\ a_7 &= 13.56059192, & \zeta_7 &= -.7253006839, \\ a_8 &= 13.52427397, & \zeta_8 &= -.7272483991, \\ a_9 &= 13.50511721. \end{aligned}$$

مراحل ۱۰، ۱۱، ۱۲ و ۱۳.

اینها مقادیر زیر را برای z_i تولید می کنند.

$$\begin{aligned} z_1 &= .0292852890, & z_6 &= .3687722609, \\ z_2 &= .0900626268, & z_7 &= .3842683589, \\ z_3 &= .1691255264, & z_8 &= .3645448400, \\ z_4 &= .2512140726, & z_9 &= .3102866742. \\ z_5 &= .3217362267, \end{aligned}$$

مراحل ۱۴، ۱۵، ۱۶ و ۱۷.

بینها ضرایب c_i را می دهند:

$$\begin{aligned} c_9 &= .3102866742, & c_4 &= .9549641893, \\ c_8 &= .5902003271, & c_3 &= .8123410598, \\ c_7 &= .8123410598, & c_2 &= .5902003271, \\ c_6 &= .9549641893, & c_1 &= .3102866742. \\ c_5 &= 1.004108771, \end{aligned}$$

تقریب قطعه قطعه خطی با

$$\phi(x) = \sum_{i=1}^9 c_i \phi_i(x)$$

داده می شود. جواب واقعی مسئلهء مقدار مرزی عبارت است از

$$f(x) = \sin(\pi x)$$

و جدول ۷.۱۰ خطای تقریب در نقاط x_i را، به ازای هر $i = 1, \dots, 9$ ، فهرست می کند.

می توان نشان داد که ماتریس هم قطری A که با معادله (۴۲.۱۰) داده شده است

مقتارن و معین مثبت است (ر.ک. تمرین ۴). لذا، دستگاه خطی پایدار است. تحت

مفروضاتی که در آغاز این بخش ارائه شدند، داریم

$$|\phi(x) - y(x)| = O(h^2), \quad 0 \leq x \leq 1.$$

برهانی از این نتیجه را می توان در صفحات ۱۰۳ و ۱۰۴ شولتز [۵۶] یافت .

انتخاب دیگری از توابع پایه شامل توابع اسپلاین درجه سوم است که در بخش ۶.۳ به منظور درونیابی مطالعه شدند . اولین مرحله در ساختن تقریب اسپلاین درجه سوم تشکیل یک افراز $[0, 1]$ ، با انتخاب عدد صحیح مثبت n و قرار دادن $h = 1/(n + 1)$ و تعریف نقاط شبکه ای به صورت $x_i = ih$ ، به ازای هر $i = 0, 1, \dots, n + 1$ ، است . یک تابع اسپلاین درجه سوم بر $[0, 1]$ نسبت به این افراز متعلق به $C^2[0, 1]$ و بر هر زیر بازه $[x_i, x_{i+1}]$ ، به ازای $i = 0, 1, \dots, n$ ، یک چند جمله ای درجه سوم خواهد بود . توابع ϕ_i ، $i = 0, 1, \dots, n + 1$ ، در این پایه اسپلاین یا اسپلاینهای زنگی شکل نامیده می شوند ، که در اصل توسط شونبرگ^۱ [۵۴] ارائه شدند . به ازای هر نقطه شبکه ای x_i ، اسپلاین مربوطه $S_i(x)$ با

$$S_i(x) = \begin{cases} 0, & x \leq x_{i-2} \\ \frac{\left[2 - \frac{(x-x_i)}{h}\right]^3}{24} - \frac{\left[1 - \frac{(x-x_i)}{h}\right]^3}{6} - \frac{\left[\frac{x-x_i}{h}\right]^3}{4} + \frac{\left[1 + \frac{(x-x_i)}{h}\right]^3}{6}, & x_{i-2} \leq x \leq x_{i-1}, \\ \frac{\left[2 - \frac{(x-x_i)}{h}\right]^3}{24} - \frac{\left[1 - \frac{(x-x_i)}{h}\right]^3}{6} - \frac{\left[\frac{x-x_i}{h}\right]^3}{4}, & x_{i-1} \leq x \leq x_i, \\ \frac{\left[2 - \frac{(x-x_i)}{h}\right]^3}{24} - \frac{\left[1 - \frac{(x-x_i)}{h}\right]^3}{6}, & x_i \leq x \leq x_{i+1}, \\ \frac{\left[2 - \frac{(x-x_i)}{h}\right]^3}{24}, & x_{i+1} \leq x \leq x_{i+2}, \\ 0, & x_{i+2} \leq x \end{cases}$$

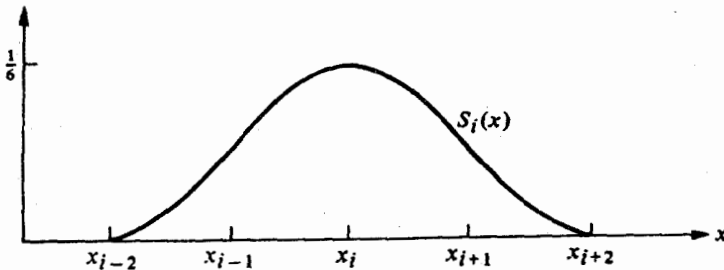
تعریف می شود ، که در آن نقاط شبکه ای اضافی مورد نیاز برای S_0 ، S_1 ، S_n ، و S_{n+1} با $x_{-2} = -2h$ ، $x_{-1} = -h$ ، $x_{n+2} = (n+2)h$ و $x_{n+3} = (n+3)h$ تعریف می شوند . یک اسپلاین نمونه در شکل ۵.۱۰ نشان داده شده است .

i	x_i	$\phi(x_i)$	$f(x_i)$	$ \phi(x_i) - f(x_i) $
1	.1	.3102866742	.3090169943	.00127
2	.2	.5902003271	.5877852522	.00242
3	.3	.8123410598	.8090169943	.00332
4	.4	.9549641896	.9510565162	.00391
5	.5	1.004108771	1.0000000000	.00411
6	.6	.9549641893	.9510565162	.00391
7	.7	.8123410598	.8090169943	.00332
8	.8	.5902003271	.5877852522	.00242
9	.9	.3102866742	.3090169943	.00127

جدول ۷.۱۰

اگر اسپلاین مکعبی، $S(x)$ را با

$$S(x) = \begin{cases} 0, & x \leq -2 \\ \frac{(2-x)^3}{24} - \frac{(1-x)^3}{6} - \frac{x^3}{4} + \frac{(1+x)^3}{6}, & -2 \leq x \leq -1, \\ \frac{(2-x)^3}{24} - \frac{(1-x)^3}{6} - \frac{x^3}{4}, & -1 \leq x \leq 0, \\ \frac{(2-x)^3}{24} - \frac{(1-x)^3}{6}, & 0 \leq x \leq 1, \\ \frac{(2-x)^3}{24}, & 1 \leq x \leq 2, \\ 0, & x \geq 2, \end{cases}$$



شکل ۵.۱۰

تعریف کنیم، می‌توانیم S_i را به‌ازای هر $i = 0, 1, \dots, n+1$ با

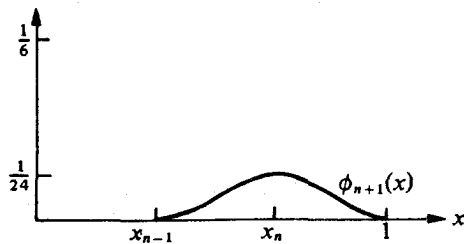
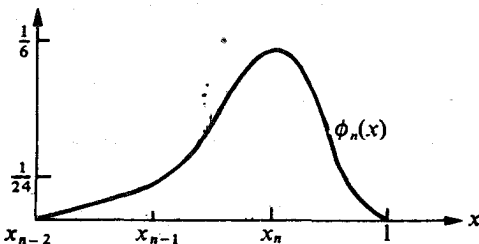
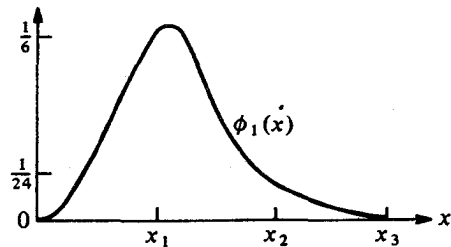
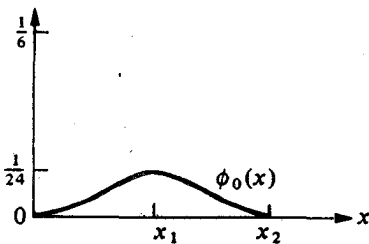
$$S_i(x) = S\left(\frac{x - x_i}{h}\right)$$

نمایش دهیم. به سادگی نشان داده می‌شود که $\{S_i\}_{i=0}^{n+1}$ یک مجموعه مستقل خطی از اسپلانیه‌های مکعبی است (ر.ک. تمرین ۱۳). برای اینکه مجموعه $\{S_i\}_{i=0}^{n+1}$ در شرایط مرزی $\phi_i(0) = \phi_i(1) = 0$ صدق کند لازم است که S_0, S_1, S_n, S_{n+1} را تغییر

دهیم. این تغییر پایه با

$$\phi_i(x) = \begin{cases} S_0(x) - 4S\left(\frac{x+h}{h}\right), & i = 0, \\ S_1(x) - S\left(\frac{x+h}{h}\right), & i = 1, \\ S_i(x), & 2 \leq i \leq n-1, \\ S_n(x) - S\left(\frac{x - (n+2)h}{h}\right), & i = n, \\ S_{n+1}(x) - 4S\left(\frac{x - (n+2)h}{h}\right), & i = n+1. \end{cases}$$

تعریف می‌شود (ر.ک. شکل ۶.۱۰).



شکل ۶.۱۰

چون $\phi_i(x)$ و $\phi'_i(x)$ فقط بازای $x_{i-2} \leq x \leq x_{i+2}$ غیر صفر هستند، ماتریس تقریب ریلی - ریتس یک ماتریس نواری با عرض نوار حداکثر هفت خواهد بود:

(۳۳۰۱۰)

$$A = \begin{bmatrix}
 a_{00} & a_{01} & a_{02} & a_{03} & 0 & \cdots & 0 \\
 a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & & \\
 a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & \\
 a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 0 & & & & & & a_{n-2, n+1} \\
 & & & & & & a_{n-1, n+1} \\
 & & & & & & a_{n, n+1} \\
 0 & & & & a_{n+1, n-2} & a_{n+1, n-1} & a_{n+1, n} & a_{n+1, n+1}
 \end{bmatrix}$$

که در آن بازای هر $i, j = 0, 1, \dots, n+1$

$$a_{ij} = \int_0^1 \{p(x)\phi'_i(x)\phi'_j(x) + q(x)\phi_i(x)\phi_j(x)\} dx$$

ماتریس A معین مثبت نیز هست (ر. ک. تمرین ۱۰). بنابراین، دستگاه (۳۳۰۱۰) را می‌توان به سادگی به وسیله الگوریتم چولسکی ۶.۷ یا به وسیله حذف گاوسی حل کرد. الگوریتم زیر ساختن تقریب اسپلاین مکعبی $\phi(x)$ را به وسیله روش ریلی - ریتس برای مسئله مقدار مرزی (۳۲۰۱۰) و (۳۳۰۱۰) ارائه شده در صفحه ۶۴۸، شرح می‌دهد.

الگوریتم اسپلاین مکعبی ریلی - ریتس ۶.۱۰
برای تقریب جواب مسئله مقدار مرزی

$$-\frac{d}{dx} \left(p(x) \frac{dy}{dx} \right) + q(x)y = f(x), \quad 0 \leq x \leq 1,$$

$$y(0) = y(1) = 0$$

با اسپلاین مکعبی

$$\phi(x) = \sum_{i=0}^{n+1} c_i \bar{\phi}_i(x)$$

- یک عدد صحیح مثبت n انتخاب می‌کنیم و قرار می‌دهیم $h = 1/(n+1)$.
- مرحله ۱. به‌ازای هر $i = 0, 1, \dots, n+1$ قرار می‌دهیم $x_i = ih$.
- مرحله ۲. پایه‌های اسپلاین درجه ۳ سوم $\{\phi_i\}_{i=0}^{n+1}$ را به‌وسیله

$$\phi_0(x) = S\left(\frac{x}{h}\right) - 4S\left(\frac{x+h}{h}\right),$$

$$\phi_1(x) = S\left(\frac{x-x_1}{h}\right) - S\left(\frac{x+h}{h}\right),$$

$$\phi_i(x) = S\left(\frac{x-x_i}{h}\right), \quad i = 2, 3, \dots, n-1$$

به‌ازای هر $i = 2, 3, \dots, n-1$

$$\phi_n(x) = S\left(\frac{x-x_n}{h}\right) - S\left(\frac{x-(n+2)h}{h}\right),$$

$$\phi_{n+1}(x) = S\left(\frac{x-x_{n+1}}{h}\right) - 4S\left(\frac{x-(n+2)h}{h}\right)$$

تعریف می‌کنیم، که در آن

$$S(t) = \begin{cases} 0, & t \leq -2, \\ \frac{1}{24}[(2-t)^3 - 4(1-t)^3 - 6t^3 + 4(1+t)^3], & -2 \leq t \leq -1, \\ \frac{1}{24}[(2-t)^3 - 4(1-t)^3 - 6t^3], & -1 \leq t \leq 0, \\ \frac{1}{24}[(2-t)^3 - 4(1-t)^3], & 0 \leq t \leq 1, \\ \frac{1}{24}(2-t)^3, & 1 \leq t \leq 2, \\ 0, & 2 \leq t. \end{cases}$$

مرحله ۳. به‌ازای هر $i = 0, 1, \dots, n+1$

$$a_{ij} = \int_{\max(x_j-2, 0)}^{\min(x_i+2, 1)} [p(x)\phi_i'(x)\phi_j'(x) + q(x)\phi_i(x)\phi_j(x)] dx$$

را به‌ازای هر $j = i, i+1, \dots, \min(i+3, n+1)$ حساب می‌کنیم.

مرحله ۴. به‌ازای هر $i = 0, 1, \dots, n+1$ و $j = i+1, \dots, \min(i+3, n+1)$

قرار می‌دهیم $a_{ri} = a_{ij}$.

مرحله ۵. به ازای هر $i = 0, 1, \dots, n+1$ و $j < i - 3$ و $j > i + 3$ ، قرار می‌دهیم
 $a_{ij} = 0$.

مرحله ۶. قرار می‌دهیم

به ازای هر $i = 0, 1, \dots, n+1$ ،

$$b_i = \int_{\max(x_i - 2, 0)}^{\min(x_i + 2, 1)} f(x) \phi_i(x) dx$$

مرحله ۷. دستگاه

$$Ac = b$$

را، که در آن $A = (a_{ij})$ ، $b = (b_0, \dots, b_{n+1})'$ و $c = (c_0, \dots, c_{n+1})'$ ، حل می‌کنیم.

مرحله ۸. روند تمام است. اسپلاین مکعبی $\phi(x) = \sum_{i=0}^{n+1} c_i \phi_i(x)$ تقریب به $y(x)$ است.

در عمل توصیه می‌شود که انتگرالگیریهای مراحل ۳ و ۶ به وسیله درونیابی p ، q ، و f با چند جمله‌ای درونیاب اسپلاین مکعبی و بعد انتگرال گرفتن انجام شوند. مفروضات ارائه شده در ابتدای این بخش جهت تضمین اینکه

$$|y(x) - \phi(x)| = O(h^4), \quad 0 \leq x \leq 1$$

کافی هستند. برای اثباتی از این نتیجه، ر.ک. شولتز [۵۶]، صفحات ۱۰۷ و ۱۰۸. پایه دیگری که غالباً در عمل بکار می‌رود چند جمله‌ایهای هرمیت^۱ قطعه قطعه مکعبی هستند. برای ارائه‌ای عالی از این روش، ر.ک. شولتز [۵۶].

مجموعه تمرینات ۴.۱۰

۱. نشان دهید که مسائل مقدار مرزی در مفروضات قضیه ۵.۱۰ صدق می‌کنند و جواب آنها را با استفاده از الگوریتم ۵.۱۰ تقریب کنید.

$$\begin{aligned} -y'' + y &= x, \quad 0 < x < 1, \quad (I) \\ y(0) &= 0, \quad y(1) = 0; \end{aligned}$$

۱. h را بکار برده جواب را با جواب واقعی زیر مقایسه کنید:

$$y = x + \left(\frac{e}{e^2 - 1} \right) (e^{-x} - e^x).$$

$$-x^2 y'' - 2xy' + 2y = -4x^2, \quad 0 < x < 1, \quad (b)$$

$$y(0) = 0, \quad y(1) = 0;$$

۰.۵ $h = 0.5$ را بکار برده و جواب را با جواب واقعی $v = x^2 - x$ مقایسه کنید.

۲. الگوریتم ۵.۱۰ را روی مسائل مقدار مرزی زیر، که در مفروضات قضیه ۵.۱۰ صدق نمی‌کنند، اعمال کنید.

$$-y'' - y = x, \quad 0 < x < 1, \quad (T)$$

$$y(0) = 0, \quad y(1) = 0;$$

۱. $h = 1$ را بکار برده و جواب را با جواب واقعی مقایسه کنید.

$$-y'' = 1, \quad 0 < x < 1, \quad (-)$$

$$y(0) = 0, \quad y(1) = 0;$$

۱. $h = 1$ را بکار برده و جواب را با جواب واقعی مقایسه کنید.

۳. نشان دهید که، برای توابع ϕ_i تعریف شده در (۴۰.۱۰)، $\{\phi_i\}_{i=1}^n$ یک مجموعه مستقل خطی بر $[0, 1]$ است.

۴. نشان دهید که ماتریس A در (۴۲.۱۰) معین مثبت است.

۵. با استفاده از $\phi(x) = \sin(j\pi x)$ ، به‌ازای هر $j = 1, 2, \dots, 5$ ، یک تقریب ریلی - ریتس برای مسئله مقدار مرزی زیر پیدا کنید:

$$-y'' + 4y = 4x(1 - e^{-2}), \quad 0 < x < 1.$$

$$y(0) = 0, \quad y(1) = 0.$$

۶. نشان دهید که ماتریس A که به‌وسیله (۴۲.۱۰) برای مسئله مقدار مرزی (۳۲.۱۰) با $p(x) = 1$ و $q(x) = 0$ داده شده است همانند ماتریس بدست آمده با روش تفاضل متناهی است.

۷. نشان دهید که مسئله مقدار مرزی

$$-\frac{d}{dx}(p(x)y') + q(x)y = f(x), \quad 0 < x < 1,$$

$$y(0) = \alpha, \quad y(1) = \beta,$$

را می‌توان با تغییر متغیر

$$z = y - \beta x - (1-x)\alpha$$

به شکل زیر تبدیل کرد:

$$-\frac{d}{dx}(p(x)z') + q(x)z = F(x), \quad 0 < x < 1,$$

$$z(0) = 0, \quad z(1) = 0.$$

۸. با استفاده از تمرین ۷ و الگوریتم ۵.۱۰، جواب مسئله مقدار مرزی زیر را تقریب کنید:

$$-y'' + y = x, \quad 0 < x < 1,$$

$$y(0) = 1, \quad y(1) = 1 + e^{-1}.$$

۹. نشان دهید که مسئله مقدار مرزی

$$-\frac{d}{dx}(p(x)y') + q(x)y = f(x), \quad a < x < b,$$

$$y(a) = \alpha, \quad y(b) = \beta,$$

را می‌توان به شکل (۳۲.۱۰) و (۳۳.۱۰) تبدیل کرد.

۱۰. نشان دهید که ماتریس A در (۴۳.۱۰) معین مثبت است.

۱۱. مسئله مثال ۱ را با استفاده از الگوریتم ۶.۱۰ حل کنید. همانگونه که پیشنهاد شده، انتگرالها را با استفاده از درونیابی تقریب کنید.

۱۲. تمرین ۱ را با استفاده از الگوریتم ۶.۱۰ تکرار کنید.

۱۳. تمرین ۳ را با استفاده از پایه اسپلاین مکعبی $\{\phi_i\}_{i=0}^{n+1}$ تکرار کنید.

۱۴. تمرین ۸ را با استفاده از الگوریتم ۶.۱۰ تکرار کنید.

حل عددی معادلات دیفرانسیل جزئی^{۱۱}

یک جسم ایزوتروپیک نامیده می‌شود اگر هدایت گرمایی در هر نقطهٔ جسم مستقل از جهت جریان گرما از آن نقطه باشد. دمای $u \equiv u(x, y, z, t)$ در یک جسم ایزو-تروپیک را می‌توان با حل معادلهٔ دیفرانسیل جزئی

$$\frac{\partial}{\partial x} \left(k \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial y} \left(k \frac{\partial u}{\partial y} \right) + \frac{\partial}{\partial z} \left(k \frac{\partial u}{\partial z} \right) = c\rho \frac{\partial u}{\partial t},$$

پیدا کرد، که در آن k ، c ، و ρ توابعی از (x, y, z) هستند و، بترتیب، هدایت گرمایی، گرمای مشخص، و چگالی جسم در نقطهٔ (x, y, z) را نمایش می‌دهند. زمانی که k ، c ، و ρ ثابت هستند، این معادله به معادلهٔ گرمای سه بعدی معروف است و می‌توان آن را به صورت زیر بیان کرد:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \frac{c\rho}{k} \frac{\partial u}{\partial t}.$$

اگر مرز جسم نسبتاً ساده باشد، جواب این معادله را می‌توان با استفاده از سری فوریه پیدا کرد.

در اکثر حالات که در آنها k ، c ، و ρ ثابت نیستند، یا وقتی مرز جسم نامنظم است، جواب معادله باید بوسیلهٔ تکنیکهای تقریبی بدست آید. مبانی تکنیکهایی از این نوع در این فصل ارائه می‌شود.

وضعیت‌های فیزیکی شامل بیش از یک متغیر را غالباً "می‌توان با معادلات حاوی مشتقات جزئی بیان کرد. در این فصل مبانی مختصری از بعضی تکنیک‌های موجود برای تقریب جواب معادلات دیفرانسیل جزئی شامل دو متغیر را، با نشان دادن اینکه چگونه این تکنیک‌ها را می‌توان روی مسائل فیزیکی متعارف معینی اعمال کرد، ارائه می‌کنیم. بررسی خود را به مسائلی از این نوع محدود می‌کنیم، زیرا اغلب تکنیک‌های پیشرفته‌تر به پیشنیازی از آنالیز نیاز دارند که از یک دانشجوی سطح لیسانس انتظار نمی‌رود.

۱۰۱۱ مسائل فیزیکی شامل معادلات دیفرانسیل جزئی

اولین نوع مسئله‌ای که در نظر می‌گیریم مثالی از یک معادله دیفرانسیل جزئی بیضوی است و به معادله پواسن^۱ معروف است:

$$(10.11) \quad \frac{\partial^2 u}{\partial x^2}(x, y) + \frac{\partial^2 u}{\partial y^2}(x, y) = f(x, y).$$

فرض کنیم در این معادله تابع f مشخص کننده داده مسئله روی یک ناحیه سطح R باشد که مرز آن را با S نمایش خواهیم داد. معادلاتی از این نوع به طور طبیعی در مطالعه مسائل فیزیکی مستقل از زمان گوناگون، نظیر حالت یکساخت توزیع گرما در یک ناحیه سطح، انرژی پتانسیل یک نقطه در صفحه که نیروهای ثقلی واقع در صفحه روی آن عمل می‌کنند، و مسائل حالت یکساخت دوبعدی شامل سیالات تراکم ناپذیر، رخ می‌دهند.

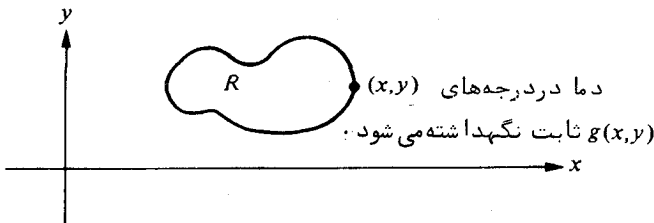
برای بدست آوردن جواب منحصر بفرد برای معادله پواسن باید قیدهای اضافی بر جواب نهاد. مثلاً، مطالعه حالت یکنواخت توزیع گرما در یک ناحیه مسطح لازم دارد که $f(x, y) \equiv 0$ ، که ساده شدن (۱.۱۱) را به

$$(۲.۱۱) \quad \frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} = 0.$$

نتیجه می دهد که معادله لاپلاس نامیده می شود. اگر دمای داخل ناحیه به وسیله توزیع دما روی مرز ناحیه معین شود، قیدها شرایط مرزی دیریکله^۱ نامیده می شوند و برای تمام (x, y) روی S ، یعنی مرز ناحیه R ، با

$$(۳.۱۱) \quad u(x, y) = g(x, y)$$

داده می شود (ر.ک. شکل ۱.۱۱).



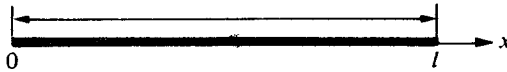
شکل ۱.۱۱

در بخش ۳.۱۱ حل عددی مسئله ای شامل یک معادله دیفرانسیل جزئی سهموی به شکل زیر بررسی خواهد شد:

$$(۴.۱۱) \quad \frac{\partial u(x, t)}{\partial t} - x^2 \frac{\partial^2 u(x, t)}{\partial x^2} = 0.$$

مسئله فیزیکی که اینجا بررسی می شود به جریان گرما در امتداد یک میله نازک به طول l مربوط است که فرض می شود در هر مقطع عرضی جسم دارای دمای یکنواخت است، بشرطی

که میله در سطح کناری اش کاملاً " عایق دار باشد. ثابت x در معادله (۴.۱۱) برای مسئله‌ای از این تسوابع بوسیله خواص هدایت گرمایی مواد تشکیل دهنده میله تعیین و فرض می‌شود که از محل استقرار در میله مستقل باشد (ر.ک. شکل ۲.۱۱).



شکل ۲.۱۱

یک دسته نوعی از این قیود برای مسئله جریان گرما آن است که توزیع اولیه گرما در میله را مشخص کرده

$$u(x, 0) = f(x),$$

و بیان کنیم که در انتهاهای میله چه رخ می‌دهد. مثلاً، اگر انتهاهای میله را در دمای ثابت U_1 و U_2 نگهداریم، شرایط مرزی به شکل زیر خواهند بود:

$$u(l, t) = U_2 \quad \text{و} \quad u(0, t) = U_1$$

و توزیع گرما در میله به توزیع دمای حدی میل می‌کند:

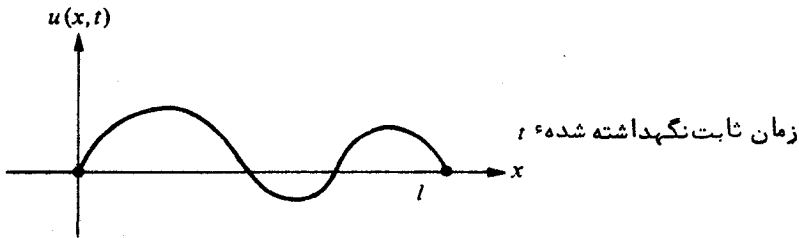
$$\lim_{t \rightarrow \infty} u(x, t) = U_1 + \frac{U_2 - U_1}{l} x$$

در عوض، اگر میله را چنان عایق دار سازیم که هیچ گرمایی از انتهاها جریان نداشته باشد، شرایط مرزی

$$\frac{\partial u}{\partial x}(l, t) = 0 \quad \text{و} \quad \frac{\partial u}{\partial x}(0, t) = 0$$

خواهند بود، و یک دمای ثابت را، به عنوان حالت حدی، در میله نتیجه می‌دهد. معادله دیفرانسیل جزئی سهموی نیز در مطالعه انتشار گاز اهمیت دارد؛ و در واقع، معادله (۴.۱۱) در بعضی از حوزه‌ها به معادله انتشار معروف است. مسئله‌ای که در بخش ۴.۱۱ مطالعه خواهیم کرد معادله موج یک بعدی نامیده می‌شود و مثالی از یک معادله دیفرانسیل جزئی هذلولوی است.

فرض کنیم یک نخ قابل ارتجاع به طول l بین دو نقطه اتکا در یک سطح افقی کشیده شده باشد (ر.ک. شکل ۳.۱۱).



شکل ۳.۱۱

هرگاه نخ چنان حرکت درآید که در یک سطح قائم نوسان کند، آنگاه تغییر مکان قائم $u(x, t)$ یک نقطه x در زمان t در معادله دیفرانسیل جزئی

$$(۵.۱۱) \quad \alpha^2 \frac{\partial^2 u}{\partial x^2}(x, t) = \frac{\partial^2}{\partial t^2} u(x, t), \quad 0 < x < l, \quad 0 < t,$$

صدق می‌کند، بشرطی که از اثرات بی‌حرکت کردن سیم صرف‌نظر شود و میدان نوسان خیلی بزرگ نباشد. برای اعمال قیود روی این مسئله، فرض کنیم محل اولیه و سرعت نخ به‌وسیله

$$\frac{\partial u}{\partial t}(x, 0) = g(x), \quad 0 \leq x \leq l \quad \text{و} \quad u(x, 0) = f(x)$$

داده شده باشند و از این امر استفاده می‌کنیم که نقاط انتهایی ثابت هستند، که نتیجه می‌دهد $u(0, t) = 0$ و $u(l, t) = 0$.

مسائل فیزیکی دیگری شامل معادلات دیفرانسیل جزئی هذلولوی (۵.۱۱) در مطالعه میله‌های نوسان کننده، که یک یا دو انتهای آن با گیره نگه‌داشته می‌شود، و انتقال الکتریسیته در یک خط انتقال طویل، که در آن مقداری نشت جریان به زمین وجود دارد، رخ می‌دهد.

مجموعه تمرینات ۱.۱۱

مسائل زیر جوابهای سری فوریه برای معادلات دیفرانسیل جزئی این بخش را خلاصه می‌کنند. فرض می‌شود که هر سری همگراست و سریهای بدست آمده از مشتگیری جمله به جمله به مشتقات جزئی مناسبی همگرا می‌باشد.

۱. معادله لا پلاس روی یک مستطیل. فرض کنید توابع $f_1(x)$ ، $f_2(x)$ ، $g_1(y)$ ، و

$g_2(y)$ سریهای فوریه زیر را داشته باشند.

$$A_n = \frac{2}{a} \int_0^a f_1(x) \sin\left(\frac{n\pi x}{a}\right) dx \quad \text{که در آن } f_1(x) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{a}\right)$$

$$B_n = \frac{2}{a} \int_0^a f_2(x) \sin\left(\frac{n\pi x}{a}\right) dx \quad \text{که در آن } f_2(x) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{a}\right)$$

$$C_n = \frac{2}{b} \int_0^b g_1(y) \sin\left(\frac{n\pi y}{b}\right) dy \quad \text{که در آن } g_1(y) = \sum_{n=1}^{\infty} C_n \sin\left(\frac{n\pi y}{b}\right)$$

$$D_n = \frac{2}{b} \int_0^b g_2(y) \sin\left(\frac{n\pi y}{b}\right) dy \quad \text{که در آن } g_2(y) = \sum_{n=1}^{\infty} D_n \sin\left(\frac{n\pi y}{b}\right)$$

نشان دهید که معادله لاپلاس روی یک مستطیل

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad 0 < x < a, \quad 0 < y < b,$$

با شرایط مرزی

$$u(x, 0) = f_1(x), \quad u(x, b) = f_2(x), \quad 0 \leq x \leq a,$$

$$u(0, y) = g_1(y), \quad u(a, y) = g_2(y), \quad 0 \leq y \leq b,$$

دارای جواب

$$u(x, y) = \sum_{n=1}^{\infty} \left\{ \left[A'_n \sinh\left(\frac{n\pi(b-y)}{a}\right) + B'_n \sinh\left(\frac{n\pi y}{a}\right) \right] \sin\left(\frac{n\pi x}{a}\right) + \left[C'_n \sinh\left(\frac{n\pi(a-x)}{b}\right) + D'_n \sinh\left(\frac{n\pi x}{b}\right) \right] \sin\left(\frac{n\pi y}{b}\right) \right\}$$

است، که در آن

$$A'_n = \frac{A_n}{\sinh\left(\frac{n\pi b}{a}\right)}, \quad B'_n = \frac{B_n}{\sinh\left(\frac{n\pi b}{a}\right)}, \quad C'_n = \frac{C_n}{\sinh\left(\frac{n\pi a}{b}\right)}$$

و

$$D'_n = \frac{D_n}{\sinh\left(\frac{n\pi a}{b}\right)}$$

۲. معادله گرما. فرض کنید تابع f دارای سری فوریه $f(x) = \sum_{n=1}^{\infty} A_n \sin \frac{n\pi x}{l}$ است،

$$A_n = \frac{2}{l} \int_0^l f(x) \sin \frac{n\pi x}{l} dx$$

(T) نشان دهید که جواب معادله گرمای

$$\frac{\partial u}{\partial t} - x^2 \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < l, \quad 0 < t,$$

با شرایط مرزی

$$u(0, t) = u(l, t) = 0, \quad 0 < t,$$

و شرط اولیه

$$u(x, 0) = f(x), \quad 0 \leq x \leq l,$$

با معادله زیر داده می‌شود:

$$u(x, t) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{l}\right) \exp\left(\frac{-x^2 n^2 \pi^2}{l^2} t\right).$$

(ب) فرض کنید $w(x) = U_1 + x(U_2 - U_1)/l$. نشان دهید هرگاه شرایط مرزی

قسمت (T) را با

$$u(0, t) = U_1, \quad u(l, t) = U_2,$$

جایگزین کنیم، آنگاه جواب به وسیله

$$u(x, t) = w(x) + \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{l}\right) \exp\left(\frac{-x^2 n^2 \pi^2}{l^2} t\right),$$

داده می‌شود، که در آن

$$B_n = \frac{2}{l} \int_0^l [f(x) - w(x)] \sin\left(\frac{n\pi x}{l}\right) dx.$$

۳. معادله گرمای . فرض کنید سری فوریه برای $f(x)$ به وسیله

$$f(x) = A_0 + \sum_{n=1}^{\infty} A_n \cos \frac{n\pi x}{l},$$

داده شده باشد، که در آن

$$A_n = \frac{2}{l} \int_0^l f(x) \cos \frac{n\pi x}{l} dx \quad \text{و} \quad A_0 = \frac{1}{l} \int_0^l f(x) dx, \quad n = 1, 2, \dots$$

نشان دهید که جواب معادله گرمای

$$\frac{\partial u}{\partial t} - x^2 \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < l, \quad 0 < t,$$

با شرایط مرزی

$$\frac{\partial u}{\partial x}(0, t) = \frac{\partial u}{\partial x}(l, t) = 0, \quad 0 < t,$$

و شرط اولیه^۴

$$u(x, 0) = f(x), \quad 0 \leq x \leq l,$$

به وسیله^۴ معادله^۴ زیر داده می شود :

$$u(x, t) = A_0 + \sum_{n=1}^{\infty} A_n \cos\left(\frac{n\pi x}{l}\right) \exp\left(-\frac{x^2 n^2 \pi^2}{l^2} t\right).$$

۴. معادله^۴ موج. فرض کنید $f(x)$ و $g(x)$ دارای سریهای فوریه^۴ زیر باشند.

$$A_n = \frac{2}{l} \int_0^l f(x) \sin\left(\frac{n\pi x}{l}\right) dx \quad \text{که در آن} \quad f(x) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{l}\right)$$

و

$$B_n = \frac{2}{l} \int_0^l g(x) \sin\left(\frac{n\pi x}{l}\right) dx \quad \text{که در آن} \quad g(x) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{l}\right)$$

نشان دهید که جواب معادله^۴ موج

$$\frac{\partial^2 u}{\partial t^2} - x^2 \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < l, \quad 0 < t,$$

$$u(0, t) = u(l, t) = 0, \quad 0 < t,$$

$$u(x, 0) = f(x), \quad 0 \leq x \leq l,$$

$$\frac{\partial u}{\partial t}(x, 0) = g(x), \quad 0 \leq x \leq l,$$

به وسیله^۴

$$u(x, t) = \sum_{n=1}^{\infty} \left[A_n \cos\left(\frac{x n \pi}{l} t\right) + B'_n \sin\left(\frac{x n \pi}{l} t\right) \right] \sin\left(\frac{n \pi x}{l}\right),$$

داده می شود، که در آن $B'_n = (l: x n \pi) B_n$

۲۰۱۱ معادلات دیفرانسیل جزئی بیضوی

معادله^۴ دیفرانسیل جزئی بیضوی که بررسی خواهیم کرد معادله^۴ پواسن است که به ازای

$$(x, y) \in R$$

$$(۶.۱۱) \quad \nabla^2 u(x, y) \equiv \frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} = f(x, y),$$

و بازای $(x, y) \in S$

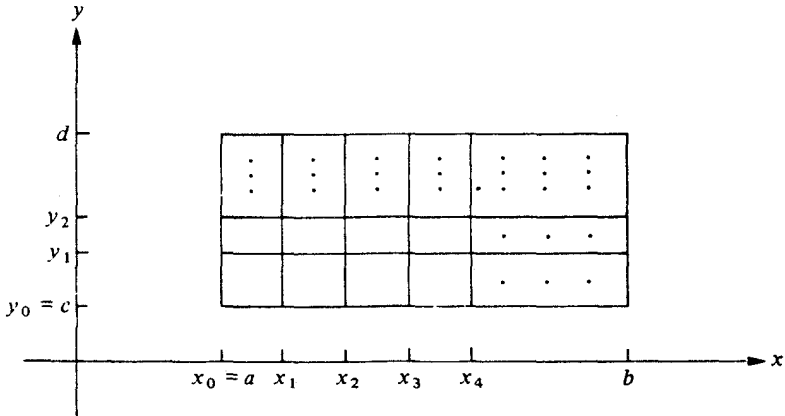
$$u(x, y) = g(x, y),$$

که در آن

$$R = \{(x, y) | a < x < b, \quad c < y < d\}$$

و S نمایانگر مرز R است. در این بحث فرض می‌کنیم f و g بر قلمروهایشان پیوسته باشند بطوری که اطمینان جواب منحصر بفرد داده شود.

روشی که بکار برده می‌شود تعدیلی از روش تفاضل متناهی برای مسائل مقدار مرزی است که در بخش ۳.۱۰ مورد بحث قرار گرفت. اولین مرحله انتخاب اعداد صحیح n و m و تعریف اندازه‌های گام h و k با $h = (b - a)/n$ و $k = (d - c)/m$ است، افزاز بازه $[a, b]$ به n قسمت مساوی به عرض h و بازه $[c, d]$ به m قسمت مساوی به عرض k (ر.ک. شکل ۴.۱۱) وسیله‌ای جهت قرار دادن شبکه‌ای روی مستطیل R ، به وسیله



شکل ۴.۱۱

کشیدن خطوط قائم و افقی ماربیر نقاط با مختصات (x_i, y_i) که در آنها

$$x_i = a + ih, \quad i = 0, 1, \dots, n$$

بمازای هر $j = 0, 1, \dots, m$ ، $y_j = c + jk$ ،

مهیما می کند .

خطوط $x = x_i$ و $y = y_j$ خطوط شبکه‌ای نامیده می شوند و محل تلاقی آنها نقاط شبکه‌ای نامیده می شوند . برای هر نقطهء شبکه‌ای که در داخل شبکه قرار دارد ، یعنی (x_i, y_j) ، $i = 1, 2, \dots, n - 1$ و $j = 1, 2, \dots, m - 1$ ، می توانیم از سری تیلور در متغیر x و حول x_i استفاده کرده فرمول تفاضل مرکزی

$$(۷.۱۱) \quad \frac{\partial^2}{\partial x^2} u(x_i, y_j) = \frac{u(x_{i+1}, y_j) - 2u(x_i, y_j) + u(x_{i-1}, y_j)}{h^2} - \frac{h^2}{12} \frac{\partial^4 u}{\partial x^4}(\xi_i, y_j),$$

را تولید کنیم ، که در آن $\xi_i \in (x_{i-1}, x_{i+1})$ ؛ و از سری تیلور در متغیر y و حول y_j جهت تولید فرمول تفاضل مرکزی

$$(۸.۱۱) \quad \frac{\partial^2}{\partial y^2} u(x_i, y_j) = \frac{u(x_i, y_{j+1}) - 2u(x_i, y_j) + u(x_i, y_{j-1}))}{k^2} - \frac{k^2}{12} \frac{\partial^4 u}{\partial y^4}(x_i, \eta_j),$$

استفاده کنیم ، که در آن $\eta_j \in (y_{j-1}, y_{j+1})$.

بکار بردن فرمولهای مذکور در معادلات (۶.۱۱) اجازه می دهند که معادلهء پواسن

را در (x_i, y_j) به صورت

$$(۹.۱۱) \quad \frac{u(x_{i+1}, y_j) - 2u(x_i, y_j) + u(x_{i-1}, y_j)}{h^2} + \frac{u(x_i, y_{j+1}) - 2u(x_i, y_j) + u(x_i, y_{j-1}))}{k^2} = f(x_i, y_j) + \frac{h^2}{12} \frac{\partial^4}{\partial x^4} u(\xi_i, y_j) + \frac{k^2}{12} \frac{\partial^4}{\partial y^4} u(x_i, \eta_j),$$

بمازای هر $i = 1, 2, \dots, (n - 1)$ و $j = 1, 2, \dots, (m - 1)$ ، و شرایط مرزی را به صورت

$$(۱۰.۱۱) \quad \begin{aligned} u(x_0, y_j) &= g(x_0, y_j) , \quad j = 0, 1, \dots, m \\ u(x_n, y_j) &= g(x_n, y_j) , \quad j = 0, 1, \dots, m \\ u(x_i, y_0) &= g(x_i, y_0) , \quad i = 1, 2, \dots, n - 1 \\ u(x_i, y_m) &= g(x_i, y_m) , \quad i = 1, 2, \dots, n - 1 \end{aligned}$$

بیان کنیم .

در قالب معادلات تفاضلی ، این معادلات روشی را نتیجه می دهند که روش تفاضل مرکزی نامیده می شود ، با خطای برشی موضعی مرتبه $O(h^2 + k^2)$ که می توان آن را به ازای هر $j = 1, 2, \dots, m - 1$ و $i = 1, 2, \dots, n - 1$ به صورت

$$(11 \cdot 11) \quad 2 \left[\left(\frac{h}{k} \right)^2 + 1 \right] w_{i,j} - (w_{i+1,j} + w_{i-1,j}) - \left(\frac{h}{k} \right)^2 (w_{i,j+1} + w_{i,j-1}) = -h^2 f(x_i, y_j),$$

نوشت و

به ازای هر $j = 0, 1, \dots, m$ $w_{0,j} = g(x_0, y_j)$

به ازای هر $j = 0, 1, \dots, m$ $w_{n,j} = g(x_n, y_j)$

(12 \cdot 11)

به ازای هر $i = 1, 2, \dots, n - 1$ $w_{i,0} = g(x_i, y_0)$

به ازای هر $i = 1, 2, \dots, n - 1$ $w_{i,m} = g(x_i, y_m)$

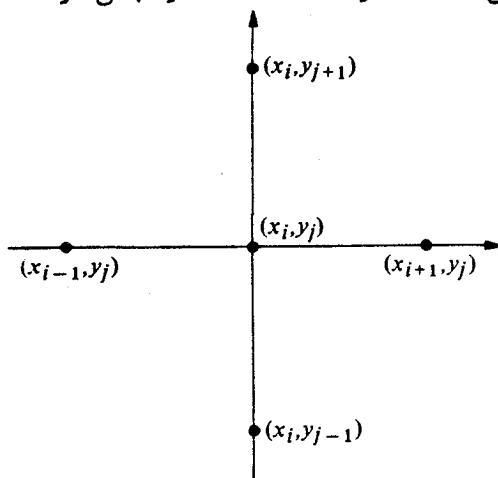
که در آنها $w_{i,j}$ مقدار $u(x_i, y_j)$ را تقریب می کند .

معادله ای نوعی در (11 \cdot 11) شامل تقریبهایی به $u(x, y)$ در نقاط

$(x_{i-1}, y_j), (x_i, y_j), (x_{i+1}, y_j), (x_i, y_{j-1}), (x_i, y_{j+1})$

است ، که قسمتی از شبکه را که این نقاط در آنها قرار دارند مجدداً " تولید می کند .

شکل ۵.۱۱ نشان می دهد که هر معادله شامل تقریبهایی در یک ناحیه ستاره ای شکل



شکل ۵.۱۱

حول (x_i, y_j) است .

اگر هر وقت مناسب است در دستگاه داده شده به وسیله (۱۱.۱۱) از اطلاعات شرایط مرزی (۱۲.۱۱) استفاده کنیم - یعنی ، در تمام نقاط (x_i, y_j) که مجاور یک نقطه شبکه‌ای مرزی هستند - دستگاهی $(m - 1) \times (n - 1)$ از معادلات خطی شامل $(n - 1) \times (m - 1)$ مجهول داریم که مجهولات تقریبهای $w_{i,j}$ به $u(x_i, y_j)$ ، برای نقاط شبکه‌ای داخلی ، هستند .

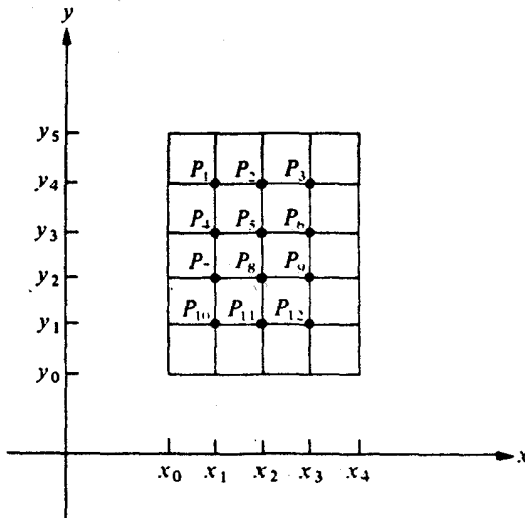
دستگاه خطی شامل این مجهولات را می‌توان برای محاسبات ماتریسی کاراتری بیان کرد ، اگر یک تجدید شماره‌گذاری برای نقاط شبکه‌ای داخلی انجام شود . شماره‌گذاری توصیه شده‌ای برای این نقاط (ر. ک. و ارکا [۶۵] ، صفحه ۱۸۷) آن است که فرض کنیم

$$w_l = w_{i,j} \quad \text{و} \quad P_l = (x_i, y_j)$$

که در آن به‌ازای هر

$$i = 1, 2, \dots, n - 1 \quad \text{و} \quad j = 1, 2, \dots, m - 1$$

$l = i + (m - 1 - j)(n - 1)$. در نتیجه ، نقاط شبکه‌ای متوالیا "از چپ به راست و از بالا به پایین شماره‌گذاری می‌شوند . مثلاً" ، با $n = 4$ و $m = 5$ ، تجدید شماره‌گذاری شبکه‌ای را نتیجه می‌دهد که نقاط آن در شکل ۶.۱۱ نشان داده شده است .



شکل ۶.۱۱

مسئله تعیین حالت یکنواخت توزیع گرما در یک صفحه فلزی نازک به شکل مربع به ابعاد ۰.۵ متر در ۰.۵ متر را در نظر می‌گیریم که دومرزمجاور آن در صفر درجه سانتیگراد نگهداشته شده در حالی که در مرزهای دیگر گرما به‌طور خطی از ۰ درجه سانتیگراد در یک گوشه تا ۱۰۰ درجه سانتیگراد به‌گوشه دیگر که همدیگر را قطع می‌کنند افزایش می‌یابد. اگر اضلاع با شرایط مرزی صفر را در امتداد محورهای x و y قرار می‌دهیم، مسئله از نظر ریاضی به صورت زیر بیان می‌شود: به‌ازای (x, y) در مجموعه

$$R = \{(x, y) | 0 < x < .5, 0 < y < .5\}$$

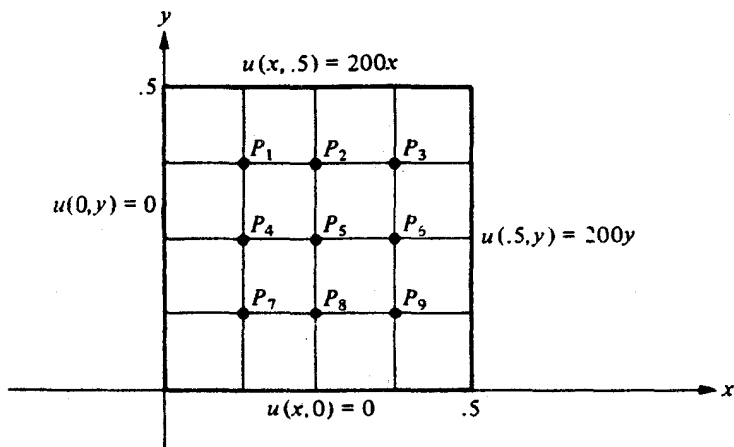
$$(13.11) \quad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

با شرایط مرزی

$$(14.11) \quad \begin{aligned} u(0, y) &= 0, & u(x, 0) &= 0. \\ u(x, .5) &= 200x, & u(.5, y) &= 200y. \end{aligned}$$

اگر $n = m = 4$ ، مسئله شبکه داده شده در شکل ۷.۱۱ را خواهد داشت و معادله تفاضلی (۱۱.۱۱) به‌ازای هر $i = 1, 2, 3$ و $j = 1, 2, 3$ خواهد شد

$$4w_{i,j} - w_{i+1,j} - w_{i-1,j} - w_{i,j-1} - w_{i,j+1} = 0.$$



شکل ۷.۱۱

بیان این معادله برحسب نقاط شبکه‌ای مجدداً " شماره‌گذاری شده $w_i = u(P_i)$

نتیجه می‌دهد که معادلات عبارتند از

$$\begin{aligned} 4w_1 - w_2 - w_4 &= w_{0.3} + w_{1.4} \\ 4w_2 - w_3 - w_1 - w_5 &= w_{2.4} \\ 4w_3 - w_2 - w_6 &= w_{4.3} + w_{3.4} \\ 4w_4 - w_5 - w_1 - w_7 &= w_{0.2} \\ 4w_5 - w_6 - w_4 - w_2 - w_8 &= 0 \\ 4w_6 - w_5 - w_3 - w_9 &= w_{4.2} \\ 4w_7 - w_8 - w_4 &= w_{0.1} + w_{1.0} \\ 4w_8 - w_9 - w_7 - w_5 &= w_{2.0} \\ 4w_9 - w_8 - w_6 &= w_{3.0} + w_{4.1} \end{aligned}$$

که در آن طرف راست معادلات از شرایط مرزی بدست می‌آیند. در واقع، شرایط داده شده در (۱۴.۱۱) نتیجه می‌دهد که

$$w_{1.0} = w_{2.0} = w_{3.0} = w_{0.1} = w_{0.2} = w_{0.3} = 0, \\ w_{1.4} = w_{4.1} = 25, \quad w_{2.4} = w_{4.2} = 50, \quad \text{and} \quad w_{3.4} = w_{4.3} = 75.$$

دستگاه خطی که به این مسئله مربوط می‌شود دارای شکل زیر است:

$$\begin{bmatrix} 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 4 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 4 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 4 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ w_8 \\ w_9 \end{bmatrix} = \begin{bmatrix} 25 \\ 50 \\ 150 \\ 0 \\ 0 \\ 50 \\ 0 \\ 0 \\ 25 \end{bmatrix}$$

مقادیر w_1, w_2, \dots, w_9 که از اعمال روش گاوس-سیدل روی این ماتریس پیدا می‌شود ذیلاً "داده شده‌اند":

i	1	2	3	4	5	6	7	8	9
w_i	18.75	37.50	56.25	12.50	25.00	37.50	6.25	12.50	18.75

این جوابها درست هستند، زیرا برای جواب $u(x, y) = 400xy$ داریم

$$\frac{\partial^4 u}{\partial x^4} = \frac{\partial^4 u}{\partial y^4} \equiv 0,$$

لذا، در هر مرحله خطای برشی صفر است.

مسئله بررسی شده در این مثال روی هر محور دارای اندازه شبکه‌ای 125 است و تنها حل یک دستگاه خطی را لازم دارد. این مطلب وضعیت را به‌طور قابل ملاحظه‌ای ساده می‌کند و مشکلات محاسباتی، که برای دستگاههای خیلی بزرگتر وجود دارد، وارد نمی‌کند. الگوریتم زیر از روش تکراری گاوس-سیدل جهت حل دستگاه خطی حاصل استفاده می‌کند و اندازه‌های شبکه‌ای غیر مساوی روی محورها را اجازه می‌دهد.

الگوریتم تفاضل متناهی معادله پواسن ۱۰۱۱

برای تقریب جواب معادله پواسن

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y), \quad a \leq x \leq b, \quad c \leq y \leq d.$$

مقید به شرایط

$$u(x, y) = g(x, y) \quad \text{اگر } x = a \text{ یا } x = b \text{ و } c \leq y \leq d$$

و

$$u(x, y) = g(x, y) \quad \text{اگر } y = c \text{ یا } y = d \text{ و } a \leq x \leq b$$

اعداد صحیح مثبت m و n را انتخاب می‌کنیم.

مرحله ۱. قرار می‌دهیم $h = (b - a) / n$ و $k = (d - c) / m$.

مرحله ۲. به‌ازای هر $i = 0, 1, \dots, n$ قرار می‌دهیم $x_i = a + ih$ و به‌ازای هر

$$j = 0, 1, \dots, m \quad y_j = c + jk$$

مرحله ۳. به‌ازای هر $i = 1, 2, \dots, (n - 1)$ و $j = 1, 2, \dots, (m - 1)$ قرار می‌دهیم

$$w_{ij}^{(0)} = 0$$

مرحله ۴. قرار می‌دهیم $\lambda = h^2 k^2$.

مرحله ۵. قرار می‌دهیم $l = 1$.

مرحله ۶. قرار می‌دهیم

$$w_{1,m-1}^{(i)} = \frac{1}{2(\lambda + 1)} [-h^2 f(x_1, y_{m-1}) + g(x_0, y_{m-1}) \\ + \lambda g(x_1, y_m) + \lambda w_{1,m-2}^{(i-1)} + w_{2,m-1}^{(i-1)}].$$

مرحله ۷. به ازای هر $i = 2, 3, \dots, (n-2)$ قرار می دهیم

$$w_{i,m-1}^{(i)} = \frac{1}{2(\lambda + 1)} [-h^2 f(x_i, y_{m-1}) + \lambda g(x_i, y_m) \\ + w_{i-1,m-1}^{(i)} + w_{i+1,m-1}^{(i-1)} + \lambda w_{i,m-2}^{(i-1)}].$$

مرحله ۸. قرار می دهیم

$$w_{n-1,m-1}^{(i)} = \frac{1}{2(\lambda + 1)} [-h^2 f(x_{n-1}, y_{m-1}) + g(x_n, y_{m-1}) \\ + \lambda g(x_{n-1}, y_m) + w_{n-2,m-1}^{(i)} + \lambda w_{n-1,m-2}^{(i-1)}].$$

قراری می دهیم $j = m - 2$.

مرحله ۹. قرار می دهیم

$$w_{1,j}^{(i)} = \frac{1}{2(\lambda + 1)} [-h^2 f(x_1, y_j) + g(x_0, y_j) \\ + \lambda w_{1,j+1}^{(i)} + \lambda w_{1,j-1}^{(i-1)} + w_{2,j}^{(i-1)}].$$

مرحله ۱۰. به ازای هر $i = 2, 3, \dots, (n-2)$ قرار می دهیم

$$w_{i,j}^{(i)} = \frac{1}{2(\lambda + 1)} [-h^2 f(x_i, y_j) + w_{i-1,j}^{(i)} + \lambda w_{i,j-1}^{(i-1)} \\ + w_{i+1,j}^{(i-1)} + \lambda w_{i,j-1}^{(i-1)}].$$

مرحله ۱۱. قرار می دهیم

$$w_{n-1,j}^{(i)} = \frac{1}{2(\lambda + 1)} [-h^2 f(x_{n-1}, y_j) + g(x_n, y_j) + w_{n-2,j}^{(i)} \\ + \lambda w_{n-1,j+1}^{(i)} + \lambda w_{n-1,j-1}^{(i-1)}].$$

مرحله ۱۲. ۱ را از z کم می کنیم.

مرحله ۱۳. اگر $z \geq 2$ ، به مرحله ۹ می رویم.

مرحله ۱۴. قرار می دهیم

$$w_{1,1}^{(i)} = \frac{1}{2(\lambda + 1)} [-h^2 f(x_1, y_1) + g(x_0, y_1) + \lambda g(x_1, y_0) + \lambda w_{1,2}^{(i)} + w_{2,1}^{(i-1)}].$$

مرحله ۱۵. به ازای $i = 2, 3, \dots, (n - 2)$ قرار می‌دهیم

$$w_{i,1}^{(i)} = \frac{1}{2(\lambda + 1)} [-h^2 f(x_i, y_1) + \lambda g(x_i, y_0) + w_{i,1}^{(i-1)} + \lambda w_{i,2}^{(i)} + w_{i+1,1}^{(i-1)}].$$

مرحله ۱۶. قرار می‌دهیم

$$w_{n-1,1}^{(i)} = \frac{1}{2(\lambda + 1)} [-h^2 f(x_{n-1}, y_1) + g(x_n, y_1) + \lambda g(x_{n-1}, y_0) + w_{n-2,1}^{(i)} + \lambda w_{n-1,2}^{(i)}].$$

مرحله ۱۷. اگر $|w_{i,j}^{(i)} - w_{i,j}^{(i-1)}|$ ، به ازای هر $i = 1, \dots, (n - 1)$ و $j = 1, \dots, (m - 1)$ به قدر کافی کوچک نباشد، ۱ را به l افزوده و به مرحله ۶ می‌رویم.

مرحله ۱۸. روند تمام است، $w_{i,j}$ ، به ازای هر $i = 1, 2, \dots, (n - 1)$ و هر $j = 1, 2, \dots, (m - 1)$ را تقریب می‌کند.

با وجود اینکه روند تکراری گاوس - سیدل برای سادگی به الگوریتم ۱۰.۱۱ تلفیق شده است، عموماً "مقتضی است، وقتی دستگاه کوچک است، از مرتبه 10^3 یا کمتر، از یک تکنیک مستقیم نظیر حذف گاوسی استفاده شود، زیرا تقارن و معین مثبت بودن پایداری نسبت به خطاهای گرد کردن را اطمینان می‌دهد. در واقع، حتی برای دستگاههای بزرگتر، تکنیکهای مستقیم معینی برای حل دستگاههایی از این نوع، بخاطر تعداد زیاد درایه‌های صفرکه وجود دارند، مناسب‌اند. بخصوص، تعمیم الگوریتم تحویل کروت ۷.۷ که در تمرین ۹، بخش ۶.۷، بررسی شد برای حل این دستگاه بسیار موثر است، زیرا ماتریس به شکل سه قطری متقارن بلوکی

$$\begin{bmatrix} A_1 & C_1 & 0 & \dots & 0 \\ C_1 & A_2 & C_2 & \dots & \vdots \\ 0 & C_2 & \dots & \dots & 0 \\ \vdots & \dots & \dots & \dots & C_{m-1} \\ 0 & \dots & 0 & C_{m-1} & A_{m-1} \end{bmatrix}$$

با بلوکهای مربع به اندازه $(n-1)$ در $(n-1)$ ، است. برای دستگاههای خیلی بزرگ توصیه می شود که از یک روش تکرار استفاده شود، بخصوص از روش SOR که در الگوریتم ۴.۸ بحث شد. انتخاب ω که در این وضعیت بهینه باشد از این واقعیت ناشی می شود که وقتی A به قطرش D و قسمتهای بالا و پایین مثلثی U و L تجزیه می شود،

$$A = D - L - U,$$

و B ماتریس ژاکوبی زیر است

$$B = D^{-1}(L + U).$$

آنگاه شعاع طیفی B عبارت است از

$$\rho(B) = \frac{1}{2} \left[\cos\left(\frac{\pi}{m}\right) + \cos\left(\frac{\pi}{n}\right) \right] \quad (\text{ر.ک. وارگا [۶۵]})$$

در نتیجه، مقدار ω که باید مورد استفاده قرار گیرد عبارت است از

$$\begin{aligned} \omega &= \frac{2}{1 + \sqrt{1 - [\rho(B)]^2}} \\ &= \frac{4}{2 + \sqrt{4 - \left[\cos\left(\frac{\pi}{m}\right) + \cos\left(\frac{\pi}{n}\right) \right]^2}} \end{aligned}$$

برای همگرایی سریعتر روند SOR، می توان یک تکنیک بلوکی در این الگوریتم وارد کرد. برای اراشهای از این تکنیک، ر.ک. وارگا [۶۵]، صفحات ۱۹۹ - ۱۹۴.

مثال ۲. معادله پواسن

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = xe^y, \quad 0 < x < 2, \quad 0 < y < 1,$$

با شرایط مرزی زیر را در نظر می گیریم:

$$u(0, y) = 0, \quad 0 \leq y \leq 1,$$

$$u(x, 0) = x, \quad 0 \leq x \leq 2,$$

$$u(2, y) = 2e^y, \quad 0 \leq y \leq 1,$$

$$u(x, 1) = ex, \quad 0 \leq x \leq 2.$$

الگوریتم ۱.۱۱ را برای تقریب جواب دقیق $u(x, y) = xe^y$ با $n = 6$ و $m = 5$ بکارمی‌بریم. معیار توقف برای مرحله ۱۷ لازم دارد که، به‌ازای هر $i = 1, \dots, 5$ و $j = 1, \dots, 4$ ،

$$|w_{i,j}^{(l)} - w_{i,j}^{(l-1)}| \leq 10^{-10}.$$

بنابراین، جواب معادلهٔ تفاضلی به‌طور دقیق بدست آمده است و روند در $l = 61$ متوقف شده است. نتایج، همراه مقادیر درست، در جدول ۱.۱۱ ارائه شده‌اند.

i	j	x_i	y_j	$w_{i,j}^{(61)}$	$u(x_i, y_j)$	$\frac{1}{2} u(x_i, y_j) - w_{i,j}^{(61)} $
1	1	.3333	.2000	.40726	.40713	1.30×10^{-4}
1	2	.3333	.4000	.49748	.49727	2.08×10^{-4}
1	3	.3333	.6000	.60760	.60737	2.23×10^{-4}
1	4	.3333	.8000	.74201	.74185	1.60×10^{-4}
2	1	.6667	.2000	.81452	.81427	2.55×10^{-4}
2	2	.6667	.4000	.99496	.99455	4.08×10^{-4}
2	3	.6667	.6000	1.2152	1.2147	4.37×10^{-4}
2	4	.6667	.8000	1.4840	1.4837	3.15×10^{-4}
3	1	1.0000	.2000	1.2218	1.2214	3.64×10^{-4}
3	2	1.0000	.4000	1.4924	1.4918	5.80×10^{-4}
3	3	1.0000	.6000	1.8227	1.8221	6.24×10^{-4}
3	4	1.0000	.8000	2.2260	2.2255	4.51×10^{-4}
4	1	1.3333	.2000	1.6290	1.6285	4.27×10^{-4}
4	2	1.3333	.4000	1.9898	1.9891	6.79×10^{-4}
4	3	1.3333	.6000	2.4302	2.4295	7.35×10^{-4}
4	4	1.3333	.8000	2.9679	2.9674	5.40×10^{-4}
5	1	1.677	.2000	2.0360	2.0357	3.71×10^{-4}
5	2	1.667	.4000	2.4870	2.4864	5.84×10^{-4}
5	3	1.667	.6000	3.0375	3.0369	6.41×10^{-4}
5	4	1.667	.8000	3.7097	3.7092	4.89×10^{-4}

جدول ۱.۱۱

مجموعه تمرینات ۲.۱۱

۱. جوابهای معادلات دیفرانسیل جزئی بیضوی زیر را با استفاده از الگوریتم ۱.۱۱ تقریب کنید.

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad 0 < x, y < 1; \quad (T)$$

$$u(x, 0) = 0, \quad u(x, 1) = x, \quad 0 \leq x \leq 1,$$

$$u(0, y) = 0, \quad u(1, y) = y, \quad 0 \leq y \leq 1.$$

از $h = k = 1$. استفاده کنید و نتایج را با جواب $u(x, y) = xy$ و با پنج جمله اول جواب سری فوریه مقایسه نمایید.

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = -2, \quad 0 < x, y < 1; \quad (-)$$

$$u(0, y) = 0, \quad u(1, y) = \sinh(\pi) \sin \pi y, \quad 0 \leq y \leq 1,$$

$$u(x, 0) = u(x, 1) = x(1 - x), \quad 0 \leq x \leq 1.$$

از $h = k = 0.5$. استفاده کنید و نتایج را با جواب زیر مقایسه نمایید

$$u(x, y) = \sinh(\pi x) \sin(\pi y) + x(1 - x).$$

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = (x^2 + y^2)e^{xy}, \quad 0 < x < 2, \quad 0 < y < 1; \quad (پ)$$

$$u(0, y) = 1, \quad u(2, y) = e^{2y}, \quad 0 \leq y \leq 1,$$

$$u(x, 0) = 1, \quad u(x, 1) = e^x, \quad 0 \leq x \leq 2.$$

از $h = 2$ و $k = 1$. استفاده کنید و نتایج را با جواب $u(x, y) = e^{xy}$ مقایسه نمایید.

۲. قسمتهای (T) و (ب) از تمرین ۱ را با استفاده از برونمایی با $h_0 = h$ ، $h_1 = h/2$ ، $h_2 = h/4$ و تکرار کنید.

۳. الگوریتمی شبیه ۱۰۱۱ بسازید، با این فرق که از روش چولسکی به جای روش تکراری گاوس - سیدل برای حل دستگاه خطی استفاده کنید.

۴. الگوریتم تمرین ۳ را روی مسائل تمرین ۱ اعمال کنید.

۵. الگوریتمی مشابه الگوریتم ۱۰۱۱ بسازید، با این فرق که برای حل دستگاه خطی از روش SOR با ω ی بهینه به جای روش گاوس - سیدل استفاده کنید.

۶. با استفاده از الگوریتم ساخته شده در تمرین ۵، تمرین ۱ را تکرار کنید.

۷. یک کابل هم محور از یک هادی داخلی ۱ اینچ مربع و یک هادی خارجی ۵ اینچ مربع ساخته شده است. فرض کنید هادی داخلی در صفر ولت نگهداشته شده باشد در حالی که هادی خارجی در ۱۱۰ ولت نگهداشته شده است. پتانسیل بین این دو هادی را با قرار دادن شبکه‌ای با فاصله شبکه‌ای افقی ۱ اینچ h و فاصله شبکه‌ای قائم ۱ اینچ k روی ناحیه

$$D = \{(x, y) | 0 \leq x, y \leq 5\}$$

پیدا کنید. جواب معادله لاپلاس را در هر نقطه شبکه تقریب کنید و دو دسته

شرایط مرزی را بکار برده دستگاهی خطی بدست آرید که به وسیله روش گاوس - سیدل حل شود.

۸. یک صفحه نقره‌ای مستطیلی ۵ سانتیمتر \times ۶ سانتیمتر در هر نقطه گرمای یکنواختی با سرعت ۱.۵ $(\text{سانتیمتر})^3 / \text{کالری}$ تولید می‌کند. فرض کنید x نمایش فاصله در امتداد لبه صفحه به طول ۶ سانتیمتر بوده و y فاصله در امتداد لبه صفحه به طول ۵ سانتیمتر باشد. همچنین، دمای u در امتداد لبه‌ها در دماهای زیر نگهداشته شده باشند:

$$u(x, 0) = x(6 - x), \quad 0 \leq x \leq 6,$$

$$u(x, 5) = 0, \quad 0 \leq x \leq 6,$$

$$u(0, y) = y(5 - y), \quad 0 \leq y \leq 5,$$

$$u(6, y) = 0, \quad 0 \leq y \leq 5.$$

که در آن مبدأ در گوشه‌ای از صفحه به مختصات $(0, 0)$ قرار دارد و لبه‌ها در امتداد محورهای مثبت x و y قرار دارند. دمای حالت یکنواخت $u = u(x, y)$ در معادله پواسن زیر صدق می‌کند:

$$\frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} = \frac{-q}{K}, \quad 0 < x < 6, \quad 0 < y < 5,$$

که در آن K ، یعنی قابلیت انتقال گرما، ۱.۰۴ ثانیه \times درجه \times سانتیمتر/کالری است. دمای $u(x, y)$ را با استفاده از الگوریتم ۱۰.۱۱، با $h = .4$ و $k = 1/3$ ، تقریب کنید.

۳.۱۱ معادلات دیفرانسیل جزئی سهموی

معادله دیفرانسیل جزئی سهموی که بررسی خواهیم کرد معادله گرما یا انتشار است:

$$(15.11) \quad \frac{\partial^2}{\partial t} u(x, t) = x^2 \frac{\partial^2}{\partial x^2} u(x, t), \quad 0 < x < l, \quad t > 0,$$

مفید به شرایط

$$u(0, t) = 0, \quad u(l, t) = 0, \quad t > 0,$$

$$u(x, 0) = f(x), \quad 0 \leq x \leq l.$$

روشی که برای تقریب جواب این مسئله بکار می‌بریم شامل تفاضلات متناهی است و شبیه‌روشی است که در بخش ۲۰.۱۱ بکار رفت. ابتدا دو ثابت شبکه‌ای h و k را، با این شرط که $m = 1/h$ صحیح باشد، انتخاب می‌کنیم. نقاط شبکه‌ای برای این وضعیت (x_i, t_j) هستند، که در آن به ازای $x_i = ih, i = 0, 1, \dots, m$ و به ازای $t_j = jk, j = 0, 1, \dots, m$ ، برای تشکیل خارج قسمت تفاضلی روش تفاضلی را با استفاده از سری تیلور در t ، جهت تشکیل خارج قسمت تفاضلی

$$(16.11) \quad \frac{\partial u}{\partial t}(x_i, t_j) = \frac{u(x_i, t_j + k) - u(x_i, t_j)}{k} - \frac{k}{2} \frac{\partial^2 u}{\partial t^2}(x_i, t_j + \theta_j k),$$

به ازای θ ای که $0 < \theta_j < 1$ و سری تیلور در x ، جهت تشکیل خارج قسمت تفاضلی

$$(17.11) \quad \frac{\partial^2 u}{\partial x^2}(x_i, t_j) = \frac{u(x_i + h, t_j) - 2u(x_i, t_j) + u(x_i - h, t_j)}{h^2} - \frac{h^2}{12} \frac{\partial^4 u}{\partial x^4}(x_i + \phi_i h, t_j),$$

که در آن $-1 < \phi_i < 1$ ، بدست می‌آوریم.

معادله دیفرانسیل جزئی (۱۵.۱۱) نتیجه می‌دهد که، در نقاط شبکه (x_i, t_j) ،

به ازای هر $i = 1, 2, \dots, m-1$ و $j = 1, 2, \dots$

$$\frac{\partial u}{\partial t}(x_i, t_j) - x^2 \frac{\partial^2 u}{\partial x^2}(x_i, t_j) = 0:$$

بنابراین، روش تفاضلی که از خارج قسمت‌های تفاضلی (۱۶.۱۱) و (۱۷.۱۱) استفاده می‌کند عبارت است از

$$(18.11) \quad \frac{w_{i,j+1} - w_{i,j}}{k} - x^2 \frac{w_{i+1,j} - 2w_{i,j} + w_{i-1,j}}{h^2} = 0,$$

که در آن مقدار w_{ij} مقدار $u(x_i, t_j)$ را تقریب می‌کند.

برای این معادله تفاضلی خطای برشی موضعی عبارت است از

به ازای θ_j و ϕ_i ای که $0 < \theta_j < 1$ و $-1 < \phi_i < 1$ ،

$$(19.11) \quad \tau_{i,j} = \frac{k}{2} \frac{\partial^2 u}{\partial t^2}(x_i, t_j + \theta_j k) - x^2 \frac{h^2}{12} \frac{\partial^4 u}{\partial x^4}(x_i + \phi_i h, t_j).$$

اگر معادله (۱۸.۱۱) نسبت به $w_{i,j+1}$ حل شود، به ازای هر $i = 1, 2, \dots, (m-1)$

و $j = 1, 2, \dots$

$$(20.11) \quad w_{i,j+1} = \left(1 - \frac{2x^2k}{h^2}\right)w_{i,j} + x^2 \frac{k}{h^2} (w_{i+1,j} + w_{i-1,j})$$

چون شرط اولیه $u(x, 0) = f(x)$ ، به ازای هر $0 \leq x \leq l$ ، نتیجه می‌دهد که به ازای هر $w_{i,0} = f(x_i)$ ، $i = 0, 1, \dots, m$ ، این مقادیر را می‌توان در معادله (۲۰.۱۱) بکار برده مقدار $w_{i,1}$ را، به ازای هر $i = 1, 2, \dots, (m-1)$ ، بدست آورد. با استفاده از شرط اضافی $u(0, t) = 0$ و $u(l, t) = 0$ نتیجه می‌شود که $w_{0,1} = w_{m,1} = 0$ ؛ تمام درایه‌ها به شکل $w_{i,1}$ را می‌توان تعیین کرد. اگر این روند یکبار دیگر اعمال شود، تمام تقریبهای $w_{i,1}$ معلوم می‌شوند، مقادیر $w_{i,2}, w_{i,3}, \dots, w_{i,m-1}$ را می‌توان به طریقی مشابه بدست آورد.

طبیعت صریح روش تفاضلی بیان شده در (۲۰.۱۱) نتیجه می‌دهد که ماتریس $(m-1) \times (m-1)$ مربوط به این دستگاه می‌تواند به شکل سه قطری نوشته شود:

$$A = \begin{bmatrix} (1-2\lambda) & \lambda & & & 0 \\ \lambda & (1-2\lambda) & & & 0 \\ 0 & & \ddots & & \\ 0 & & & \ddots & \\ 0 & & & & \lambda & (1-2\lambda) \end{bmatrix}$$

که در آن $\lambda = x^2(k/h^2)$. هرگاه قرار دهیم

$$w^{(j)} = \begin{bmatrix} w_{1,j} \\ w_{2,j} \\ \vdots \\ w_{m-1,j} \end{bmatrix}, \quad j = 1, 2, \dots$$

به ازای هر $j = 1, 2, \dots$

$$w^{(0)} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_{m-1}) \end{bmatrix},$$

نگاه جواب تقریبی چنین داده می‌شود:

$$w^{(j)} = Aw^{(j-1)}, \quad j = 1, 2, \dots$$

به ازای هر $j = 1, 2, \dots$

این روش تفاضلی به روش تفاضلی پیشرو معروف است ، و هرگاه جواب معادله دیفرانسیل جزئی دارای چهار مشتق جزئی پیوسته نسبت به x و دو مشتق جزئی پیوسته نسبت به t باشد ، آنگاه (۱۹.۱۱) نتیجه می‌دهد که روش از مرتبه $O(k + h^2)$ است .

مثال ۱. معادله گرما

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 1, \quad 0 < t,$$

را با شرایط مرزی

$$u(0, t) = u(1, t) = 0, \quad 0 < t,$$

و شرایط اولیه

$$u(x, 0) = \sin(\pi x), \quad 0 \leq x \leq 1.$$

در نظر می‌گیریم . به سادگی تحقیق می‌شود که جواب این مسئله عبارت است از

$$u(x, t) = e^{-\pi^2 t} \sin(\pi x).$$

جواب در $t = .5$ با استفاده از روش تفاضلی پیشرو ، ابتدا با $h = .1$ ، $k = .0005$ ، و $\lambda = .05$ و بعد با $h = .1$ ، $k = .01$ ، و $\lambda = 1$ ، تقریب شده است . نتایج در جدول ۲.۱۱ ارائه شده است .

x_i	$u(x_i, .5)$	$w_{i, 1000}$ $k = .005$	$ u(x_i, .5) - w_{i, 1000} $	$w_{i, 50}$ $k = .01$	$ u(x_i, .5) - w_{i, 50} $
0	0	0	-	0	-
.1	.00222241	0.0228652	6.411×10^{-5}	8.19876×10^{-7}	8.20×10^{-7}
.2	.00422728	.00434922	1.219×10^{-4}	-1.55719×10^{-6}	1.557×10^{-6}
.3	.00581836	.00598619	1.678×10^{-4}	2.13833×10^{-6}	2.138×10^{-6}
.4	.00683989	.00703719	1.973×10^{-4}	-2.50642×10^{-6}	2.506×10^{-6}
.5	.00719 88	.00739934	2.075×10^{-4}	2.62685×10^{-6}	2.627×10^{-6}
.6	.00683989	.00703719	1.973×10^{-4}	-2.49015×10^{-6}	2.490×10^{-6}
.7	.00581836	.00598619	1.678×10^{-4}	2.11200×10^{-6}	2.112×10^{-6}
.8	.00422728	.00434922	1.219×10^{-4}	-1.53086×10^{-6}	1.531×10^{-6}
.9	.00222241	.00228652	6.511×10^{-5}	8.03604×10^{-7}	8.036×10^{-7}
1.0	0	0	-	0	-

جدول ۲.۱۱

با وجود اینکه خطای برشی موضعی از مرتبه $O(k + h^2)$ قابل انتظار است، یقیناً در مورد مثال ما، وقتی $h = 0.1$ و $k = 0.1$ ، چنین نیست. برای توضیح این مشکل باید به پایداری روش تفاضلی پیشرو توجه کرد. اگر خطای $e^{(0)} = (e_1^{(0)}, e_2^{(0)}, \dots, e_m^{(0)})^T$ در ارائه داده اولیه $w^{(0)} = (f(x_1), f(x_2), \dots, f(x_{m-1}))^T$ انجام شده باشد یا، بخاطر آن، در هر مرحله خاص (انتخاب گام اولیه صرفاً "برای راحتی است)، خطایی چون $Ae^{(0)}$ در $w^{(1)}$ انتشار می یابد، زیرا

$$w^{(1)} = A(w^{(0)} + e^{(0)}) = Aw^{(0)} + Ae^{(0)}.$$

این فرایند با این فرض که خطاهای دیگری وارد نمی شوند ادامه خواهد یافت، در n مین مرحله خطا در $w^{(n)}$ برابر $A^n e^{(0)}$ خواهد بود. در نتیجه، روش پایدار خواهد بود اگر و فقط اگر این خطاها، وقتی n زیاد می شود، افزایش پیدا نکنند - یعنی، اگر و فقط اگر به ازای تمام n ها، $\|A^n e^{(0)}\| \leq \|e^{(0)}\|$. اما این نتیجه می دهد که $\|A^n\| \leq 1$ ، شرطی که، بنا بر قضیه ۱۵.۸، نتیجه می دهد که شعاع طیفی $\rho(A^n) = (\rho(A))^n \leq 1$ بنا بر این، روش تفاضلی پیشرو فقط اگر $\rho(A) \leq 1$ پایدار است. برای محاسبه شعاع طیفی A ، ابتدا توجه می کنیم که (ر. ک. تمرین ۵) مقادیر ویژه A عبارتند از:

$$\mu_i = 1 - 4\lambda \sin^2\left(\frac{i\pi}{2m}\right), \quad i = 1, 2, \dots, (m-1)$$

شرط پایداری با تعیین اینکه آیا

$$\rho(A) = \max_{1 \leq i \leq m-1} 1 - 4\lambda \sin^2\left(\frac{i\pi}{2m}\right) \leq 1,$$

تحویل می شود که چنین ساده می شود:

$$0 \leq \lambda \sin^2\left(\frac{i\pi}{2m}\right) \leq \frac{1}{2}, \quad i = 1, 2, \dots, m-1$$

چون پایداری لازم دارد که این شرط نامساوی وقتی $h \rightarrow 0$ ، یا به طور معادل وقتی $m \rightarrow +\infty$ ، برقرار باشد، این حقیقت که

$$\lim_{m \rightarrow \infty} \sin^2\left(\frac{(m-1)\pi}{2m}\right) = 1$$

بدین معنی است که پایداری فقط اگر $0 \leq \lambda \leq \frac{1}{2}$ رخ می دهد. چون $\lambda = \tau^2(k h^2)$ یقیناً

مثبت است، این نامساوی اساساً لازم دارد که h و k چنان انتخاب شوند که

$$\alpha^2 \frac{k}{h^2} \leq \frac{1}{2}.$$

برای مثال ما این شرط، وقتی $h = 1$ و $k = 0.005$ ، برقرار بود؛ اما، وقتی k ، بدون افزایش متناظر h ، به 0.1 افزایش یافت، نسبت چنین شد

$$\frac{.01}{(.1)^2} = 1 > \frac{1}{2}$$

و مسائل ناپایداری ظاهر گشت. این روش را، به طور سازگار با جمله بندی فصل ۶، روش پایدار مشروط می نامیم و متذکر می شویم که روش با سرعت همگرایی $O(k + h^2)$ به جواب معادله (۱۵.۱۱) همگراست اگر

$$\alpha^2 \frac{k}{h^2} \leq \frac{1}{2}$$

و اگر شرایط پیوستگی لازم روی جواب برقرار باشند. (برای برهانی مفصل از این نتیجه، ر. ک. ایزاکسون و کلر [۳۹]، صفحات ۵۰۵-۵۰۲).

برای بدست آوردن روشی که پایدار غیرمشروط باشد یک روش تفاضلی ضمنی را بررسی خواهیم کرد که از کاربرد خارج قسمت تفاضلی پسر و برای $(\partial u / \partial t)(x_i, t_j)$ به شکل

$$\frac{\partial u}{\partial t}(x_i, t_j) = \frac{u(x_i, t_j) - u(x_i, t_{j-1})}{k} + \frac{k}{2} \frac{\partial^2 u}{\partial t^2}(x_i, t_j + \theta_j k),$$

که در آن $-1 < \theta_j < 0$ ، نتیجه می شود. جایگذاری این معادله، به همراه معادله (۱۷.۱۱) برای $(\partial^2 u / \partial x^2)$ ، در معادله دیفرانسیل جزئی می دهد: به ازای θ_j و ϕ_i که $-1 < \phi_i < 1$ و $-1 < \theta_j < 0$

$$\begin{aligned} & \frac{u(x_i, t_j) - u(x_i, t_{j-1})}{k} - \alpha^2 \frac{u(x_{i+1}, t_j) - 2u(x_i, t_j) + u(x_{i-1}, t_j))}{h^2} \\ & = -\frac{k}{2} \frac{\partial^2 u}{\partial t^2}(x_i, t_j + \theta_j k) - \frac{h^2}{12} \frac{\partial^4 u}{\partial x^4}(x_i + \phi_i h, t_j); \end{aligned}$$

و روش تفاضلی پسر و که نتیجه می شود عبارت است از

به ازای هر $i = 1, 2, \dots, m-1$ و $j = 1, 2, \dots$

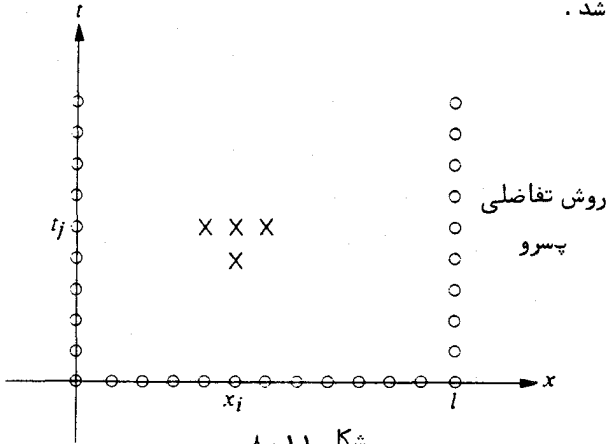
$$(21.11) \quad \frac{w_{i,j} - w_{i,j-1}}{k} - \alpha^2 \frac{w_{i+1,j} - 2w_{i,j} + w_{i-1,j}}{h^2} = 0,$$

این روش، در یک مرحله نوعی، شامل نقاط شبکه‌ای

$$(x_i, t_j), (x_i, t_{j-1}), (x_{i-1}, t_j), (x_{i+1}, t_j),$$

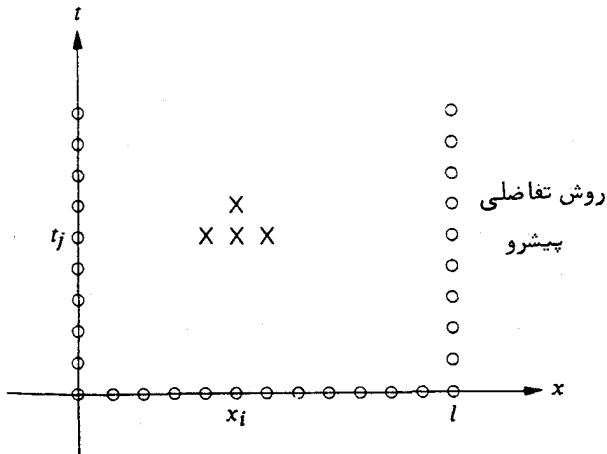
است یا، در شکل شبکه‌ای، شامل تقریبهایی در نقاط علامتگذاری شده با x ها، در شکل

۸.۱۱، می‌باشد.



شکل ۸.۱۱

چون شرایط مرزی و اولیه مربوط به مسئله اطلاعاتی در نقاط شبکه‌ای دایره زده شده بدست می‌دهند، از شکل پیدا است که از هیچ روند صریحی نمی‌توان برای حل معادله (۲۱.۱۱) استفاده کرد. بخاطر آورید که در روش تفاضلی پیشرو (ر.ک. شکل ۹.۱۱)



شکل ۹.۱۱

تقریبها در

$$(x_{i-1}, t_j), (x_i, t_j), (x_i, t_{j+1}), \text{ و } (x_{i+1}, t_j)$$

بکار می‌رفتند؛ لذا، یک روش صریح برای تعیین تقریبها، که متکی بر آگاهی از شرایط مرزی و اولیه باشد، موجود بود.

اگر مجدداً λ نمایانگر کمیت $\gamma^2(k, h^2)$ باشد، روش تفاضلی پسرو، به‌ازای هر

$$i = 1, 2, \dots, m-1 \text{ و } j = 1, 2, \dots, \text{ می‌شود}$$

$$(23.11) \quad (1 + 2\lambda)w_{i,j} - \lambda w_{i+1,j} - \lambda w_{i-1,j} = w_{i,j-1}$$

با استفاده از این مطلب که به‌ازای هر $i = 1, 2, \dots, m-1$ و $w_{i,0} = f(x_i)$ ، به‌ازای

$$\text{هر } j = 1, 2, \dots, w_{m,j} = w_{0,j} = 0, \text{ روش تفاضلی نمایش ماتریسی}$$

$$(23.11) \quad \begin{bmatrix} (1+2\lambda) & -\lambda & & 0 & \cdots & 0 \\ -\lambda & (1+2\lambda) & -\lambda & & & \\ & & & & & \\ 0 & & -\lambda & & & \\ & & & & & \\ & & 0 & & & 0 \\ & & & & -\lambda & \\ & & & & & (1+2\lambda) \end{bmatrix} \begin{bmatrix} w_{1,j} \\ w_{2,j} \\ \vdots \\ w_{m-1,j} \end{bmatrix} = \begin{bmatrix} w_{1,j-1} \\ w_{2,j-1} \\ \vdots \\ w_{m-1,j-1} \end{bmatrix}$$

با $Aw^{(j)} = w^{(j-1)}$ را، به‌ازای هر $j = 1, 2, \dots$ دارد. چون $\lambda > 0$ ، ماتریس A ، علاوه بر سه قطری و متقارن بودن، معین مثبت و قطر غالب است. بنابراین، می‌توانیم یا از تجزیه کروت برای دستگاههای خطی سه‌قطری (الگوریتم ۷.۷) یا از روش SOR (الگوریتم ۴.۸) برای حل این دستگاه استفاده کنیم. الگوریتم زیر (۲۳.۱۱) را با استفاده از تجزیه کروت، که روشی قابل قبول است مگر آنکه m بسیار بزرگ باشد، حل می‌کند. در این الگوریتم، به‌منظور توقف، فرض می‌کنیم کرانی برای ϵ داده شده است.

الگوریتم تفاضلی پسرو معادله ۲۰.۱۱ گرما

برای تقریب جواب معادله دیفرانسیل جزئی سهموی

$$\frac{\partial u}{\partial t} - \alpha^2 \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < l, \quad 0 < t < T,$$

تحت شرایط

$$u(0, t) = u(l, t) = 0, \quad 0 < t < T,$$

$$u(x, 0) = f(x), \quad 0 \leq x \leq l,$$

یک عدد صحیح $m > 0$ و یک گام زمانی $k > 0$ را طوری انتخاب می‌کنیم که برای عدد صحیحی چون $T = kN$ ، $N > 0$.

مرحله ۱. قرار می‌دهیم $h = l/m$ و، به‌ازای هر $i = 0, 1, \dots, m$ ، $x_i = ih$.

مرحله ۲. به‌ازای هر $j = 0, 1, \dots, N$ قرار می‌دهیم $t_j = jk$.

مرحله ۳. قرار می‌دهیم $\lambda = \alpha^2(k/h^2)$.

مرحله ۴. به‌ازای هر $i = 1, 2, \dots, m-1$ قرار می‌دهیم $w_{i,0} = f(x_i)$.

مرحله ۵. به‌ازای هر $j = 1, 2, \dots, N$ قرار می‌دهیم $w_{0,j} = w_{m,j} = 0$.

مرحله ۶. قرار می‌دهیم $l_1 = 1 + 2\lambda$.

$$u_1 = -\lambda/l_1.$$

مرحله ۷. قرار می‌دهیم $n = 2$.

مرحله ۸. قرار می‌دهیم $l_n = 1 + 2\lambda + \lambda\mu_{n-1}$.

$$u_n = -\lambda/l_n.$$

مرحله ۹. l را به n می‌افزاییم.

مرحله ۱۰. اگر $n < m-1$ ، به مرحله ۸ می‌رویم.

مرحله ۱۱. قرار می‌دهیم $l_{m-1} = 1 + 2\lambda + \lambda\mu_{m-2}$.

مرحله ۱۲. قرار می‌دهیم $j = 0$.

مرحله ۱۳. قرار می‌دهیم $z_1 = w_{1,j}/l_1$.

و به‌ازای هر $n = 2, \dots, (m-1)$ ، $z_n = [w_{n,j} + \lambda z_{n-1}]/l_n$.

مرحله ۱۴. قرار می‌دهیم $w_{m-1,j+1} = z_{m-1}$.

و به‌ازای هر $n = (m-2), (m-3), \dots, 1$ ، $w_{n,j+1} = z_n - u_n w_{n+1,j+1}$.

مرحله ۱۵. l را به j می‌افزاییم.

مرحله ۱۶. اگر $j \leq N-1$ ، به مرحله ۱۳ می‌رویم.

مرحله ۱۷. روند تمام است؛ $w_{i,j}$ ، به‌ازای هر $i = 0, 1, \dots, m$ و $j = 0, 1, \dots, N$.

$u(x_i, t_j)$ را تقریب می‌کند.

مثال ۲. روش تفاضلی پسرو (الگوریتم ۲.۱۱)، با $h = .1$ و $k = .01$ ، برای تقریب جواب معادله گرمای زیر، که در مثال ۱ بررسی شد،

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 1, \quad 0 < t,$$

تحت قیود

$$u(0, t) = u(1, t) = 0, \quad 0 < t,$$

و

$$u(x, 0) = \sin \pi x, \quad 0 \leq x \leq 1,$$

بکار خواهد رفت. برای نشان دادن پایداری غیر مشروط روش تفاضلی پسرو مجدداً

با $w_{i, .50}$ را با $u(x_i, .5)$ مقایسه می‌کنیم که در آن $i = 0, 1, \dots, 10$.

نتایج فهرست شده در جدول ۳.۱۱ باید با ستونهای پنجم و ششم جدول ۳.۱۱

مقایسه شوند.

x_i	$w_{i, .50}$	$u(x_i, .5)$	$ w_{i, .50} - u(x_i, .5) $
0	0	0	—
.1	.00289802	.00222241	6.756×10^{-4}
.2	.00551236	.00422728	1.285×10^{-3}
.3	.00758711	.00581836	1.769×10^{-3}
.4	.00891918	.00683989	2.079×10^{-3}
.5	.00937818	.00719188	2.186×10^{-3}
.6	.00891918	.00683989	2.079×10^{-3}
.7	.00758711	.00581836	1.769×10^{-3}
.8	.00551236	.00422728	1.285×10^{-3}
.9	.00289802	.00222241	6.756×10^{-4}
1.0	0	0	—

جدول ۳.۱۱

روش تفاضلی پسرو مشکلات پایداری روش تفاضلی پیشرو را ندارد. دلیل این مطلب

را می‌توان مجدداً " با تحلیل مقادیر ویژه ماتریس A دریافت. در این حالت (ر.ک.

تمرین ۶) مقادیر ویژه، به‌ازای هر $i = 1, 2, \dots, (m - 1)$ ، به شکل

$$\mu_i = 1 + 4\lambda \sin^2\left(\frac{i\pi}{2m}\right)$$

هستند، و چون $\lambda > 0$ ، به ازای تمام $i = 1, 2, \dots, (m-1)$ ، $\mu_i > 1$ ، این نتیجه می‌دهد (بنابر تمرین ۱۳ در بخش ۱۰.۸) که A^{-1} وجود دارد. خطای $e^{(0)}$ در داده‌های اولیه خطای $e^{(0)}$ را در مرحله n م تولید می‌کند. چون مقادیر ویژه A^{-1} معکوس مقادیر ویژه A هستند، شعاع طیفی A^{-1} به ۱ کرانداراست و این روش، مستقل از انتخاب $\lambda = \alpha^2(k/h^2)$ ، پایدار است. با اصطلاحات فصل ۶، روش تفاضلی را یک روش پایدار غیر مشروط می‌نامیم. خطای برشی موضعی برای این روش از مرتبه $O(k+h^2)$ است، بشرط آنکه جواب معادله دیفرانسیل در شرایط عادی مشتق‌پذیری صدق کند؛ و در این حالت روش با همین سرعت همگرایی به جواب معادله دیفرانسیل همگراست (ر.ک. ایزاکسون و کلر [۳۹]، صفحه ۵۰۸).

ضعف روش تفاضلی پسرو از این حقیقت ناشی می‌شود که خطای برشی موضعی دارای قسمتی با مرتبه $O(k)$ است، که نیازمند آن است که بازه‌های زمانی خیلی کوچکتر از بازه‌های فاصله‌ای باشند. آشکارا طرح روندی با خطای برشی موضعی از مرتبه $O(k^2+h^2)$ مطلوب می‌باشد. اولین مرحله در این جهت احتمالاً "کاربرد یک معادله تفاضلی برای $(\partial u/\partial t)(x, t)$ با خطای $O(k^2)$ به جای آنهایی است که قبلاً"، با خطای $O(k)$ ، استفاده کردیم. این را می‌توان با استفاده از سری تیلور در t برای تابع $u(x, t)$ در نقطه (x_i, t_j) ، و محاسبه آن در (x_i, t_{j+1}) و (x_i, t_{j-1}) جهت بدست آوردن فرمول تفاضل مرکزی

$$\frac{\partial u}{\partial t}(x_i, t_j) = \frac{u(x_i, t_{j+1}) - u(x_i, t_{j-1})}{2k} - \frac{k^2}{6} \frac{\partial^3 u(x_i, t_j + \theta_j k)}{\partial t^3},$$

که در آن $-1 < \theta_j < 1$ ، انجام داد. روش تفاضلی که از جایگذاری این و خارج قسمت تفاضلی معمول برای $(\partial^2 u/\partial x^2)$ ، یعنی معادله (۱۷.۱۱)، در معادله دیفرانسیل نتیجه می‌شود روش ریچاردسون نامیده شده و چنین ارائه می‌شود:

$$\frac{w_{i,j+1} - w_{i,j-1}}{2k} - \alpha^2 \frac{w_{i+1,j} - 2w_{i,j} + w_{i-1,j}}{h^2} = 0.$$

این روش حتماً "خطای برشی موضعی مرتبه $O(k^2+h^2)$ دارد، اما متأسفانه مسائل پایداری جدی نیز دارد (ر.ک. تمرین ۴).

می‌توان روش با ارزش‌تری با میانگین‌گیری از روش تفاضل پیشرو در مرحله j م در

t ، یعنی

$$\frac{w_{i,j+1} - w_{i,j}}{k} - \alpha^2 \frac{w_{i+1,j} - 2w_{i,j} + w_{i-1,j}}{h^2} = 0,$$

که خطای برشی موضعی

$$\tau_F = \frac{k}{2} \frac{\partial^2}{\partial t^2} u(x_i, t_j + \theta_j k) + O(h^2),$$

دارد و فرمول تفاضل پیشرو در مرحله $(j+1)$ م در t ، یعنی

$$\frac{w_{i,j+1} - w_{i,j}}{k} - \alpha^2 \frac{w_{i+1,j+1} - 2w_{i,j+1} + w_{i-1,j+1}}{h^2} = 0,$$

که خطای برشی موضعی

$$\tau_B = -\frac{k}{2} \frac{\partial^2}{\partial t^2} u(x_i, t_{j+1} + \theta_j k) + O(h^2)$$

دارد بدست آورد. چون $0 < \theta_j < 1$ درحالی‌که $-1 < \theta_j < 0$ ، موجه است فرض کنیم

$t_j + \theta_j k \approx t_{j-1} + \theta_j k$ ، درواقع، هرگاه این مقادیر مساوی باشند، آنگاه روش تفاضلی

میانگین شده، یعنی

$$\frac{w_{i,j+1} - w_{i,j}}{k} - \frac{\alpha^2}{2} \left[\frac{w_{i+1,j} - 2w_{i,j} + w_{i-1,j}}{h^2} + \frac{w_{i+1,j+1} - 2w_{i,j+1} + w_{i-1,j+1}}{h^2} \right] = 0,$$

خطای برشی موضعی از مرتبه $O(k^2 + h^2)$ خواهد داشت، البته شرایط مشتق‌پذیری معمول

برقرارند.

این روش به‌عنوان روش گرانگ^۱ - نیگلسون^۲ شناخته شده است، و می‌تواند به شکل

ماتریسی $Aw^{(j+1)} = Bw^{(j)}$ ، به‌زای هر $j = 0, 1, 2, \dots$ ، نمایش داده شود که در آن

$$\lambda = \alpha^2 \frac{k}{h^2}, \quad w^{(j)} = (w_{1,j}, w_{2,j}, \dots, w_{m-1,j})^t,$$

و ماتریسهای A و B چنین داده می‌شوند:

$$A = \begin{bmatrix} (1 + \lambda) & -\frac{\lambda}{2} & 0 & 0 \\ -\frac{\lambda}{2} & (1 + \lambda) & -\frac{\lambda}{2} & 0 \\ 0 & -\frac{\lambda}{2} & 0 & -\frac{\lambda}{2} \\ 0 & 0 & -\frac{\lambda}{2} & (1 + \lambda) \end{bmatrix}$$

$$B = \begin{bmatrix} (1 - \lambda) & \frac{\lambda}{2} & 0 & 0 \\ \frac{\lambda}{2} & (1 - \lambda) & \frac{\lambda}{2} & 0 \\ 0 & \frac{\lambda}{2} & 0 & \frac{\lambda}{2} \\ 0 & 0 & \frac{\lambda}{2} & (1 - \lambda) \end{bmatrix}$$

چون A یک ماتریس معین مثبت، متقارن، اکیدا "قطر غالب" و سه‌قطری است، نامنفرد است، و یکی از روشهای تحویل کروت برای دستگاه خطی سه‌قطری (الگوریتم ۷.۷) یا روش SOR (الگوریتم ۴.۸) را می‌توان جهت تعیین $w^{(j+1)}$ از $w^{(j)}$ ، به‌ازای هر $j = 0, 1, 2, \dots$ بکار برد.

الگوریتم زیر تحویل کروت به تکنیک کرانک - نیکلسون را تشکیل می‌دهد. همانند الگوریتم ۲.۱۱، برای تعیین روند توقف، باید طولی متناهی برای بازه زمانی مشخص شود.

الگوریتم کرانک - نیکلسون ۳.۱۱

برای تقریب جواب معادله دیفرانسیل جزئی سهمی

$$\frac{\partial u}{\partial t} - x^2 \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < l, \quad 0 < t < T,$$

تحت شرایط

$$u(0, t) = u(l, t) = 0, \quad 0 < t < T,$$

$$u(x, 0) = f(x), \quad 0 \leq x \leq l.$$

عدد صحیح $m > 0$ و گام زمانی k را چنان انتخاب می‌کنیم، که به‌ازای عدد صحیح

$$\cdot T = kN, \quad N > 0$$

مرحله ۱. قرار می‌دهیم $h = l/m$ و، به‌ازای هر $i = 0, 1, \dots, m$ ، $x_i = ih$.

مرحله ۲. به‌ازای هر $j = 0, 1, \dots, N$ قرار می‌دهیم $t_j = jk$.

مرحله ۳. قرار می‌دهیم $\lambda = x^2(k/h^2)$.

مرحله ۴. به‌ازای هر $i = 1, 2, \dots, m-1$ قرار می‌دهیم $w_{i,0} = f(x_i)$.

مرحله ۵. به‌ازای هر $j = 1, 2, \dots, N$ قرار می‌دهیم $w_{0,j} = w_{m,j} = 0$.

مرحله ۶. قرار می‌دهیم $l_1 = 1 + \lambda$.

$$u_1 = -\lambda/(2l_1).$$

مرحله ۷. قرار می‌دهیم $n = 2$.

مرحله ۸. قرار می‌دهیم

$$l_n = 1 + \lambda + \lambda \frac{u_{n-1}}{2},$$

$$u_n = -\frac{\lambda}{2l_n}.$$

مرحله ۹. ۱ را به n می‌افزاییم.

مرحله ۱۰. اگر $n < m-1$ ، به مرحله ۸ می‌رویم.

مرحله ۱۱. قرار می‌دهیم $l_{m-1} = 1 + \lambda + \lambda \frac{u_{m-2}}{2}$.

مرحله ۱۲. قرار می‌دهیم $j = 0$.

مرحله ۱۳. قرار می‌دهیم $z_1 = \left[(1 - \lambda)w_{1,j} + \frac{\lambda}{2} w_{2,j} \right] / l_1$ و به‌ازای هر

$$n = 2, 3, \dots, m-1$$

$$z_n = \left[(1 - \lambda)w_{n,j} + \frac{\lambda}{2} w_{n+1,j} + \frac{\lambda}{2} w_{n-1,j} + \frac{\lambda}{2} z_{n-1} \right] / l_n$$

مرحله ۱۴. قرار می‌دهیم $w_{m-1,j+1} = z_{m-1}$ ، و به‌ازای هر $1, 3, \dots, m-2, m$

$$w_{n,j+1} = z_n - u_n w_{n+1,j+1}$$

مرحله ۱۵. ۱ را به z می‌افزاییم.

مرحله ۱۶. اگر $z \leq N-1$ ، به مرحله ۱۳ می‌رویم.

مرحله ۱۷. روند تمام است، $w_{i,j}$ ، به‌ازای هر $i = 0, 1, \dots, m$ و $j = 0, 1, \dots, N$ ، $u(x_i, t_j)$ را تقریب می‌کند.

روش کرانک - نیکلسون. پایدار غیر مشروط است و مرتبه همگرایی $O(k^2 + h^2)$ دارد.

اثبات این مطالب را می‌توان در کتاب ایزاکسون و کلر [۳۹]، صفحات ۵۱۲-۵۰۸، که قبلاً ذکر شد، یافت.

مثال ۳. روش کرانک - نیکلسون را می‌توان برای تقریب جواب مسطه مثالهای ۱ و ۲ متشکل از معادله

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 1, \quad 0 < t,$$

تحت شرایط

$$u(0, t) = u(1, t) = 0, \quad 0 < t,$$

و

$$u(x, 0) = \sin(\pi x), \quad 0 \leq x \leq 1.$$

بکار برد. انتخابهای $m = 10$ ، $h = .1$ ، $N = 50$ ، $k = .01$ ، و $\lambda = 1$ در الگوریتم ۳.۱۱ بکار برده شده‌اند، همانگونه که در مثالهای قبل بودند. نتایج مندرج در جدول ۴.۱۱

x_i	$w_{i,50}$	$u(x_i, .5)$	$ w_{i,50} - u(x_i, .5) $
0	0	0	—
.1	.00230512	.00222241	8.271×10^{-5}
.2	.00438461	.00422728	1.573×10^{-4}
.3	.00603489	.00581836	2.165×10^{-4}
.4	.00709444	.00683989	2.546×10^{-4}
.5	.00745954	.00719188	2.677×10^{-4}
.6	.00709444	.00683989	2.546×10^{-4}
.7	.00603489	.00581836	2.165×10^{-4}
.8	.00438461	.00422728	1.573×10^{-4}
.9	.00230512	.00222241	8.271×10^{-5}
1.0	0	0	—

افزایش در دقت را برای روش کرانک - نیکلسون در مقایسه با روش تفاضلی پسرو، بهترین دو تکنیکی که قبلاً مورد بحث قرار گرفتند، نشان می دهد.

مجموعه تمرینات ۳.۱۱

۱. جوابهای معادلات دیفرانسیل جزئی سهموی زیر را با استفاده از روش تفاضلی پیشرو تقریب کنید.

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 2, \quad 0 < t: \quad (T)$$

$$u(0, t) = u(2, t) = 0, \quad 0 < t,$$

$$u(x, 0) = \sin 2\pi x, \quad 0 \leq x \leq 2.$$

از $h = .01$ و $k = .01$ استفاده کنید و جوابهای خود را در $t = .5$ با جواب واقعی مقایسه کنید. سپس از $h = .1$ و $k = .005$ استفاده و جوابها را مقایسه کنید

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 2, \quad 0 < t: \quad (B)$$

$$u(0, t) = u(2, t) = 0, \quad 0 < t,$$

$$u(x, 0) = x(2 - x), \quad 0 \leq x \leq 2.$$

از $h = .1$ استفاده کنید و مقدار مناسبی برای k انتخاب کنید. جوابهای خود را با پنج جمله اول جواب سری فوریه در $t = .5$ مقایسه کنید.

۲. با استفاده از الگوریتم تفاضلی پسرو ۳.۱۱، تمرین ۱ را تکرار کنید.

۳. با استفاده از الگوریتم کرانک - نیکلسون، تمرین ۱ را تکرار کنید.

۴. با استفاده از روش ریچاردسون، تمرین ۱ را تکرار کنید.

۵. نشان دهید که مقادیر ویژه ماتریس سه قطری $(m-1) \times (m-1)$ ، A که به وسیله

$$a_{ij} = \begin{cases} \lambda & , j = i - 1 \text{ یا } j = i + 1 \\ (1 - 2\lambda) & , j = i \\ 0 & \text{در غیر این صورت} \end{cases}$$

داده شده است عبارتند از

$$\mu_i = 1 - 4\lambda \sin^2\left(\frac{i\pi}{2m}\right), \quad i = 1, 2, \dots, m-1$$

[راهنمایی . شکل بردارهای ویژه را بررسی کنید .]
 ۶ . نشان دهید که ماتریس سه قطری $(m-1) \times (m-1)$ ، A که به وسیله

$$a_{ij} = \begin{cases} -\lambda & , j = i - 1 \text{ یا } j = i + 1 \\ 1 + 2\lambda & j = i \\ 0 & , \text{در غیر این صورت} \end{cases}$$

داده شده است ، که در آن $\lambda > 0$ ، معین مثبت و قطر غالب بوده و دارای مقادیر ویژه زیر است .

$$\mu_i = 1 + 4\lambda \sin^2\left(\frac{i\pi}{2m}\right) , i = 1, 2, \dots, m-1$$

[راهنمایی . شکل بردارهای ویژه را بررسی کنید .]
 ۷ . الگوریتمهای ۲۰۱۱ و ۳۰۱۱ را تغییر دهید تا معادله دیفرانسیل جزئی سهموی زیر را دربرگیرد :

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = F(x), \quad 0 < x < l, \quad 0 < t,$$

$$u(0, t) = u(l, t) = 0, \quad 0 < t,$$

$$u(x, 0) = f(x), \quad 0 \leq x \leq l.$$

۸ . از نتایج تمرین ۷ استفاده کرده جواب

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 2, \quad 0 < x < 1, \quad 0 < t$$

$$u(0, t) = u(1, t) = 0, \quad 0 < t,$$

$$u(x, 0) = \sin \pi x + x(1-x),$$

را با $h = .1$ و $k = .01$ تقریب کنید و جوابهای خود را با جواب واقعی

$$u(x, t) = e^{-\pi^2 t} \sin \pi x + x(1-x)$$

۹ . الگوریتمهای ۲۰۱۱ و ۳۰۱۱ را تغییر دهید تا معادله دیفرانسیل جزئی

$$\frac{\partial u}{\partial t} - x^2 \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < l, \quad 0 < t,$$

$$u(0, t) = \phi(t), \quad u(l, t) = \Psi(t), \quad 0 < t,$$

$$u(x, 0) = f(x), \quad 0 \leq x \leq l,$$

را، که در آن $f(0) = \phi(0)$ و $f(l) = \Psi(0)$ ، دربرگیرد.

۱۰. دمای $u(x, t)$ در یک میله نازک طویل با مقطع عرضی ثابت و ماده هدایت کننده همگن از معادله گرمای یک بعدی تبعیت می کند. اگر در ماده گرما تولید شود، مثلاً " به وسیله مقاومت در مقابل جریان یا عکس العمل هسته ای، معادله گرما چنین می شود:

$$\frac{\partial^2 u(x, t)}{\partial x^2} + \frac{Kr}{\rho C} = K \frac{\partial u(x, t)}{\partial t}, \quad 0 < x < l, \quad 0 < t,$$

که در آن l طول، ρ چگالی، C گرمای مشخص، و K انتشار گرمایی میله است. تابع $r = r(x, t, u)$ گرمای تولید شده در هر واحد حجم را نمایش می دهد. فرض کنیم

$$l = 1.5 \text{ cm},$$

$$\rho = 10.6 \text{ g/cm}^3,$$

$$K = 1.04 \text{ cal/cm} \cdot \text{deg} \cdot \text{sec},$$

$$C = .056 \text{ cal/g} \cdot \text{deg},$$

و

$$r(x, t, u) = 5.0 \text{ cal/sec} \cdot \text{cm}^3.$$

هرگاه انتهای میله در صفر درجه سانتیگراد نگهداشته شده باشد، آنگاه

$$u(0, t) = u(l, t) = 0, \quad t > 0.$$

فرض کنید توزیع دمای اولیه با

$$u(x, 0) = \sin \frac{\pi x}{l}, \quad 0 \leq x \leq l$$

داده شده باشد. نتایج تمرین ۷ را بکار برده توزیع دما را با $h = .15$ و $K = .0225$ تقریب کنید.

۱۱. وی. ساگار^۱ و دی. ج. پین^۲، در تحلیل رابطه های فشار - کشش و خواص مواد یک سیلندر، متناوباً " تحت گرم شدن و خنک شدن، معادله

$$\frac{\partial^2 T}{\partial r^2} + \frac{1}{r} \frac{\partial T}{\partial r} = \frac{1}{4K} \frac{\partial T}{\partial t}, \quad \frac{1}{2} < r < 1, \quad 0 < T.$$

را در نظر گرفتند که در آن $T = T(r, t)$ دما، r فاصله شعاعی از مرکز سیلندر، t

زمان، و K یک ضریب انتشار است.

(\bar{T}) تقریبهایی به $T(r, 10)$ برای یک سیلندر با شعاع خروجی یک پیدا کنید، شرایط اولیه و مرزی چنین داده شده اند:

$$T(1, t) = 100 + 40t, \quad 0 \leq t \leq 10,$$

$$T(\frac{1}{2}, t) = t, \quad 0 \leq t \leq 10,$$

$$T(r, 0) = 200(r - .5), \quad .5 \leq r \leq 1,$$

از روش تغییر یافته تفاضلی پسر ویا $K \equiv .1$ ، $k = .5$ ، و $h = \Delta r = .1$ استفاده کنید.

(ب) با استفاده از توزیع دمای قسمت (\bar{T})، کشش را با تقریب انتگرال

$$I \equiv \int_{.5}^1 x T(r, t) r dr,$$

حساب کنید، که در آن $x = 10.7$ و $t = 10$. از روش دوزنقه‌ای مرکب که در قضیه ۶.۵ (ص ۲۷۷) داده شده است با $m = 5$ استفاده کنید.

۱۲. معادله جریان یک بعدی، تک فاز، و کمی فشارپذیر در تهریه مخزن مواد نفتی، برای $0 < x < 1000$ و $0 < t$ ، با معادله زیر داده می شود:

$$\frac{\phi \mu C}{K} \frac{\partial p(x, t)}{\partial t} = \frac{\partial^2 p(x, t)}{\partial x^2} - \begin{cases} 0, & x \neq 500 \\ 1000, & x = 500 \end{cases}$$

که در آن فرض شده است که خلل و فرج وسایل و مخزن همگن هستند، مایع ایده‌آل است، و اثرات جاذبه قابل اغماض می باشد. نمادهایی که تعریف شده اند عبارتند از x : برای نمایش فاصله (به فوت)، t : زمان (به روز)، p : فشار (به پوند در هر اینچ مربع)، ϕ : ثابت بدون بعد، μ : چسبندگی (به سانتی پواز)، K : نفوذپذیری وسایل (به میلی داری)، و C : قابلیت تراکم (به $^{-1}$ پوند در هر اینچ مربع). با فرض اینکه $\phi \mu C / K = .00004$ (فوت)² روز و α و اینکه شرایط زیر برقرارند:

$$p(x, 0) = 2.5 \times 10^7, \quad 0 \leq x \leq 1000,$$

$$\frac{\partial p(0, t)}{\partial Kx} = \frac{\partial p(1000, t)}{\partial x} = 0, \quad 0 < t,$$

فشار p را در $t = 5$ ، با استفاده از روش کرانک - نیکلسون، با $k = \Delta t = .5$ و $h = \Delta x = 100$ پیدا کنید.

۴.۱۱ معادلات دیفرانسیل جزئی هذلولوی

در این بخش حل عددی معادله موج، مثالی از یک معادله دیفرانسیل جزئی هذلولوی، را بررسی می‌کنیم. معادله موج با معادله دیفرانسیل

$$(۲۴.۱۱) \quad \frac{\partial^2 u}{\partial t^2}(x, t) - \alpha^2 \frac{\partial^2 u}{\partial x^2}(x, t) = 0, \quad 0 < x < l, \quad t > 0,$$

تحت شرایط

$$\begin{aligned} u(0, t) &= u(l, t) = 0, \quad t > 0, \\ u(x, 0) &= f(x), \quad 0 \leq x \leq l, \end{aligned}$$

و

$$\frac{\partial u}{\partial t}(x, 0) = g(x), \quad 0 \leq x \leq l,$$

داده می‌شود که در آن x یک ثابت است. برای بدست آوردن روش تفاضل متناهی، یک عدد صحیح $m > 0$ و اندازه گام زمانی $k > 0$ را انتخاب می‌کنیم. با $h = l/m$ نقاط شبکه‌ای (x_i, t_j) با

$$x_i = ih, \quad i = 0, 1, \dots, m$$

و

$$t_j = jk, \quad j = 0, 1, \dots$$

تعریف می‌شوند. در هر نقطه شبکه‌ای داخلی (x_i, t_j) ، معادله موج می‌شود

$$(۲۵.۱۱) \quad \frac{\partial^2 u}{\partial t^2}(x_i, t_j) - \alpha^2 \frac{\partial^2 u}{\partial x^2}(x_i, t_j) = 0.$$

روش تفاضلی با استفاده از خارج قسمت تفاضل مرکزی برای مشتقات جزئی مرتبه دوم که با فرمولهای زیر داده می‌شوند بدست می‌آید.

$$(۲۶.۱۱) \quad \frac{\partial^2 u}{\partial t^2}(x_i, t_j) = \frac{u(x_i, t_{j+1}) - 2u(x_i, t_j) + u(x_i, t_{j-1}))}{k^2} - \frac{k^2}{12} \frac{\partial^4 u(x_i, t_j + \theta_j k)}{\partial t^4},$$

که در آن $-1 < \theta_j < 1$ ، و

$$(۲۷.۱۱) \quad \frac{\partial^2 u}{\partial x^2}(x_i, t_j) = \frac{u(x_{i+1}, t_j) - 2u(x_i, t_j) + u(x_{i-1}, t_j))}{h^2} - \frac{h^2}{12} \frac{\partial^4 u(x_i + \phi_i h, t_j)}{\partial x^4}$$

ر. آن $-1 < \phi_i < 1$. جایگذاری اینها در (۲۵.۱۱) می‌دهد

$$(28.11) \quad \frac{u(x_i, t_{j+1}) - 2u(x_i, t_j) + u(x_i, t_{j-1}))}{k^2} - \alpha^2 \frac{u(x_{i+1}, t_j) - 2u(x_i, t_j) + u(x_{i-1}, t_j))}{h^2} = \frac{1}{12} \left[k^2 \frac{\partial^4 u(x_i, t_j + \theta_j k)}{\partial t^4} - \alpha^2 h^2 \frac{\partial^4 u(x_i + \phi_i h, t_j)}{\partial x^4} \right]$$

و اغماض خطای برشی

$$(29.11) \quad \tau_{i,j} = \frac{1}{12} \left[k^2 \frac{\partial^4 u}{\partial t^4}(x_i, t_j + \theta_j k) - \alpha^2 h^2 \frac{\partial^4 u}{\partial x^4}(x_i + \phi_i h, t_j) \right]$$

به معادله تفاضلی زیر منجر می شود:

$$(30.11) \quad \frac{w_{i,j+1} - 2w_{i,j} + w_{i,j-1}}{k^2} - \alpha^2 \frac{w_{i+1,j} - 2w_{i,j} + w_{i-1,j}}{h^2} = 0.$$

با فرض استفاده از λ جهت نمایش $(\alpha k, h)$ می توان معادله تفاضلی را به صورت زیر نوشت:

$$w_{i,j+1} - 2w_{i,j} + w_{i,j-1} - \lambda^2 w_{i+1,j} + 2\lambda^2 w_{i,j} - \lambda^2 w_{i-1,j} = 0$$

و آن را نسبت به $w_{i,j+1}$ ، یعنی پیشرفته ترین گام زمانی تقریب، حل کرده بدست آورد

$$(31.11) \quad w_{i,j+1} = 2(1 - \lambda^2)w_{i,j} + \lambda^2(w_{i+1,j} + w_{i-1,j}) - w_{i,j-1}.$$

این معادله به ازای هر $i = 1, 2, \dots, (m-1)$ و $j = 1, 2, \dots$ برقرار است. شرایط مرزی

می دهد

$$(32.11) \quad \cdot w_{0,j} = w_{m,j} = 0, \quad j = 1, 2, 3, \dots$$

و شرط اولیه نتیجه می دهد که

$$(33.11) \quad \cdot w_{i,0} = f(x_i), \quad i = 1, 2, \dots, m-1$$

نوشتن این مجموعه معادلات به شکل ماتریسی نتیجه می دهد که

$$(34.11) \quad \begin{bmatrix} w_{1,j+1} \\ w_{2,j+1} \\ \vdots \\ w_{m-1,j+1} \end{bmatrix} = \begin{bmatrix} 2(1-\lambda^2) & \lambda^2 & 0 & \dots & 0 \\ \lambda^2 & 2(1-\lambda^2) & \lambda^2 & \dots & 0 \\ 0 & \lambda^2 & 2(1-\lambda^2) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \lambda^2 & 0 & \dots & 2(1-\lambda^2) \end{bmatrix} \begin{bmatrix} w_{1,j} \\ w_{2,j} \\ \vdots \\ w_{m-1,j} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\times \begin{bmatrix} w_{1,j} \\ w_{2,j} \\ \vdots \\ w_{m-1,j} \end{bmatrix} - \begin{bmatrix} w_{1,j-1} \\ w_{2,j-1} \\ \vdots \\ w_{m-1,j-1} \end{bmatrix}$$

از (۳۱.۱۱) و (۳۲.۱۱) واضح است که $(j+1)$ مین گام زمانی نیازمند مقادیر گامهای زمانی j و $(j-1)$ م است. این مطلب مشکل آغازی جزئی تولید می کند، که در آن مقادیر برای $j=0$ با (۳۳.۱۱) داده شده اند، اما مقادیر برای $j=1$ ، که در (۳۱.۱۱) جهت محاسبه $w_{i,2}$ لازمند، باید از شرط سرعت اولیه زیر بدست آیند:

$$\frac{\partial u}{\partial t}(x, 0) = g(x), \quad 0 \leq x \leq l.$$

اولین طریقه جایگزینی $(\partial u / \partial t)$ به وسیله تقریب تفاضلی پسرور زیر است:

$$\frac{\partial u}{\partial t}(x_i, 0) = \frac{u(x_i, t_1) - u(x_i, 0)}{k} - \frac{k}{2} \frac{\partial^2 u}{\partial t^2}(x_i, \theta_i k), \quad 0 < \theta_i < 1,$$

سپس $w_{i,1}$ با

$$(۳۵.۱۱) \quad w_{i,1} = w_{i,0} + kg(x_i), \quad i = 1, \dots, m-1$$

داده خواهد شد. اما معادله (۳۵.۱۱) تقریبی از مرتبه $O(k)$ به داده های اولیه می دهد، در حالی که خطای برشی موضعی برای (۳۱.۱۱) عبارت است از $O(k^2 + h^2)$. یک تقریب $O(k^2)$ برای مشتق را می توان به صورت زیر بدست آورد؛ چون به ازای

θ ای با $0 < \theta < 1$

$$\frac{u(x_i, t_1) - u(x_i, 0)}{k} = \frac{\partial u}{\partial t}(x_i, 0) + \frac{k}{2} \frac{\partial^2 u}{\partial t^2}(x_i, 0) + \frac{k^2}{6} \frac{\partial^3 u}{\partial t^3}(x_i, \theta k).$$

فرض می کنیم معادله موج نیز روی خط اولیه صدق کند؛ یعنی،

$$\frac{\partial^2 u}{\partial t^2}(x_i, 0) - x^2 \frac{\partial^2 u}{\partial x^2}(x_i, 0) = 0, \quad i = 0, 1, \dots, m$$

هرگاه f'' وجود داشته باشد، آنگاه

$$\frac{\partial^2 u}{\partial t^2}(x_i, 0) = x^2 \frac{\partial^2 u}{\partial x^2}(x_i, 0) = x^2 \frac{d^2}{dx^2} f(x_i) = x^2 f''(x_i).$$

اما می توانیم بنویسیم: اگر $f \in C^4[0, l]$

بمازای θ_i ای با $-1 < \theta_i < 1$ ،

$$f''(x_i) = \frac{f(x_{i+1}) - 2f(x_i) + f(x_{i-1}))}{h^2} - \frac{h^2}{12} f^{(4)}(x_i + \theta_i h)$$

که در این حالت نتیجه می دهد که تقریب چنین شود

$$\frac{u(x_i, t_1) - u(x_i, 0)}{k} = g(x_i) + \frac{kx^2}{2h^2} [f(x_{i+1}) - 2f(x_i) + f(x_{i-1}))] + O(k^2 + h^2k)$$

یا

$$\begin{aligned} u(x_i, t_1) &= u(x_i, 0) + kg(x_i) \\ &+ \frac{\lambda^2}{2} [f(x_{i+1}) - 2f(x_i) + f(x_{i-1}))] + O(k^3 + h^2k^2) \\ &= (1 - \lambda^2)f(x_i) + \frac{\lambda^2}{2} f(x_{i+1}) \\ &+ \frac{\lambda^2}{2} f(x_{i-1}) + kg(x_i) + O(k^3 + h^2k^2). \end{aligned}$$

بنابراین، معادله تفاضلی

$$(۳۶.۱۱) \quad w_{i,1} = (1 - \lambda^2)f(x_i) + \frac{\lambda^2}{2} f(x_{i+1}) + \frac{\lambda^2}{2} f(x_{i-1}) + kg(x_i).$$

را می توان برای تعیین $w_{i,1}$ ، بمازای هر $i = 1, 2, \dots, m - 1$ ، بکار برد .

الگوریتم زیر معادله (۳۶.۱۱) را جهت تقریب $w_{i,1}$ بکار می برد، گرچه از

(۳۵.۱۱) نیز می توان استفاده کرد . فرض شده که کران بالایی برای مقدار t موجود است ، که در تکنیک توقف از آن استفاده شود .

الگوریتم تفاضل منتهای معادله موج ۴.۱۱

برای تقریب جواب معادله موج

$$\frac{\partial^2 u}{\partial t^2} - x^2 \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < l, \quad 0 < t < T,$$

تحت شرایط

$$u(0, t) = u(l, t) = 0, \quad 0 < t,$$

$$u(x, 0) = f(x), \quad 0 \leq x \leq l.$$

و

$$\frac{\partial u}{\partial t}(x, 0) = g(x), \quad 0 \leq x \leq l.$$

اعداد صحیح $m > 0$ و $N > 0$ را انتخاب می‌کنیم .

مرحله ۱ . قرار می‌دهیم $h = l/m$ و $k = T/N$. قرار می‌دهیم $\lambda = kh$

مرحله ۲ . به‌ازای هر $i = 0, 1, \dots, m$ قرار می‌دهیم $x_i = ih$.

به‌ازای هر $j = 0, 1, \dots, N$ قرار می‌دهیم $t_j = jk$.

مرحله ۳ . به‌ازای هر $j = 1, 2, \dots, N$ قرار می‌دهیم $w_{0,j} = w_{m,j} = 0$.

به‌ازای هر $i = 0, 1, 2, \dots, m-1$ قرار می‌دهیم $w_{i,0} = f(x_i)$

، به‌ازای هر $i = 1, 2, \dots, m-1$

$$w_{i,1} = (1 - \lambda^2)f(x_i) + \frac{\lambda^2}{2} [f(x_{i+1}) + f(x_{i-1})] + kg(x_i) .$$

مرحله ۴ . قرار می‌دهیم $j = 1$.

مرحله ۵ . به‌ازای هر $i = 1, 2, \dots, m-1$ قرار می‌دهیم

$$w_{i,j+1} = 2(1 - \lambda^2)w_{i,j} + \lambda^2(w_{i+1,j} + w_{i-1,j}) - w_{i,j-1} .$$

مرحله ۶ . 1 را به j می‌افزاییم .

مرحله ۷ . اگر $j \leq N-1$ ، به مرحله ۵ می‌رویم .

مرحله ۸ . روند تمام است ؛ $w_{i,j}$ ، به‌ازای هر $i = 0, 1, \dots, m$ و $j = 0, 1, \dots, N$ ،

$u(x_i, t_j)$ را تقریب می‌کند .

مثال ۱ . مسئله هذلولوی

$$\frac{\partial^2 u}{\partial t^2} - 4 \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 1, \quad 0 < t,$$

را با شرایط مرزی $u(0, t) = u(1, t) = 0, 0 < t$ و شرایط اولیه

$$u(x, 0) = \sin(\pi x), \quad 0 \leq x \leq 1,$$

و

$$\frac{\partial u}{\partial t}(x, 0) = 0, \quad 0 \leq x \leq 1$$

در نظر می‌گیریم . به سادگی تحقیق می‌شود که جواب این مسئله عبارت است از

$$u(x, t) = \sin(\pi x) \cos(2\pi t).$$

در این مثال از روش تفاضل متناهی (الگوریتم ۴.۱۱) با $T = 1$ ، $m = 10$ و $N = 20$ ، که نتیجه می‌دهند $h = .1$ ، $k = .05$ ، و $\lambda = 1$ استفاده می‌کنیم . جدول زیر نتایج تقریب $w_{i,N}$ ، به ازای $i = 0, 1, \dots, 10$ ، رافهرست می‌کند . مقادیر فهرست شده در جدول ۵.۱۱ تا ارقام داده شده درست هستند .

x_i	$w_{i,20}$
.0	.0000000000
.1	.3090169944
.2	.5877852523
.3	.8090169944
.4	.9510565163
.5	1.0000000000
.6	.9510565163
.7	.8090169944
.8	.5877852523
.9	.3090169944
1.0	.0000000000

جدول ۵.۱۱

نتایج این مثال ، بیش از آنکه خطای برشی $O(k^2 + h^2)$ باعث باور ما شود ، بسیار دقیق هستند . توضیح این پدیده در این امر نهفته است که جواب حقیقی این معادله بی‌نهایت بار مشتق‌پذیر است . وقتی چنین حالتی باشد ، ساده است که با استفاده از سری تیلور نشان دهیم که

$$\frac{u(x_{i+1}, t_j) - 2u(x_i, t_j) + u(x_{i-1}, t_j))}{h^2} = \frac{\partial^2 u}{\partial x^2}(x_i, t_j) + 2 \left[\frac{h^2}{4!} \frac{\partial^4 u(x_i, t_j)}{\partial x^4} + \frac{h^4}{6!} \frac{\partial^6 u(x_i, t_j)}{\partial x^6} + \dots \right]$$

$$\frac{u(x_i, t_{j+1}) - 2u(x_i, t_j) + u(x_i, t_{j-1}))}{k^2} = \frac{\partial^2 u}{\partial t^2}(x_i, t_j) + 2 \left[\frac{k^2}{4!} \frac{\partial^4 u(x_i, t_j)}{\partial t^4} + \frac{k^4}{6!} \frac{\partial^6 u(x_i, t_j)}{\partial t^6} + \dots \right].$$

بنابراین ،

(۳۷.۱۱)

$$\frac{u(x_i, t_{j+1}) - 2u(x_i, t_j) + u(x_i, t_{j-1}))}{k^2} - x^2 \frac{u(x_{i+1}, t_j) - 2u(x_i, t_j) + u(x_{i-1}, t_j))}{h^2} = 2 \left[\frac{1}{4!} \left(k^2 \frac{\partial^4 u(x_i, t_j)}{\partial t^4} - x^2 h^2 \frac{\partial^4 u(x_i, t_j)}{\partial x^4} \right) + \frac{1}{6!} \left(k^4 \frac{\partial^6 u(x_i, t_j)}{\partial t^6} - x^2 h^4 \frac{\partial^6 u(x_i, t_j)}{\partial x^6} \right) + \dots \right].$$

اما ، با مشتگیری از معادله موج ،

$$\begin{aligned} k^2 \frac{\partial^4 u(x_i, t_j)}{\partial t^4} &= k^2 \frac{\partial^2}{\partial t^2} \left[x^2 \frac{\partial^2 u(x_i, t_j)}{\partial x^2} \right] \\ &= x^2 k^2 \frac{\partial^2}{\partial x^2} \left[\frac{\partial^2 u(x_i, t_j)}{\partial t^2} \right] \\ &= x^2 k^2 \frac{\partial^2}{\partial x^2} \left[x^2 \frac{\partial^2 u(x_i, t_j)}{\partial x^2} \right] \\ &= x^4 k^2 \frac{\partial^4 u(x_i, t_j)}{\partial x^4} \end{aligned}$$

ملاحظه می‌کنیم که

$$\begin{aligned} \frac{1}{4!} \left[k^2 \frac{\partial^4 u(x_i, t_j)}{\partial t^4} - x^2 h^2 \frac{\partial^4 u(x_i, t_j)}{\partial x^4} \right] \\ = \frac{\alpha^2}{4!} [\alpha^2 k^2 - h^2] \frac{\partial^4 u(x_i, t_j)}{\partial x^4} = 0. \end{aligned}$$

زیرا

$$\lambda^2 = \frac{x^2 k^2}{h^2} = 1$$

را انتخاب کردیم . به این طریق ، تمام جملات سمت راست (۳۷.۱۱) صفر هستند ، که

خطای برشی موضعی صفر را نتیجه می دهند. تنها خطاها در مثال ۱ آنهایی هستند که از تقریب $w_{i,1}$ و گرد کردن ناشی می شوند.

همانند حالت روش تفاضلی پیشرو برای معادله گرما، روش تفاضل متناهی صریح برای معادله موج مشکلات پایداری دارد. در واقع، برای اینکه روش پایدار باشد، لازم است که $\lambda = \alpha(k/h) \leq 1$ (ر.ک. ایزاکسون و کلر [۳۹]، صفحه ۰۴۸۹) روشهای ضمنی که پایدار غیر مشروط هستند وجود دارند، هر چند که آنها را مورد بحث قرار نمی دهیم. بحثی درباره این روشها را می توان در ایتم [۴]، صفحه ۱۹۹، یادگر کتابهای نوشته شده توسط میچل^۲ [۴۳] یا اسمیت^۳ [۶]، یافت. اگر f و g به قدر کافی مشتقپذیر باشند، روش صریح داده شده در الگوریتم ۴.۱۱ با $\lambda \leq 1$ همگرایی $O(h^2 + k^2)$ خواهد بود. برای آزمایش این مطلب، ر.ک. ایزاکسون و کلر [۳۹]، صفحه ۴۹۱.

در مسائل کاربردی، بخصوص در زمینه مکانیک جسم صلب، معادلات دیفرانسیل جزئی هذلولوی غالباً " داده های اولیه پیوسته ندارند. در چنین مسائلی از روشهای تفاضلی استفاده نمی شود، زیرا این روشها به هموار کردن ناپیوستگیها تمایل دارند. روشی که این ناپیوستگیها را بحساب می آورد روش مشخصه نامیده می شود. این روش را می توان روی معادلات هذلولوی کلیتر نیز اعمال کرد.

برای بررسی روش مشخصه، در نظر گرفتن معادله هذلولوی کلیتر به شکل

$$(38.11) \quad a_1 \frac{\partial^2 u}{\partial t^2} + a_2 \frac{\partial^2 u}{\partial x \partial t} + a_3 \frac{\partial^2 u}{\partial x^2} = a_4, \quad 0 < x < l, \quad t > 0,$$

که در آن a_i ، $i = 1, 2, 3, 4$ ، نمایانگر تابعی از متغیرهای x ، t ، u ، $(\partial u / \partial x)$ ، و $(\partial u / \partial t)$ است. یک مسئله از این نوع شبه خطی نامیده می شود، زیرا نسبت به مشتقات جزئی دوم خطی است. شرایط مرزی و اولیه مجدداً " به صورت زیر داده می شوند:

$$u(x, 0) = f(x), \quad 0 \leq x \leq l,$$

$$\frac{\partial u}{\partial t}(x, 0) = g(x), \quad 0 \leq x \leq l,$$

$$u(0, t) = u(l, t) = 0, \quad t > 0.$$

همچنین، فرض کنیم $f'(x)$ وجود داشته باشد بطوری که هر وقت $0 \leq x \leq l$ ،

$$(\partial u / \partial x)(x, 0) = f'(x)$$

چون u ، $(\partial u / \partial t)$ ، و $(\partial u / \partial x)$ به ازای $0 \leq x \leq l$ و $t = 0$ معلومند، شرایطی

را جستجو می‌کنیم که تحت آنها

$$\frac{\partial^2 u}{\partial t^2}, \quad \frac{\partial^2 u}{\partial x \partial t}, \quad \frac{\partial^2 u}{\partial x^2}$$

را بتوان به‌طور منحصر بفرد بدست آورد تا در (۳۸.۱۱) صدق کند. هرگاه مشتقهای

جزئی اول پیوسته باشند، آنگاه دیفرانسیلهای

$$d\left(\frac{\partial u}{\partial t}\right) \quad \text{و} \quad d\left(\frac{\partial u}{\partial x}\right)$$

باید وجود داشته باشند. معادلات

$$d\left(\frac{\partial u}{\partial x}\right) = \frac{\partial^2 u}{\partial x^2} dx + \frac{\partial^2 u}{\partial x \partial t} dt,$$

$$d\left(\frac{\partial u}{\partial t}\right) = \frac{\partial^2 u}{\partial x \partial t} dx + \frac{\partial^2 u}{\partial t^2} dt,$$

بهمراه (۳۸.۱۱) دستگاه خطی زیر را می‌دهد:

$$(39.11) \quad \begin{bmatrix} a_1 & a_2 & a_3 \\ 0 & dt & dx \\ dt & dx & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial^2 u}{\partial x^2} \\ \frac{\partial^2 u}{\partial x \partial t} \\ \frac{\partial^2 u}{\partial t^2} \end{bmatrix} = \begin{bmatrix} a_4 \\ d\left(\frac{\partial u}{\partial x}\right) \\ d\left(\frac{\partial u}{\partial t}\right) \end{bmatrix}.$$

این دستگاه را می‌توان برای تعیین

$$\frac{\partial^2 u}{\partial x^2}, \quad \frac{\partial^2 u}{\partial x \partial t}, \quad \frac{\partial^2 u}{\partial t^2}$$

به‌طور منحصر بفرد حل کرد، مگر اینکه

$$(40.11) \quad \det \begin{bmatrix} a_1 & a_2 & a_3 \\ 0 & dt & dx \\ dt & dx & 0 \end{bmatrix} = -a_1(dx)^2 + a_2 dt dx - a_3(dt)^2 = 0.$$

یعنی، مگر اینکه

$$a_3 \left(\frac{dt}{dx} \right)^2 - a_2 \frac{at}{dx} + a_1 = 0.$$

حل این معادله نسبت به (dt/dx) نتیجه می‌دهد که جواب منحصر بفرد خواهد بود مگر اینکه

$$\frac{dt}{dx} = \frac{a_2 \pm \sqrt{a_2^2 - 4a_1a_3}}{2a_3}.$$

این معادله نسبت به (dt/dx) نه فقط به نقطه (x, t) بلکه به مقادیر u ، $(\partial u / \partial x)$ ، و $(\partial u / \partial t)$ در (x, t) نیز بستگی دارد و شیب منحنیهایی را که منحنیهای مشخص نامیده می‌شوند بیان می‌کند. مبین این معادله، یعنی

$$\Delta = a_2^2 - 4a_1a_3,$$

نشان می‌دهد که صفر، یک، یا دو منحنی مشخص ماربر هر نقطه وجود دارد. به عنوان وسیله‌ای برای دسته بندی معادلات دیفرانسیل جزئی مرتبه دوم، معادلات بیضوی دارای $\Delta < 0$ و هیچ مشخص حقیقی اند، معادلات سهموی دارای $\Delta = 0$ و یک منحنی مشخص حقیقی ماربر هر نقطه است، درحالی که معادلات هذلولوی که در این بخش بررسی می‌کنیم دارای $\Delta > 0$ می‌باشند؛ و در نتیجه، دو منحنی مشخص دارند که از هر نقطه می‌گذرد.

برای معادلات هذلولوی، جهت‌های مشخص در یک نقطه P با مختصات (x, t) را با

$$(41.11) \quad m_1(P) = m_1(x, t) = \frac{a_2 + \sqrt{a_2^2 - 4a_1a_3}}{2a_3}$$

و

$$(42.11) \quad m_2(P) = m_2(x, t) = \frac{a_2 - \sqrt{a_2^2 - 4a_1a_3}}{2a_3}.$$

نمایش می‌دهیم. اگرچه نمادگذاری در (41.11) و (42.11) آن را صریحا بیان نمی‌کند، باید در ذهن داشته باشیم که این مقادیر به

$$\frac{\partial u}{\partial t}(x, t) \text{ و } \frac{\partial u}{\partial x}(x, t) \text{ و } u(x, t)$$

نیز بستگی دارند.

روش مشخصها با امتحان اینکه چه اتفاقی برای دیگر دترمینانهای مربوط به دستگاه خطی (۳۹.۱۱) می افتد، زمانی که این مشخصها پذیرفته می شوند، یا به طور معادل، وقتی معادله (۴۰.۱۱) برقرار می شود، بدست می آید. وقتی دترمینان واقع در (۴۰.۱۱) صفر می شود، قاعده کرامر (ر.ک. تمرین ۱۸ در بخش ۳.۷) نتیجه می دهد که یک جواب برای دستگاه نمی تواند اتفاق افتد مگر اینکه نیز داشته باشیم:

$$\det \begin{bmatrix} a_4 & a_2 & a_3 \\ d\left(\frac{\partial u}{\partial x}\right) & dt & dx \\ d\left(\frac{\partial u}{\partial t}\right) & dx & 0 \end{bmatrix}$$

$$= a_2 d\left(\frac{\partial u}{\partial t}\right) dx + a_3 \left(d\left(\frac{\partial u}{\partial x}\right) dx - d\left(\frac{\partial u}{\partial t}\right) dt \right) - a_4 (dx)^2 = 0,$$

$$\det \begin{bmatrix} a_1 & a_2 & a_4 \\ 0 & dt & d\left(\frac{\partial u}{\partial x}\right) \\ dt & dx & d\left(\frac{\partial u}{\partial t}\right) \end{bmatrix}$$

$$= a_1 \left(d\left(\frac{\partial u}{\partial t}\right) dt - d\left(\frac{\partial u}{\partial x}\right) dx \right) + a_2 d\left(\frac{\partial u}{\partial x}\right) dt - a_4 (dt)^2 = 0,$$

$$\det \begin{bmatrix} a_1 & a_4 & a_3 \\ 0 & d\left(\frac{\partial u}{\partial x}\right) & dx \\ dt & d\left(\frac{\partial u}{\partial t}\right) & 0 \end{bmatrix} = -a_1 d\left(\frac{\partial u}{\partial t}\right) dx - a_3 d\left(\frac{\partial u}{\partial x}\right) dt + a_4 dx dt = 0.$$

مخصوصاً، معادله سوم را می توان مجدداً به صورت زیر نوشت:

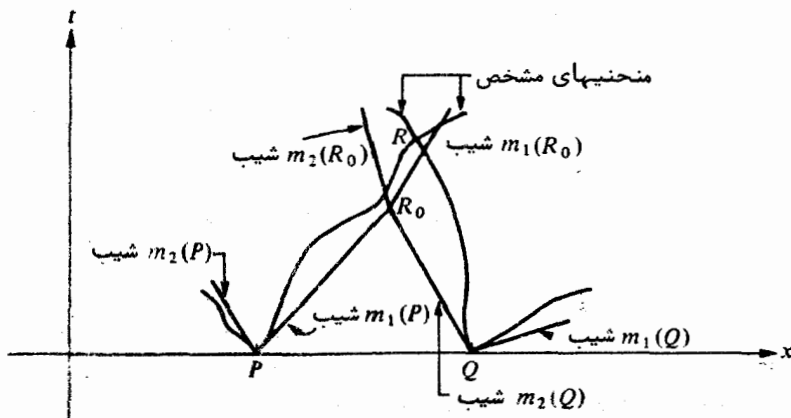
$$(۴۳.۱۱) \quad a_1 d\left(\frac{\partial u}{\partial t}\right) + a_3 d\left(\frac{\partial u}{\partial x}\right) \frac{dt}{dx} - a_4 dt = 0.$$

روش مشخصها از این معادله، به همراه معادلات (۴۱.۱۱) و (۴۲.۱۱) و این امر که دیفرانسیل کامل u می تواند به صورت

$$(۴۴.۱۱) \quad du = \frac{\partial u}{\partial t} dt + \frac{\partial u}{\partial x} dx.$$

نوشته شود، استفاده می کند.

برای تشریح عمل این روش، فرض کنیم $P = (x_1, t_1)$ و $Q = (x_2, t_2)$ ، با $t_1 = t_2 = 0$ ، دو نقطه در امتداد محور x بین 0 و l باشند. از P خطوط مستقیمی که بر منحنیهای مشخص در P مماس باشند رسم می کنیم، یعنی، خطوطی با شیبهای $m_1(P)$ و $m_2(P)$ ؛ همین طور از Q خطوطی با شیبهای $m_1(Q)$ و $m_2(Q)$ رسم می کنیم. انتظار می رود که دوتا از این خطوط، مثلاً "خطوط با شیبهای $m_1(P)$ و $m_2(Q)$ " در نقطه ای مانند $R_0 = (x^{(0)}, t^{(0)})$ به مختص t مثبت همدیگر را قطع کنند. این نقطه تقریب اولیای برای R ، محل تلاقی دو منحنی مشخص ماربر P و Q ، خواهد بود (ر. ک. شکل ۱۰.۱۱).



شکل ۱۰.۱۱

برای بهتر کردن تقریب به R ، ابتدا تقریبهایی به

$$\frac{\partial u}{\partial t}(R_0) \text{ و } \frac{\partial u}{\partial x}(R_0) \text{ ، } u(R_0)$$

پیدا می کنیم. روشی که بکار می رود عبارت است از میانگین کردن مقادیر a_1 ، m_1

a_3 ، و a_4 در P و R_0 جهت تشکیل یک قسمت تقریب ، و مقادیر m_2 ، a_1 ، a_3 ، و a_4 در Q و R_0 جهت تشکیل جزء باقیمانده . مثلاً ، با استفاده از میانگینها و تفاضلات در نقاط P و R_0 از معادله (۴۳.۱۱) تقریبهایی برای

$$\frac{\partial u}{\partial x}(R_0) \quad \text{و} \quad \frac{\partial u}{\partial t}(R_0)$$

بدست آمده تا حاصل شود

$$\begin{aligned} & \frac{1}{4}[a_3(R_0) + a_3(P)][m_1(R_0) + m_1(P)] \left[\frac{\partial u}{\partial x}(R_0) - \frac{\partial u}{\partial x}(P) \right] \\ (45.11) \quad & + \frac{1}{2}[a_1(R_0) + a_1(P)] \left[\frac{\partial u}{\partial t}(R_0) - \frac{\partial u}{\partial t}(P) \right] \\ & = \frac{1}{2}[a_4(R_0) + a_4(P)](t^{(0)} - t_1). \end{aligned}$$

و میانگینها و تفاضلات در Q و R_0 در همان معادله بدست می دهد

$$\begin{aligned} & \frac{1}{4}[a_3(R_0) + a_3(Q)][m_2(R_0) + m_2(Q)] \left[\frac{\partial u}{\partial x}(R_0) - \frac{\partial u}{\partial x}(Q) \right] \\ (46.11) \quad & + \frac{1}{2}[a_1(R_0) + a_1(Q)] \left[\frac{\partial u}{\partial t}(R_0) - \frac{\partial u}{\partial t}(Q) \right] \\ & = \frac{1}{2}[a_4(R_0) + a_4(Q)](t^{(0)} - t_2). \end{aligned}$$

این معادلات را می توان به طور همزمان برای تقریبهایی به

$$\frac{\partial u}{\partial t}(R_0) \quad \text{و} \quad \frac{\partial u}{\partial x}(R_0)$$

حل کرد .

پس از انجام این محاسبات می توانیم یک تقریب ، یعنی $u^{(1)}(R_0)$ ، به $u(R_0)$ پیدا

کنیم ، با فرض اینکه $du = u^{(1)}(R_0) - u(P)$ و با استفاده از میانگینها و معادله (۴۴.۱۱) بدست می آید

$$\begin{aligned} u^{(1)}(R_0) - u(P) &= \frac{1}{2} \left[\frac{\partial u}{\partial t}(R_0) + \frac{\partial u}{\partial t}(P) \right] (t^{(0)} - t_1) \\ &+ \frac{1}{2} \left[\frac{\partial u}{\partial x}(R_0) + \frac{\partial u}{\partial x}(P) \right] (x^{(0)} - x_1) \end{aligned}$$

$$u^{(1)}(R_0) = u(P) + \frac{1}{2} \left[\frac{\partial u}{\partial t}(R_0) + \frac{\partial u}{\partial t}(P) \right] (t^{(0)} - t_1) \\ + \frac{1}{2} \left[\frac{\partial u}{\partial x}(R_0) + \frac{\partial u}{\partial x}(P) \right] (x^{(0)} - x_1).$$

به طریقی مشابه، می‌توانیم، با استفاده از نقاط R_0 و Q ، تقریب دیگری، یعنی $u^{(2)}(R_0)$ ، برای بدست آوردیم

$$u^{(2)}(R_0) = u(Q) + \frac{1}{2} \left[\frac{\partial u}{\partial t}(R_0) + \frac{\partial u}{\partial t}(Q) \right] (t^{(0)} - t_2) \\ + \frac{1}{2} \left[\frac{\partial u}{\partial x}(R_0) + \frac{\partial u}{\partial x}(Q) \right] (x^{(0)} - x_2).$$

انتظار می‌رود که میانگین کردن این تقریبات تقریب بهتری دهد، لذا، فرض می‌کنیم

$$u(R_0) = \frac{1}{2} [u^{(1)}(R_0) + u^{(2)}(R_0)]$$

بعد از اینکه همه تقریبات لازم برای R_0 محاسبه شدند، R_1 به‌عنوان محل تلاقی خط

ماربر P با شیب $\frac{1}{2} [m_1(R_0) + m_1(P)]$ و خط ماربر Q با شیب $\frac{1}{2} [m_2(R_0) + m_2(Q)]$

تعیین می‌شود؛ یعنی، مختصات $(x^{(1)}, t^{(1)}) = R_1$ با حل همزمان معادلات

$$\frac{1}{2} [m_1(R_0) + m_1(P)] = \frac{t^{(1)} - t_1}{x^{(1)} - x_1}$$

و

$$\frac{1}{2} [m_2(R_0) + m_2(Q)] = \frac{t^{(1)} - t_2}{x^{(1)} - x_2}$$

نسبت به $x^{(1)}$ و $t^{(1)}$ بدست می‌آید. (ر.ک. شکل ۱۱-۱۱)

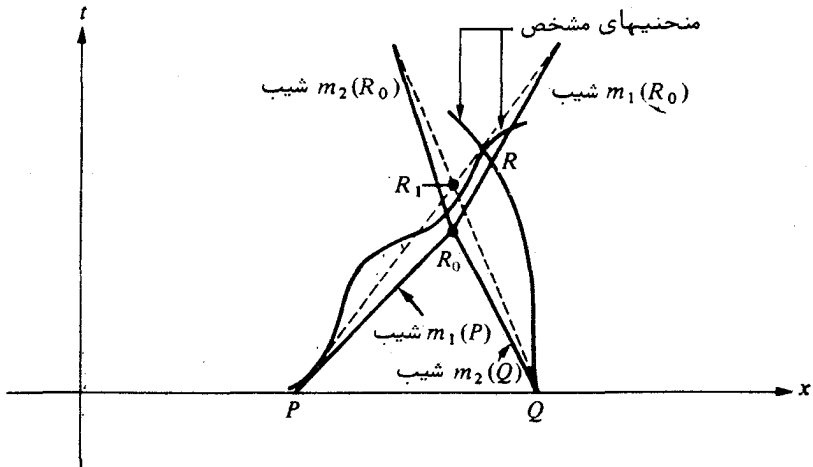
تقریبات به

$$\frac{\partial u}{\partial t}(R_1) \text{ و } \frac{\partial u}{\partial x}(R_1), u(R_1)$$

را می‌توان با همان روشهایی که برای تعیین تقریبات به

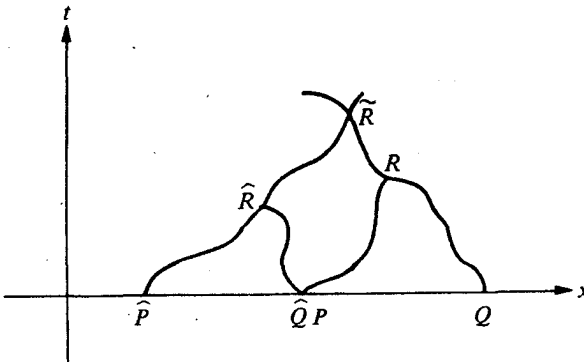
$$\frac{\partial u}{\partial t}(R_0) \text{ و } \frac{\partial u}{\partial x}(R_0), u(R_0)$$

بگر رفت بدست آورد. ادامه دادن به این طریق دنباله‌ای از نقاط بدست خواهد داد که فرض می‌کنیم به R ، محل برخورد منحنیهای مشخص، همگرا خواهد بود، و تقریبهایی نیز به مقادیر u ، $(\partial u / \partial x)$ ، و در آن نقطه مهیا خواهد کرد.



شکل ۱۱.۱۱

تمام روند برای نقاط جدید P و Q تکرار می‌شود و بعد دوباره با مقادیر نقاط جدید تکرار می‌شود، تا اینکه در جهت محور زمان بالا برویم. به این طریق شبکه‌ای از نقاط بدست می‌آید که در آن u و مشتقات جزئی آن تقریب می‌شوند، همانطور که در شکل ۱۲.۱۱ نشان داده شده است. الگوریتم زیر روند کامل را شرح می‌دهد.



شکل ۱۲.۱۱

الگوریتم روش مشخصها ۵.۱۱

بازای تقریب جواب معادله هذلولوی

$$a_1 u_{tt} + a_2 u_{xt} + a_3 u_{xx} = a_4, \quad 0 < x < l, \quad t > 0,$$

تحت شرایط

$$u(x, 0) = f(x), \quad 0 \leq x \leq l,$$

$$u_t(x, 0) = g(x), \quad 0 \leq x \leq l,$$

$$u(0, t) = u(l, t) = 0, \quad t > 0,$$

اعداد صحیح $N, M > 0$ و میزان تحمل خطای $\varepsilon > 0$ را انتخاب می‌کنیم.

مرحله ۱. قرار می‌دهیم $h = l/M$ و بازای هر $i = 0, 1, \dots, M$ $x_{i,0} = ih$ و $t_{i,0} = 0$

مرحله ۲. قرار می‌دهیم

$$w_{i,0} = f(x_i),$$

$$w'_{i,0} = g(x_i),$$

بازای هر $i = 0, 1, \dots, M$ $w'_{i,0} = f'(x_i)$

مرحله ۳. بازای هر $i = 0, 1, \dots, M$ قرار می‌دهیم $P_{i,0} = (x_{i,0}, t_{i,0}, w_{i,0}, w'_{i,0})$

مرحله ۴. قرار می‌دهیم $j = 0$

مرحله ۵. قرار می‌دهیم $i = 0$. اگر j فرد باشد قرار می‌دهیم $s = 1$; و اگر j

زوج باشد، قرار می‌دهیم $s = 2$. قرار می‌دهیم $r = i + s - 1$

مرحله ۶. اگر $\frac{a_2(P_{i,j}) + \sqrt{a_2^2(P_{i,j}) - 4a_1(P_{i,j})a_3(P_{i,j})}}{2a_3(P_{i,j})} > 0$ قرار می‌دهیم

$$m_1(P_{i,j}) = \frac{a_2(P_{i,j}) + \sqrt{a_2^2(P_{i,j}) - 4a_1(P_{i,j})a_3(P_{i,j})}}{2a_3(P_{i,j})}$$

در غیر این صورت، قرار می‌دهیم

$$m_1(P_{i,j}) = \frac{a_2(P_{i,j}) - \sqrt{a_2^2(P_{i,j}) - 4a_1(P_{i,j})a_3(P_{i,j})}}{2a_3(P_{i,j})}$$

مرحله ۷. اگر $\frac{a_2(P_{i+1,j}) - \sqrt{a_2^2(P_{i+1,j}) - 4a_1(P_{i+1,j})a_3(P_{i+1,j})}}{2a_3(P_{i+1,j})} < 0$ قرار

می‌دهیم

$$m_2(P_{i+1,j}) = \frac{a_2(P_{i+1,j}) - \sqrt{a_2^2(P_{i+1,j}) - 4a_1(P_{i+1,j})a_3(P_{i+1,j})}}{2a_3(P_{i+1,j})}$$

در غیر این صورت ، قرار می دهیم

$$m_2(P_{i+1,j}) = \frac{a_2(P_{i+1,j}) + \sqrt{a_2^2(P_{i+1,j}) - 4a_1(P_{i+1,j})a_3(P_{i+1,j})}}{2a_3(P_{i+1,j})}$$

مرحله ۸ . قرار می دهیم $k = 0$ ،

$$x_{r,j+1}^{(0)} = \frac{t_{i+1,j} - t_{i,j} + m_1(P_{i,j})x_{i,j} - m_2(P_{i+1,j})x_{i+1,j}}{m_1(P_{i,j}) - m_2(P_{i+1,j})}$$

$$t_{r,j-1}^{(0)} = m_1(P_{i,j})(x_{r,j+1}^{(0)} - x_{i,j}) + t_{i,j}$$

$$P_{r,j+1} = (x_{r,j+1}^{(0)}, t_{r,j+1}^{(0)}, w_{i,j}, \dot{w}_{i,j}, w'_{i,j})$$

مرحله ۹ . قرار می دهیم

$$A = \frac{1}{2}[a_1(P_{r,j+1}) + a_1(P_{i,j})],$$

$$B = \frac{1}{2}[a_3(P_{r,j+1}) + a_3(P_{i,j})][m_1(P_{r,j+1}) + m_1(P_{i,j})],$$

$$C = \frac{1}{2}[a_1(P_{r,j+1}) + a_1(P_{i+1,j})],$$

$$D = \frac{1}{2}[a_3(P_{r,j+1}) + a_3(P_{i+1,j})][m_2(P_{r,j+1}) + m_2(P_{i+1,j})],$$

$$E = \frac{1}{2}[a_4(P_{r,j+1}) + a_4(P_{i,j})][t_{r,j+1}^{(k)} - t_{i,j}] + A\dot{w}_{i,j} + Bw'_{i,j}$$

$$F = \frac{1}{2}[a_4(P_{r,j+1}) + a_4(P_{i+1,j})][t_{r,j+1}^{(k)} - t_{i+1,j}] + C\dot{w}_{i+1,j} + Dw'_{i+1,j}$$

$$\dot{w}_{r,j+1} = \frac{ED - FB}{AD - BC}$$

$$w'_{r,j+1} = \frac{AF - CE}{AD - BC}$$

$$H = \frac{1}{2}[\dot{w}_{r,j+1} + \dot{w}_{i,j}][t_{r,j+1}^{(k)} - t_{i,j}] + \frac{1}{2}[w'_{r,j+1} + w'_{i,j}][x_{r,j+1}^{(k)} - x_{i,j}]$$

$$I = \frac{1}{2}[\dot{w}_{r,j+1} + \dot{w}_{i+1,j}][t_{r,j+1}^{(k)} - t_{i+1,j}] + \frac{1}{2}[w'_{r,j+1} + w'_{i+1,j}][x_{r,j+1}^{(k)} - x_{i+1,j}]$$

$$w_{r,j+1} = \frac{1}{2}[w_{i,j} + H + w_{i+1,j} + I]$$

$$P_{r,j+1} = (x_{r,j+1}^{(k)}, t_{r,j+1}^{(k)}, w_{r,j+1}, \dot{w}_{r,j+1}, w'_{r,j+1})$$

$$G = \frac{1}{2}[m_1(P_{r,j+1}) + m_1(P_{i,j})]x_{i,j} - \frac{1}{2}[m_2(P_{r,j+1}) + m_2(P_{i+1,j})]x_{i+1,j} + t_{i+1,j} - t_{i,j}$$

$$x_{r,j+1}^{(k+1)} = \frac{G}{\frac{1}{2}[m_1(P_{i,j}) - m_2(P_{i+1,j}) + m_1(P_{r,j+1}) - m_2(P_{r,j-1})]},$$

$$t_{r,j+1}^{(k+1)} = \frac{1}{2}[m_1(P_{r,j+1}) + m_1(P_{i,j})](x_{r,j+1}^{(k+1)} - x_{i,j}) + t_{i,j}.$$

مرحله ۱۰. ۱ را به k می‌افزاییم.

مرحله ۱۱. اگر $|x_{r,j+1}^{(k)} - x_{r,j+1}^{(k-1)}| \geq \varepsilon$ یا $|t_{r,j+1}^{(k)} - t_{r,j+1}^{(k-1)}| \geq \varepsilon$ ، به مرحله ۹

می‌رویم. در غیر این صورت، قرار می‌دهیم $x_{r,j-1} = x_{r,j+1}^{(k)}$ و $t_{r,j+1} = t_{r,j+1}^{(k)}$.

مرحله ۱۲. ۱ را به i می‌افزاییم.

مرحله ۱۳. اگر $i \leq M - 1$ و $s = 1$ ، به مرحله ۶ می‌رویم. اگر $i \leq M$ و $s = 2$ ،

به مرحله ۶ می‌رویم.

مرحله ۱۴. اگر $s = 1$ ، به مرحله ۱۶ می‌رویم.

مرحله ۱۵. قرار می‌دهیم

$$x_{0,j} = 0,$$

$$t_{0,j} = t_{1,j},$$

$$w_{0,j} = 0,$$

$$\dot{w}_{0,j} = g(0),$$

$$w'_{0,j} = f'(0),$$

$$P_{0,j} = (x_{0,j}, t_{0,j}, w_{0,j}, \dot{w}_{0,j}, w'_{0,j}),$$

و به مرحله ۱۷ می‌رویم.

مرحله ۱۶. قرار می‌دهیم

$$x_{M,j} = l,$$

$$t_{M,j} = t_{M-1,j},$$

$$w_{M,j} = 0,$$

$$\dot{w}_{M,j} = g(l),$$

$$w'_{M,j} = f'(l),$$

$$P_{M,j} = (x_{M,j}, t_{M,j}, w_{M,j}, \dot{w}_{M,j}, w'_{M,j}).$$

مرحله ۱۷. ۱ را به z می‌افزاییم.

مرحله ۱۸. اگر $z \leq N - 1$ ، به مرحله ۵ می‌رویم.

مرحله ۱۹. روند تمام است؛ شبکه مشخص

$$\{(x_{i,j}, t_{i,j}) : 0 \leq i \leq M, 0 \leq j \leq N\}$$

$$w_{i,j} \approx u(x_{i,j}, t_{i,j}),$$

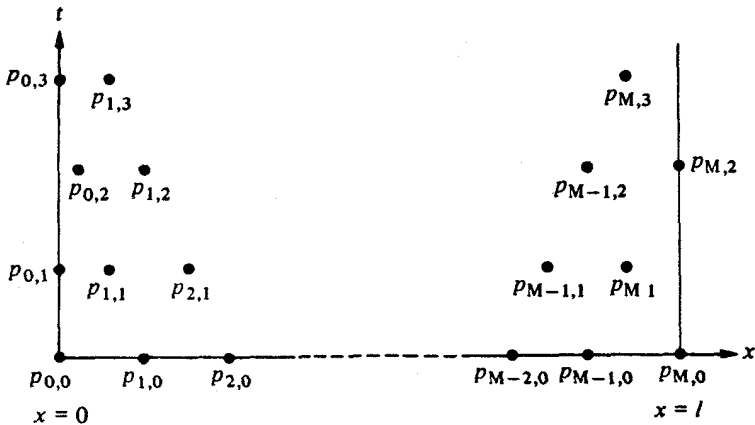
$$w'_{i,j} \approx \frac{\partial u}{\partial x}(x_{i,j}, t_{i,j}),$$

9

$$\dot{w}_{i,j} \approx \frac{\partial u}{\partial t}(x_{i,j}, t_{i,j}),$$

بمازای هر $i = 0, 1, \dots, M$ و $j = 0, 1, \dots, N$ ساخته شده است.

اگر مقادیر $t_{i,N}$ بمازای هر $i = 0, 1, \dots, M$ به قدر کافی بزرگ نباشند، پیشنهاد می‌شود که روند تا وقتی مقادیر t بزرگ بدست آیند ادامه یابد. دلیل کاربرد متغیر "کلیدی" s در مرحله ۵ آن است که اطمینان حاصل کنیم شبکه مشخص در سراسر $\{(x, t) | 0 \leq x \leq l, 0 \leq t \leq t_N\}$ پراکنده بماند (ر.ک. شکل ۱۳.۱۱).



شکل ۱۳.۱۱

مثال ۲. روش مشخصها (الگوریتمها ۵.۱۱) با $M = 4$ و $N = 2$ برای تقریب جواب معادله دیفرانسیل جزئی هذلولوی

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} - 50\pi^2 t \frac{\partial u}{\partial t} - 50\pi^2 u = 0, \quad 0 < x < .2, \quad 0 < t,$$

تحت قیود

$$u(0, t) = u(2, t) = 0, \quad 0 < t,$$

$$u(x, 0) = 0, \quad 0 \leq x \leq 2,$$

$$u_t(x, 0) = \sin 10\pi x, \quad 0 \leq x \leq 2.$$

بکار خواهد رفت . برای این مسئله

$$a_1(x, t, u, u_x, u_t) = 1$$

$$a_2(x, t, u, u_x, u_t) = 0$$

$$a_3(x, t, u, u_x, u_t) = -1$$

$$a_4(x, t, u, u_x, u_t) = 50\pi^2(u + tu_t),$$

و به سادگی تحقیق می شود که $u(x, t) = t \sin 10\pi x$ جواب است . نتایج الگوریتم مذکور در جدول ۶.۱۱ فهرست شده است .

j	s	i	r	$x_{r,j+1}$	$t_{r,j+1}$	$w_{r,j+1}$	$\hat{w}_{r,j+1}$	$w'_{r,j+1}$	$u(x_{r,j+1}, t_{r,j+1})$	$ w_{r,j+1} - u(x_{r,j+1}, t_{r,j+1}) $
0	2	0	1	.025	.025	.012500	.50000	.50000	.017678	.0051
0	2	1	2	.075	.025	.014428	.65421	-.50000	.017678	.0032
0	2	2	3	.125	.025	-.012500	-.50000	-.50000	-.017678	.0051
0	2	3	4	.175	.025	-.014428	-.65421	.50000	-.017678	.0032
0		0	0		.025	0	0	0	0	0
1	1	0	0	.0125	.0375	.0096160	.53855	.53855	.014351	.0047
1	1	1	1	.05	.05	.032934	.48047	.09494	.050000	.0171
1	1	2	2	.1	.05	.0067491	.38572	-.01249	0	.0067
1	1	3	3	.15	.05	-.032934	-.48047	-.09494	-.050000	.0171
1		4	2		.05	0	0	0	0	0
2	2	0	1	.0375	.0625	.036149	.66770	.21686	.057742	.0216
2	2	1	2	.075	.075	.039186	.44245	-.71986	.053033	.0138
2	2	2	3	.125	.075	-.012718	.65443	-1.3611	-.053033	.0403
2	2	3	4	.175	.075	-.029987	-.79387	.36844	-.053033	.0230
2		0	0		.0625	0	0	0	0	0

جدول ۶.۱۱

مجموعه تمرینات ۴.۱۱

۱ . جواب $u(x, t) = \sin \pi x \cos \pi t$ برای معادله موج

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 1, \quad 0 < t.$$

$$u(0, t) = u(1, t) = 0, \quad 0 < t,$$

$$u(x, 0) = \sin \pi x, \quad 0 \leq x \leq 1,$$

$$\frac{\partial u}{\partial t}(x, 0) = 0, \quad 0 \leq x \leq 1,$$

را با استفاده از الگوریتم ۴.۱۱ با $h = .1$ و $k = .05$ و با $h = .05$ و $k = .1$ و با $h = .05$ و $k = .1$ مقایسه کنید. نتایج خود را با جواب دقیق در $t = .5$ مقایسه کنید.

۲. تمرین ۱ را با استفاده از تقریب

$$w_{i,1} = w_{i,0} + kg(x_i), \quad i = 0, 1, \dots, m$$

تکرار کنید. آیا می‌توانید نتایج را توضیح دهید؟

۳. جواب معادله موج

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 1, \quad 0 < t,$$

$$u(0, t) = u(1, t) = 0, \quad 0 < t,$$

$$u(x, 0) = \sin 2\pi x, \quad 0 \leq x \leq 1,$$

$$\frac{\partial u}{\partial t}(x, 0) = 2\pi \sin 2\pi x, \quad 0 \leq x \leq 1,$$

را با استفاده از الگوریتم ۴.۱۱ با $h = .1$ و $k = .1$ تقریب کنید. نتایج خود را

با جواب واقعی $u(x, t) = \sin 2\pi x [\cos 2\pi t + \sin 2\pi t]$ در $t = .3$ مقایسه کنید.

۴. جواب معادله موج زیر را تقریب کنید

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 1, \quad 0 < t,$$

$$u(0, t) = u(1, t) = 0, \quad 0 < t,$$

$$u(x, 0) = \begin{cases} 1, & 0 \leq x \leq \frac{1}{2} \\ -1, & \frac{1}{2} < x \leq 1 \end{cases}$$

$$\frac{\partial u}{\partial t}(x, 0) = 0, \quad 0 \leq x \leq 1.$$

از $h = .1$ و $k = .1$ استفاده کنید و جواب خود را با جواب سری فوریه در $t = .5$

مقایسه نمایید.

۵. جوابهای معادلات دیفرانسیل جزئی هذلولوی زیر را با استفاده از روش مشخصها تقریب کنید.

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 1, \quad 0 < t, \quad (T)$$

$$u(0, t) = u(1, t) = 0, \quad 0 < t,$$

$$u(x, 0) = \begin{cases} 1, & 0 \leq x \leq \frac{1}{2}, \\ -1, & \frac{1}{2} < x \leq 1, \end{cases}$$

$$\frac{\partial u}{\partial t}(x, 0) = 0, \quad 0 \leq x \leq 1.$$

از ۱. h استفاده کنید و جواب خود را با جواب سری فوریه و نتایج بدست آمده در تمرین ۴ مقایسه کنید.

$$\frac{\partial^2 u}{\partial t^2} + \frac{\partial^2 u}{\partial x \partial t} - \frac{\partial^2 u}{\partial x^2} + 2t - 2x + 1 = 0, \quad 0 < x < 1, \quad 0 < t. \quad (B)$$

$$u(0, t) = u(1, t) = 0, \quad 0 < t,$$

$$u(x, 0) = 0, \quad 0 \leq x \leq 1,$$

$$\frac{\partial u}{\partial t}(x, 0) = x^2 - x, \quad 0 \leq x \leq 1.$$

از ۱. h استفاده کنید و جوابهای خود را با جواب $u(x, t) = tx(x-1)$ مقایسه کنید.

۶. در یک سیم انتقال برق به طول l ، که جریان متناوب با فرکانس بالا انتقال می دهد (به نام خط "بی افت")، ولتاژ V و شدت جریان i به وسیله

$$\frac{\partial^2 V}{\partial x^2} = LC \frac{\partial^2 V}{\partial t^2}, \quad 0 < x < l, \quad 0 < t,$$

$$\frac{\partial^2 i}{\partial x^2} = LC \frac{\partial^2 i}{\partial t^2}, \quad 0 < x < l, \quad 0 < t,$$

بیان می شود، که در آن L ضریب سلف در واحد طول و C ظرفیت در واحد طول است. فرض کنید طول سیم ۲ فوت بوده و C و L با

$$C = \text{فوت/فاراد} = 1$$

$$L = \text{فوت/هانری} = 1$$

داده شده باشند. همچنین، ولتاژ و شدت جریان در

$$V(0, t) = V(2, t) = 0, \quad 0 < t$$

$$V(x, 0) = 110 \sin \frac{\pi x}{2}, \quad 0 \leq x \leq 2,$$

$$\frac{\partial V}{\partial t}(x, 0) = 0, \quad 0 \leq x \leq 2,$$

$$i(0, t) = i(2, t) = 0, \quad 0 < t,$$

$$i(x, 0) = 5.5 \cos \frac{\pi x}{2}, \quad 0 \leq x \leq 2,$$

$$\frac{\partial i}{\partial t}(x, 0) = 0, \quad 0 \leq x \leq 2.$$

صدق کنند.

(T) با استفاده از الگوریتم ۴.۱۱ با $h = k = 1$ ، ولتاژ و شدت جریان را در $t = 2$.

و $t = 5$ تقریب کنید.

(ب) سه جمله اول جواب سری فوریه برای V و i را پیدا کنید و مقادیر $V(5, 2)$ ،

$V(5, 5)$ ، $i(5, 2)$ ، و $i(5, 5)$ را با تقریبهای (T) مقایسه نمایید.

۷. فشار هوای $p(x, t)$ در یک لوله آرگ از معادله موج

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2}, \quad 0 < x < l, \quad 0 < t,$$

تبعیت می‌کند، که در آن l طول لوله و c یک ثابت فیزیکی است. اگر لوله باز باشد، شرایط مرزی با

$$p(l, t) = p_0 \quad \text{و} \quad p(0, t) = p_0$$

داده می‌شوند؛ و اگر لوله در انتها بسته باشد، جایی که $x = l$ ، شرایط مرزی عبارتند از

$$\frac{\partial p(l, t)}{\partial x} = 0 \quad \text{و} \quad p(0, t) = p_0$$

فرض کنید $c = 1$ و $l = 1$ و شرایط اولیه عبارت باشند از

$$p(x, 0) = p_0 \cos 2\pi x, \quad 0 \leq x \leq l,$$

$$p_t(x, 0) = 0, \quad 0 \leq x \leq l.$$

(T) با استفاده از الگوریتم ۴.۱۱، با $h = k = 1$ ، فشار را برای یک لوله باز

$p_0 = .9$ در $x = \frac{1}{2}$ با زای $t = .5$ و $t = 1$ تقریب کنید .
 (ب) الگوریتم ۴.۱۱ را برای مسئله لوله بسته با $p_0 = .9$ تغییر دهید و با استفاده از $h = k = .1$ ، $p(5, .5)$ و $p(5, 1)$ را تقریب کنید .

۵.۱۱ آشنایی با روش عناصر متناهی

روش دیگری که غالباً "برای حل معادلات دیفرانسیل جزئی بیضوی بکار می رود روش عناصر متناهی نامیده می شود . این روش مشابه روند ریلی - ریتس است که در بخش ۴.۱۰ بحث شد ، اما به ابعاد بالاتر تعمیم داده می شود .

یک مزیت روش عناصر متناهی بر روشهای تفاضل متناهی سادگی نسبی آن در پرداختن به شرایط مرزی مسئله است . بسیاری از مسائل فیزیکی دارای شرایط مرزی هستند که شامل مشتقات است ، و ، در حالت کلی ، مرز ناحیه شکلی نامنظم دارد . پرداختن به شرایط مرزی اینها با استفاده از تکنیکهای تفاضل متناهی بسیار مشکل است ، زیرا هر شرط مرزی شامل یک مشتق باید به وسیله یک خارج قسمت تفاضلی در نقاط شبکه ای تقریب شود ، و شکل نامنظم مرز جا دادن نقاط شبکه ای رامشکل می کند . چون روش عناصر متناهی شامل شرایط مرزی به صورت انتگرالهایی در یک تابعی است که مینیمم می شود ، اساس ساخت روند مستقل از شرایط مرزی خاص مسئله است .

در این بحث ، معادله دیفرانسیل جزئی

$$(47.11) \quad \frac{\partial}{\partial x} \left(p(x, y) \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial y} \left(q(x, y) \frac{\partial u}{\partial y} \right) + r(x, y)u(x, y) = f(x, y),$$

را ، با $(x, y) \in D$ ، که در آن D ناحیه ای با مرز \mathcal{S} است ، در نظر می گیریم .
 شرایط مرزی به شکل

$$(48.11) \quad u(x, y) = g(x, y)$$

بر یک جزء \mathcal{S}_1 از مرز اعمال خواهد شد ؛ و بر باقیمانده مرز ، یعنی \mathcal{S}_2 ، لازم خواهد بود که $u(x, y)$ در

$$(49.11) \quad p(x, y) \frac{\partial u}{\partial x} \cos \theta_1 + q(x, y) \frac{\partial u}{\partial y} \cos \theta_2 + g_1(x, y)u(x, y) = g_2(x, y),$$

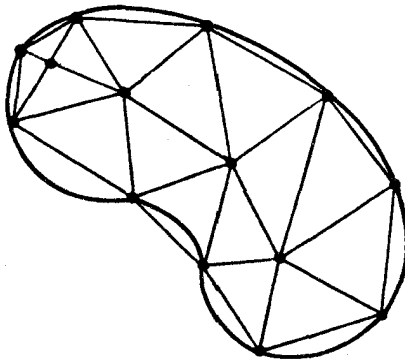
صدق کند ، که در آن θ_1 و θ_2 زوایای جهتی قائم روبرو خارج بر مرز در نقطه (x, y) است .
 مسائل فیزیکی در زمینه های مکانیک صلب و الاستیسیته معادلات دیفرانسیل جزئی

مربوطه مشابه (۴۷.۱۱) دارند. حل یک مسئله از این نوع، نوعاً "مینیم سازی یک تابعی معینسی، شامل انتگرالهایی روی ردهای از توابع است که به وسیله مسئله مشخص می شود. هرگاه p, q, r ، و f همه در $\mathcal{S} \cup D$ پیوسته باشند، p و q مشتقات جزئی اول پیوسته داشته باشند، و g_1 و g_2 بر \mathcal{S}_1 پیوسته باشند، آنگاه جواب (۴۷.۱۱) به طور منحصر بفرد تابعی

$$I[w] = \iint_D \left\{ \frac{1}{2} \left[p(x, y) \left(\frac{\partial w}{\partial x} \right)^2 + q(x, y) \left(\frac{\partial w}{\partial y} \right)^2 - r(x, y) w^2 \right] + f(x, y) w \right\} dx dy + \int_{\mathcal{S}_2} \left\{ -g_2(x, y) w + \frac{1}{2} g_1(x, y) w^2 \right\} dS \quad (50.11)$$

را روی تمام توابع w ، که بر \mathcal{S}_1 در (۴۷.۱۱) صادقند و دوبار به طور پیوسته مشتق پذیر می باشند، مینیم می سازد. روش عناصر متناهی این جواب را با مینیم سازی تابعی I روی ردهای کوچکتر از توابع تقریب می کند.

اولین مرحله در این روند تقسیم کردن ناحیه به تعدادی متناهی قسمت، یا عنصر، به شکل منتظم، یا مستطیل یا، به طور معمولتر، مثلث است. (ر. ک. شکل ۱۴.۱۱).



شکل ۴.۱۲

مجموعه توابعی که برای تقریب بکار می رود در حالت کلی یک مجموعه از قطعه قطعه چندجمله ایهای با درجه ثابت از x و y است و برای تقریب لازم است چند جمله ایها به گونه ای سر هم شوند که تابع حاصل پیوسته باشد یا مشتق اول یا دوم پیوسته یا انتگرال پذیر روی

تمام ناحیه. چند جمله‌ایهای از نوع خطی در x و y

$$\phi(x, y) = a + bx + cy$$

معمولا" با عناصر مثلثی بکار می‌روند در حالی که چند جمله‌ایهای از نوع دوخطی از x و y

$$\phi(x, y) = a + bx + cy + dxy,$$

با عناصر مستطیلی بکار می‌روند.

برای این بحث، فرض کنیم ناحیه D به عناصر مثلثی تقسیم شده باشد. رئوس

این عناصر گرهها نامیده می‌شوند. فرض کنیم S_1 مجموعه گرههای واقع بر \mathcal{S}_1 و S_2

مجموعه گرههای واقع بر \mathcal{S}_2 باشد. روش تقریبی به شکل

$$\phi(x, y) = \sum_{i=1}^m \gamma_i \phi_i(x, y),$$

را جستجو می‌کنیم، که در آن $\phi_1, \phi_2, \dots, \phi_m$ قطعه قطعه چند جمله‌ایهای درجه اول

مستقل خطی هستند و $\gamma_1, \gamma_2, \dots, \gamma_m$ مقادیر ثابت هستند. بعضی از این ثابتها، مثلا"

$\gamma_{n+1}, \gamma_{n+2}, \dots, \gamma_m$ ، برای اطمینان از اینکه شرط مرزی

$$\phi(x, y) = g(x, y)$$

بر \mathcal{S}_1 برقرار باشد بکار می‌روند در حالی که ثابتهای باقیمانده $\gamma_1, \gamma_2, \dots, \gamma_n$ برای

مینیمم سازی تابعی $I[\sum_{i=1}^m \gamma_i \phi_i]$ بکار می‌روند.

از معادله (۵۰.۱۱) تابعی به شکل زیر است:

$$\begin{aligned} (51.11) \quad I[\phi] &= I\left[\sum_{i=1}^m \gamma_i \phi_i\right] \\ &= \iint_D \left\{ \frac{1}{2} \left[p(x, y) \left[\sum_{i=1}^m \gamma_i \frac{\partial \phi_i}{\partial x}(x, y) \right]^2 \right. \right. \\ &\quad \left. \left. + q(x, y) \left[\sum_{i=1}^m \gamma_i \frac{\partial \phi_i}{\partial y}(x, y) \right]^2 - r(x, y) \left[\sum_{i=1}^m \gamma_i \phi_i(x, y) \right]^2 \right\} \right. \\ &\quad \left. + f(x, y) \sum_{i=1}^m \gamma_i \phi_i(x, y) \right\} dy dx \\ &+ \int_{\mathcal{S}_2} \left\{ -g_2(x, y) \sum_{i=1}^m \gamma_i \phi_i(x, y) \right. \\ &\quad \left. + \frac{1}{2} g_1(x, y) \left[\sum_{i=1}^m \gamma_i \phi_i(x, y) \right]^2 \right\} dS. \end{aligned}$$

برای اینکه مینیممی اتفاق افتد، با در نظر گرفتن I به عنوان تابعی از $\gamma_1, \dots, \gamma_n$ لازم است داشته باشیم

$$(52.11) \quad \frac{\partial I}{\partial \gamma_j} = 0, \quad j = 1, 2, \dots, n$$

مشتگیری از (51.11) نتیجه می دهد

$$\begin{aligned} \frac{\partial I}{\partial \gamma_j} = & \iint_D \left\{ p(x, y) \sum_{i=1}^m \gamma_i \frac{\partial \phi_i}{\partial x}(x, y) \frac{\partial \phi_j}{\partial x}(x, y) \right. \\ & + q(x, y) \sum_{i=1}^m \gamma_i \frac{\partial \phi_i}{\partial y}(x, y) \frac{\partial \phi_j}{\partial y}(x, y) \\ & - r(x, y) \sum_{i=1}^m \gamma_i \phi_i(x, y) \phi_j(x, y) + f(x, y) \phi_j(x, y) \left. \right\} dx dy \\ & + \int_{S_2} \left\{ -g_2(x, y) \phi_j(x, y) \right. \\ & \left. + g_1(x, y) \sum_{i=1}^m \gamma_i \phi_i(x, y) \phi_j(x, y) \right\} dS: \end{aligned}$$

لذا، باید داشته باشیم

$$, \quad j = 1, 2, \dots, n$$

$$(53.11) \quad 0 = \sum_{i=1}^m \left[\iint_D \left\{ p(x, y) \frac{\partial \phi_i(x, y)}{\partial x} \frac{\partial \phi_j(x, y)}{\partial x} \right. \right. \\ \left. \left. + q(x, y) \frac{\partial \phi_i(x, y)}{\partial y} \frac{\partial \phi_j(x, y)}{\partial y} \right. \right. \\ \left. \left. - r(x, y) \phi_i(x, y) \phi_j(x, y) \right\} dx dy \right. \\ \left. + \int_{S_2} g_1(x, y) \phi_i(x, y) \phi_j(x, y) dS \right] \gamma_i \\ + \iint_D f(x, y) \phi_j(x, y) dx dy - \int_{S_2} g_2(x, y) \phi_j(x, y) dS$$

این دسته از معادلات را می توان به صورت یک دستگاه خطی نوشت:

$$(54.11) \quad A\mathbf{c} = \mathbf{b},$$

که در آن $A = (a_{ij})$ ، $\mathbf{c} = (\gamma_1, \dots, \gamma_n)'$ و $\mathbf{b} = (\beta_1, \dots, \beta_n)'$ به صورت زیر تعریف می شوند:

$$i = 1, 2, \dots, n, j = 1, 2, \dots, m$$

$$(55.11) \quad \alpha_{ij} = \iint_D \left[p(x, y) \frac{\partial \phi_i(x, y)}{\partial x} \frac{\partial \phi_j(x, y)}{\partial x} + q(x, y) \frac{\partial \phi_i(x, y)}{\partial y} \frac{\partial \phi_j(x, y)}{\partial y} - r(x, y) \phi_i(x, y) \phi_j(x, y) \right] dx dy + \int_{S_2} g_1(x, y) \phi_i(x, y) \phi_j(x, y) dS$$

و به ازای هر $i = 1, \dots, n$

$$(56.11) \quad \beta_i = - \iint_D f(x, y) \phi_i(x, y) dx dy + \int_{S_2} g_2(x, y) \phi_i(x, y) dS - \sum_{k=n+1}^m \alpha_{ik} \gamma_k.$$

انتخاب خاص توابع پایه بسیار مهم است، زیرا انتخاب مناسب اغلب می تواند ماتریس A را معین مثبت و کراندار سازد. برای مسئله مرتبه دوم (47.11) فرض خواهیم کرد که D به صورت چندضلعی و S_1 مجموعه ای پیوسته از خطوط مستقیم است. برای شروع روند ناحیه D را به گردآیهای از مثلثهای T_1, T_2, \dots, T_M تقسیم می کنیم که مثلث i م که دارای سه راس، یا سه گره است چنین نمایش داده می شوند:

$$V_j^{(i)} = (x_j^{(i)}, y_j^{(i)}), \quad j = 1, 2, 3$$

برای ساده کردن نمادگذاری، $V_j^{(i)}$ را، وقتی با مثلث ثابت T_i کار می کنیم، مختصراً "به صورت $(x_j, y_j) = V_j$ می نویسیم. به هر راس V_j یک چندجمله ای خطی مربوط می کنیم:

$$N_j^{(i)} \equiv N_j = a_j + b_j x + c_j y,$$

که در آن

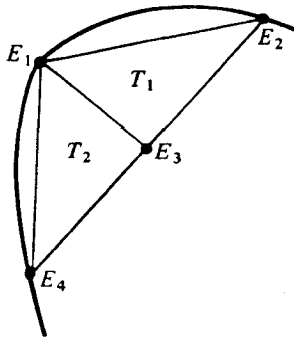
$$N_j^{(i)}(x_k, y_k) = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases} \text{ اگر}$$

این امر دستگانهایی خطی به شکل زیر را حاصل می کند:

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} \begin{bmatrix} a_j \\ b_j \\ c_j \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

که مولفه یک در سطر z م بردار طرف راست اتفاق می افتد.

فرض کنیم E_1, \dots, E_n شماره گذاری گرههای واقع در $\mathcal{S} \cup D$ از چپ به راست و از بالا به پایین باشد. به هر گره E_k یک تابع ϕ_k مربوط می کنیم که بر هر مثلث خطی است، و در E_k مقدار یک دارد و در هر یک از گرههای دیگر صفر است. این انتخاب، زمانی که گره E_k راس نمایش داده شده با $V_j^{(i)}$ باشد، ϕ_k را بر T_i با $N_j^{(i)}$ یکی می کند. برای تولید درایه های ماتریس A به عنوان مثال شکل ۱۵.۱۱ را در نظر می گیریم که جزء چپ بالایی ناحیهء نموده شده در شکل ۱۴.۱۱ است. برای سادگی فرض می کنیم



شکل ۱۵.۱۱

E_1 یکی از گرههای واقع بر \mathcal{S}_2 نباشد. برای این جزء رابطهء بین گرهها و رعوس مثلثها عبارت است از

$$E_1 = V_3^{(1)} = V_1^{(2)}, \quad E_2 = V_2^{(2)},$$

$$E_3 = V_2^{(1)} = V_3^{(2)}, \quad E_4 = V_1^{(1)}$$

چون ϕ_1 و ϕ_3 هر دو بر T_1 و T_2 ناصفر هستند، درایه های $\alpha_{1,3} = \alpha_{3,1}$ چنین محاسبه خواهند شد:

$$\alpha_{1,3} = \iint_D \left[p \frac{\partial \phi_1}{\partial x} \frac{\partial \phi_3}{\partial x} + q \frac{\partial \phi_1}{\partial y} \frac{\partial \phi_3}{\partial y} - r \phi_1 \phi_3 \right] dx dy$$

$$\begin{aligned}
&= \iint_{T_1} \left[p \frac{\partial \phi_1}{\partial x} \frac{\partial \phi_3}{\partial x} + q \frac{\partial \phi_1}{\partial y} \frac{\partial \phi_3}{\partial y} - r \phi_1 \phi_3 \right] dx dy \\
&+ \iint_{T_2} \left[p \frac{\partial \phi_1}{\partial x} \frac{\partial \phi_3}{\partial x} + q \frac{\partial \phi_1}{\partial y} \frac{\partial \phi_3}{\partial y} - r \phi_1 \phi_3 \right] dx dy \\
&= b_3^{(1)} b_2^{(1)} \iint_{T_1} p dx dy + c_3^{(1)} c_2^{(1)} \iint_{T_1} q dx dy \\
&- \iint_{T_1} r (a_3^{(1)} + b_3^{(1)} x + c_3^{(1)} y) (a_2^{(1)} + b_2^{(1)} x + c_2^{(1)} y) dx dy \\
&+ b_1^{(2)} b_3^{(2)} \iint_{T_2} p dx dy + c_1^{(2)} c_3^{(2)} \iint_{T_2} q dx dy \\
&- \iint_{T_2} r (a_1^{(2)} + b_1^{(2)} x + c_1^{(2)} y) (a_3^{(2)} + b_3^{(2)} x + c_3^{(2)} y) dx dy.
\end{aligned}$$

قسمتی از درایه β_1 از تحدید ϕ_1 به T_1 و بقیه به وسیله تحدید ϕ_1 به T_2 حاصل می شود. در حالت کلی، یک درایه β_k سهمهایی از تحدید ϕ_k به هریک از مثلثهایی E_k یک راس آنهاست خواهد داشت. بعلاوه، گرههایی که بر \mathcal{S}_2 قرار دارند انتگرالهایی خطی دارند که به درایه های آنها در A و b اضافه می شوند.

الگوریتم زیر روش عناصر متناهی روی یک معادله دیفرانسیل بیضوی مرتبه دوم را اجرا می کند. الگوریتم در ابتدا تمام مقادیر ماتریس A و بردار b را صفر قرار می دهد و بعد از اینکه انتگرالگیریها بر تمام مثلثها انجام شدند، این مقادیر را به درایه های مناسب در A و b اضافه می کند.

الگوریتم عناصر متناهی ۶.۱۱

برای حل معادله دیفرانسیل جزئی

$$\frac{\partial}{\partial x} \left(p(x, y) \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial y} \left(q(x, y) \frac{\partial u}{\partial y} \right) - r(x, y) u(x, y) = f(x, y), \quad (x, y) \in D$$

تحت شرایط

$$u(x, y) = g(x, y)$$

بر \mathcal{S}_1 ، و

$$p(x, y) \frac{\partial u}{\partial x} \cos \theta_1 + q(x, y) \frac{\partial u}{\partial y} \cos \theta_2 + g_1(x, y)u(x, y) = g_2(x, y)$$

بر \mathcal{S}_2 ، که در آن $\mathcal{S}_1 \cup \mathcal{S}_2$ مرز D است و θ_1 و θ_2 زوایای جهتی قائم بر مرز است ، مرحله ۱ . ناحیه D را به مثلثهای T_1, T_2, \dots, T_M که رئوس آنها در $\mathcal{S}_2 \cup D$ با E_1, E_2, \dots, E_n بر \mathcal{S}_1 با $E_{n+1}, E_{n+2}, \dots, E_m$ شماره گذاری شده اند ، تقسیم می کنیم . فرض کنیم T_1, T_2, \dots, T_K نمایانگر آن مثلثهایی باشد که تمام رئوس آنها در داخل D باشند ، $T_{K+1}, T_{K+2}, \dots, T_N$ آنهایی که حداقل یک ضلعشان در امتداد \mathcal{S}_2 باشد ، و $T_{N+1}, T_{N+2}, \dots, T_M$ نمایانگر بقیه مثلثها باشند .

سه رأس مثلث T_i را با

$$(x_1^{(i)}, y_1^{(i)}) , (x_2^{(i)}, y_2^{(i)}) , \text{ و } (x_3^{(i)}, y_3^{(i)})$$

شماره گذاری می کنیم .

مرحله ۲ . به ازای هر $l = n + 1, n + 2, \dots, m$ قرار می دهیم $\gamma_l = g(x_l, y_l)$ که در آن $E_l = (x_l, y_l)$.

مرحله ۳ . به ازای هر $i, j = 1, 2, \dots, n$ قرار می دهیم $x_{i,j} = 0$.
 به ازای هر $i = 1, 2, \dots, n$ قرار می دهیم $\beta_i = 0$.

مرحله ۴ . به ازای هر $i = 1, 2, \dots, M$ محاسبه می کنیم

$$\Delta_i = \frac{1}{2} \det \begin{vmatrix} 1 & x_1^{(i)} & y_1^{(i)} \\ 1 & x_2^{(i)} & y_2^{(i)} \\ 1 & x_3^{(i)} & y_3^{(i)} \end{vmatrix}$$

$$a_1^{(i)} = \frac{x_2^{(i)}y_3^{(i)} - y_2^{(i)}x_3^{(i)}}{2\Delta_i}, \quad b_1^{(i)} = \frac{y_2^{(i)} - y_3^{(i)}}{2\Delta_i}, \quad c_1^{(i)} = \frac{x_3^{(i)} - x_2^{(i)}}{2\Delta_i}$$

$$a_2^{(i)} = \frac{x_3^{(i)}y_1^{(i)} - y_3^{(i)}x_1^{(i)}}{2\Delta_i}, \quad b_2^{(i)} = \frac{y_3^{(i)} - y_1^{(i)}}{2\Delta_i}, \quad c_2^{(i)} = \frac{x_1^{(i)} - x_3^{(i)}}{2\Delta_i}$$

$$a_3^{(i)} = \frac{x_1^{(i)}y_2^{(i)} - y_1^{(i)}x_2^{(i)}}{2\Delta_i}, \quad b_3^{(i)} = \frac{y_1^{(i)} - y_2^{(i)}}{2\Delta_i}, \quad c_3^{(i)} = \frac{x_2^{(i)} - x_1^{(i)}}{2\Delta_i}$$

مرحله ۵ . به ازای هر $i = 1, 2, \dots, M$ و $j = 1, 2, 3$ قرار می دهیم

$$N_j^{(i)}(x, y) = a_j^{(i)} + b_j^{(i)}x + c_j^{(i)}y$$

مرحله ۶. بهازای هر $i = 1, 2, \dots, M$ ، $j = 1, 2, 3$ ، و $k = 1, 2, 3$ قرار می‌دهیم

$$Z_{j,k}^{(i)} = b_j^{(i)} b_k^{(i)} \iint_{T_i} p(x, y) dx dy + c_j^{(i)} c_k^{(i)} \iint_{T_i} g(x, y) dx dy - \iint_{T_i} r(x, y) N_j^{(i)}(x, y) N_k^{(i)}(x, y) dx dy$$

و

$$H_j^{(i)} = - \iint_{T_i} f(x, y) N_j^{(i)}(x, y) dx dy.$$

مرحله ۷. بهازای هر $i = K + 1, K + 2, \dots, N$ ، $j = 1, 2, 3$ ، و $k = 1, 2, 3$ قرار می‌دهیم

$$J_{j,k}^{(i)} = \int_{\mathcal{S}_2} g_1(x, y) N_j^{(i)}(x, y) N_k^{(i)}(x, y) dS$$

و

$$I_j^{(i)} = \int_{\mathcal{S}_2} g_2(x, y) N_j^{(i)}(x, y) dS,$$

که در آن انتگرالهای خطی روی تمام ضلعهای T_i که رئوس واقع در \mathcal{S}_2 را بهم وصل می‌کنند در جهت مثبت گرفته می‌شوند.

مرحله ۸. بهازای هر $i = 1, 2, \dots, M$ ، $l = 1, 2, \dots, n$ ، $j = 1, 2, 3$ ، و $k = 1, 2, 3$ (یک) بهازای هر $n, p = 1, 2, \dots, n$ ، وقتی z مین راس T_i عبارت است از E_p و k امین راس T_i عبارت است از E_l ، $Z_{k,j}^{(i)}$ را به $\alpha_{l,p}$ می‌افزاییم.

(دو) بهازای هر $p = n + 1, \dots, m$ ، $Z_{k,j}^{(i)}$ را در γ_p ضرب می‌کنیم و مقدار حاصل را، وقتی z مین راس T_i عبارت است از E_p و k امین راس T_i عبارت است از E_l ، از β_l کسر می‌کنیم.

مرحله ۹. بهازای هر $i = 1, 2, \dots, M$ ، $l = 1, 2, \dots, n$ ، و $k = 1, 2, 3$ ، وقتی k امین راس T_i عبارت است از E_l ، $H_k^{(i)}$ را به β_l می‌افزاییم.

مرحله ۱۰. بهازای هر $i = K + 1, \dots, N$ ، $l = 1, 2, \dots, n$ ، $j = 1, 2, 3$ ، و $k = 1, 2, 3$

(یک) به ازای هر $p = 1, \dots, n$ ، وقتی z مین راس T_i عبارت است از E_p و k مین راس T_i عبارت است از E_l ، $J_{k,j}^{(i)}$ را به $\alpha_{l,p}$ می افزاییم
 (دو) به ازای هر $p = n+1, \dots, m$ ، $J_{k,j}^{(i)}$ را در γ_p ضرب می کنیم و این مقدار را، وقتی z مین راس T_i عبارت است از E_p و k امین راس T_i عبارت است از E_l ، از β_l کسر می کنیم.

مرحله ۱۱. به ازای هر $i = K+1, \dots, N$ ، $l = 1, 2, \dots, n$ ، و $k = 1, 2, 3$ ، وقتی k مین راس T_i عبارت است از E_l ، $I_k^{(i)}$ را به β_l می افزاییم.
 مرحله ۱۲. دستگاه

$$Ac = b$$

را، که در آن $A = (\alpha_{l,p})$ ، $b = (\beta_l)$ ، و $c = (\gamma_l)$ ، به ازای $l = 1, 2, \dots, n$ و $p = 1, 2, \dots, n$ ، حل می کنیم.

مرحله ۱۳. به ازای هر $k = 1, 2, \dots, m$ ، اگر E_k عبارت باشد از z مین راس T_i ، بر T_i تعریف می کنیم $\phi_k = N_j^{(i)}$.

مرحله ۱۴. روند تمام است، تابع قطعه قطعه خطی

$$\phi(x, y) = \sum_{i=1}^m \gamma_i \phi_i(x, y)$$

$u(x, y)$ را بر $\mathcal{S}_1 \cup \mathcal{S}_2 \cup D$ تقریب می کند.

مثال ۱. دمای $u(x, y)$ در ناحیه دو بعدی D در معادله لاپلاس زیر صدق می کند:

$$D \text{ بر } \frac{\partial^2 u}{\partial x^2}(x, y) + \frac{\partial^2 u}{\partial y^2}(x, y) = 0.$$

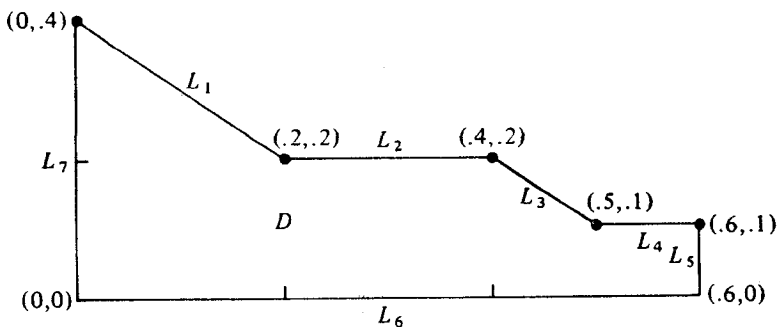
ناحیه D نموده شده در شکل ۱۶.۱۱ را در نظر می گیریم و فرض می کنیم شرایط مرزی زیر داده شده باشند.

$$\text{برای } (x, y) \in L_6 \text{ و } (x, y) \in L_7 \text{، } u(x, y) = 4$$

$$\text{برای } (x, y) \in L_2 \text{ و } (x, y) \in L_4 \text{، } \frac{\partial u}{\partial n}(x, y) = x$$

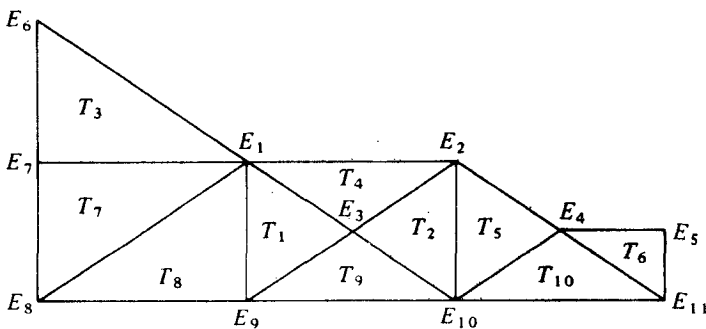
$$\text{برای } (x, y) \in L_5 \text{، } \frac{\partial u}{\partial n}(x, y) = y$$

$$\frac{\partial u}{\partial n}(x, y) = \frac{x+y}{\sqrt{2}}, \quad (x, y) \in L_3 \text{ و } (x, y) \in L_1 \text{ برای}$$



شکل ۱۶.۱۱

که در آن $(\partial u / \partial n)$ نمایانگر مشتق قائم بر مرز ناحیه D در نقطه (x, y) است. ابتدا D را به مثلثهایی، با شمارهگذاری پیشنهاد شده در مرحله اول الگوریتم، تقسیم می‌کنیم. برای این مثال، $\mathcal{S}_1 = L_6 \cup L_7$ و $\mathcal{S}_2 = L_1 \cup L_2 \cup L_3 \cup L_4 \cup L_5$ شمارهگذاری مثلثها در شکل ۱۷.۱۱ نموده شده است.



شکل ۱۷.۱۱

شرط مرزی $u(x, y) = 4$ بر L_6 و L_7 نتیجه می‌دهد که وقتی $l = 6, 7, \dots, 11$ ، $\tau_l = 4$. برای تعیین مقادیر τ_l ، به‌ازای $l = 1, 2, \dots, 5$ ، بقیه مراحل الگوریتم را اجرا کرده ماتریس و بردار زیر را تولید می‌کنیم:

$$A = \begin{bmatrix} 2.5 & 0 & -1 & 0 & 0 \\ 0 & 1.5 & -1 & -.5 & 0 \\ -1 & -1 & 4 & 0 & 0 \\ 0 & -.5 & 0 & 2.5 & -.5 \\ 0 & 0 & 0 & -.5 & 1 \end{bmatrix}$$

$$b = \begin{bmatrix} 6.0\overline{666} \\ 0.06\overline{33} \\ 8.0000 \\ 6.05\overline{66} \\ 2.031\overline{6} \end{bmatrix}$$

جواب معادله $Ac = b$ عبارت است از

$$c = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \end{bmatrix} = \begin{bmatrix} 4.0383 \\ 4.0782 \\ 4.0291 \\ 4.0496 \\ 4.0565 \end{bmatrix}$$

که تقریبهای زیر را برای جواب معادله لاپلاس و شرایط مرزی بر مثلثهای متناظر می دهد:

$$T_1 : \phi(x, y) = 4.0383(1 - 5x + 5y) + 4.0291(-2 + 10x) + 4(2 - 5x - 5y),$$

$$T_2 : \phi(x, y) = 4.0782(-2 + 5x + 5y) + 4.0291(4 - 10x) + 4(-1 + 5x - 5y),$$

$$T_3 : \phi(x, y) = 4(-1 + 5y) + 4(2 - 5x - 5y) + 4.0383(5x),$$

$$T_4 : \phi(x, y) = 4.0383(1 - 5x + 5y) + 4.0782(-2 + 5x + 5y) + 4.0291(2 - 10y),$$

$$T_5 : \phi(x, y) = 4.0782(2 - 5x + 5y) + 4.0496(-4 + 10x) + 4(3 - 5x - 5y),$$

$$T_6 : \phi(x, y) = 4.0496(6 - 10x) + 4.0565(-6 + 10x + 10y) + 4(1 - 10y),$$

$$T_7 : \phi(x, y) = 4(-5x + 5y) + 4.0383(5x) + 4(1 - 5y),$$

$$T_8 : \phi(x, y) = 4.0383(5y) + 4(1 - 5x) + 4(5x - 5y),$$

$$T_9 : \phi(x, y) = 4.0291(10y) + 4(2 - 5x - 5y) + 4(-1 + 5x - 5y),$$

$$T_{10} : \phi(x, y) = 4.0496(10y) + 4(3 - 5x - 5y) + 4(-2 + 5x - 5y).$$

جواب واقعی مسئله مقدار مرزی عبارت است از $u(x, y) = xy + 4$. جدول ۷.۱۱ مقدار u را با مقدار ϕ در E_i ، به ازای $i = 1, \dots, 5$ ، مقایسه می کند .

x	y	$\phi(x, y)$	$u(x, y)$	$ \phi(x, y) - u(x, y) $
.2	.2	4.0383	4.04	.0017
.4	.2	4.0782	4.08	.0018
.3	.1	4.0291	4.03	.0009
.5	.1	4.0496	4.05	.0004
.6	.1	4.0565	4.06	.0035

جدول ۷.۱۱

خطا برای مسائل مرتبه دوم با توابع ضریبی هموار نوعاً " $O(h^2)$ " است ، که در آن h ماکزیم وتر عناصر مثلثی است . انتظار می رود که توابع پایه قطعه قطعه دوخطی بر عناصر مستطیلی نیز نتایج $O(h^2)$ بدهند ، که در آن h ماکزیم طول عناصر مستطیلی است . رده های دیگری از توابع پایه را می توان بکار برد که نتایج $O(h^4)$ بدهند ، اما ساختن آنها مشکلتر است .

برای روشهای عناصر متناهی مشکل است که قضایای خطای کارا وضع و اعمال کرد ، زیرا دقت تقریب به خواص پیوستگی جواب و منظم بودن مرز وابسته است .

روش عناصر متناهی را می توان برای معادلات دیفرانسیل جزئی هذلولوی و سهموی نیز بکار برد ، اما روند مینیم سازی مشکلتر است و روندهای تفاضل متناهی بیشتر قابل رقابت هستند .

مروری خوب از مزیتها و تکنیکهای روش عناصر متناهی بکار رفته در مسائل فیزیکی گوناگون را می توان در مقاله های توسط فیکس^۱ [۲۷] یافت . برای بحث گسترده تر به کتاب استرانگ^۲ و فیکس [۶۲] یا زینکوویچ^۳ [۷۱] مراجعه کنید .

مجموعه تمرینات ۵.۱۱

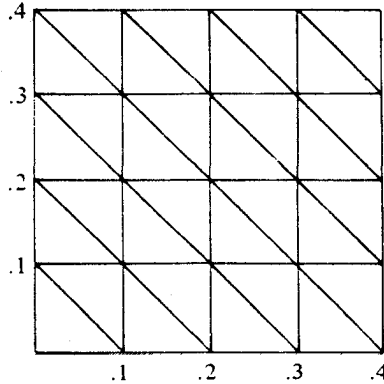
۱ . جواب معادله دیفرانسیل جزئی

$$\frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} - 12.5\pi^2 u(x, y) = -25\pi^2 \sin \frac{5\pi}{2} x \sin \frac{5\pi}{2} y, 0 < x, y < .4,$$

تحت شرایط مرزی دیریکله

$$u(x, y) = 0$$

را با استفاده از الگوریتم عناصر متناهی ۶۰۱۱، با عناصر داده شده در شکل ۱۸۰۱۱،



شکل ۱۸۰۱۱

تقریب کنید و جواب تقریبی را با جواب واقعی

$$u(x, y) = \sin \frac{5\pi}{2} x \sin \frac{5\pi}{2} y$$

در رئوس داخلی و در نقاط

$$(.125, .125), (.125, .25), (.25, .125), \text{ and } (.25, .25)$$

مقایسه نمایید.

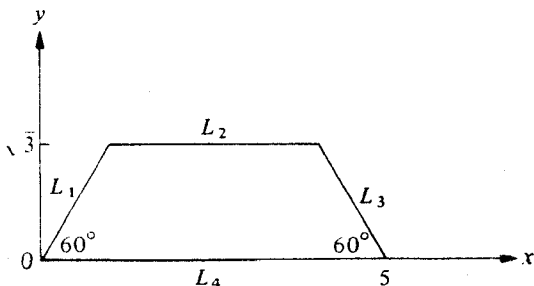
۲. تمرین ۱ را با $f(x, y) = -25\pi^2 \cos \frac{5\pi}{2} x \cos \frac{5\pi}{2} y$ و با استفاده از شرط مرزی نیومن

$$\frac{\partial u}{\partial n}(x, y) = 0$$

تکرار کنید. جواب واقعی برای این مسئله عبارت است از

$$u(x, y) = \cos \frac{5\pi}{2} x \cos \frac{5\pi}{2} y.$$

۳. الگوریتمی برای روش عناصر منتهای، با استفاده از عناصر مستطیلی و توابع پایه شکل $a + bx + cy + dxy$ ، بسازید.
۴. با استفاده از الگوریتم وضع شده در تمرین ۳، مسئله بحث شده در (آ) تمرین ۱ (ب) تمرین ۲ را حل کنید.
۵. صفحهای نقره‌ای به شکل دوزنقه (ر.ک. شکل ۱۹.۱۱) به‌طور یکنواخت در هر



شکل ۱۹.۱۱

نقطه گرمایی با سرعت ۱.۵ ثانیه. سانتیمتر مکعب/کالری q تولید می‌کند. دمای صفحہ در حالت یکنواخت در معادله پواسن صدق می‌کند.

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{-q}{k}$$

که در آن k ، یعنی هدایت گرمایی، ۱.۰۴ ثانیه. درجه. سانتیمتر/کالری است. فرض کنید دما بر L_2 در ۱۵ سانتیگراد نگهداشته شده باشد، که گرما در لبه‌های اریب L_1 و L_3 برحسب شرط مرزی $(\partial u / \partial n) = 4$ از دست می‌رود، و اینکه هیچ گرمایی بر L_4 از دست نمی‌رود؛ یعنی $(\partial u / \partial n) = 0$. دمای صفحہ را، با استفاده از الگوریتم ۶.۱۱، در $(1, 0)$ ، $(4, 0)$ ، و $(\frac{5}{2}, \sqrt{3})$ تقریب کنید.

کتابنامه

- [1.] ABBOTT, M. B. (1966), *The method of characteristics*. Thames and Hudson, London; 243 pp.
- [2.] ACTON, F. S. (1970), *Numerical methods that work*. Harper & Row, New York; 541 pp.
- [3.] AHLFORS, L. V. (1966), *Complex analysis* (Second edition). McGraw-Hill, New York; 317 pp.
- [4.] AMES, W. F. (1969), *Numerical methods for partial differential equations*. Barnes & Noble, New York; 291 pp.
- [5.] BAILEY, N. T. J. (1967), *The mathematical approach to biology and medicine*. John Wiley & Sons, London; 296 pp.
- [6.] BAILEY, N. T. J. (1957), *The mathematical theory of epidemics*. C. Griffin, London; 194 pp.
- [7.] BAILEY, P. B., L. F. SHAMPINE, and P. E. WALTMAN (1968), *Nonlinear two-point boundary-value problems*. Academic Press, New York; 171 pp.
- [8.] BARTLE, R. G. (1976), *The elements of real analysis* (Second edition). John Wiley & Sons, New York; 480 pp.
- [9.] BEKKER, M. G. (1969), *Introduction to terrain vehicle systems*. University of Michigan Press, Ann Arbor, Mich.; 846 pp.
- [10.] BERNADELLI, H. (1941), "Population Waves." *Journal of the Burma Research Society*, 31, 1-18.
- [11.] BIRKHOFF, G., and G. ROTA (1962), *Ordinary differential equations*. Blaisdell, Waltman, Mass.; 317 pp.
- [12.] BROGAN, W. L. (1974), *Modern control theory*. Quantum Publishers, New York; 393 pp.
- [13.] BULIRSCH, R., and J. STOER (1966), "Numerical treatment of ordinary differential equations by extrapolation methods." *Numerische Mathematik*, 8, 1-13.
- [14.] BULIRSCH, R., and J. STOER (1968), "Fehlerabschätzungen und extrapolation mit rationalen Funktionen bei Verfahren von Richardson-typus." *Numerische Mathematik*, 6, 413-427.
- [15.] BULIRSCH, R., and J. STOER (1968), "Asymptotic upper and lower bounds for

- results of extrapolation methods." *Numerische Mathematik*, **8**, 93–104.
- [16.] BUTCHER, J. C. (1965), "On the attainable order of Runge-Kutta methods." *Mathematics of Computation*, **19**, 408–417.
- [17.] CANTONI, A., and P. BUTLER (1976), "Properties of the eigenvectors of per-symmetric matrices with applications to communication theory." *IEEE Transactions on Communications*, Vol. Com-24, **8**, 804–809.
- [18.] CHEN, B. H. (1976), "Holdup and axial mixing in bubble columns containing screen cylinders." *Industrial and Engineering Chemistry, Process Design and Development*, **15**, No. 1, 20–24.
- [19.] CHIARELLA, C., W. CHARLTON, and A. W. ROBERTS (1975), "Optimum chute profiles in gravity flow of granular materials: A discrete segment solution method." *Transactions of the ASME, Journal of Engineering for Industry, Series B*, **97**, 10–13.
- [20.] CHU, Y. (1970). *Introduction to computer organization*. Prentice-Hall, Englewood Cliffs, N.J.: 376 pp.
- [21.] DORN, G. L. and A. B. BURDICK (1962), "On the recombinational structure of complementation relationships in the *m-dy* complex of the *Drosophila melanogaster*." *Genetics*, **47**, 503–518.
- [22.] DETTMAN, J. W. (1965), *Applied complex variables*. Macmillan, New York; 481 pp.
- [23.] ENRIGHT, W. H., T. E. HULL, and B. LINDBERG (1975), "Comparing numerical methods for stiff systems of O.D.E.'s." *BIT*, **15**, 10–48.
- [24.] FEHLBERG, E. (1964). "New high-order Runge-Kutta formulas with step-size control for systems of first- and second-order differential equations." *ZAMM*, **44**, 17–29.
- [25.] FEHLBERG, E. (1966), "New high-order Runge-Kutta formulas with an arbitrarily small truncation error." *Zeitschrift für Angewandte Mathematik und Mechanik*, **46**, 1–16.
- [26.] FEHLBERG, E. (1970), "Klassische Runge-Kutta Formeln vierter und niedrigerer Ordnung mit Schrittweiten-Kontrolle und ihre Anwendung auf Wärmeleitungsprobleme." *Computing*, **6**, 61–71.
- [27.] FIX, G. (1975), "A survey of numerical methods for selected problems in continuum mechanics." *Proceedings of a Conference on Numerical Methods of Ocean Circulation, National Academy of Sciences, Durham, N. H.*, October 17–20, 1972, 268–283.
- [28.] FORSYTHE, G. E., and C. B. MOLER (1967), *Computer solution of linear algebraic systems*. Prentice-Hall, Englewood Cliffs, N.J.; 148 pp.
- [29.] FULKS, W. (1969), *Advanced calculus* (Second edition). John Wiley & Sons, New York; 597 pp.
- [30.] GEAR, C. W. (1971), *Numerical initial-value problems in ordinary differential equations*. Prentice-Hall, Englewood Cliffs, N.J.; 253 pp.
- [31.] GEORGE, J. A. (1973), "Nested dissection of a regular finite-element mesh." *SIAM Journal of Numerical Analysis*, **10**, No. 2, 345–362.
- [32.] GRAGG, W. B. (1965), "On extrapolation algorithms for ordinary initial-value problems," *SIAM Journal of Numerical Analysis*, **2**, 384–403.

- [33.] HENRICI, P. (1962), *Discrete variable methods in ordinary differential equations*. John Wiley & Sons, New York; 407 pp.
- [34.] HENRICI, P. (1964), *Elements of numerical analysis*. John Wiley & Sons, New York; 328 pp.
- [35.] HENRICI, P. (1963), *Error propagation for difference methods*. John Wiley & Sons, New York; 73 pp.
- [36.] HILDEBRAND, F. B. (1974), *Introduction to numerical analysis* (Second edition). McGraw-Hill, New York; 511 pp.
- [37.] HULL, T. E., and W. H. ENRIGHT (1976), "Test results on initial-value methods for nonstiff ordinary differential equations." *SIAM Journal of Numerical Analysis*, 13, No. 6, 944-961.
- [38.] HULL, T. E., W. H. ENRIGHT, B. M. FELLEN, and A. E. SEDGEWICK (1972), "Comparing numerical methods for ordinary differential equations." *SIAM Journal of Numerical Analysis*, 9, No. 4, 603-637.
- [39.] ISAACSON, E., and H. B. KELLER (1966), *Analysis of numerical methods*. John Wiley & Sons, New York; 541 pp.
- [40.] KELLER, H. B. (1968), *Numerical methods for two-point boundary-value problems*. Blaisdell, London; 184 pp.
- [41.] KAMMERER, W. J., G. W. REDDIEN, and R. S. VARGA (1974), "Quadratic splines." *Numerische Mathematik*, 22, 241-259.
- [42.] LARSON, H. J. (1974), *Introduction to probability theory and statistical inference* (Second edition). John Wiley & Sons, New York; 430 pp.
- [43.] MITCHELL, A. R. (1969), *Computational methods for partial-differential equations*. John Wiley & Sons, London; 255 pp.
- [44.] NA, T. Y., and G. M. KURAJIAN (1976), "Initial-curvature and lateral-load effects on thin struts with large elastic displacements." *Transactions of the ASME, Journal of Engineering for Industry, Series B*, 98, 34-38.
- [45.] NOBLE, B. (1969), *Applied linear algebra*. Prentice-Hall, Englewood Cliffs, N.J.; 523 pp.
- [46.] ORTEGA, J. M. (1972), *Numerical analysis—A second course*. Academic Press, New York; 193 pp.
- [47.] ORTEGA, J. M., and W. C. RHEINBOLDT (1970), *Iterative solution of nonlinear equations in several variables*. Academic Press, New York; 572 pp.
- [48.] PANDIT, S. M., T. L. SUBRAMANIAN, and S. M. WU (1975), "Modeling machine-tool chatter by time series." *Transactions of the ASME, Journal of Engineering for Industry, Series B*, 97, 211-215.
- [49.] RALSTON, A. (1965), *A first course in numerical analysis*. McGraw-Hill, New York; 541 pp.
- [50.] RALSTON, A., and H. S. WILF, ed. (1960 and 1967), *Numerical methods for digital computers*. Vols. 1 and 2. John Wiley & Sons, New York; 293 + 287 pp.
- [51.] RASHEVSKY, N. (1968), *Looking at history through mathematics*. Massachusetts Institute of Technology Press, Cambridge, Mass.; 199 pp.
- [52.] SALE, P. F., and R. DYBDALH (1975), "Determinants of community structure for coral-reef fishes in experimental habitat." *Ecology*, 56, 1345-1355.

- [53.] SAGAR, V., and D. J. PAYNE (1975), "Incremental collapse of thick-walled circular cylinders under steady axial tension and torsion loads and cyclic transient heating." *Journal of the Mechanics and Physics of Solids*, 21, No. 1, 39-54.
- [54.] SCHOENBERG, I. J. (1946), "Contributions to the problem of approximation of equidistant data by analytic functions." *Quarterly of Applied Mathematics*, 4, Part A, 45-99; Part B, 112-141.
- [55.] SCHROEDER, L. A. (1973), "Energy budget of the larvae of the moth *Pachysphinx modesta*," *Oikos*, 24, 278-281.
- [56.] SCHULTZ, M. H. (1966), *Spline analysis*. Prentice-Hall, Englewood Cliffs, N.J.: 156 pp.
- [57.] SEARLE, S. R. (1966), *Matrix algebra for the biological sciences*. John Wiley & Sons, New York; 296 pp.
- [58.] SECRIST, D. A., and R. W. HORNBECK (1976), "An analysis of heat transfer and fade in disk brakes." *Transactions of the ASME, Journal of Engineering for Industry, Series B*, 98, No. 2, 385-390.
- [59.] SINGH, V. P. (1976) "Investigations of attenuation and internal friction of rocks by ultrasonics." *International Journal of Rock Mechanics and Mining Sciences*, 69-72.
- [60.] SMITH, G. D. (1965), *Numerical solution of partial-differential equations*. Oxford University Press, Oxford; 179 pp.
- [61.] STETTER, H. J. (1973), *Analysis of discretization methods for ordinary differential equations from Tracts in natural philosophy*. Springer-Verlag, Berlin: 388 pp.
- [62.] STRANG, W. G., and G. J. FIX (1973), *An analysis of the finite element method*. Prentice-Hall, Englewood Cliffs, N.J.: 306 pp.
- [63.] STROUD, A. H., and D. SECRIST (1966), *Gaussian quadrature formulas*. Prentice-Hall, Englewood Cliffs, N.J.: 384 pp.
- [64.] THOMAS, G. B. (1971), *Calculus and analytic geometry* (Alternate edition). Addison-Wesley, Reading, Massachusetts: 1034 pp.
- [65.] VARGA, R. S. (1962), *Matrix iterative analysis*. Prentice-Hall, Englewood Cliffs, N.J.; 322 pp.
- [66.] WANG, H. T. (1975), "Determination of the accuracy of segmented representations of cable shape." *Transactions of the ASME, Journal of Engineering in Industry, Series B*, 97, No. 2, 472-478.
- [67.] WENDROFF, B. (1966), *Theoretical numerical analysis*. Academic Press, New York: 239 pp.
- [68.] WILKINSON, J. H. (1963), *Rounding errors in algebraic processes*. H.M. Stationery Office, London; 161 pp.
- [69.] WILKINSON, J. H. (1965), *The algebraic eigenvalue problem*. Clarendon Press, Oxford; 662 pp.
- [70.] YOUNG, D. M. (1971), *Iterative solution of large linear systems*. Academic Press, New York; 570 pp.
- [71.] ZIENKIEWICZ, O. (1971), *The finite-element method in engineering science*. McGraw-Hill, London; 521 pp.

جواب تمرینات منتخب

فصل ۱

مجموعه تمرینات ۱.۱ (صفحه ۱۱)

۱. $f'(0)$ و $f'(1)$ را در نظر بگیرید.

$$P_3(x) = e^{-1}[1 - (x-1) + (x-1)^2/2 - (x-1)^3/6], \quad 3 \quad e^{-.99} \approx P_3(.99) = .3715766909, \quad |R_3(.99)| < 5 \times 10^{-10}$$

۵. (ت) خیر، $f'(0)$ وجود ندارد. (ب) بلی، $c = \frac{64}{27}$

(پ) خیر، $f'(0)$ وجود ندارد. (ت) بلی، $c = \frac{1}{2}$

مجموعه تمرینات ۲.۱ (صفحه ۲۵)

$$3.14002 < x < 3.14316 \quad (T) \quad 1$$

$$14.2 \quad (T) \quad 3 \quad \text{خطای نسبی } 1.3 \times 10^{-4}$$

$$20 \quad (P) \quad \text{خطای نسبی } 4.4 \times 10^{-3}$$

$$-1.827 \quad (T) \quad 7 \quad 6.959 \times 10^{-3} \quad (B)$$

$$1.0000, .33333, .11111, .037037, .012346, .0041153, .0013718, .00045727, .00015242 \quad 9$$

فصل ۲

مجموعه تمرینات ۱.۲ (صفحه ۳۵)

۱. $f'(x)$ را بر $[1, 2]$ در نظر بگیرید.

مجموعه تمرینات ۲.۲ (صفحه ۴۵)

۳. توابع مذکور در (T) و (ب) دارای نقاط ثابت منحصر به فردی بر [0, 2] اند. تابع آمده در (پ) نیز نقطه ثابت منحصر به فرد دارد مگر آنکه $m = 0$.

مجموعه تمرینات ۳.۲ (صفحه ۵۱)

۱. ۵۵. n با استفاده از نتیجه ۵.۲ به ازای $p_0 = 5$. در حقیقت پنج تکرار لازم است.

۲. ۱.32472

۷. $(T) \left(\frac{e^x}{3}\right)^{1/2}$ بر $[0, 1]$: ۰.91001

مجموعه تمرینات ۴.۲ (صفحه ۶۳)

۵. ۰.91001 (ب) ۱.96887 (ت)

۷. ۱.89549 (ب) ۰.۱۹ 33.2^2

۲۱. (T) 2.10639, 4.51469, 7.22928, 10.12546

مجموعه تمرینات ۵.۲ (صفحه ۸۱)

۱. ۱.3247 ۲. ۱.732 ۵. ۱.89549

تکرار ۱۰ م

m	دقت ساده	دقت مضاعف
1	-0.5664420	-0.5664420
2	-0.5671433	-0.5705472
3	-0.5667686	-0.5664667

جواب درست عبارت است از -0.5671433.

مجموعه تمرینات ۶.۲ (صفحه ۱۰۷)

۹. با استفاده از تکنیک توقف 10^{-6} برای k ای بسا وقتی $i = 30$ ، نتایج زیر به دست می آیند:

(T) ریشه های تقریبی: $-4.12310, -4.12307, 1.32297i, \pm 2.49998$

(پ) ریشه های تقریبی: $1.41420i, \pm 7.5 \times 10^{-6}, -0.866030i, \pm 0.500008$

فصل ۳

مجموعه تمرینات ۱.۳ (صفحه ۱۱۹)

- ۱ . $(x-1)^2 + 2(x-1) - 2$ (T)
 ۳ . 04761875. کران خطا 3.1×10^{-7} ، خطای واقعی 3.0×10^{-7} .
 ۵ . 8.86×10^{-7} $0.0953083 \cdot \gamma$ $2 \times 10^6 \cdot 11$

مجموعه تمرینات ۲.۳ (صفحه ۱۳۱)

- ۵ . 3.05048. کران خطا 1.19×10^{-4} ، خطای واقعی 1.14×10^{-4} .
 ۷ . 1.32436 (T) 2.18350 (ب) $1.15278, 2.01191$ (پ)
 ۱۳ . پیش‌بینیها عبارتند از: بهازای ۱۹۱۰، ۳۱۷۶۴؛ بهازای ۱۹۶۵، ۱۹۳۲۷۱؛ بهازای ۲۰۰۰، ۵۵۳۹۴.

مجموعه تمرینات ۳.۳ (صفحه ۱۳۹)

- ۱ . $P_{0.1.2.3.4} = 4980705$
 ۳ . 198269 $0.8095 \cdot \gamma$ $0.567142 \cdot 9$

مجموعه تمرینات ۴.۳ (صفحه ۱۴۷)

- ۳ . 1.05126 1.56830 . ۵

مجموعه تمرینات ۵.۳ (صفحه ۱۵۸)

- ۱ . 4980704 $0.80903236 \cdot ۳$
 ۵ . 1.2840246 (T) 8.3×10^{-6} (ب)

مجموعه تمرینات ۶.۳ (صفحه ۱۷۴)

- ۱ . 492174, 497627 () 498070. با استفاده از مشتقات از بخش ۵.۳.
 ۳ . 1.1070136 (ب) 1.59×10^{-7} (پ) $1.04113925, 1.04081053$ (ت)
 ۵ . معادله اسپلاین عبارت است از

$$S(x) \equiv S_i(x) = f_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$$

بر بازه $[x_i, x_{i+1}]$ ، که در آن نتایج جدول ض. ۱ به دست آمده‌اند.

جدول ض ۱۰

x_i	f_i	b_i	c_i	d_i
.9	.7	3.541214	-23.11823	77.06073
.8	.3	2.605763	32.47269	-185.3031
.7	.2	1.035753	-16.77264	164.1511
.6	0	2.251228	4.617891	-71.3018
.5	-.2	1.959329	-1.698893	21.05594
.4	-.4	1.911449	2.177681	-12.92191
.3	-.5	-.6051273	22.98808	-69.36810
.2	-.3	-2.490925	-4.130146	90.39406
.1	-.1	-1.431164	-6.467466	7.79107
0	0	-.7844177	0	21.55821

۱۳. اسپلاین مکعبی آزاد جمعیت‌های 88,219؛ 205,013؛ 524,575 را پیش‌بینی می‌کند.

فصل ۴

مجموعه تمرینات ۱۰۴ (صفحه ۱۸۵)

۳. حالت $0 \equiv f(x) \equiv 0$ یا $g(x) \equiv 0$ را جدا در نظر بگیرید. در غیر این صورت، از راهنمایی

$$\lambda = - \frac{\{\int_a^b |f|^2\}^{1/2}}{\{\int_a^b |g|^2\}^{1/2}}$$

استفاده کنید.

۵. در $C[0, 1]$ ، $\|f\|_1 \leq \|f\|_2 \leq \|f\|_\infty$

$$9. \quad \|e^x\|_1 = 31.19287, \quad \|e^x\|_2 = 21.59667, \quad \|e^x\|_\infty = 20.08554 \quad (\text{T})$$

$$(\text{ت}) \quad \|\sin x\|_1 = 1.89189, \quad \|\sin x\|_2 = 1.24692, \quad \|\sin x\|_\infty = .90930$$

مجموعه تمرینات ۲۰۴ (صفحه ۲۰۰)

$$.841679x^2 + .876603x + 1 \cdot 3 \quad . 1 \quad 1.70784x + .89968$$

$$3.87x + 25.70 \cdot 7 \quad . 5 \quad .1795x + 8.2084$$

مجموعه تمرینات ۳۰۴ (صفحه ۲۱۴)

$$P_1(x) = -x + 2.8333 \quad . 1$$

$$\psi_0(x) = 1, \quad \psi_1(x) = x, \quad \psi_2(x) = x^2 - \frac{1}{2} \quad . 3$$

$$S_2(x) = 2 \sin x - \sin 2x \quad . 9$$

مجموعه تمرینات ۴.۴ (صفحه ۲۲۵)

۳. (آ) چند جمله‌ایهای متعامد عبارتند از

$$\phi_0(x) = 1, \quad \phi_1(x) = x - .5, \quad \phi_2(x) = x^2 - x + \frac{1}{6}.$$

$$\phi_3(x) = x^3 - 1.5x^2 + .6x - .05$$

چند جمله‌ای کمترین مربعات وزندار از درجه سه عبارت است از

$$P(x) = .841471\phi_0(x) - .467546\phi_1(x) - .431010\phi_2(x) + .078824\phi_3(x)$$

با خطای L_2 ، 1.722×10^{-4} .

(پ) چند جمله‌ایهای متعامد عبارتند از

$$\phi_0(x) = 1, \quad \phi_1(x) = x - .581977,$$

$$\phi_2(x) = (x - .484843)\phi_1(x) - .0793264,$$

$$\phi_3(x) = (x - .497483)\phi_2(x) - .0671167.$$

چند جمله‌ای کمترین مربعات وزندار عبارت است از

$$P(x) = .270349\phi_0(x) + .918895\phi_1(x) + 1.84493\phi_2(x) + 2.06350\phi_3(x)$$

با خطای L_2 ، 6.114×10^{-3} .

$$Q_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3), \quad Q_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x) \quad \cdot \quad ۷$$

مجموعه تمرینات ۵.۴ (صفحه ۲۳۶)

$$P_6(x) = x + x^2 + \frac{1}{2}x^3 + \frac{1}{6}x^4 + \frac{1}{24}x^5 + \frac{1}{120}x^6 \quad \cdot \quad ۳$$

$$= .00026 + x + .99531x^2 + .5x^3 + .17917x^4 + .04167x^5 + \frac{1}{3840}T_6$$

۱.04873 \cdot ۵

فصل ۵

مجموعه تمرینات ۱.۵ (صفحه ۲۵۲)

۱. با استفاده از معادله (۱۲.۵) $f'(1.8) \approx 6.049444$ ، با خطای $10^{-4} \times 2.04$

با استفاده از معادله (۱۳.۵) $f'(1.8) \approx 6.049748$ ، با خطای $10^{-4} \times 1.01$

۲.27403, 2.27510 \cdot ۳

$f'(1.005) \approx 5.0, f(1.015) \approx 6.0$ (آ) \cdot ۵

$f''(1.01) \approx 100$ (ب) \cdot

(پ) f' به اندازه 1.0، و f'' به اندازه 200 دقیق است.

$$f'(x_0) = \frac{1}{6h} [-8f(x_0) + 9f(x_0 + h) - f(x_0 + 3h)] \quad \cdot \gamma$$

$$f'(1.8) = 6.049342 \text{ با خطای } 3.05 \times 10^{-4}$$

۱۱. (ب) معادله اسپلاین عبارت است از $s(x) = S_i(x) \equiv f_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$ بر بازه $[x_i, x_{i+1}]$ ، که نتایج جدول ض. ۲۰ در آن به دست آمده اند.

جدول ض. ۲۰

x_i	f_i	b_i	c_i	d_i
9	.7	4.652147	-3.876282	-226.4519
8	.3	2.308093	27.31677	-103.9769
.7	.2	1.115502	-15.39093	142.3590
.6	0	2.229899	4.246964	-65.45966
.5	-2	1.964892	-1.596888	19.47949
.4	-4	1.910522	2.140592	-12.45826
.3	-5	-6069813	23.03444	-69.64616
.2	-3	-2.48258	-4.278495	91.04306
.1	-1	-1.462688	-5.920434	5.473132
0	0	-6666667	-2.039782	-12.93550

مجموعه تمرینات ۲۰۵ (صفحه ۲۵۷)

۱. 6.049680 ، خطا 3.3×10^{-5}

۳. 6.0496675 ، خطا 2.0×10^{-5}

۵. $6.05, 000$ ، خطای 3.5×10^{-4} اصولاً "ناشی از خطای گرد شدن است"

۷. 6.0508125 ، خطا 1.165×10^{-3}

مجموعه تمرینات ۳۰۵ (صفحه ۲۶۴)

۱. قاعده دوزنقه 785398 . را با کران خطای 6.5×10^{-1} نتیجه می دهد، ولی بدون خطای واقعی. قاعده سیمپسون همین نتیجه را با کران خطای 1.3×10^{-1} به ما می دهد.

۳. قاعده دوزنقه 023208 . را با خطای 01160 . نتیجه می دهد. قاعده سیمپسون 032296 . را با خطای 00252 . به ما می دهد. کرانهای خطا را نمی توان محاسبه کرد، زیرا $\lim_{x \rightarrow 0} f''(x)$ و $\lim_{x \rightarrow 0} f^{(4)}(x)$ هر دو نامتناهی اند.

مجموعه تمرینات ۴۰۵ (صفحه ۲۷۱)

۱. شکل بسته 76692 را نتیجه می دهد . شکل باز 76658 را نتیجه می دهد .
۳. (T) نتایج به ترتیب خواسته شده عبارتند از: 102459819 ، 102459821 ، 102459822 ، 102466280 و 102459824 . مقدار واقعی عبارت است از 102459822 .
- (پ) نتایج به ترتیب خواسته شده عبارتند از: 1.477536117 ، 1.477528858 ، 1.477523049 ، 1.470981472 و 1.477511615 . مقدار واقعی عبارت است از 1.477523046 .
۵. خطاها بترتیب عبارتند از -1.8×10^{-6} ، 1.6×10^{-6} ، -6.7×10^{-7} ، 2.7×10^{-8} و 1.6×10^{-6} .

مجموعه تمرینات ۵۰۵ (صفحه ۲۷۹)

۱. $m = 25$ ، $h = .04$. تقریب با این مقادیر عبارت است از 10.95017032 ، که تا ارقام ذکر شده درست است .
۳. با استفاده از $h = 10.95017045$ نتیجه می شود ، اگرچه $h < .00039$ کفایت خواهد کرد .
۵. 907937
۷. (T) $f^{(3)}$ در $x = .1$ ناپیوسته است .
(ب) 302506 . کران خطا 1.9×10^{-4}
(پ) 302625 . مقدار واقعی 302425 .
۹. (T) 2.6209 ثانیه (ب) 2.6197 ثانیه (پ) 2.6197 ثانیه
۱۱. (T) خیر . در $x = .5$ ناپیوسته (ب) 2792223171
(پ) 3048961993 (ث) 2792222147
۱۳. (ب) 63391 (ت) 3.19095

مجموعه تمرینات ۶۰۵ (صفحه ۲۹۰)

۱. $R_{5,5}$ مساوی است با 10.9501703
۳. $R_{10,10}$ عبارت است از 14.02585093 ، که تا ارقام ذکر شده دقیق است .

مجموعه تمرینات ۷۰۵ (صفحه ۲۹۸)

۱. 11.141495 ، 10.948403 ، 10.950140

n	ریشه‌ها	ضرایب
2	.5857864	.8535534
	3.4142136	.1464466
3	.4157746	.7110930
	2.2942804	.2785177
	6.2899451	.0103893

وقتی $n = 2$ ، تقریب عبارت است از .43246 . وقتی $n = 3$ ، تقریب عبارت است از .49603 . مقدار دقیق مساوی 5 می باشد .

فصل ۶

مجموعه تمرینات ۱۰۶ (صفحه ۳۰۷)

۱. $(T) = 1$ ثابت لیپ‌شیتس L ؛ مسئله خوش‌وضع

(پ) $\frac{3}{8}\sqrt{3}$ ثابت لیپ‌شیتس L ؛ مسئله خوش‌وضع

$$y(t) = t^2(e^t - e) \quad . ۳$$

مجموعه تمرینات ۲۰۶ (صفحه ۳۱۸)

۱. مقادیر مربوطه :

i	t_i	w_i
2	1.2	.684755
5	1.5	3.18744
7	1.7	6.46640
10	2.0	15.3982

بهازای خطای مطلق $h \approx .00064$ ، 1

(پ) مقادیر مربوطه :

بهازای $h = .1$

بهازای $h = .2$

i	t_i	w_i
5	1.0	1.32768
15	3.0	3.03518
25	5.0	5.00378

i	t_i	w_i
10	1.0	1.34868
30	3.0	3.14055
50	5.0	5.03050

۳. مقادیر مربوطه :

i	t_i	w_i
10	1.0	1.20319
20	2.0	2.14982

$$w_{50} = .10430 \approx p(50) \quad (\text{ب}) \quad \cdot ۵$$

$$p(t) = 1 - .99e^{-.002t}, \quad p(50) = .10421 \quad (\text{پ})$$

مجموعه تمرینات ۳.۶ (صفحه ۳۲۷)

۱. (T) مقادیر مربوطه:

روش تیلور از مرتبه ۴

i	t_i	w_i
10	1.0	-2.0
20	2.0	-3.0

روش تیلور از مرتبه ۲

i	t_i	w_i
10	1.0	-2.0
20	2.0	-3.0

$$r(5) \approx w_{25} = 5.0134730 \quad \cdot ۳$$

۵. (T) مقادیر مربوطه:

روش تیلور از مرتبه ۴

i	t_i	w_i
2	.2	5.86356
5	.5	2.81820
7	.7	.846005
10	1.0	-2.09829

روش تیلور از مرتبه ۲

i	t_i	w_i
2	.2	5.86227
5	.5	2.81515
7	.7	.841809
10	1.0	-2.10427

مجموعه تمرینات ۴.۶ (صفحه ۳۲۸)

۱. مقادیر مربوطه:

i	t_i	روش نقطه میانی	روش پیراسته اولیور	روش هیون
10	1.0	1.2530068	1.2536129	1.2532123
20	2.0	2.0963063	2.0958129	2.0961403

$$r(5) = 5.0134759, \quad w_{25} = 5.0134720 \quad \cdot ۳$$

۵. (T) مقادیر مربوطه:

t	روش اولیور	روش هیون	رونک - کوتای ۴
1.5	-.6578870	-.6664112	-.6666640
2.0	-.4855158	-.4996025	-.4999962

(پ) مقادیر مربوطه:

t	روش اوپلر	روش هیون	رونک - کوتای 4
.5	.6004670	.6190949	.6178911
1.0	1.1176186	1.1365805	1.1353475

۷. (ب) 7.9787 فوت

مجموعه تمرینات ۵.۶ (صفحه ۳۴۹)

۱. (ت) به ازای $\varepsilon = 10^{-6}$

t_i	w_i
0	0
.0316228	.0311280
.158114	.146247
.297977	.257682
.442162	.357355
.591421	.446460
.746087	.525782
.906557	.596088

(پ) به ازای $\varepsilon = 10^{-6}$ ، نتایج منتخب عبارتند از:

t_i	w_i
1.0000000	0
1.0316228	.092943976
1.0718555	.23266112
1.1123888	.40009595
1.1530315	.59760166
⋮	⋮
1.5221141	4.3176931
1.5633087	5.0254885
1.6045036	5.8107688
1.6456892	6.6794630
⋮	⋮
1.8511599	12.504088
1.8921030	14.015999
1.9329773	15.662141
1.9737757	17.451272

(ث) به ازای $\varepsilon = 10^{-4}$ ، مقادیر به دست آمده به ازای $0 \leq t \leq 3$ عبارتند از:

t_i	w_i
0	.0183156
.100000	.027052
.329271	.061339
.568316	.128771
.797946	.235765
1.029553	.389947
1.282833	.597920
1.599636	.851924
1.924669	.994363
1.995247	.999999
2.277558	1.080728
2.529462	1.324359
2.798438	1.892909
2.982842	2.628975

(چ) به ازای $\varepsilon = 10^{-3}$ ، مقادیر منتخب به ازای t_i و w_i عبارتند از:

t_i	w_i
.007267	.231827
.014831	.159003
.023103	.105532
.032164	.067774
.042175	.042232
.053333	.025996
.065899	.016689
.080218	.012464
.096747	.011992
.116067	.014466
.138791	.019573
.165192	.027359
.194538	.037844
⋮	⋮
.822826	.677014
.854286	.729776
.885746	.784517
.917206	.841238
.948666	.899938
.980126	.960618

$\lim_{t \rightarrow \infty} y(t) = 100,000$ (ب)

t_i	w_i	$y(t_i)$
0	1000	1000
1	1219	1219
5	2672	2672
11.28	8798	8793
15.87	19,435	19,449
20.36	37,242	37,212
24.99	59,931	59,938
31.78	85,323	85,332

(ت) . ۵

یک درونیایی خطی براین نتایج تقریب 78666 را در $t = 30$ می دهد. یک اسپلاین مکعبی طبیعی 79,353 را می دهد. نتیجه دقیق عبارت است از 80,296.

مجموعه تمرینات ۶.۶ (صفحه ۳۷۵)

۱. (آ) نتایج مربوطه عبارتند از: (پ) نتایج مربوطه عبارتند از:

i	t_i	w_i	i	t_i	w_i
2	.2	.1812691	5	1.25	1.206333
5	.5	.3934668	10	1.50	3.967653
7	.7	.5034131	15	1.75	9.298721
10	1.0	.6321199	20	2.00	18.683089

۳. (آ) جواب منحصر به فرد به ازای $0 < h < \frac{24}{9}(\max_{0 < t < 2} e^{9t})^{-1}$ (ب) نتایج مربوطه عبارتند از:

i	t_i	w_i
10	.1	1.317218
20	.2	1.784511

(پ) روش نیوتن، با استفاده از محکهای توقف $10^{-6} \leq |w_i^{(k)} - w_i^{(k-1)}|$ ، تعداد

تکرارها بر مرحله را از 4 به 3 تقلیل می دهد.

۵. نتایج مربوطه:

i	t_i	w_i	h	مبدأ
1	.0037397	.1816072	.0037397	RK4
4	.0149589	.6178580	.0037397	PRED-CORR
7	.0261781	.9656008	.0037397	PRED-CORR
8	.0318039	1.119802	.0056258	RK4
12	.0574415	1.717981	.0087603	RK4
16	.0970496	2.447690	.0133271	RK4
20	.1570653	3.323621	.0200343	RK4
27	.3361041	5.202989	.0297340	PRED-CORR
31	.5105646	6.511292	.04361513	PRED-CORR
41	1.322574	9.363833	.0991144	PRED-CORR
45	1.942664	8.856295	.155022	PRED-CORR

مجموعه تمرینات ۷.۶ (صفحه ۳۸۱)

۱. (آ) نتایج مربوطه عبارتند از: (پ) نتایج مربوطه عبارتند از:

i	t_i	w_i
2	.2	.18126925
4	.4	.32967995
6	.6	.45118836
8	.8	.55067104
10	1.0	.63212056

i	t_i	w_i
2	1.2	.86664254
4	1.4	2.6203596
6	1.6	5.7209615
8	1.8	10.793625
10	2.0	18.683097

(ث) نتایج مربوطه عبارتند از: (چ) نتایج مربوطه عبارتند از:

i	t_i	w_i
5	1.0	.3678944
10	2.0	1.0000000
15	3.0	2.7182818
20	4.0	54.598150
25	5.0	8103.0839

i	t_i	w_i
2	.2	.04001537
5	.5	.25000003
7	.7	.49000003
10	1.0	1.00000000

مجموعه تمرینات ۸.۶ (صفحه ۳۸۹)

۱. (آ) نتایج مربوطه عبارتند از: (پ) نتایج مربوطه عبارتند از:

i	t_i	$w_{1,i}$	$w_{2,i}$
2	.2	.632872	1.45160
5	.5	3.85385	4.46038
7	.7	10.8525	11.3490
10	1.0	49.2066	49.5744

i	t_i	$w_{1,i}$	$w_{2,i}$
2	.2	-.525560	-.640149
5	.5	-.693561	-.388708
7	.7	-.718129	.229078
10	1.0	-.353308	2.57897

(ث) نتایج مربوطه عبارتند از:

i	t_i	w_{1i}	w_{2i}	w_{3i}
5	1	3.73186	4.18087	4.45810
10	2	11.3146	12.5027	13.7536
15	3	34.0455	37.3715	40.7377

(چ) نتایج مربوطه عبارتند از:

i	t_i	w_{1i}	w_{2i}
5	.25	2.99512	-.0572923
10	.50	2.96354	-.208374
15	.75	2.88569	-.422317
20	1.0	2.74965	-.669048

۵. مقادیر مربوطه به ازای $h = .5$ عبارتند از:

i	t_i	$w_{1,i}$	$w_{2,i}$
2	1	8716	1435
4	2	7907	2120
6	3	6666	2813

مجموعه تمرینات ۹.۶ (صفحه ۴۰۵)

$$\begin{array}{ll}
 w_{20} = -50.983 \approx y(.2), & (\varphi) \\
 w_{50} = -1.16 \times 10^{17} \approx y(.5), & \\
 w_{-0} = -4.32 \times 10^{26} \approx y(.7) & \\
 w_2 = .18065 & (\psi) \\
 w_5 = .35785 & \\
 w_{-} = .15342 & \\
 w_{10} = -9.7414 &
 \end{array}$$

$$\begin{array}{l}
 \gamma \quad \text{به ازای } h = .1 : w_{10} = .367883 \approx y(1) = .3678794, \\
 w_{100} = 3.75800 \approx y(10) = .0000454;
 \end{array}$$

$$\begin{array}{l}
 \text{به ازای } h = .01 : w_{100} = .367879 \approx y(1) = .3678794, \\
 w_{1000} = .677194 \approx y(10) = .0000454
 \end{array}$$

$$\left(1 - \frac{9h\lambda}{24}\right)\beta^3 - \left(1 + \frac{19h\lambda}{24}\right)\beta^2 + \frac{5h\lambda}{24}\beta - \frac{h\lambda}{24} \quad (T) \quad . 11$$

۱۳. (T) مقادیر مربوطه عبارتند از:

i	t_i	h	دوزنقه	R-K4
5	.05	.01	.124715	.125129
10	.1	.01	.0548101	.0551133
20	.2	.075	.046024	.046106
25	.51	.075	.260097	.260669
30	.885	.075	.783220	.783769

(P) مقادیر مربوطه عبارتند از:

i	t_i	دوزنقه	R-K4
5	.05	.4166256	.4178647
10	.1	.2342621	.2351730
20	.2	.2167388	.2169858
25	.51	.4881578	.4881463
30	.885	.7139033	.7737179

فصل ۷

مجموعه تمرینات ۱.۷ (صفحه ۴۱۸)

۱. $x_1 = -4, x_2 = 2, x_3 = 0, x_4 = 3; (\bar{T})$

$x_1 = 1, x_2 = 2, x_3 = 0; (\bar{\Psi})$

$x_1 = 1, x_2 = 1, x_3 = 1 (\bar{\Theta})$

۳. $x_1 = -1, x_2 = 2, x_3 = 0, x_4 = 1; (\bar{T})$

($\bar{\Psi}$) روند از کار می‌افتد، زیرا جوابی وجود ندارد.

مجموعه تمرینات ۲.۷ (صفحه ۴۳۱)

۱. $x_1 = 1.1875$ و $x_2 = 1.8125$ و $x_3 = .875 (\bar{T})$ ؛ تعویض سطرها لازم است.

($\bar{\Psi}$) $x_1 = -.0317460016$ و $x_2 = .5952376656$ و $x_3 = -2.380951197$ و $x_4 = 2.777776938$

بدون تعویض سطرها

مجموعه مسائل ۳.۷ (صفحه ۴۵۲)

۳. $\det A = -5\frac{1}{2}, \det B = -6, \det AB = 33, \det BA = 33$

۱۳. $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -1 & -3 & 0 & 1 \\ 3 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix}$

$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix}$

$A^2 = A^{-1} (\bar{\Psi})$

$A^{3n} = I (\bar{T}) \cdot 17$

مجموعه تمرینات ۴.۷ (صفحه ۴۶۶)

۱. $x_1 = .9260, x_2 = 14.29; x_1 = .8275, x_2 = 20.00; (\bar{T})$
 $x_1 = .9260, x_2 = 14.29$

($\bar{\Psi}$) $x_1 = 1.196, x_2 = 1.152; x_1 = 1.153, x_2 = 1.153;$
 $x_1 = 1.153, x_2 = 1.153$

۵. برای (\bar{T}) $x_1 = .8275, x_2 = 20.00$

$$\text{برای ۱ (پ)} \quad x_1 = 1.153, x_2 = 1.153$$

۷. با استفاده از الگوریتم ۱۰۷،

$$P(x) = .999984 - .0998317x + .00460069x^2 + .00009429x^3$$

مجموعه تمرینات ۵۰۷ (صفحه ۴۷۴)

۹. (\bar{T}) حاصل تزویج نر i با ماده z ، یا نر z با ماده i ، همان ویژگیهای بال را دارد.

$$\text{ب) خیر، } x = (1, 0, -1)'$$

مجموعه تمرینات ۶۰۷ (صفحه ۴۹۱)

$$x_1 = x_2 = x_3 = 1 \quad \text{۱. ۰}$$

$$x_i = 1 \quad \text{بازای هر } i = 1, 2, \dots, 10 \quad \text{۲. ۰}$$

$$x_i = 1 \quad \text{بازای هر } i = 1, 2, \dots, 9 \quad \text{۱۱. ۰}$$

۱۳. n ریشه دوم، تقسیم/ضرب $(n^3 + 9n^2 + 2n)$ و $\frac{1}{8}(n^3 + 6n^2 - 7n)$ /تفریق/جمع

فصل ۸

مجموعه تمرینات ۱۰۸ (صفحه ۵۱۳)

$$\text{۵. } (0, 0, 0)' \quad (\bar{T})$$

$$\text{پ) } (1, -1, 1)'$$

$$\text{ت) } 3$$

$$\text{پ) } 7$$

$$\text{۱۵. } 3 \quad (\bar{T})$$

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{۱۹. ۰}$$

$$x = (6, 3, 1)' \quad \text{و } \lambda = 1 \quad (\bar{T}) \quad \text{۲۷. ۰}$$

$$\text{۲۱. } (ج)$$

مجموعه تمرینات ۲۰۸ (صفحه ۵۲۶)

$$\text{۱. } \bar{x}_1^{(5)} = 1.09, \bar{x}_2^{(5)} = .488, \bar{x}_3^{(5)} = 1.49, K_r(4) \approx 5.22 \quad (\bar{T})$$

$$\text{ب) } \bar{x}_1^{(6)} = 1.00, \bar{x}_2^{(6)} = 1.01, K_r(4) \approx 18.7$$

$$\text{ت) } \bar{x}_1^{(3)} = -227.0781, \bar{x}_2^{(3)} = 47.69253, \bar{x}_3^{(3)} = -17.76931, K_r(4) \approx 75.5$$

$$\text{۵. } \bar{x} = (-1.0000, 2.00000) \quad \text{بلی}$$

$$\text{۷. } 1.01(n^3 + 3n^2) \quad \text{نزدیکتر است}$$

مجموعه تمرینات ۳۰۸ (صفحه ۵۴۴)

۱. (T) ژاکوبی واگراست، گاوس - سیدل واگراست

(پ) ژاکوبی:

$$x^{(8)} = (98497, 1.9575, .98497, 1.96994, .97874, 1.96994)^t$$

گاوس - سیدل:

$$x^{(5)} = (.97408, 1.9778, .99054, 1.98434, .98662, 1.99429)^t$$

مجموعه تمرینات ۴۰۸ (صفحه ۵۶۱)

۳. (T) مقدار ویژه λ_1 ، مقادیر ویژه λ_2 و λ_3 با $-2 \leq \lambda_2, \lambda_3 \leq 4$

(پ) به ازای جميع مقادیر ویژه λ ، $0 \leq \lambda \leq 6$

$$\lambda \approx 2.53648 = \mu^{(10)} \quad (پ) \quad ۷$$

$$x^{(10)} = (.533179, .463621, .707655)^t$$

۱۱. (ب) قسمت (T) $\lambda_3 = 1$

قسمت (پ) $\lambda_3 = -.01664728$

۱۳. (T) $\lambda_2 = 2$

$$\lambda_2 = 1.480148 \quad (پ)$$

$$B = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix} \quad (T) \quad ۱۵$$

$$\lambda_2 = 2$$

$$v^{(2)} = (0, 5, -1)^t$$

$$B = \begin{bmatrix} .13172577 & -.18413712 \\ -1.0865033 & 1.3317484 \end{bmatrix} \quad (پ)$$

$$\lambda_2 = 1.480121$$

$$v^{(2)} = (.3634398, .4598282, -.5706659)^t$$

$$\lambda = .3820240 \quad (پ) \quad ۱۷$$

مجموعه تمرینات ۵۰۸ (صفحه ۵۸۴)

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & -1 & -2 \\ 0 & -2 & 3 \end{bmatrix} \quad ۱$$

$$2, 6, 6, 7, 6 (\bar{T}) \quad \cdot 5$$

$$2.26853, 6.41228, 1.08641 + .172172i, 1.08641 - .172172i (\bar{\Psi})$$

فصل ۹

مجموعه تمرینات ۱۰۹ (صفحه ۵۹۹)

$$x_1 = x_2 = 0 \quad \cdot 5$$

$$x_1 = 8000, x_2 = 4000 \quad \cdot 9 \quad \text{بلی:}$$

مجموعه تمرینات ۲۰۹ (صفحه ۶۰۷)

$$x_1 = 1, x_2 = 1 (\bar{T}) \quad \cdot 1 \quad \text{روش نیوتن در 4 تکرار همگراست.}$$

$$x_1 = 0, x_2 = .1, x_3 = 1 (\bar{\theta}) \quad \cdot 1 \quad x_1 = .5, x_2 = .8660254 (\bar{\Psi})$$

$$x_1 = 1, x_2 = 2, x_3 = 0 (\bar{\chi}) \quad \cdot 1 \quad \text{همگرایی در 2 تکرار}$$

$$k_1 = 8.77129, k_2 = .259695, k_3 = -1.37228 (\bar{T}) \quad \cdot 5$$

$$3.18 \text{ اینج (ب)}$$

فصل ۱۰

مجموعه تمرینات ۱۰۱۰ (صفحه ۶۲۰)

$$(\bar{T}), (\bar{\theta}), (\bar{\chi}), (\bar{\Psi}), (\bar{\theta}), (\bar{\chi}), (\bar{\theta}), (\bar{\chi}) \quad \cdot 1$$

$$(\bar{T}) \text{ مقادیر مربوطه:} \quad \cdot 3 \quad \cdot 7 \quad \text{مقادیر مربوطه بدازای } h = .05:$$

i	x_i	w_{1i}
5	.25	.08224765
10	.50	.006764673
15	.75	.000556349

i	x_i	w_{1i}
5	1.25	.6431423
10	1.50	.6832421
15	1.75	.6922685

مجموعه تمرینات ۲۰۱۰ (صفحه ۶۳۱)

$$(\bar{T}) \text{ مقادیر مربوطه:} \quad \cdot 1 \quad (\bar{\Psi}) \text{ مقادیر مربوطه:}$$

i	x_i	w_{1i}
2	1.2	.4545455
5	1.5	.4000000
7	1.7	.3703704

i	x_i	w_{1i}
5	1.25	-.7272908
10	1.50	-.8000371
15	1.75	-.8889473

مجموعه تمرینات ۳۰۱۰ (صفحه ۶۴۵)

۱. (\bar{T}) مقادیر مربوطه: (پ) مقادیر مربوطه:

i	x_i	w_i
2	.2	1.3642
5	.5	.95450
7	.7	.90014

i	x_i	w_i
5	1.25	.16797186
10	1.50	.45842388
15	1.75	.60787334

۳. (\bar{T}) مقادیر مربوطه:

i	x_i	w_i
5	1.25	.643282
10	1.50	.683328
15	1.75	.692302

(پ) مقادیر مربوطه:

i	x_i	w_i
2	1.2	.4545563
5	1.5	.4000130
7	1.7	.3703798

۷. (\bar{T}) مقادیر مربوطه:

i	x_i	w_i
5	1.25	-.7272810
10	1.50	-.8000128
15	1.75	-.8889005

مجموعه تمرینات ۴۰۱۰ (صفحه ۶۶۲)

$$\begin{aligned} c_1 &= 1.4777309 \times 10^{-2} \\ c_2 &= 2.8700968 \times 10^{-2} \\ c_3 &= 4.0908778 \times 10^{-2} \\ c_4 &= 5.0521349 \times 10^{-2} \\ c_5 &= 5.6633300 \times 10^{-2} \end{aligned}$$

$$\begin{aligned} c_6 &= 5.8304182 \times 10^{-2} \quad (\bar{T}) \quad ۱ \\ c_7 &= 5.4549063 \times 10^{-2} \\ c_8 &= 4.4328659 \times 10^{-2} \\ c_9 &= 2.6538926 \times 10^{-2} \end{aligned}$$

$$.045900355 \sin \pi x - .007321101 \sin 2\pi x + .002286058 \sin 3\pi x - .0009829617 \sin 4\pi x \cdot \Delta + .0005077925 \sin 5\pi x$$

فصل ۱۱

مجموعه تمرینات ۲۰۱۱ (صفحه ۶۸۳)

۱. (\bar{T}) مقادیر مربوطه: (پ) مقادیر مربوطه:

i	j	x_i	y_j	$w_{i,j}$
4	3	.8	.3	1.27136
4	7	.8	.7	1.75084
8	3	1.6	.3	1.61675
8	7	1.6	.7	3.06587

i	j	x_i	y_j	$w_{i,j}$
3	3	.3	.3	.089941
3	7	.3	.7	.209928
7	3	.7	.3	.209951
7	7	.7	.7	.489941

۷. مقادیر مربوطه:

i	j	x_i	y_j	$w_{i,j}$
1	4	1.0	4.0	88
2	1	2.0	1.0	66
4	2	4.0	2.0	66

مجموعه تمرینات ۳.۱۱ (صفحه ۷۰۰)

۱. (\bar{T}) مقادیر مربوطه به ازای $h = .1$ و $k = .01$:

i	j	x_i	t_j	$w_{i,j}$
4	50	.4	.5	-9.3352×10^8
10	50	1.0	.5	-9.1860×10^8
17	50	1.7	.5	2.6047×10^8

۳. (\bar{T}) مقادیر مربوطه به ازای $h = .1$ و $k = .01$:

i	j	x_i	t_i	$w_{i,j}$
4	50	.4	.5	2.3541×10^{-9}
10	50	1.0	.5	6.2567×10^{-17}
17	50	1.7	.5	-3.8090×10^{-9}

۱۱. مقادیر مربوطه:

i	j	r_i	t_j	$w_{i,j}$
6	50	.6	10	700.23
7	50	.7	10	450.42
8	50	.8	10	325.78
9	50	.9	10	417.30

مجموعه تمرینات ۴.۱۱ (صفحه ۷۲۳)

۱. مقادیر مربوطه به ازای $h = .1$ و $k = .05$:

i	j	x_i	t_j	w_{ij}
2	10	.2	.5	.00285202
5	10	.5	.5	.00485216
7	10	.7	.5	.00392548

مقادیر مربوطه به ازای $h = .05$ و $k = .1$:

i	j	x_i	t_j	w_{ij}
4	5	.2	.5	-.0028789
10	5	.5	.5	-.0048977
14	5	.7	.5	-.0039623

مقادیر مربوطه به ازای $h = .05$ و $k = .05$:

i	j	x_i	t_j	w_{ij}
4	10	.2	.5	-1.6792×10^{-15}
10	10	.5	.5	-3.1780×10^{-15}
14	10	.7	.5	-2.8588×10^{-15}

۳ . مقادیر مربوطه :

i	j	x_i	t_j	w_{ij}
2	3	.2	.3	.6729902
5	3	.5	.3	6.317389×10^{-16}
7	3	.7	.3	-.6729902

۵ . (\bar{T}) مقادیر مربوطه :

i	j	x_{ij}	t_{ij}	w_{ij}
3	2	.3	.1	1.00
7	2	.7	.1	-1.00
3	4	.3	.2	1.00
7	4	.7	.2	-.875

۷ . $p(5, 1.0) \approx .9$ و $p(5, .5) \approx .9(\bar{T})$

واژه‌نامه فارسی به انگلیسی

	ت		الف
function	تابع	spline	اسپلاین
continuous	پیوسته	cubic	مکعبی
weight	وزن	independence	استقلال
factorization	تجزیه	linear	خطی
direct	مستقیم	algorithm	الگوریتم
of a matrix	یک ماتریس	quadrature	انتگرالگیری
transpose	ترانهاده	numerical	عددی
of a matrix	یک ماتریس	step size	اندازه گام
difference	تفاضل		
forward	پیشرو	vector	بردار
approximation	تقریب	column	ستونی
least square	کمترین مربعات	row	سطری
division	تقسیم	characteristic	مشخص
synthetic	ترکیبی	residual	مانده‌ای
reduction	تحویل	eigen	ویژه
iteration	تکرار	extrapolation	برونبایی
functional	تابعی		
shooting	تیراندازی	stable	پایدار
nonlinear	غیرخطی	weakly	به طور ضعیف
		strongly	به طور قوی
constant	ثابت	conditionally	به طور مشروط
asymptotic error	خطای مجانبی	unconditionally	غیرشرطی
		stability	پایداری
		underflow	پاریز
chopping	جدا کردن	basis	پایه

	و		چ
method	روش	polynomial	چند جمله‌ای
of characteristic	الگوریتم	osculating	بوسان
algorithm	مشخص	backward	پسرو
modified	پیراسته	piecewise	قطعه قطعه
predictor-corrector	پیشگو - اصلاحگر	trigonometric	مثلثاتی
binary-search	تحقیق دویی	characteristic	مشخص
relaxation	تخفیف		
difference	تفاضل		ح
forward	پیشرو	limit	حد
finite	متناهی	of a function	یک تابع
central	مرکزی	of a sequence	یک دنباله
divided	منقسم	arithmetic	حساب
one-step	تک گامی	finite -digits	ارقام متناهی
consistent	سازگار	double -precision	دو دقتی
convergent	همگرا	computer	کامپیوتر
bisection	تصویف		
multistep	چند گامی		خ
consistent	سازگار	property	خاصیت
convergent	همگرا	variational	تغییراتی
trapezoidal	دوزنقه‌ای	error	خطا
explicit	صریح	truncation	برشی
implicit	ضمنی	local	موضعی
of false position	ناپجایی	global	جامع
midpoint	نقطه میانی	round-off	گرد کردن
	س	absolute	مطلق
overflow	سرریز	local	موضعی
		relative	نسبی
	ش		
quasilinear	شبه خطی		د
condition	شرط	degree	درجه
root	ریشه‌ای	of precision	دقت
boundary	مرزی	interpolation	درونیابی
free	آزاد	linear	خطی
clamped	مقید	system	دستگاه
radius	شعاع	linear	خطی
spectral	طیفی	nonlinear	غیر خطی

ill-conditioned	بد وضع	ع
lower -triangular	پایین مثلثی	عدد
persymmetric	تمام متقارن	شرطی
sparse	تنک	عرض
permutation	جایگشتی	نوار
tridiagonal	سه قطری	عنصر
condition number of	عدد شرطی	محوری
diagonally dominant	قطر غالب	
similar	متشابه	ف
orthogonal	متعامد	فاصله
positive definite	معین مثبت	بین بردارها
singular	منفرد	فرمول
band	نواری	تفاضل مرکزی
cofactor of	همعامل	درجه دو
set	مجموعه	فضا
convex	محدب	برداری
multiplicity	مرتبه تکرار	
of a root	یک ریشه	ق
problem	مسئله	قانون
perturbed	پریشیده	گازهای کامل
well -posed	خوش وضع	قضیه
differentiation	مشتقگیری	fundamental ...of algebra
numerical	عددی	اساسی جبر
characteristic	مشخص	extreme value
of a real number	یک عدد حقیقی	مقدار اکستریم
equation	معادله	mean value
test	آزمون	intermediate value
diffusion	انتشار	مقدار میانی
partial differential	دیفرانسیل جزئی	گ
elliptic	بیضوی	گرد کردن
parabolic	سه‌موی	م
hyperbolic	هذلولوی	ماتریس (ها)
heat	گرما	افزوده
wave	موج	اکیداً قطری
normal	نرمال	غالب
inverse	معکوس	الگوریتم
of a matrix	یک ماتریس	تجزیه مستقیم
		بالا مثلثی
number		distance
condition		between vectors
width		formula
band		central-difference
element		quadratic
pivot		space
		vector
		law
		ideal gas
		theorem
		fundamental ...of algebra
		extreme value
		mean value
		intermediate value
		matrix (ces)
		augmented
		strictly diagonally
		dominant
		direct factorization
		algorithm
		upper -triangular

fixed	ثابت	average value	مقدار متوسط
mesh	شبکه‌ای	of a function	یک تابع
representation	نمایش	curve	منحنی
floating-point	با ممیز سیار	characteristic	مشخص
binary	دودویی		
seminorm	نیم نرم		ن
discrete	گسسته	region	ناحیه
		of absolute stability	پایداری مطلق
	ه	norm	نرم
convergence	همگرایی	natural	طبیعی
quadratic	درجه دو	of a vector	یک بردار
accelerated	شتابدار	of a matrix	یک ماتریس
order of	مرتبه	theory	نظریه
rate of	میزان	approximation	تقریب
		point	نقطه

واژه‌نامه انگلیسی به فارسی

A			
absolute error	مطلق خطای	boundary-value problem	مقدار مرزی مسئله
algorithm	الگوریتم		C
annihilation technique	صفرسازی تکنیک	centered-difference formula	تفاضل مرکزی فرمول
approximation theory	تقریب نظریه	method	روش
asymptotic error constant	مجاانبی ثابت خطای	characteristic curve	مشخص منحنی
augmented matrix	افزوده ماتریس	of a real number polynomial	یک عدد حقیقی چند جمله‌ای
average value	متوسط مقدار	value	مقدار
		vector	بردار
B		chopping	جدا کردن
band	نوار (ی)	clamped	مقید
matrix	ماتریس	boundary condition	شرط مرزی
width	عرض	cubic	الگوریتم
basis	پایه	spline algorithm	اسپلاین مکعبی
bell-shaped spline	زنگی شکل اسپلاین	closed	بسته
binary representation	دودویی نمایش	method	روش
search method	روش تحقیق	cofactor	همعامل
bisection algorithm	تقسیم الگوریتم	of a matrix	یک ماتریس
method	روش	column	ستون (ی)
		vector	بردار
		composite	مرکب
		quadrature	انتگرالگیری
		computer	کامپیوتر
		arithmetic	حساب

condition	شرط (ی)	differential	دیفرانسیل
number	عدد	equation	معادله
conditionally	به‌طور مشروط	differentiation	مشتق‌گیری
stable	پایدار	diffusion	انتشار
consistent	سازگار	equation	معادله
multistep method	روش چندگامی	direct	مستقیم
one-step method	روش تک‌گامی	factorization algorithm	الگوریتم تجزیه
continuous	پیوسته	discrete	گسسته
function	تابع	least-squares	تقریب کمترین
least-squares	تقریب کمترین	approximation	مربعات
approximation	مربعات	seminorm	نیم‌نرم
convergence	همگرایی	distance	فاصله
accelerated	شتابدار	between vectors	بین بردارها
linear	خطی		
order of	مرتبه	E	
quadratic	درجه دو	eigenvalue	مقدار ویژه
rate of	میزان	bound for	کران برای
convergent	همگرا	dominant	غالب
matrix	ماتریس	invariance of	پایایی
multistep method	روش چندگامی	eigenvector	بردار ویژه
one-step method	روش تک‌گامی	elliptic	بیضوی
convex	محدب	partial differential	معادله دیفرانسیل
set	مجموعه	equation	جزئی
cubic	مکعبی	error	خطا
spline	اسپلاین	absolute	مطلق
		global	جامع
		local	موضعی
D		relative	نسبی
deflation	تقلیل	round-off	گرد کردن
degree	درجه	truncation	برشی
of precision	دقت	explicit	صریح
derivative	مشتق	method	روش
determinant	دترمینان	extrapolation	برونمایی
of matrix	ماتریس	algorithm	الگوریتم
diagonal	قطر (ی)	extreme	اکستریم
matrix	ماتریس	value theorem	قضیه مقدار
diagonally	قطری		
dominant	غالب	F	
difference	تفاضل (ی)	factorization	تجزیه
equation	معادله	of a matrix	یک ماتریس

finite	متناهی	intermediate value theorem	مقدار میانی قضیه
difference method	روش تفاضل	interpolation of a matrix	درونیابی یک ماتریس
digit arithmetic element algorithm	حساب با ارقام الگوریتم عنصر		
fixed	ثابت	L	
point algorithm	الگوریتم نقطه	least squares approximation	کمترین مربعات تقریب
forward difference	پیشرو تفاضلی	limit of a function of a sequence	حد یک تابع یک دنباله
free	آزاد	linear differential equation	خطی معادله دیفرانسیل
boundary condition	شرطی مرزی	finite-difference algorithm	الگوریتم تفاضل متناهی
functional iteration	تابعی تکرار	independence	استقلال
fundamental theorem of algebra	اساسی قضیه... جبر	interpolation system	درونیابی دستگاه
G		local error truncation	موضعی خطای برشی
global error	جامع خطای	lower triangular matrix	پایین مثلثی ماتریس
growth of error	رشد خطا		
H		M	
heat equation	گرما معادله	mantissa of a real number	مانتیس یک عدد حقیقی
hyperbolic	هذلولوی	matrix addition	ماتریس (ی) جمع
I		block tridiagonal condition number of factorization	سه قطری بلوکی عدد شرطی تجزیه
ideal	کامل	minor of multiplication	مینور ضرب
gas law	قانون گازهای	norm	نرم
identity matrix	همانی ماتریس	null	پوچ
ill-conditioned matrix	بد وضع ماتریس	orthogonal	متعامد
implicit method	ضمنی روش	permutation	جابجستگی
initial value problem	مقدار اولیه مسئله		
integration	انتگرالگیری		

persymmetric	تمام متقارن	O	
positive definite	معین مثبت	open	باز
rotation	دورانی	method	روش
singular	منفرد	orthogonal	متعامد
sparse	تنک	function	تابع
symmetric	متقارن	vector	بردار
transpose	ترانهاده	orthonormal	متعامد نرمال
upper triangular	بالا مثلثی	vector	بردار
well conditioned	خوش وضع	osculating	بوسان
mean value	مقدار میانگین	polynomial	چند جمله‌ای
theorem	قضیه	overflow	سرریز
mesh	شبکه (ای)	P	
point	نقطه	parabolic	سه‌موی
method	روش	differential equation	معادله دیفرانسیل
of false position	نابجایی	partial	جزئی
midpoint	نقطه میانی	differential equation	معادله دیفرانسیل
method	روش	perturbed	پریشیده
multiplicity	مرتبه تکرار	problem	مسئله
of a root	یک ریشه	piecewise	قطعه قطعه
multistep	چند گامی	polynomial	چند جمله‌ای
method	روش	pivot	محور (ی)
N		element	عنصر
natural	طبیعی	pivoting	محورگیری
cubic spline	الگوریتم	maximal	ماکزیمال
algorithm	اسپلاین مکعبی	partial	جزئی
norm	نرم	scaled column	ستونی مقیاس شده
spline	اسپلاین	total	کل
nonlinear	غیرخطی	polynomial	چند جمله‌ای
finite-difference	الگوریتم	piecewise	قطعه قطعه
algorithm	تفاضل متناهی	trigonometric	مثلثاتی
system of equations	دستگاه معادلات	power method	روش توانی
norm	نرم	algorithm	الگوریتم
of a vector	یک بردار	predictor - corrector	پیشگو - اصلاحگر
normal	نرمال	method	روش
equation	معادله	Q	
numerical	عددی	quadratic	درجه دو
differentiation	مشتقگیری	formula	فرمول
integration	انتگرالگیری	spline interpolation	درونیابی اسپلاین

quasilinear شبه خطی
 quotient -difference خارج قسمت - تفاضل
 algorithm الگوریتم

R
 region ناحیه
 of absolute stability پایداری مطلق
 relative نسبی
 error خطای
 relaxation تخفیف
 method روش
 residual مانده‌ای
 vector بردار
 root ریشه (ای)
 condition شرط
 rounding گرد کردن
 round -off گرد کردن
 error خطای
 row سطر (ی)
 vector بردار

S
 significant بامعنی
 digit رقم
 spectral طیفی
 radius شعاع
 stability پایداری
 condition number عدد شرطی
 strongly به‌طور قوی
 stable پایدار
 submatrix زیر ماتریس
 leading principle اصلی پیشرو
 synthetic ترکیبی
 division تقسیم

T
 test آزمون (امتحانی)
 equation معادله
 trapezoidal ذوزنقه‌ای
 method روش
 rule قاعده
 composite مرکب
 interpolatory درونیابی
 trigonometric مثلثاتی
 polynomial چندجمله‌ای

U
 unconditionally به‌طور غیرمشروط
 stable پایدار
 underflow پاریز
 uppertriangular بالا مثلثی
 matrix ماتریس

V
 variational تغییراتی
 property خاصیت
 vector بردار (ی)
 notation نماد
 space فضای

W
 wave موج
 equation معادله
 weakly به‌طور ضعیف
 stable پایدار
 weight وزن
 function تابع
 well-posed خوش وضع
 problem مسئله

فهرست راهنما

- الف**
- ازقام با معنی، ۱۵
 اسپلاین طبیعی (ر.ک. اسپلاین مکعبی، شرط مرزی آزاد)
 اسپلاینهای مکعبی
 بکار رفته برای مسائل مقدار مرزی، ۶۵۶
 زینگی شکل، ۶۵۷
 شرایط مرزی آزاد
 توصیف، ۱۶۱
 وجود و یکتایی، ۱۶۴
 شرایط مرزی مقید
 تقریب مشتق در نقاط انتهایی، ۲۵۴
 توصیف، ۱۶۱
 کران خطا، ۱۷۴
 وجود و یکتایی، ۱۶۶
 استقلال خطی
 بردارها، ۵۴۸
 بردارهای متعامد، ۵۶۱
 بردارهای ویژه، ۵۴۹
 توابع، ۲۰۷
 اقتصادی کردن سریهای توانی، ۲۳۴
 الگوریتم
 اس او آر، ۵۴۳
 اسپلاین مکعبی طبیعی، ۱۶۴
 اسپلاین مکعبی مقید، ۱۶۷
 استفسنسن، ۷۹
 تقلیل ویلاند، ۵۶۰
 اوپلر، ۳۱۰
 پرونیایی، ۳۷۸
 تجزیه مستقیم، ۴۷۸
- تصفیه تکراری، ۵۲۴
 تفاضل منتهای خطی، ۶۳۵
 تفاضل منتهای غیرخطی، ۶۴۲
 تفاضل منتهای معادله پواسن، ۶۷۹
 تفاضلی پسر و معادله گرما، ۶۹۲
 تقسیم ترکیبی درجه دو، ۹۸
 تکنیک تکراری زاکوبی، ۵۳۱
 تکنیک تکراری گاوس - سیدل، ۵۳۴
 تصنیف، ۳۰
 تیراندازی خطی، ۶۱۶
 چولسکی، ۴۸۳
 خارج قسمت - تفاضل، ۱۰۴
 خط قاطع، ۶۱
 درونیایی تکراری نویل، ۱۳۸
 رامبرگ، ۲۸۹
 روشن توانی، ۵۵۳
 زونگ - کوتا - فلیبرگ، ۳۴۷
 عنصر منتهای، ۷۲۷
 قطعه قطعه خطی ریلی - ریتس، ۵۵۳
 کرانک - نیکلسون، ۶۹۷
 کیو آر، ۵۷۷
 محاسبه کمترین مربعات، ۲۲۳
 مرکب سیمپسون، ۲۷۵
 نقطه ثابت، ۴۱
 نیوتن - رفسون، ۵۳
 هاوس هولدر، ۵۶۹
 هورنر، ۸۴
 انتگرالگیری
 عددی
 پایداری، ۲۷۱

نایبوستگیها با جهش مناهمی، ۲۸۱
خطاهای گرد کردن، ۲۷۰
درجه دقت، ۲۵۸
نایبوستگیهای نامتناهی، ۲۸۲
گاوسی

پاریز، ۱۳
پایه
برای مجموعه‌ای از توابع، ۲۰۸

ت

تابع
پیوسته، ۳
وزن، ۲۱۶
تجزیه مستقیم با الگوریتم محورگیری ستونی
ماکزیمال، ۴۸۰

توصیف، ۲۹۲
درجه دقت، ۲۹۲
اندازه گام متغیر آدامز
الگوریتم پیشگو - اصلاحگر، ۳۶۷

ب

ترانهادۀ یک ماتریس، ۴۶۹
تفاضل پیشرو، ۷۸
تقریب کمترین مربعات پیوسته
توصیف، ۲۰۳
معادلات نرمال، ۲۰۴
تقسیم ترکیبی، ۸۶
درجه دو، ۹۷
تقلیل، ۵۶۳، ۵۵۷
هتلینگ، ۵۶۳
تحویل کروت برای الگوریتم دستگاههای خطی سه
قطری، ۴۸۷

بردار (های)
تعریف، ۴۱۶
ستونی، ۴۱۶
سطری، ۴۱۶
مانده‌ای، ۵۱۸
متعامد، ۵۴۹
نماد، ۴۱۷، ۴۹۶، ۵۸۹
ویژه
استقلال خطی، ۵۴۹
تعریف، ۵۰۸

تکرار تابعی در R ، ۴۱
تکرار تابعی در R^n ، ۵۹۶
تکرار نقطه ثابت در R
تعریف، ۴۱
کران خطا برای، ۴۵
همگرایی، ۴۴
همگرایی درجه دو، ۷۱
توابع متعامد، ۲۱۶، ۲۰۹
تیراندازی غیرخطی با الگوریتم روشن نیوتن، ۶۲۷

ماتریس مقارن، ۵۴۹
برونیابی
انتگرالگیری عددی، ۲۸۴
مسائل مقدار اولیه، ۳۷۳
مسائل مقدار مرزی خطی، ۶۳۸
مسائل مقدار مرزی غیرخطی، ۶۴۳
مشتقگیری عددی، ۲۵۴
گراگ، ۳۷۳

پ

ث
ثابت خطای مجانبی، ۶۸

پایدار

به طور ضعیف، ۴۰۱
به طور قوی، ۴۰۱
به طور مشروط، ۶۹۰
غیرشرطی، ۶۹۰
پایداری

ج

جدا کردن، ۱۴

چ

چند جمله‌ای (های)
بوسان، ۱۴۹
پسرونیوتن، ۳۵۵

تعریف، ۲۰
حذف گاوسی، ۴۷۴
روشهای مقدار اولیه، ۳۹۳، ۴۰۰
عدد شرطی، ۵۲۵

درجهٔ دقت، ۲۵۸
 درونیابی
 اسپلاین درجهٔ دو، ۱۶۱، ۱۷۶
 خطی، ۱۲۲
 معکوس، ۱۴۰
 دستگاه
 خطی، ۴۱۲
 معادلات غیرخطی، ۵۸۷

و

روش

اس او آر، ۵۴۰
 الگوریتم مشخص، ۷۱۹
 اویلر، ۳۰۸
 پیراسته، ۳۳۳ و ۳۹۴
 پیشگو - اصلاحگر آدامز، ۳۶۳
 تحقیق دویی، ۳۰
 تفاضل پیشرو، ۶۸۷
 تفاضل مرکزی، ۶۷۵
 تفاضل منقسم، ۱۴۱
 تک‌گامی سازگار، ۳۲۲ و ۳۹۴
 تک‌گامی همگرا، ۳۲۲، ۳۹۴
 تنصیف، ۳۰
 چندگامی، ۳۵۴
 سازگار، ۳۹۸
 همگرا، ۳۹۷
 چولسکی، ۴۷۷
 Δ^2 ایتنکن، ۷۶
 دولیتل، ۴۷۷
 دوزنقه‌ای، ۴۰۳
 رونگ - کوتا

الگوریتم مرتبهٔ چهار، ۳۳۶
 مرتبهٔ دو، ۳۳۲
 ریچارد سون، ۶۹۵
 ریلی - ریتس، ۶۴۸
 سیمپسون، ۳۷۰
 ضمنی، ۳۵۴
 کروت، ۴۷۷
 گاوس - ژردان، ۴۵۳
 میلن، ۳۷۰
 میلن، سیمپسون، ۳۷۰

تیلور

برای درونیابی، ۱۱۲
 تعریف، ۱۰
 برشی، ۱۰
 چیشف، ۲۲۰
 قطعه قطعه، ۱۵۹
 لاگرانژ، ۱۲۳
 لاگور، ۲۲۶
 لزاندر، ۲۲۰
 مثلثاتی، ۲۱۰
 مشخص
 روش چندگامی، ۳۹۸
 ماتریس، ۵۰۸
 هرمیت، ۱۴۹

ح

حد

یک تابع، ۳
 یک دنباله، ۴
 حساب
 ارقام متناهی، ۱۷
 دودقتی، ۲۲
 کامپیوتر، ۱۵
 حذف گاوسی، ۴۲۰
 با الگوریتم جایگذاری پسرو، ۴۲۵
 با الگوریتم محورگیری ستونی ماکزیمال، ۴۶۴
 با الگوریتم محورگیری ستونی مقیاس شده، ۴۶۶
 خاصیت تغییراتی، ۶۴۸

خ

خطا

برشی، ۱۰
 برشی موضعی، ۳۴۱
 جامع، ۳۴۲
 گرد کردن، ۱۲
 مطلق، ۱۴
 موضعی، ۳۴۲
 نسبی، ۱۴

د

دترمینان ماتریس، ۴۴۷

- نایجایی، ۶۵
 نقطهٔ میانی، ۳۳۲
 نیول، ۱۳۷
 هیون، ۳۳۳
 روشهای
 آدامز - بشفورث، ۳۵۵، ۳۷۰، ۳۹۶
 آدامز - مولتون، ۳۹۶
 پیشگو - اصلاحگر، ۳۶۲، ۳۶۷
 تحفیف، ۵۴۰
 تفاضل متناهی، ۶۳۲
 تیلور برای مسائل مقدار اولیه، ۳۲۷
 صریح، ۳۵۴
- ز**
- زیر ماتریس، ۴۴۷
- ژ**
- ژاکوبین، ۶۰۴
- س**
- سرریز، ۱۳
 سری
 تیلور، ۹
 فوریه، ۲۱۰
- ش**
- شبه خطی، ۷۱۱
 شرایط
 فوریه، ۶۰
 مرزی
 دیریکله، ۶۶۷
 ریشه‌ای، ۳۹۹
 شعاع طیفی یک ماتریس، ۵۱۰
- ع**
- عرض نوار، ۴۷۰
 عنصر محوری، ۴۲۳، ۴۶۲
- ف**
- فاصلهٔ بین بردارها، ۵۰۱
- فرمول
 تفاضل مرکزی، ۶۳۳
 درجهٔ دو، ۲۶
 فرمولهای
 باز نیوتن - کاتس، ۲۶۸
 بستهٔ نیوتن - کاتس، ۲۶۴
 فرم هسنبِرگ، ۵۶۷
 فضای برداری، ۴۳۴
- ق**
- قاعده
 درونیابی
 سیمپسون، ۲۶۰، ۲۶۶
 دوزنقه‌ای، ۲۶۲، ۲۶۶
 دوزنقه‌ای در انتگرالگیری رامبرگ، ۲۸۴
 کرامر، ۴۶۰
 مرکب
 دوزنقه‌ای، ۲۷۶
 سیمپسون، ۲۷۵
 قانون گازهای کامل، ۱
 قضیه
 اساسی جبر، ۸۳
 استرووسکی، ریج، ۵۴۲
 اشتاین - روزنبرگ، ۵۳۹
 تقریب وایراشتراس، ۱۱۲
 تیلور، ۹
 رل، ۵
 سیلورمن - تیلیتز، ۲۸۹
 شور، ۵۶۷
 کاهان، ۵۴۲
 گرد کردن، ۱۴
 مقدار اکستریم، ۶
 مقدار میانگین، ۵
 برای انتگرالها، ۷
 مقدار میانگین وزندار برای انتگرالها، ۷
 مقدار میانی، ۹
- م**
- ماتریس (ها)
 افزوده، ۴۱۷
 اکیداً قطر غالب، ۴۷۱

- الگوریتم تجزیه مستقیم، ۴۷۸
 بالا مثلثی، ۴۳۹
 بد وضع، ۵۲۰
 پایین مثلثی، ۴۳۹
 تجزیه، ۴۴۰، ۴۷۶، ۵۳۲
 ترانهاده، ۳۶۹
 تساوی، ۴۳۲
 تعریف، ۴۱۵
 تمام مقارن، ۵۶۵
 تنک، ۴۹۵
 جایگشتی، ۴۵۷
 جمع، ۴۳۲
 دورانی، ۵۸۶
 زاکوبین، ۶۰۴
 سنه قطری، ۴۷۰
 ضرب، ۴۳۴
 اسکالر، ۴۳۲
 عدد شرطی، ۵۲۰
 فرم بالا هسنبرگی، ۵۶۷
 قطری، ۴۳۹
 قطر غالب، ۴۷۱
 منشابه، ۵۴۹
 معامد، ۵۴۹
 متعین مثبت، ۴۷۲
 منفرد، ۴۴۱
 مینور، ۴۴۷
 نامنفرد، ۴۴۱
 نرم، ۵۰۵
 نوزاری، ۴۷۰
 همانی، ۴۳۶
 همگرا، ۵۱۲
 همعامل، ۴۴۷
 ماتریس یک عدد حقیقی، ۱۲
 مجموعه محدب، ۳۰۴
 مرتبه تکرار یک ریشه، ۷۳
 مسئله
 پربشیده، ۳۰۶، ۳۹۹
 خوش وضع، ۳۰۶
 مشتقگیری عددی
 برونیایی ریچاردسون، ۲۵۵
 خطای گرد کردن، ۲۴۹
- فرمول تفاضل
 پسر، ۲۴۰
 پیشرو، ۲۴۰
 فرمولهای پنج نقطه‌ای، ۲۴۹
 مشتقگیریهای بالاتر، ۲۵۴
 مشخصه یک عدد حقیقی، ۱۲
 معادلات دیفرانسیل جزئی
 بیضوی، ۶۶۶، ۶۷۲، ۷۲۷
 سهموی، ۶۶۷، ۶۸۵
 هذلولوی، ۶۶۸، ۷۰۴
 معادلات نرمال
 کمترین مربعات
 پیوسته، ۲۰۵
 گسسته خطی، ۱۹۲
 گسسته کلی، ۱۹۶
 معادله
 آزمون، ۴۰۲
 انتشار، ۶۶۸، ۶۸۵
 برنولی، ۳۵۲
 پواسن، ۶۶۶، ۶۷۲
 تفاضلی، ۳۱۰
 دیفرانسیل خطی، ۶۱۴
 گرما، ۶۶۶، ۶۸۷
 لاپلاس، ۶۶۷
 لوجستیکی، ۳۸۲
 موج، ۶۶۸، ۷۰۴
 الگوریتم تفاضل متناهی، ۷۰۷
 معکوس یک ماتریس، ۴۴۱
 مقدار متوسط یک تابع، ۷
 منحنیهای مشخص، ۷۱۳
 مینور یک ماتریس، ۴۴۷
- ن
 ناحیه پایداری مطلق، ۴۰۳
 نامساوی
 شوارتز، ۱۸۶
 کشی، ۱۸۶
 کشی - بونیاکوفسکی - شوارتز
 برای آنگرالها، ۱۸۶
 برای مجموعهها، ۱۸۷

شبکه‌ای، ۳۰۸	نرم
نمایش	اقلیدسی، ۴۹۷
با ممیز سیار، ۱۳	طبیعی، ۵۰۶
دودویی یک عدد حقیقی، ۱۲	یک بردار
نیم نرم گسسته، ۱۸۷	نرم l_1 ، ۵۱۴
	نرم l_2 ، ۴۹۷
	نرم l_∞ ، ۴۹۷
ه	یک ماتریس
همعامل یک ماتریس، ۴۴۷	نرم l_1 ، ۵۱۴
همگرایی	نرم l_2 ، ۵۰۶
در R ، ۴	نرم l_∞ ، ۵۰۶
در R^n ، ۵۰۲	نظریه تقریب، ۱۸۲
درجه دو، ۶۹	نقطه
شتابدار، ۷۶	ثابت
مرتبه، ۶۸	در R ، ۳۷
میزان، ۲۲	در R^n ، ۵۹۳

فهرست نمادها

$C(X)$	مجموعه تمام توابع پیوسته بر مجموعه X ، ۳
$C^n(X)$	مجموعه تمام توابعی که بر مجموعه X ، n مشتق پیوسته دارند، ۵
$C^\infty(X)$	مجموعه تمام توابعی که بر مجموعه X از هر مرتبه مشتق دارند، ۵
R	مجموعه تمام اعداد حقیقی، ۵
$f(y)$	شکل ممیز سیار عدد حقیقی y ، ۱۳
$O(\cdot)$	مرتبه همگرایی، ۲۲
Δ	تفاضل پیشرو، ۷۸
\bar{z}	مزدوج عدد مختلط z ، ۸۴
$f[\cdot]$	تفاضل منقسم تابع f ، ۱۴۱
$\binom{s}{k}$	ضریب دو جمله‌ای، $\frac{s!}{k!(s-k)!}$ ، ۱۴۴
$\ f\ _2$	نرم L_2 تابع f ، ۱۸۳
$\ f\ _\infty$	نرم L_∞ تابع f ، ۱۸۳
$\ f\ _1$	نرم L_1 تابع f ، ۱۸۵
R^n	مجموعه تمام n تاییهای مرتب از اعداد حقیقی، ۳۰۳
τ_i	خطای برشی موضعی در مرحله i م، ۳۲۱
\rightarrow	جایگزینی، ۴۱۴
\leftrightarrow	تعویض، ۴۱۳
(a_{ij})	ماتریس با درایه a_{ij} در سطر i م و ستون j م، ۴۱۵
x	بردار، یا عنصری از R^n ، ۴۱۶
$[A, b]$	ماتریس افزوده، ۴۱۷

ماتریس با تمام درایه‌های صفر، ۴۳۳	O
دلتای کرونکر، ۴۳۶	δ_{ij}
ماتریس همانی $n \times n$ ، ۴۳۶	I_n
ماتریس معکوس برای ماتریس A ، ۴۴۱	A^{-1}
مینور یک ماتریس، ۴۴۷	M_{ij}
همعامل یک ماتریس، ۴۴۷	A_{ij}
دترمینان ماتریس A ، ۴۴۷	$\det A$
بردار با تمام مختصات صفر، ۴۵۲	\circ
ترانژاده ماتریس A ، ۴۶۹	A'
نرم بردار x ، ۴۹۶	$\ x\ $
نرم l_2 بردار x ، ۴۹۷	$\ x\ _2$
نرم l_∞ بردار x ، ۴۹۷	$\ x\ _\infty$
نرم ماتریس A ، ۵۰۵	$\ A\ $
نرم l_∞ ماتریس A ، ۵۰۶	$\ A\ _\infty$
نرم l_2 ماتریس A ، ۵۰۶	$\ A\ _2$
شعاع طیفی ماتریس A ، ۵۱۰	$\rho(A)$
عدد شرطی ماتریس A ، ۵۲۰	$K(A)$
صفحه مختلط، مجموعه تمام اعداد مختلط، ۵۴۶	\mathcal{C}
تابعی که R^n را بتوی R^m می‌نگارد، ۵۸۹	F
ماتریسی که درایه‌هایش توابعی از R^n بتوی R هستند، ۶۰۲	$A(x)$
ماتریس ژاکوبین، ۶۰۴	$J(x)$

فهرست الگوریتمها

چولسکی، ۴۸۳	تصنیف، ۳۱
تحویل کروت برای دستگاههای خطی سه قطری، ۴۸۸	نقطه ثابت، ۴۱
تصفیه تکراری، ۵۲۴	نیوتن - رفسون، ۵۳
ژاکوبی، ۵۳۲	خط قاطع، ۶۲
گاوس - سیدل، ۵۳۵	استفسن، ۷۹
فوق تخفیف متوالی (SOR)، ۵۴۳	هورنر، ۸۵
توانی، ۵۵۳	تقسیم ترکیبی درجه دو، ۹۸
توانی برای ماتریس های متقارن، ۵۵۵	خارج قسمت - تفاضل (QD)، ۱۰۴
تقلیل ویلاند، ۵۶۰	درونیایی تکراری نویل، ۱۳۸
هاوس هولدر، ۵۶۹	اسپلاین مکعبی طبیعی، ۱۶۴
QR، ۵۷۷	اسپلاین مکعبی مقید، ۱۶۷
نیوتن برای دستگاهها، ۶۰۴	محاسبه کمترین مربعات، ۲۲۳
تیراندازی خطی، ۶۱۶	قاعده مرکب سیمپسون، ۲۷۶
تیراندازی غیر خطی با روش نیوتن، ۶۲۷	رامبرگ، ۲۸۹
تفاضل متناهی خطی، ۶۳۵	اویلر، ۳۱۰
تفاضل متناهی غیرخطی، ۶۴۲	رونگ - کوتا (مرتبه چهار)، ۳۳۶
قطعه قطعه خطی ریلی - ریتس، ۶۵۳	رونگ - کوتا - فلبرگ، ۳۴۷
اسپلاین مکعبی ریلی - ریتس، ۶۶۰	پیشگو - اصلاحگر مرتبه چهار آدامز، ۳۶۳
معادله تفاضل متناهی پواسن، ۶۷۹	پیشگو - اصلاحگر با اندازه گام متغیر آدامز، ۳۶۷
معادله تفاضلی پسر و معادله گرما، ۶۹۲	برونیایی، ۳۷۸
کرانک - نیکلسون، ۶۹۷	حذف گاوسی با جایگزینی پسر و، ۴۲۵
معادله تفاضل متناهی موج، ۷۰۷	حذف گاوسی با محورگیری ستونی ماکزیمال، ۴۶۴
مشخصها، ۷۱۹	حذف گاوسی با محورگیری ستونی مقیاس شده، ۴۶۶
عنصر متناهی، ۷۳۳	تجزیه مستقیم، ۴۷۸
	تجزیه مستقیم با محورگیری ستونی ماکزیمال، ۴۸۰