به نام پروردگار یگانه و یکتا

دانشگاه صنعتی اصفهان-دانشکده علوم ریاضی

محاسبات عددی

پدیدآورنده : رضا مختاری

ویرایش پنجم : زمستان ۹۸ و بهار ۹۹

فهرست مطالب

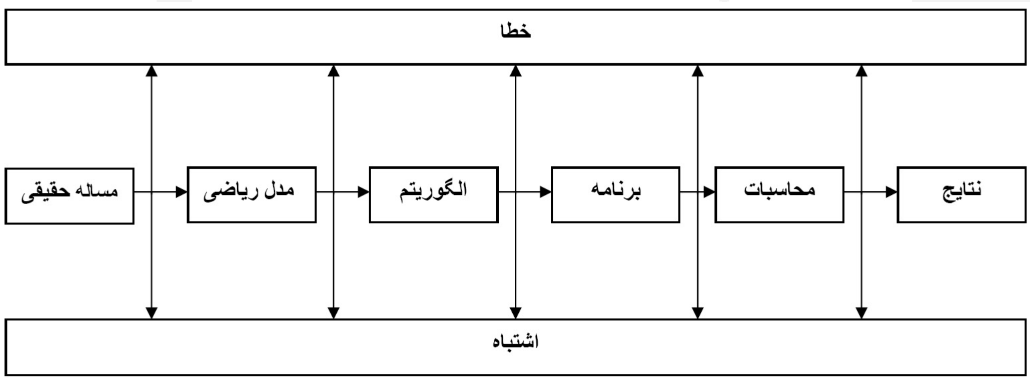
فصل ۱

خطاها

زمانی که به دست آوردن جواب دقیق (واقعی یا تحلیلی) یک مسئله به سادگی امکان‌پذیر نیست و یا مقرون به صرفه نیست، به کمک روش‌های عددی، یک جواب تقریبی برای مسئله پیدا کنیم. این فرایند به تولید خطا منجر می‌شود. در این فصل قصد داریم منابع تولید خطا و انواع خطا را شناسایی کرده و تا حدی از انتشار خطا جلوگیری کنیم.

۱.۱ منابع تولید خطا

بیشتر مواقع در عمل با یک مسئله حقیقی (فیزیکی) مواجه هستیم و بنابر دلایلی، جواب تقریبی (عددی) آن را جستجو می‌کنیم. مراحل یافتن جواب تقریبی یک مسئله حقیقی در روندنمای آمده در شکل ۱.۱ خلاصه می‌شود. این روندنما مکان‌های احتمالی بروز خطا و همچنین اشتباهات را نیز نشان می‌دهد.



شکل ۱.۱: فرایند تولید جواب عددی (تقریبی)

تذکر ۱.۱ *اختلافی که بین جواب دقیق و تقریبی وجود دارد ممکن است از اشتباهات و خطاها ناشی شده باشد. اشتباه را می‌توان برطرف کرد ولی خطا بیشتر اوقات اجتناب‌ناپذیر است. به عنوان مثال قراردادن ۲۳۲۲ به جای ۲۲۳۲ یک اشتباه است و استفاده از ۳/۱۴ به جای عدد π که بسط ده‌دهی نامختوم دارد، موجب بروز خطا می‌شود.*

مراحل مختلف روندنمای آمده در شکل ۱.۱ را در مثال بعد دنبال می‌کنیم.

مثال ۱.۱ **(مسئله حقیقی)** *می‌خواهیم دوره تناوب حرکت نوسانی و متناوب یک آونگ ساده به جرم m و طول l را به دست آوریم. فرض کنید $\theta(t)$ جابجایی زاویه‌ای آونگ در زمان t باشد. به کمک برخی از قوانین و اصول فیزیک و ریاضیات، صرف نظر از مقاومت هوا و اصطکاک در لولا، مدل این مسئله به صورت*

$$ml\frac{d^۲\theta}{dt^۲} = -mg\sin\theta,$$

یا $\frac{d^۲\theta}{dt^۲} = -\frac{g}{l}\sin\theta$ به دست می‌آید که یک معادله دیفرانسیل عادی غیرخطی است. با فرض کوچک بودن θ یعنی

$$\theta = ۶^\circ \simeq ۰/۱۰۴۷^{\rm rad}, \quad \sin\theta \simeq ۰/۱۰۴۵,$$
$$\theta = ۱۵^\circ \simeq ۰/۲۶۲^{\rm rad}, \quad \sin\theta \simeq ۰/۲۵۹,$$

می‌توان فرض کرد $\sin\theta \simeq \theta^{\rm rad}$ و معادله دیفرانسیل غیرخطی را به صورت زیر نوشت

$$\frac{d^۲\theta}{dt^۲} + \omega^۲\theta = ۰, \qquad \omega^۲ = \frac{g}{l}.$$

این معادله دیفرانسیل خطی جوابی متناوب به صورت $\theta(t) = c_۱\sin\omega t + c_۲\cos\omega t$ دارد و بنابراین دوره تناوب آونگ ساده عبارت است از $T = \frac{۲\pi}{\omega} = ۲\pi\sqrt{\frac{l}{g}}$. $\triangle$

با توجه به این مثال، می‌توان خطاها را از نقطه نظر منبع تولید به صورت زیر تقسیم‌بندی کرد.

انواع خطا

۱. ذاتی

- مدل (ناشی از صرف‌نظرها، چشم‌پوشی‌ها و ساده‌سازی‌ها مانند فرض $\sin\theta \simeq \theta^{rad}$)
- داده‌های مدل (ناشی از آزمایشات و اندازه‌گیری‌ها مثل g, l)

۲. محاسباتی

- نمایش اعداد (مانند $\pi = ۳/۱۴$)
- اعمال ریاضی (به عنوان مثال $\frac{l}{g}$)
- روش‌های (الگوریتم‌های) عددی (محاسباتی) (مثل خطای روش محاسبه $\sqrt{\frac{l}{g}}$)

تذکر ۲.۱ *خطاهای ذاتی به محاسبات عددی مربوط نمی‌شود اما برای پرهیز از خطاهای محاسباتی و کنترل آن‌ها باید راه چاره‌ای پیدا کرد.*

تذکر ۳.۱ *برای اطلاعات بیشتر در خصوص اثرات مخرب خطاها و اشتباهات به آدرس‌های زیر مراجعه کنید.*

en.wikipedia.org/wiki/Computer_bug, www.devtopics.com/20-famous-software-disasters/

در اینجا فقط به دو مورد زیر اشاره می‌شود.

- *عدم موفقیت موشک پاتریوت در جنگ خلیج فارس سال ۱۹۹۱ (۲۸ کشته و ۱۰۰ زخمی) به دلیل وقوع خطای گرد کردن در محاسبات مسیر*
- *شکست ماموریت موشک آریان ۵ فرانسه در سال ۱۹۹۶ (۵۰۰ میلیون دلار خسارت مادی) به دلیل وقوع پاریز[۱] در رایانه آن*

۲.۱ نمایش اعداد

در این بخش به بررسی نمایش اعداد حقیقی می‌پردازیم. اثبات برخی از قضایا را می‌توان در مراجع آنالیز عددی یافت.

قضیه ۱.۱ *هر عدد حقیقی مثبت x نمایشی به صورت*

$$\begin{aligned} x &= a_m\beta^m + a_{m-1}\beta^{m-1} + \cdots + a_1\beta^1 + a_0\beta^0 + a_{-1}\beta^{-1} + a_{-2}\beta^{-2} + \cdots \\ &= (a_m a_{m-1}\ldots a_1 a_0 / a_{-1}a_{-2}\ldots)_\beta, \end{aligned} \tag{1.1}$$

دارد که در آن $m \in \mathbb{Z}$، $a_i \in \{0, 1, 2, \ldots, \beta - 1\}$ و $a_m \neq 0$.

تذکر ۴.۱ *رابطه (۱.۱) به نمایش (بسط) عدد x در مبنای β معروف است و اگر $\beta = 10$ اختیار شود به آن، نمایش (بسط) ده‌دهی (اعشاری) گویند (متداول در زندگی روزمره) و در حالتی که $\beta = 2$ در نظر گرفته شود به آن، نمایش دودویی (باینری) گفته می‌شود (مبنای کار رایانه).*

قضیه ۲.۱ *نمایش یک عدد گویا (در هر مبنایی) یا مختوم است یا نامختوم متناوب.*

نتیجه ۱.۲.۱ *بسط یک عدد گنگ، نامختوم نامتناوب است.*

مثال ۲.۱ *به موارد زیر در مبناهای متفاوت توجه کنید*

$$\begin{aligned} &\tfrac{2}{3} = 0/666\cdots = 0/\bar{6} = (0/2)_3, \quad \tfrac{3}{8} = 0/375 = (0/3)_8, \quad \sqrt{2} = 1/4142\cdots, \\ &0/1 = (0/0\overline{0011})_2, \qquad \tfrac{1}{4} = 0/25 = (0/01)_2, \quad \pi = 3/141592\cdots. \end{aligned}$$

$\triangle$

---

[۱] Underflow

تذکر ۵.۱ *اگر چه فرض $a_m \neq ۰$ برای یکتایی نمایش (۱.۱) در نظر گرفته شده است، برای منحصر به فرد بودن نمایش (۱.۱) به فرض‌های دیگری نیز نیاز است. به مثال زیر توجه کنید*

$$\begin{aligned} ۳/۴۷۹۹۹\cdots &= ۳/۴۷\bar{۹} = ۳ \times ۱۰^{۰} + ۴ \times ۱۰^{-۱} + ۷ \times ۱۰^{-۲} + ۹ \times ۱۰^{-۳} + ۹ \times ۱۰^{-۴} + \cdots \\ &= ۳/۴۷ + \frac{۹ \times ۱۰^{-۳}}{۱-۱۰^{-۱}} = ۳/۴۷ + ۰/۰۱ = ۳/۴۸. \end{aligned}$$

یعنی برای اعداد $۳/۴۷\bar{۹}$ و $۳/۴۸$ یک نمایش وجود دارد. اگر فرض کنیم عدد صحیح j چنان وجود داشته باشد که $\cdots = a_{j-۱} = a_j = ۰$ (به عبارتی فرض کنیم بسط مختوم باشد) این مشکل برطرف می‌شود.

تذکر ۶.۱ *هنگام کار با رایانه (ماشین حساب) اعداد را در مبنای ۱۰ وارد کرده و انتظار داریم نتایج (خروجی) نیز در همین مبنا نمایش داده شود ولی این وسایل با مبنای دیگری (امروزه مبنای ۲ و در قدیم مبناهای دیگری مانند ۱۶) کار می‌کنند. بنابراین مسئله تغییر مبنا مطرح می‌شود که ممکن است خطایی به دنبال داشته باشد که در اینجا از بررسی آن صرف نظر می‌کنیم.*

۳.۱ نمایش اعداد در رایانه

برای نمایش اعداد در ماشین، ابتدا نمایشی به نام ممیز ثابت[۲] در نظر گرفته شد که در آن هر عدد حقیقی x به صورت زیر نمایش داده می‌شود

$$x = \pm(a_n a_{n-۱} \ldots a_۱ a_۰ / a_{-۱} a_{-۲} \ldots a_{-m})_\beta,$$

که در آن m و n اعداد مشخص و ثابتی هستند. در اصل در این نمایش مکان ممیز مشخص و ثابت است. برای نمایش اعداد بسیار بزرگ (کوچک) در این نمایش با مشکل مواجه می‌شویم و در نتیجه این نمایش برای محاسبات علمی مناسب نیست ولی برای بسیاری از کاربردها مانند حسابداری، این نمایش سودمند است و هنوز هم ماشین‌هایی بر این اساس ساخته می‌شوند.
یک روش جدید و متداول برای نمایش اعداد در رایانه، نمایش ممیز (نقطه) شناور (سیار)[۳] است که از بدو پیدایش مورد توجه سازندگان سخت‌افزار رایانه قرار گرفت و تا حدودی به طور سلیقه‌ای با آن برخورد شد تا زمانی که استانداردی توسط $IEEE$[۴] وضع شد.

۱.۳.۱ نمایش ۶۴-بیتی ممیز شناور

این نمایش پیش از این به دقت دو برابر (مضاعف)[5] معروف بوده و متناظر با نوع double در زبان C است. در این نمایش برای نمایش هر عدد در مبنای ۲، ابتدا یک ساختار به طول ۶۴ بیت در نظر گرفته می‌شود. اولین بیت به بیت علامت

[۲] Fixed point
[۳] Floating point
[۴] IEEE standard 754-1985
[5] Double precision

معروف است و با s نمایش داده شود و بلافاصله بعد از آن ۱۱ بیت برای مشخصه[۶] در نظر گرفته می‌شود و با c نمایش داده می‌شود و ۵۲ بیت باقی‌مانده به نام مانتیس منظور می‌شود که آن را با f نشان می‌دهند. اگرچه برای مانتیس ۵۲ بیت در نظر گرفته شده است ولی در واقع ساختار $۱ + f$ موجب می‌شود که ۵۳ بیت داشته باشیم که آن یک بیت اضافه به بیت پنهان معروف است و برای یکتایی نمایش لازم است. برای هر عدد نمایشی به صورت $(-۱)^s \times ۲^e \times (۱ + f)$ در نظر گرفته می‌شود که در آن $e = c - ۱۰۲۳$ توان (نما) است. محدودیت $۰ \le c \le (۱۱\cdots۱)_۲ = ۲^{۱۱} - ۱ = ۲۰۴۷$ موجب می‌شود که

$$-۱۰۲۳ \le e \le ۱۰۲۴.$$

از طرف دیگر محدودیت $۰ \le f \le (۰/\overbrace{۱۱\cdots۱}^{۵۲ times})_۲$ باعث می‌گردد که $۱ \le (۱ + f)_۲ \le (۱/\overbrace{۱۱\cdots۱}^{۵۲ times})_۲ < ۲$. بنابراین اگر کوچک‌ترین و بزرگ‌ترین عدد مثبت قابل نمایش را به ترتیب با mN (realmin در محیط MATLAB) و MN (realmax در محیط MATLAB) نشان دهیم آنگاه

$$mN = ۲^{-۱۰۲۲} \times (۱/۰\cdots۰)_۲ \simeq ۲/۲۲۵۰۷ \times ۱۰^{-۳۰۸},$$
$$MN = ۲^{۱۰۲۴} \times (۱/۰\cdots۰)_۲ \simeq ۱/۷۹۷۶۹ \times ۱۰^{۳۰۸}.$$

تذکر ۷.۱ *باید توجه داشت که توان ۱۰۲۳– برای نمایش صفر و توان ۱۰۲۴ با مانتیس مثبت برای نمایش* ∞ *(*inf *در محیط* MATLAB*) و صور مبهم (*NaN *در محیط* MATLAB*) مورد استفاده قرار می‌گیرد.*

تعریف ۱.۱ *در نمایش اعداد ماشینی، کوچک‌ترین عدد مثبت ماشینی که اگر به ۱ اضافه شود عددی بزرگ‌تر از ۱ به دست می‌آید به اپسیلون ماشین[۷] معروف است و با* eps *نمایش داده می‌شود.*

چون در نمایش ۶۴-بیتی بعد از ۱ عدد $(۱/\underbrace{۰\cdots۰}_{۵۱ times}۱)_۲$ قرار می‌گیرد، پس

$$\mathrm{eps} = (۱/\overbrace{۰\cdots۰}^{۵۱ times}۱)_۲ - ۱ = (۰/\overbrace{۰\cdots۰}^{۵۱ times}۱)_۲ = ۲^{-۵۲} \simeq ۲/۲۲۰۴۴۶ \times ۱۰^{-۱۶}.$$

حال بزرگترین عدد صحیح مثبت M را تعیین می‌کنیم که هر عدد صحیح x با شرط $۰ \le x \le M$ در این نمایش به طور دقیق قابل نمایش باشد. به وضوح، تمام اعداد صحیح نامنفی که بزرگتر از $۲^{۵۳} - ۱ = (۱/۱۱\cdots۱)_۲ \times ۲^{۵۲}$ نباشند به طور دقیق قابل نمایش هستند و به علاوه $۲^{۵۳}$ نیز به صورت $(۱/\underbrace{۰\cdots۰}_{۵۲ times})_۲ \times ۲^{۵۳}$ قابل نمایش است. اما تعداد ارقام در مانتیس جهت نمایش $۲^{۵۳}+۱$ کافی نیست (۵۳ رقم در مانتیس لازم است). بنابراین $M = ۲^{۵۳} \simeq ۹/۰۰۷۱۹۹ \times ۱۰^{۱۵}$ و در نتیجه تمام اعداد صحیح ۱۵ رقمی و بسیاری از اعداد ۱۶ رقمی در این نمایش به طور دقیق قابل نمایش هستند (گفته می‌شود دقت در نمایش ۶۴-بیتی ۱۵ الی ۱۶ رقم است). البته بسیاری از توان‌های ۲ نیز به طور دقیق قابل نمایش هستند.

[۶] Characteristic

[۷] Machine epsilon

تذکر ۸.۱ *مجموعه اعداد با ممیز شناور را با $\mathbb{FP}$ نمایش می‌دهیم. محور اعداد با ممیز شناور بر خلاف محور اعداد حقیقی، متناهی و گسسته است و بر خلاف آنچه به نظر می‌رسد، نقاط روی این محور هم‌فاصله نیستند. اگر در هنگام انجام محاسبات، عددی در فاصله‌ی $(-mN, mN)$ تولید شود پیام پاریز[۸] و اگر عدد تولید شده از MN بزرگ‌تر یا از $-MN$ کوچک‌تر شود پیام سرریز[۹] صادر می‌شود. از توان‌های ۱۰۲۳– و ۱۰۲۴ برای نمایش پیام‌های پاریز و سرریز نیز استفاده می‌شود.*

تذکر ۹.۱ *بسیاری از نرم‌افزارها مانند* MATLAB *از محدودیت‌های سخت‌افزار پیروی می‌کنند و بعضی از نرم‌افزارها مانند* Mathematica*، محدودیت‌های سخت‌افزار را از طریق برنامه‌های نرم‌افزاری برطرف می‌کنند و به اصطلاح دقت را بالا می‌برند که ممکن است با کاهش سرعت انجام محاسبات همراه باشد.*

۲.۳.۱ اعداد ماشینی

به منظور سادگی نوشتار، برای هر عدد ماشینی[۱۰] (یا عناصر مجموعه اعداد $\mathbb{FP}$)، نمایشی به صورت

$$\pm \underbrace{0/d_1 \cdots d_k}_{\text{مانتیس}} \times 10^{n} \longleftarrow \text{نما (توان)}$$

در نظر گرفته می‌شود که در آن n به زیرمجموعه‌ای از $\mathbb{Z}$ مانند $[L, U]$ تعلق دارد و به ازای $i = 1, 2, \ldots, k$ داریم $d_i \in \{0, 1, \ldots, 9\}$ و برای یکتایی نمایش فرض می‌شود $d_1 \neq 0$. به این نمایش، نمایش ممیز شناور ده‌دهی نرمال‌شده گفته می‌شود و چنین اعدادی را، اعداد ماشینی ده‌دهی k–رقمی نیز می‌نامند. ایده چنین نمایشی، از نمایش علمی نرمال‌شده اعداد ناشی شده است. در نمایش علمی نرمال‌شده، هر عدد حقیقی مخالف صفر x را می‌توان به صورت

$$x = \pm \underbrace{0/d_1 d_2 \cdots d_k d_{k+1} d_{k+2} \cdots}_{\text{مانتیس}} \times 10^{n} \longleftarrow \text{نما (توان)}$$

نمایش داد که در آن $n \in \mathbb{Z}$ و به ازای $i = 1, 2, \ldots, k$ داریم $d_i \in \{0, 1, \ldots, 9\}$ و $d_1 \neq 0$. واضح است که برای نمایش x به صورت ممیز شناور ده‌دهی نرمال‌شده در یک ماشین (k–رقمی)، باید k رقم از مانتیس آن را حفظ کرده و بقیه را کنار گذاشت، که برای این کار روش‌های زیر موجود است

۱. روش قطع کردن (برش)[۱۱]

۲. روش گرد کردن معمولی[۱۲]

۳. روش گرد کردن به زوج[۱۳]

---

[۸] Underflow

[۹] Overflow

[۱۰] Machine number

[۱۱] Chopping

[۱۲] Rounding

[۱۳] Rounding to even

در روش قطع کردن، k رقم از مانتیس حفظ و بقیه کنار گذاشته می‌شود در حالی که در روش گرد کردن معمولی، ابتدا روش قطع کردن اِعمال شده، سپس اگر $d_{k+1} \geq ۵$ یک واحد به d_k اضافه می‌گردد. اما در روش گرد کردن به زوج، ابتدا روش قطع کردن اِعمال شده و در هر یک از حالت‌های زیر یک واحد به d_k اضافه می‌گردد

- $d_{k+1} > ۵$
- $d_{k+1} = ۵$ و رقم مخالف صفری در سمت راست d_{k+1} مشاهده شود
- $d_{k+1} = ۵$ و رقم مخالف صفری در سمت راست d_{k+1} مشاهده نشود و d_k فرد باشد.

قضیه ۳.۱ *اگر $x = ۰/d_۱d_۲d_۳\cdots d_kd_{k+۱}\cdots \times ۱۰^n$ یک عدد حقیقی ناصفر در نمایش علمی نرمال‌شده باشد، $\widetilde{fl}(x) = ۰/d_۱d_۲d_۳\cdots d_k \times ۱۰^n$ عدد ماشینی k_رقمی متناظر با x است که از روش قطع کردن به دست می‌آید و*

$$\left|x - \widetilde{fl}(x)\right| \leq ۱ \times ۱۰^{n-k}.$$

هم‌چنین $fl(x) = ۰/\delta_۱\delta_۲\cdots\delta_k \times ۱۰^m$ ($m = n$ یا $m = n + ۱$) عدد ماشینی k رقمی متناظر با x است که از روش گرد کردن معمولی به دست می‌آید و خواهیم داشت

$$|x - fl(x)| \leq ۵ \times ۱۰^{n-k-۱} = ۰/۵ \times ۱۰^{n-k}.$$

پرسش ۱.۱ *تفاوت* ۴۷km *با* ۴۷۰۰۰m *یا تفاوت ۳/۷ با ۳/۷۰ یا ۳/۷۰۰ در چیست؟*

تعریف ۲.۱ *منظور از ارقام بامعنای یک عدد مخالف صفر، ارقام مخالف صفر، صفرهای بین دو رقم مخالف صفر و صفرهایی است که در سمت راست عدد به منظور نشان دادن نوعی دقت قرار داده می‌شوند است (تمام ارقام مانتیس در نمایش علمی نرمال‌شده).*

مثال ۳.۱ عدد ۰/۰۰۰۷۰۴۵۰۰۰ *حداقل ۴ رقم بامعنا و حداکثر ۷ رقم بامعنا دارد.* △

قرارداد ۱.۱ *rD یعنی r رقم اعشار (تنظیم ماشین حساب روی Mode fix r) و rS یعنی r رقم بامعنا (تنظیم ماشین حساب روی Mode sci r)*[۱۴].

مثال ۴.۱ *در جدول ۱.۱ ، چند عدد و مقدارهای تقریبی متناظر با آن‌ها با روش‌های قطع کردن ($\widetilde{x}$) و گرد کردن معمولی ($\widetilde{x}$) با دقت $۳S$ و $۳D$ داده شده است.* △

تذکر ۱۰.۱ *از این به بعد، گرد کردن معمولی را به کار می‌بریم.*

---

[۱۴] D حرف اول کلمه Decimal و S حرف اول کلمه Significant

x	۲۸/۶۴۲۴	۰/۰۰۵۷۶۷۱	۴/۹۸۵۰	−۲۱۷۵/۳۴۵۱۲
$\widetilde{x}$(۳S)	۲۸/۶	۰/۰۰۵۷۷	۴/۹۹	−۲۱۸۰
$\widetilde{x}$(۳D)	۲۸/۶۴۲	۰/۰۰۶	۴/۹۸۵	−۲۱۷۵/۳۴۵
$\widehat{x}$(۳S)	۲۸/۶	۰/۰۰۵۷۶	۴/۹۸	−۲۱۷۰
$\widehat{x}$(۳D)	۲۸/۶۴۲	۰/۰۰۵	۴/۹۸۵	−۲۱۷۵/۳۴۵

جدول ۱.۱: مثال‌هایی از گرد کردن و قطع کردن معمولی

۴.۱ انواع خطا

تعریف ۳.۱ *اگر a تقریبی از A باشد $\Delta a = |A - a|$ را خطای مطلق a نسبت به A نامند. Δa منحصر به فرد است و در عمل بیشتر مواقع قابل تعیین نیست و به جای آن از هر عدد b_a استفاده می‌شود که کمتر از Δa نباشد. b_a منحصر به فرد نیست و به آن کران خطای مطلق گویند. بنابراین $\Delta a \leq b_a$ و در نتیجه $a - b_a \leq A \leq a + b_a$. بعضی مواقع از نمایش $A = a \pm b_a$ استفاده می‌شود.*

مثال ۵.۱ *اگر عدد $a = ۱/۷۳۲$ را به عنوان تقریبی از $A = \sqrt{۳}$ در نظر بگیریم، آنگاه*

$$\Delta a = \left|\sqrt{۳} - ۱/۷۳۲\right| = ۱/۷۳۲۰۵۰۸۰۷۵\cdots - ۱/۷۳۲ = ۰/۰۰۰۰۵۰۸۰۷۵\cdots,$$

از طرفی می‌دانیم $۱/۷۳۲۱ > \sqrt{۳} > ۱/۷۳۲۰$ بنابراین $۰/۰۰۰۱ > \sqrt{۳} - ۱/۷۳۲ > ۰$ پس $b_a = ۰/۰۰۰۱$ که معیاری برای نزدیکی $۱/۷۳۲$ به $\sqrt{۳}$ است. $\triangle$

پرسش ۲.۱ *آیا خطای مطلق معیار مناسبی برای مقایسه خطاها است؟*
پاسخ. *خیر. به عنوان مثال خطای مطلق یک صندوق‌دار بانک، تایپیست و دروازه‌بان را در نظر بگیرید.*

تعریف ۴.۱ *اگر $a \neq ۰$ تقریبی از A باشد $\delta a = \frac{|A-a|}{|A|} = \frac{\Delta a}{|A|}$ خطای نسبی a نسبت به A نامیده می‌شود و مشابه خطای مطلق منحصر به فرد است و بیشتر مواقع در عمل قابل تعیین نیست و از کران خطای نسبی استفاده می‌شود. $\delta a \times ۱۰۰$ به درصد خطا معروف است.*

قضیه ۴.۱ *اگر a تقریبی از A باشد و b_a یک کران خطای مطلق برای این تقریب باشد آنگاه*

$$\delta a \leq \frac{b_a}{|a| - b_a}.$$

به علاوه اگر b_a نسبت به $|a|$ خیلی کوچک باشد

$$\delta a \leq \frac{b_a}{|a|}.$$

مثال ۶.۱ *اگر عدد $a = ۱/۷۳۲$ را به عنوان تقریبی از $A = \sqrt{۳}$ در نظر بگیریم، آنگاه*

$$\delta a = \frac{\left|\sqrt{۳} - ۱/۷۳۲\right|}{\sqrt{۳}} = \frac{۱/۷۳۲۰۵۰۸۰۷۵\cdots - ۱/۷۳۲}{۱/۷۳۲۰۵۰۸۰۷۵\cdots} = \frac{۰/۰۰۰۰۵۰۸۰۷۵\cdots}{۱/۷۳۲۰۵۰۸۰۷۵\cdots}.$$

پس

$$\delta a = ۰/۰۰۰۰۲۹۳۳۳۷\cdots < ۰/۰۰۰۰۳.$$

اما با توجه به قضیه ۴.۱ و $b_a = ۰/۰۰۰۱$ می‌توان نوشت

$$\delta a \leq \frac{۰/۰۰۰۱}{۱/۷۳۲ - ۰/۰۰۰۱} = \frac{۰/۰۰۰۱}{۱/۷۳۱۹} = ۰/۰۰۰۰۵۷۷۴۰۰۵\cdots < ۰/۰۰۰۰۶$$

و یا

$$\delta a \leq \frac{۰/۰۰۰۱}{۱/۷۳۲} = ۰/۰۰۰۰۵۷۷۳۶۷۲\cdots < ۰/۰۰۰۰۶.$$

$\triangle$

قضیه ۵.۱ *اگر a تقریبی از A با n رقم بامعنای درست باشد و $b = ۱۰^t \times a$ و $B = ۱۰^t \times A$ که در آن t عددی صحیح است آنگاه b نیز تقریبی از B با n رقم بامعنای درست است و خطای نسبی b و a برابر است.*

قضیه ۶.۱ *اگر a گردشده A تا n رقم بامعنا باشد آنگاه a دارای n رقم بامعنای درست است.*

قضیه ۷.۱ ارتباط دقت (خطای نسبی) با تعداد ارقام بامعنای درست
اگر a دارای n رقم بامعنای درست باشد، آنگاه $\delta a < ۵ \times ۱۰^{-n}$ به شرط آن که ارقام بامعنای درست a از یک رقم ۱ و $n - ۱$ رقم صفر در جلوی آن تشکیل نشده باشد. برعکس اگر $\delta a \leq ۵ \times ۱۰^{-n-۱} = ۰/۵ \times ۱۰^{-n}$ آنگاه a دارای دست کم n رقم بامعنای درست است.

مثال ۷.۱ *تقریبی از $\sqrt{۳}$ ارایه دهید که خطای نسبی آن از $۱۰^{-۴}$ کمتر باشد. بنابر قضیه ۷.۱ اگر a تقریبی از $\sqrt{۳}$ باشد که ۵ رقم بامعنای درست داشته باشد آنگاه $e_r(a) < ۵ \times ۱۰^{-۵} < ۱۰^{-۴}$. از این‌رو، با توجه به قضیه ۶.۱ کافی است a گرد شده $\sqrt{۳}$ تا ۵ رقم بامعنا باشد، یعنی $a = ۱/۷۳۲۱$.* $\triangle$

۵.۱ خطای محاسبات (فرمول)

فرض کنید $z = f(x_۱, \ldots, x_n)$ تابعی باشد که می‌خواهیم آن را در نقطه $(x_۱, \ldots, x_n)$ ارزیابی کنیم. بدون کاستن از کلیت فرض کنید $\Delta x_i = x_i - \widehat{x}_i$ که در آن $\widehat{x}_i$ مقدار تقریبی x_i است. پس می‌توان نوشت

$$z = f(x_۱, \ldots, x_n) = f(\widehat{x}_۱ + \Delta x_۱, \ldots, \widehat{x}_n + \Delta x_n).$$

بنابر بسط تیلور توابع n متغیره داریم

$$z = f(\widehat{x}_1, \dots, \widehat{x}_n) + \left(\Delta x_1 \frac{\partial f}{\partial x_1} + \dots + \Delta x_n \frac{\partial f}{\partial x_n}\right)(\widehat{x}_1, \dots, \widehat{x}_n) + R,$$

که در آن R جمله خطا بوده و شامل حاصل‌ضرب‌ها و توان‌های Δx_iها است و چون Δx_iها کوچک هستند از R چشم‌پوشی کرده، خواهیم داشت

$$f(x_1, \dots, x_n) - f(\widehat{x}_1, \dots, \widehat{x}_n) \simeq \left(\Delta x_1 \frac{\partial f}{\partial x_1} + \dots + \Delta x_n \frac{\partial f}{\partial x_n}\right)(\widehat{x}_1, \dots, \widehat{x}_n),$$

و اگر $\widehat{z} = f(\widehat{x}_1, \dots, \widehat{x}_n)$ آنگاه

$$\Delta z = |z - \widehat{z}| \simeq \left|\left(\Delta x_1 \frac{\partial f}{\partial x_1} + \dots + \Delta x_n \frac{\partial f}{\partial x_n}\right)(\widehat{x}_1, \dots, \widehat{x}_n)\right|,$$

و بلافاصله داریم

$$\delta z = \frac{\Delta z}{|z|} \simeq \left|\frac{\left(\Delta x_1 \frac{\partial f}{\partial x_1} + \dots + \Delta x_n \frac{\partial f}{\partial x_n}\right)(\widehat{x}_1, \dots, \widehat{x}_n)}{f(\widehat{x}_1, \dots, \widehat{x}_n)}\right|.$$

مثال ۸.۱ (مستقیم) *یک استوانه به شعاع قاعده $\frac{4}{3}$ و ارتفاع $\sqrt{2}$ را در نظر بگیرید. اگر شعاع و ارتفاع استوانه و عدد π را با دقت $4D$ وارد محاسبات کنیم، حجم این استوانه با چه خطایی به دست می‌آید؟*

می‌دانیم حجم یک استوانه از قاعده $\pi r^2 h$ تعیین می‌شود که در آن r شعاع قاعده و h ارتفاع استوانه است. اگر تعریف کنیم $z = V(p, r, h) = pr^2 h$، آنگاه باید V به ازای $p = \pi$، $r = \frac{4}{3}$ و $h = \sqrt{2}$ ارزیابی شود. تمام محاسبات را با دقت $4D$ دنبال می‌کنیم. بنابراین

$$p = \pi \quad \rightarrow \quad \widehat{p} = 3/1416$$

$$r = \tfrac{4}{3} \quad \rightarrow \quad \widehat{r} = 1/3333$$

$$h = \sqrt{2} \quad \rightarrow \quad \widehat{h} = 1/4142$$

و داریم $\Delta p, \Delta r, \Delta h \le 0/5 \times 10^{-4}$. بنابراین

$$\widehat{z} = V(\widehat{p}, \widehat{r}, \widehat{h}) = 3/4116 \times 1/3333^2 \times 1/4142 = 7/8980.$$

از طرفی

$$\Delta z \simeq \Delta p \widehat{r}^2 \widehat{h} + 2 \Delta r \widehat{p} \widehat{r} \widehat{h} + \Delta h \widehat{p} \widehat{r}^2 \le (\widehat{r}^2 \widehat{h} + 2 \widehat{p} \widehat{r} \widehat{h} + \widehat{p} \widehat{r}^2) \times 0/5 \times 10^{-4} < 10^{-3}.$$

در نتیجه حجم استوانه برابر است با $۷/۸۹۸ \pm ۰/۰۰۱$. به کمک یک ماشین حساب ۱۰ رقمی به دست می‌آوریم $z = ۷/۸۹۸۴۵۸۵۵۵$ و بنابراین خطای واقعی عبارت است از

$$\Delta z = |z - \hat{z}| = ۴/۵۸۵۵۵۴ \times ۱۰^{-۴} < ۰/۰۰۱.$$

$\triangle$

مثال ۹.۱ **(معکوس)** اعداد $x = \sqrt{۵}$ و $y = \frac{\pi}{۱۱}$ را با چه دقتی در نظر بگیریم تا مقدار $۶x^۲(\ln x + \sin ۲y)$ با دقت $۲D$ حساب شود؟ اگر فرض کنیم

$$z = f(x, y) = ۶x^۲(\ln x + \sin ۲y),$$

آنگاه

$$\frac{\partial f}{\partial x} = ۱۲x(\ln x + \sin ۲y) + ۶x, \qquad \frac{\partial f}{\partial y} = ۱۲x^۲ \cos ۲y,$$

و در نتیجه با فرض $x = ۲/۲$ و $y = ۰/۳$ (x و y را با دقت $۱S$ یا $۲S$ در نظر می‌گیریم) داریم

$$\Delta z \simeq \left| \Delta x \frac{\partial f(۲/۲, ۰/۳)}{\partial x} + \Delta y \frac{\partial f(۲/۲, ۰/۳)}{\partial y} \right| \simeq |۴۸/۹\Delta x + ۴۷/۹\Delta y|,$$

و برای برقراری نابرابری $\Delta z \leq ۰/۵ \times ۱۰^{-۲}$ باید داشته باشیم

$$|۴۸/۹\Delta x + ۴۷/۹\Delta y| \leq ۰/۵ \times ۱۰^{-۲}.$$

یک جواب نامعادله اخیر عبارت است از

$$\Delta x \leq ۰/۵ \times ۱۰^{-۴}, \qquad \Delta y \leq ۰/۵ \times ۱۰^{-۴}.$$

بنابراین برای رسیدن به نتیجه مطلوب، x و y را می‌توان با دقت $۴D$ در نظر گرفت. $\triangle$

مثال ۱۰.۱ اگر در محاسبه $z = ab^۲c^۳$، خطای نسبی a، b و c حداکثر $۰/۰۱$ باشد، بیشترین خطای نسبی قابل انتظار برای z چقدر است؟ با فرض $z = f(a, b, c) = ab^۲c^۳$ داریم

$$\delta z \simeq \frac{\left|\Delta a \frac{\partial f}{\partial a} + \Delta b \frac{\partial f}{\partial b} + \Delta c \frac{\partial f}{\partial c}\right|}{|ab^۲c^۳|} = \frac{|b^۲c^۳\Delta a + ۲abc^۳\Delta b + ۳ab^۲c^۲\Delta c|}{|ab^۲c^۳|} \leq \delta a + ۲\delta b + ۳\delta c = ۰/۰۶.$$

$\triangle$

۱.۵.۱ خطای اعمال ریاضی

هنگام کار با اعداد ماشینی (ممیز شناور) دقت شود که بعضی از اصول میدان $\mathbb{R}$ نظیر شرکت پذیری، منحصر به فرد بودن عضو خنثی و غیره برقرار نیست. به طور کلی، عبارت‌هایی که از نظر ریاضی معادل هستند ممکن است از نظر محاسباتی معادل نباشند. علت اصلی بروز این مشکلات، خطای گرد کردن است.

تعریف ۵.۱ *فرض کنید A و B دو عدد حقیقی، a و b تقریب‌هایی از آن‌ها و $\otimes$ بیان‌گر یک عمل دوتایی باشد. متناظر با $A \otimes B$ در ماشین عمل $a \otimes^* b$ انجام می‌شود و داریم*

$$|A \otimes B - a \otimes^* b| = |(A \otimes B - a \otimes b) + (a \otimes b - a \otimes^* b)| \leq \underbrace{|A \otimes B - a \otimes b|}_{\text{خطای منتشر شده}} + \underbrace{|a \otimes b - a \otimes^* b|}_{\text{خطای تولید شده}}$$

قضیه ۸.۱ *اگر a و b تقریب‌هایی از A و B بوده و همه این اعداد مثبت باشند، آنگاه*

۱. $\Delta(a \pm b) \leq \Delta a + \Delta b$، $\delta(a \pm b) \leq \frac{A}{|A\pm B|}\delta a + \frac{B}{|A\pm B|}\delta b$، $\delta(a+b) \leq \max\{\delta a, \delta b\}$،

۲. $\Delta(ab) \leq a\Delta b + b\Delta a$، $\delta(ab) \leq \delta a + \delta b$،

۳. $\Delta\left(\frac{a}{b}\right) \leq \dfrac{b\Delta a + a\Delta b}{b^{۲}}$، $\delta\left(\frac{a}{b}\right) \leq \delta a + \delta b$.

مثال ۱۱.۱ *اگر a و b هر یک n رقم بامعنای درست داشته باشند، حداقل تعداد ارقام بامعنای درست $a+b$ و ab را تعیین کنید.* بنابر قضیه ۷.۱، $\delta a < ۵ \times ۱۰^{-n}$ و $\delta b < ۵ \times ۱۰^{-n}$. هم‌چنین داریم

$$\delta(a+b) \leq \max\{\delta a, \delta b\} < ۵ \times ۱۰^{-n} = ۵ \times ۱۰^{-(n-۱)-۱}.$$

پس $a+b$ حداقل $n-۱$ رقم بامعنای درست دارد. به علاوه می‌توان نوشت

$$\delta(ab) \leq \delta a + \delta b < ۵ \times ۱۰^{-n} + ۵ \times ۱۰^{-n} = ۱۰ \times ۱۰^{-n} = ۱۰^{-(n-۲)-۱} < ۵ \times ۱۰^{-(n-۲)-۱}.$$

بنابراین ab دست کم $n-۲$ رقم بامعنای درست دارد. $\triangle$

تذکر ۱۱.۱ *با توجه به خطای نسبی $\delta(a-b) \simeq \frac{\Delta(a-b)}{|a-b|}$ واضح است که اگر a و b دو عدد هم‌علامت نزدیک به هم باشند $|a-b|$ کوچک و در نتیجه $\delta(a-b)$ بزرگ خواهد شد و در نتیجه تعداد ارقام بامعنای $a-b$ کم خواهد بود. بنابراین در عمل بهتر است از تفاضل دو عدد هم‌علامت نزدیک به هم جلوگیری شود (تفاضل دو عدد هم‌علامت نزدیک به هم موجب از بین رفتن ارقام بامعنا[۱۵] می‌شود مانند $۱/۴۲ - ۱/۴۱ = ۰/۰۱$). اگر تفاضل اجتناب‌ناپذیر است باید عمل با دقت دو برابر (یا بیشتر) انجام شود.*

---
[۱۵] Cancellation

مثال ۱۲.۱ *جهت اجتناب از تفاضل در محاسبات می‌توان از اتحادها کمک گرفت. به عنوان مثال*

$$e^{a-b} = \frac{e^a}{e^b}, \qquad ۱ - \cos x = ۲\sin^۲ \frac{x}{۲}, \qquad \ln a - \ln b = \ln \frac{a}{b}.$$

$\triangle$

مثال ۱۳.۱ *با فرض $g(x) = x\left(\sqrt[۳]{۱ + \frac{۱}{x}} - ۱\right)$، مقدار $g(۱۰^۹)$ را با دقت یک ماشین حساب ۱۰ رقمی به دست آورید. تمام محاسبات را با دقت $۹D$ انجام می‌دهیم. پس*

$$۱ + \frac{۱}{x} = ۱/۰۰۰۰۰۰۰۰۱, \qquad \sqrt[۳]{۱ + \frac{۱}{x}} = ۱/۰۰۰۰۰۰۰۰۰.$$

بنابراین با دقت $۹D$، ماشین حساب نتیجه زیر را به دست می‌دهد

$$g(۱۰^۹) = ۱۰^۹(۱/۰۰۰۰۰۰۰۰۰ - ۱) = ۰/۰۰۰۰۰۰۰۰۰.$$

به این ترتیب با یک ماشین حساب ۱۰ رقمی $g(۱۰^۹) = ۰/۰۰۰۰۰۰۰۰۰$، که نتیجه‌ای نادرست است. در حالی که اگر از یک رایانه استفاده شود، نتیجه درست $g(۱۰^۹) = ۰/۳۳۳۳۳۳۳۳۳$ به دست می‌آید. دلیل این خطای فاحش، تفاضل دو عدد نزدیک به هم در محاسبات است که منجر به از بین رفتن ارقام بامعنا می‌شود. برای به دست آوردن تقریب بهتر، روش محاسبه را به صورت زیر تغییر می‌دهیم

$$g(x) = x\left(\sqrt[۳]{۱ + \frac{۱}{x}} - ۱\right) \times \frac{\sqrt[۳]{(۱ + \frac{۱}{x})^۲} + \sqrt[۳]{۱ + \frac{۱}{x}} + ۱}{\sqrt[۳]{(۱ + \frac{۱}{x})^۲} + \sqrt[۳]{۱ + \frac{۱}{x}} + ۱} = \frac{۱}{\sqrt[۳]{(۱ + \frac{۱}{x})^۲} + \sqrt[۳]{۱ + \frac{۱}{x}} + ۱}.$$

در این صورت با توجه به

$$\left(۱ + \frac{۱}{x}\right)^۲ = ۱/۰۰۰۰۰۰۰۰۲, \quad \sqrt[۳]{\left(۱ + \frac{۱}{x}\right)^۲} = ۱/۰۰۰۰۰۰۰۰۰,$$

با دقت $۹D$ نتیجه زیر به دست می‌آید

$$g(۱۰^۹) = \frac{۱}{۱/۰۰۰۰۰۰۰۰۰ + ۱/۰۰۰۰۰۰۰۰۰ + ۱} = ۰/۳۳۳۳۳۳۳۳۳.$$

$\triangle$

این مثال نه تنها نشان می‌دهد که ممکن است یک ماشین حساب هم نتایج نادرستی تولید کند بلکه به خوبی نشان می‌دهد که به دلیل خطای گرد کردن، ممکن است محاسبه با دو روش مختلف که از نظر ریاضی هم‌ارز هستند به نتایج متفاوتی منجر شوند. از این رو باید از نظر عددی بین الگوریتم‌هایی که از نظر ریاضی هم‌ارز هستند تفاوت قایل شویم.

تذکر ۱۲.۱ *با توجه به قضیه ۸.۱، در محاسبات باید از ضرب اعداد بزرگ در اعداد تقریبی (تقسیم اعداد تقریبی به*

اعداد کوچک) پرهیز کرد.

مثال ۱۴.۱ *در محاسبه* $x = 10000\pi$ *داریم*

$$\begin{aligned} \pi &= 3/14 && \to \quad \tilde{x} = 31400, \\ \pi &= 3/142 && \to \quad \tilde{x} = 31420, \\ \pi &= 3/1416 && \to \quad \tilde{x} = 31416. \end{aligned}$$

در تقریب اول خطایی به اندازه ۱۶ واحد، در تقریب دوم خطایی نزدیک به ۴ واحد و در تقریب سوم خطایی کمتر از ۰/۱ مرتکب شده‌ایم. △

تذکر ۱۳.۱ *چون هر عمل محاسباتی خطایی به همراه دارد، یک قاعده کلی دیگر آن است که از حجم محاسبات تا آنجا که ممکن است کاسته شود.*

مثال ۱۵.۱ *به جای عبارت* $ax^3 + bx^2 + cx + d$ *از عبارت* $((ax + b)x + c)x + d$ *استفاده شود.* △

۲.۵.۱ تقریب توابع یک متغیره

قضیه ۹.۱ (تیلور با باقیمانده لاگرانژ) فرض کنید $f \in C^n[a,b]$ *(یعنی تابع* f *و مشتقات تا مرتبه* n *آن روی بازه* $[a,b]$ *پیوسته هستند)* و $f^{(n+1)}$ بر (a,b) موجود باشد و $x_0 \in [a,b]$. *در این صورت، به ازای هر* $x \in [a,b]$، *نقطه‌ای مانند* $y(x)$ *بین* x_0 *و* x *وجود دارد که*

$$f(x) = p_n(x) + R_n(x, x_0)$$

که در آن

$$p_n(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n = \sum_{i=0}^{n} \frac{f^{(i)}(x_0)}{i!}(x - x_0)^i$$

و

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

در اینجا p_n *چندجمله‌ای تیلور مرتبه* n*اُم* f *حول* x_0 *و* $R_n(x, x_0)$ *جمله باقی‌مانده (یا خطای برش[۱۶]) متناظر با* p_n *نامیده می‌شود. اگر* $n \to \infty$ *آنگاه* p_n *به یک سری بی‌پایان تبدیل می‌شود که به آن سری تیلور* f *حول نقطه* x_0 *گویند. در این حالت، شرط بی‌نهایت بار مشتق‌پذیر بودن* f *در* x_0 *لازم است.*

تذکر ۱۴.۱ *در قضیه قبل اگر* $x_0 = 0$ *آنگاه واژه تیلور به مکلورن تبدیل می‌شود.*

---
[۱۶] Truncation error

تذکر ۱۵.۱ *$R_n(x, x_0)$ مقدار خطا در استفاده از p_n به جای f را نشان می‌دهد. در عمل با یافتن کرانی برای جمله باقی‌مانده، در واقع برای خطای تقریب f با p_n کرانی پیدا می‌کنیم.*

تذکر ۱۶.۱ *ویژگی مهم چندجمله‌ای تیلور مرتبه nاُم آن است که p_n و مشتقات تا مرتبه‌ی nاُم آن با f و مشتقات تا مرتبه nاُم آن در نقطه x_0 برابر هستند.*

تذکر ۱۷.۱ *شکل دیگر (کاربردی) قضیه تیلور به صورت زیر است*

$$f(x_0+h) = f(x_0) + f'(x_0)h + \cdots + \frac{f^{(n)}(x_0)}{n!}h^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1} = \sum_{i=0}^{n} \frac{f^{(i)}(x_0)}{i!}h^i + \frac{f^{(n+1)}(y)}{(n+1)!}h^{n+1}$$

و یا

$$f(x) = f(x_0+h) = f(x_0) + f'(x_0)h + \cdots + \frac{f^{(n)}(x_0)}{n!}h^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1} = \sum_{i=0}^{n} \frac{f^{(i)}(x_0)}{i!}h^i + \frac{f^{(n+1)}(y)}{(n+1)!}h^{n+1},$$

که در آن $h = x - x_0$.

مثال ۱۶.۱ *فرض کنید $f(x) = 2 + 4x - x^3$. آنگاه به وضوح برای $n \geq 3$ داریم*

$$p_n(x) = 2 + 4x - x^3, \qquad R_n(x, 0) = 0,$$

و برای $n = 2$ خواهیم داشت

$$p_2(x) = 2 + 4x, \qquad R_2(x, 0) = -x^3,$$

و برای $n = 1$ می‌توان نوشت

$$p_1(x) = 2 + 4x, \qquad R_1(x, 0) = -3x^2 y(x),$$

که در آن $y(x)$ نقطه‌ای بین 0 و x است. $\triangle$

بنابر قضیه تیلور، می‌توان به جای کار کردن با یک تابع پیچیده از چندجمله‌ای تیلور نظیر آن استفاده کرد. مثال‌هایی که در ادامه خواهند آمد چگونگی این تقریب را نشان می‌دهند.

مثال ۱۷.۱ *مطلوب است محاسبه مقدار $e^{\frac{\pi}{10}}$ با دقت $2D$.*
روش اول- به کمک قضیه تیلور داریم

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \frac{x^{n+1}}{(n+1)!}e^{y(x)},$$

که در آن $0 < y(x) < x$. در نتیجه

$$e^{\frac{\pi}{10}} = 1 + \frac{\pi}{10} + \frac{(\frac{\pi}{10})^2}{2!} + \cdots + \frac{(\frac{\pi}{10})^n}{n!} + \frac{(\frac{\pi}{10})^{n+1}}{(n+1)!}e^y,$$

که در آن $0 < y < \frac{\pi}{10} < 1$. چون e^x تابعی صعودی است پس $1 = e^0 < e^y < e^1 < 3$. از طرفی $\frac{\pi}{10} < 0/315$ (دقت $3D$ منظور می‌شود) و بنابراین

$$\text{خطا} \leq \left|\frac{(\frac{\pi}{10})^{n+1}}{(n+1)!}e^{\xi}\right| < \frac{3 \times (0/315)^{n+1}}{(n+1)!}.$$

حال باید داشته باشیم $\frac{3\times(0/315)^{n+1}}{(n+1)!} < 0/5 \times 10^{-2}$ که نتیجه می‌دهد $n \geq 3$. پس

$$e^{\frac{\pi}{10}} \simeq 1 + 0/315 + \frac{(0/315)^2}{2!} + \frac{(0/315)^3}{3!} = 1/370,$$

و با دقت $2D$ داریم $e^{\frac{\pi}{10}} \simeq 1/37$ که با جواب ماشین حساب یعنی ۱/۳۶۹۱۰۷۷۷۱ کمتر از ۰/۰۰۱ اختلاف دارد.

روش دوم- به کمک سری تیلور داریم

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots.$$

در نتیجه

$$e^{\frac{\pi}{10}} = 1 + \frac{\pi}{10} + \frac{(\frac{\pi}{10})^2}{2!} + \cdots + \frac{(\frac{\pi}{10})^n}{n!} + \cdots.$$

با توجه به $\frac{\pi}{10} < 0/315$ (دقت $3D$ منظور می‌شود) از $\left|\frac{(0/315)^n}{n!}\right| < 0/5 \times 10^{-2}$ نتیجه می‌شود $n \geq 4$. پس

$$\begin{aligned} e^{\frac{\pi}{10}} &\simeq 1 + 0/315 + \tfrac{(0/315)^2}{2!} + \tfrac{(0/315)^3}{3!} + \tfrac{(0/315)^4}{4!} \\ &= 1/315 + 0/050 + 0/005 + 0/000 = 1/370. \end{aligned}$$

$\triangle$

تذکر ۱.۱۸ بعضی مواقع ممکن است تعداد جملاتی که از روش دوم به دست می‌آید کافی نباشد و بهتر است با جملات بیشتر هم مقایسه کرد.

مثال ۱.۱۸ (همگرایی سریع) می‌خواهیم تابع $\cos x$ را به ازای مقادیر $|x| < \frac{\pi}{2}$ با دقت $5D$ ارزیابی کنیم. با توجه به سری تیلور

$$\cos x = \sum_{i=0}^{\infty}(-1)^i \frac{x^{2i}}{(2i)!},$$

و این که

$$\left|\frac{x^{2n}}{(2n)!}\right| < \frac{(\frac{\pi}{2})^{2n}}{(2n)!} < \frac{1/6^{2n}}{(2n)!},$$

باید داشته باشیم

$$\frac{1/6^{2n}}{(2n)!} < 0/5 \times 10^{-5},$$

که نتیجه می‌دهد $n \geq 6$. △

مثال ۱.۱۹ *(همگرایی کند)* با توجه به

$$\frac{1}{1+t} = 1 - t + t^2 - t^3 + - \cdots,$$

داریم

$$\int_0^x \frac{dt}{1+t} = \int_0^x (1 - t + t^2 - t^3 + - \cdots)dt,$$

و یا

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + - \cdots.$$

برای ارزیابی $\ln(1+x)$ با دقت $5D$، باید داشته باشیم

$$\left|\frac{x^n}{n}\right| < 0/5 \times 10^{-5},$$

که برای $x = 0/99$ نتیجه می‌دهد $n \geq 582$. △

مثال ۱.۲۰ تابع $Si(x) = \int_0^x \frac{\sin t}{t}dt$ برای $x \geq 0$ را در نظر بگیرید. دست کم چند جمله از بسط مک‌لورن تابع $f(t) = \sin t$ لازم است تا $Si(1)$ با دقت $6D$ مشخص شود؟
با اِعمال قضیه تیلور برای تابع f داریم

$$Si(1) = \int_0^1 \frac{1}{t}\left(f(0) + \frac{f'(0)}{1!}t + \frac{f''(0)}{2!}t^2 + \cdots + \frac{f^{(k)}(0)}{k!}t^k + \frac{f^{(k+1)}(y(t))}{(k+1)!}t^{k+1}\right)dt,$$

و یا

$$Si(1) = \int_0^1 \frac{1}{t}\left(t - \frac{t^3}{3!} + \frac{t^5}{5!} + \cdots + (-1)^n\frac{t^{2n+1}}{(2n+1)!} + (-1)^{n+1}\frac{t^{2n+2}}{(2n+2)!}\sin(y(t))\right)dt.$$

در نتیجه

$$Si(1) = \int_0^1 \left(1 - \frac{t^2}{3!} + \frac{t^4}{5!} + \cdots + (-1)^n\frac{t^{2n}}{(2n+1)!}\right)dt + \frac{(-1)^{n+1}}{(2n+2)!}\int_0^1 t^{2n+1}\sin(y(t))dt.$$

بنابراین

$$Si(1) = 1 - \frac{1}{3\times 3!} + \frac{1}{5\times 5!} + \cdots + (-1)^n\frac{1}{(2n+1)\times(2n+1)!} + E,$$

که در آن

$$E = \frac{(-1)^{n+1}}{(2n+2)!}\int_0^1 t^{2n+1}\sin(y(t))dt.$$

حال می‌توان نوشت

$$|E| = \frac{1}{(2n+2)!}\left|\int_0^1 t^{2n+1}\sin(y(t))dt\right| \le \frac{1}{(2n+2)!}\int_0^1 t^{2n+1}\left|\sin(y(t))\right|dt.$$

پس

$$|E| \le \frac{1}{(2n+2)!}\int_0^1 t^{2n+1}dt = \frac{1}{(2n+2)\times(2n+2)!}.$$

از $\frac{1}{(2n+2)\times(2n+2)!} < 0/5\times 10^{-6}$ *نتیجه می‌شود* $n \ge 4$*. بنابراین*

$$Si(1) \simeq 1 - \frac{1}{3\times 3!} + \frac{1}{5\times 5!} - \frac{1}{7\times 7!} + \frac{1}{9\times 9!},$$

و یا

$$Si(1) \simeq 1 - 0/0555556 + 0/0016667 - 0/0000283 + 0/0000003 = 0/9460831.$$

پس با دقت $6D$ *داریم* $Si(1) \simeq 0/946083$. $\triangle$

تمرین

۱. یک استوانه به شعاع قاعده $\frac{4}{3}$ و ارتفاع $\sqrt{2}$ در نظر بگیرید. اگر بخواهیم حجم این استوانه را با دقت $3D$ به دست آوریم، شعاع قاعده و ارتفاع استوانه و حتی عدد π را با چه دقتی وارد محاسبات کنیم؟

۲. مطلوب است تعیین تقریبی از عدد π با دقت $3D$ به کمک بسط مکلورن تابع $f(x) = \tan^{-1} x$.

۳. حداقل چند جمله از بسط مک‌لورن تابع $f(x) = e^{-\frac{x}{2}}$ نیاز است تا $f(\ln(2))$ با دقت $3D$ به دست آید؟

۴. حداکثر خطای نسبی محاسبه حجم یک مخروط از فرمول $V = \frac{\pi}{3}r^2h$ چقدر است، هرگاه تقریبی از عدد π با خطای نسبی حداکثر $0/003$ داشته باشیم و کمیت‌های r و h را بتوان با حداکثر خطای نسبی $0/002$ اندازه‌گیری کرد.

۵. برای محاسبه عبارت $e^{1-\ln(2/73)}$ در یک ماشین‌حساب با سه رقم بامعنا کدام گزینه مناسب‌تر است؟

الف) $e^{1-\ln(2/73)}$ ب) $\frac{e}{2/73}$ ج) $e\times e^{-\ln(2/73)}$ د) $\frac{e}{e^{\ln(2/73)}}$

۶. برای همه گزینه‌ها a_0 تقریبی از عدد π با دقت $3S$ است. در کدام مورد خطای مطلق در محاسبه a_{10} کمتر است؟

الف) $a_n = 1 + 2a_{n-1}$ ب) $a_n = 1 + na_{n-1}$ ج) $a_n = 1 + a_{n-1}$ د) $a_n = 1 + \frac{1}{2}a_{n-1}$

۷. برای محاسبه عبارت $\frac{1-\cos(x)}{x}$ در کامپیوتر، وقتی x عدد مثبت کوچکی است، کدام گزینه مناسب‌تر است؟

الف) $\frac{1-\cos(x)}{x}$ ب) $\frac{2\sin^2(\frac{x}{2})}{x}$ ج) $\frac{\sin^2(x)}{x(1+\cos(x))}$ د) $\frac{1}{x} - \frac{\cos(x)}{x}$

۸. برای محاسبه $\sqrt{x^2+\frac{1}{x^2}}-\sqrt{x^2-\frac{1}{x^2}}$ برای xهای خیلی بزرگ در ماشین کدام گزینه مناسب‌تر است؟

الف) $\frac{1}{x^3}$ ب) $\frac{2}{x\left(\sqrt{x^4+1}+\sqrt{x^4-1}\right)}$ ج) $\sqrt{x^2+\frac{1}{x^2}}-\sqrt{x^2-\frac{1}{x^2}}$ د) $\frac{1}{x}\left(\sqrt{x^4+1}-\sqrt{x^4-1}\right)$

۹. در یک ماشین با دقت $4S$ و گرد کردن، کوچکترین عدد طبیعی x که در تساوی $x+1=x$ صدق می‌کند، کدام گزینه است؟ الف) $0/5000\times 10^4$ ب) $0/1000\times 10^5$ ج) $0/5000\times 10^5$ د) $0/1000\times 10^6$

۱۰. برای زوایای کمتر از ۶ درجه، $\sin(x)$ را با مقدار x بر حسب رادیان تقریب می‌زنند. کران بالای خطای مطلق این کار چقدر است؟

۱۱. حداقل چند جمله از بسط مک‌لورن تابع $f(x)=\sin(\frac{x}{2})$ لازم است تا $\sin(0/1)$ با دقت $4D$ مشخص شود؟

۱۲. درصد خطای محاسبه عدد π و اندازه‌گیری شعاع دایره به ترتیب در هر یک از گزینه‌های زیر داده شده است. با انتخاب کدام گزینه می‌توان مساحت دایره را با حداکثر خطای نسبی 10^{-2} مشخص کرد؟

الف) ۰٫۵٪ و ۰٫۲۵٪ ب) ۰٫۶٪ و ۰٫۳٪ ج) ۰٫۵٪ و ۰٫۳۵٪ د) ۰٫۵۵٪ و ۰٫۲۵٪

۱۳. می‌خواهیم کمیت‌های U و V را از فرمول‌های $U=x^2\div y$ و $V=xy^3$ محاسبه کنیم. درصد خطای x و y به ترتیب در هر یک از گزینه‌های زیر داده شده است. با انتخاب کدام گزینه می‌توان U و V را با حداکثر خطای نسبی 10^{-2} مشخص کرد؟

الف) ۰٫۲٪ و ۰٫۳۵٪ ب) ۰٫۲۵٪ و ۰٫۳٪ ج) ۰٫۳٪ و ۰٫۲۵٪ د) ۰٫۳۵٪ و ۰٫۲٪

۱۴. کدام یک از گزینه‌های زیر در حساب ممیز شناور درست است؟

الف) عمل جمع دارای خاصیت شرکت پذیری است. ب) عضو خنثای عمل جمع یکتا است.

ج) عمل جمع خاصیت جابجایی دارد. د) جمع دو عدد مثبت نزدیک به هم دارای خطای زیادی است.

۱۵. حداکثر خطای نسبی محاسبه حجم یک کره از فرمول $V=\frac{4}{3}\pi r^3$ چقدر است، هرگاه تقریبی از اعداد π و $\frac{1}{3}$ با خطای نسبی حداکثر ۰٫۰۰۲ داشته باشیم و شعاع را بتوان با حداکثر خطای نسبی ۰٫۰۰۲ اندازه‌گیری کرد؟

الف) ۰٫۰۰۱ ب) ۰٫۰۱۰ ج) ۰٫۰۲۰ د) ۰٫۰۰۲

۱۶. کدام یک از گزینه‌های زیر در حساب ممیز شناور درست نیست؟

الف) در محاسبه تقریبی $\frac{\sqrt{2}}{1000000}$ قطعا خطای زیادی داریم.

ب) در محاسبه تقریبی $1000000\sqrt{2}$ قطعا خطای زیادی داریم.

ج) در تفاضل دو عدد منفی نزدیک به هم، انتشار خطا رخ می‌دهد.

د) در تفاضل دو عدد مثبت نزدیک به هم، انتشار خطا رخ می‌دهد.

۱۷. در محاسبه عبارت $\dfrac{e^{x^2}-\cos x}{x^2}$ کدام گزینه صحیح است؟

الف) به ازای تمام مقادیر x انتشار خطا رخ نمی‌دهد. ب) برای مقادیر بزرگ $|x|$ انتشار خطا رخ می‌دهد.

ج) برای مقادیر کوچک $|x|$ انتشار خطا رخ می‌دهد. د) به ازای همه مقادیر $|x|$ انتشار خطا رخ می‌دهد.

۱۸. در یک دستگاه ممیز شناور که در آن اعداد به صورت $\pm ۰/d_۱d_۲d_۳d_۴ \times ۱۰^n$ با $d_۱ \neq ۰$ و $-۷ \leq n \leq ۸$ و $۰ \leq d_i \leq ۹$ نمایش داده می‌شوند فاصله بین عدد ۱۰۰۰۰ و اولین عدد قابل نمایش بزرگتر از ۱۰۰۰۰ کدام است؟

الف) ۱۰ ب) ۰٫۱ ج) ۰٫۰۱ د) ۰٫۰۰۵

۱۹. اگر مقدار $\ln(۱+x)$ را با استفاده از دو جمله اول بسط مکلورن آن تقریب بزنیم، مقدار تقریبی و حداکثر خطای محاسبه $\ln(۱٫۱)$ به ترتیب کدام است؟

الف) $(۰٫۰۵, ۰٫۰۰۱)$ ب) $(۰٫۰۵, ۰٫۱)$ ج) $(۰٫۰۰۵, ۰٫۰۵)$ د) $(۰٫۰۰۵, ۰٫۱)$

۲۰. در محاسبه دوره تناوب آونگ ساده $T = ۲\pi\sqrt{\frac{l}{g}}$، درباره اثر خطای نسبی مقادیر π, l, g در خطای نسبی T چه می‌توان گفت؟

الف) اثر خطای نسبی π بیشتر است. ب) اثر خطای نسبی l بیشتر است.

ج) اثر خطای نسبی g بیشتر است. د) π, l, g اثر یکسانی دارند.

۲۱. در یک دستگاه ممیز شناور که در آن اعداد به صورت $\pm ۰/d_۱d_۲d_۳d_۴ \times ۱۰^n$ با $d_۱ \neq ۰$ و $۰ \leq d_i \leq ۹$ و $-۷ \leq n \leq ۸$ با گرد کردن نمایش داده می‌شوند، بزرگترین عدد x که در معادله $۵۰۰ + x = ۵۰۰$ صدق می‌کند، کدام است؟

الف) ۰٫۰۴۹۹۹ ب) ۰٫۰۹۹۹۹ ج) ۰٫۰۵۰۰۰ د) ۰٫۱۰۰۰

۲۲. اگر مقدار دقیق و تقریبی در توانی از ده ضرب شوند، آنگاه

الف) خطای نسبی به توانی از ده تقسیم می‌شود. ب) خطای مطلق به توانی از ده تقسیم می‌شود.

ج) خطای نسبی تغییری نمی‌کند. د) خطای مطلق تغییری نمی‌کند.

۲۳. گزینه مناسب برای محاسبه $T = \frac{۱ - \cos x}{\sin x}$ به ازای مقادیر x نزدیک صفر کدام است؟

الف) ۰ ب) $\frac{x}{۲}$ ج) $\frac{\pi}{۲}$ د) $\frac{۱}{\sin x} - \cot x$

۲۴. در محاسبه تقریبی $\ln(۱٫۰۰۱)$ با سه جمله از بسط تیلور، حداکثر خطا چقدر است؟

الف) $\frac{۱۰^{-۶}}{۳}$ ب) $\frac{۱۰^{-۹}}{۳}$ ج) $\frac{۱۰^{-۴}}{۳}$ د) $\frac{۱۰^{-۶}}{۲}$

۲۵. در محاسبه حجم $(V = \frac{۴}{۳}\pi r^۳)$ یک کره به شعاع r، درباره اثر خطای نسبی داده‌ها در خطای نسبی V چه می‌توان گفت؟

الف) اثر خطای نسبی r بیشتر است. ب) اثر خطای نسبی π بیشتر است.

ج) اثر خطای نسبی $\frac{۴}{۳}$ بیشتر است. د) هر سه اثر خطای نسبی یکسانی دارند.

۲۶. گزینه مناسب برای محاسبه $T = \frac{\cos(۲x) - ۱ + ۲x^۲}{x^۴}$ به ازای مقادیر x نزدیک صفر کدام است؟

الف) $\frac{۳}{۲}$ ب) $\frac{۲}{۳}$ ج) $\frac{۲x}{۳}$ د) $\frac{۳x}{۲}$

۲۷. برای محاسبه $f(3)$ با خطایی کمتر از ۰/۰۰۰۷ چند جمله از بسط تیلور تابع f حول نقطه $x_0 = 2/5$ لازم است (می‌دانیم $|f^{(n)}(x)| \leq \frac{1}{n}$)؟

الف) دو جمله ب) سه جمله ج) پنج جمله د) چهار جمله

۲۸. در یک دستگاه ممیز شناور نرمال‌شده که هر عدد حقیقی به صورت $\pm 0/d_1d_2d_3d_4 \times 10^e$ با $d_1 \neq 0$ و $0 \leq d_i \leq 9$ برای $i = 1, 2, 3, 4$ و $-6 \leq e \leq 7$، نمایش داده می‌شود، اعداد قابل نمایش قبل و بعد از ۱۰۰۰۰ کدام هستند؟

الف) ۹۹۹۹, ۱۰۰۰۱ ب) ۹۹۹۰, ۱۰۰۱۰ ج) ۹۹۹۹, ۱۰۰۱۰ د) ۹۹۹۰, ۱۰۰۰۱

۲۹. در دستگاه ممیز شناور نرمال‌شده سوال قبل، تعداد اعداد قابل نمایش در بازه $[1, 10]$ کدام است؟

الف) ۹۰۰۱ ب) ۹۰۰۰ ج) ۸۹۹۹ د) ۹۰۰۲

فصل ۲

ریشه‌یابی (حل معادلات غیرخطی)

هدف از این فصل یافتن ریشه معادله $f(x) = 0$ یا صفر تابع f است، یعنی α را به گونه‌ای می‌یابیم که داشته باشیم $f(\alpha) = 0$. در حالت کلی هیچ روش تحلیلی برای یافتن α زمانی که f یک چندجمله‌ای درجه پنج یا بالاتر باشد وجود ندارد. هم‌چنین برای دسته بزرگی از معادلات که f شامل توابع متعالی باشد یا روش حل تحلیلی موجود نیست یا پیچیده است. به عنوان مثال برای معادلات ساده‌ای مانند $x^5 - x^3 - 1 = 0$ و $x + \cos x = 0$ روش حل تحلیلی وجود ندارد. پس در هر یک از این حالات، روش‌های عددی را به کار گرفته و به کمک آن‌ها جوابی تقریبی برای معادله می‌یابیم.

۱.۲ بررسی کمّی ریشه‌ها

در این بخش قصد داریم وجود و تعداد ریشه‌های یک معادله داده‌شده را مورد بررسی قرار دهیم. در این راستا نه تنها می‌توان از آن دسته از قضایای ریاضیات عمومی که به بررسی رفتار تابع می‌پردازند (قضایای مربوط به اکسترمم‌ها) بهره برد بلکه قضیه بولزانو (حالت خاصی از قضیه مقدار میانی) یک قضیه کلیدی است.

قضیه ۱.۲ **(بولزانو)** *اگر f تابعی پیوسته در $[a, b]$ باشد و $f(a)f(b) < 0$، آن‌گاه معادله $f(x) = 0$ دست کم یک ریشه در (a, b) دارد و اگر f بر $[a, b]$ یکنوا (صعودی یا نزولی اکید) باشد آن ریشه منحصر به فرد است.*

در این راستا با دو مسئله اساسی مواجه هستیم که عبارتند از

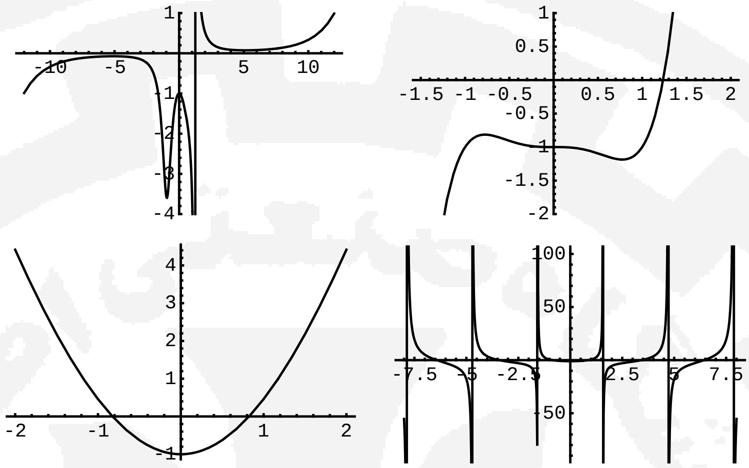
۱. یافتن بازه $[a, b]$ (تا آن جا که ممکن است کوچک) شامل فقط یک ریشه؛

۲. یافتن ریشه با دقت مطلوب.

برای بررسی مورد اول از ابزارهای زیر استفاده می‌کنیم.

- بررسی رفتار تابع (رسم نمودار تابع)؛
- جدول‌بندی مقادیر تابع؛
- تلفیق دو مورد قبل؛
- استفاده از قضایای ریاضی.

مثال ۱.۲ *معادله* $۳\cosh x - ۲\cos x = ۰$ *ریشه ندارد (شکل ۱.۲ سمت چپ بالا) در حالی که معادله* $x - \cos x = ۰$ *فقط یک ریشه در بازه* $[۰, \frac{\pi}{۲}]$ *دارد (شکل ۱.۲ سمت راست بالا). همچنین معادله* $x^۲ - \cos x = ۰$ *دو ریشه قرینه (شکل ۱.۲ سمت چپ پایین) و معادله* $x\tan x - ۱ = ۰$ *بی‌نهایت ریشه مثبت و منفی دارد (شکل ۱.۲ سمت راست پایین).* △



شکل ۱.۲: بررسی تعدد ریشه‌ها به کمک رسم

مثال ۲.۲ *با انتخاب* a *و* b *مناسب و* $n \in \mathbb{N}$*، قرار می‌دهیم* $h = \frac{b-a}{n}$ *و با* $x_۰ = a$ *برای* $i = ۰, \ldots, n$ *قرار می‌دهیم* $x_i = x_۰ + ih$ *و جدولی از مقادیر تابع* f *در نقاط* $x_۰, \ldots, x_n$ *ساخته، سپس از قضیه بولزانو کمک گرفته و از* $f(x_i)f(x_{i+۱}) < ۰$ *نتیجه می‌گیریم* f *دست کم یک ریشه در بازه* $[x_i, x_{i+۱}]$ *دارد. با انتخاب* $a = -۱۰$*،* $b = ۱۰$ *و* $n = ۱۰$ *جدول ۱.۲ را برای تابع* $f(x) = x^۵ - x^۳ - ۱$ *می‌سازیم.*

x	$f(x)$	x	$f(x)$	x	$f(x)$	x	$f(x)$
−۱۰	−۹۹۰۰۱	۰	−۱	۱٫۲۰	−۰٫۲۳۹۶۸	۱٫۲۲۰	−۰٫۱۱۳۱۴
−۸	−۳۲۲۵۷	۰٫۲	−۱٫۰۰۷۶۸	۱٫۲۲	−۰٫۱۱۳۱۴	۱٫۲۲۲	−۰٫۰۹۹۸۵۹
−۶	−۷۵۶۱	۰٫۴	−۱٫۰۵۳۷۶	۱٫۲۴	۰٫۰۲۵۰۰۱	۱٫۲۲۴	−۰٫۰۸۶۴۶۱۱
−۴	−۹۶۱	۰٫۶	−۱٫۱۳۸۲۴	۱٫۲۶	۰٫۱۷۵۴۲۱	۱٫۲۲۶	−۰٫۰۷۲۹۴۶
−۲	−۲۵	۰٫۸	−۱٫۱۸۴۳۲	۱٫۲۸	۰٫۳۳۸۸۲۲	۱٫۲۲۸	−۰٫۰۵۹۳۱۳
۰	−۱	۱٫۰	−۱	۱٫۳۰	۰٫۵۱۵۹۳	۱٫۲۳۰	−۰٫۰۴۵۵۶۱۳
۲	۲۳	۱٫۲	−۰٫۲۳۹۶۸	۱٫۳۲	۰٫۷۰۷۴۹۶	۱٫۲۳۲	−۰٫۰۳۱۶۹۰۳
۴	۹۵۹	۱٫۴	۱٫۶۳۴۲۴	۱٫۳۴	۰٫۹۱۴۲۹۶	۱٫۲۳۴	−۰٫۰۱۷۶۹۹۲
۶	۷۵۵۹	۱٫۶	۵٫۳۸۹۷۶	۱٫۳۶	۱٫۱۳۷۶۸۳	۱٫۲۳۶	−۰٫۰۰۳۵۸۷۴
۸	۳۲۲۵۵	۱٫۸	۱۲٫۰۶۳۷۰	۱٫۳۸	۱٫۳۷۶۸۳	۱٫۲۳۸	۰٫۰۱۰۶۴۵۸
۱۰	۹۸۹۹۹	۲٫۰	۲۳٫۰۰۰۰۰	۱٫۴۰	۱٫۶۳۴۲۴	۱٫۲۴۰	۰٫۰۲۵۰۰۱۱

جدول ۱.۲: جدول‌بندی مقادیر تابع $f(x) = x^۵ - x^۳ - ۱$

بنابراین $۱٫۲۳۶ \leq \alpha \leq ۱٫۲۳۸$ *و در نتیجه با دقت* $۲D$ *داریم* $\alpha = ۱٫۲۴$. △

تذکر ۱.۲ *روش به کار رفته در مثال اخیر کارایی ندارد.*

مثال ۳.۲ *بدون رسم و جدول‌بندی ثابت می‌کنیم $f(x) = x^{۲} - (۱ - x)^{۵}$ فقط یک ریشه دارد. به کمک آزمون سعی و خطا داریم $f(۰) = -۱$ و $f(۱) = ۱$. پس دست کم یک ریشه در بازه $[۰, ۱]$ موجود است. از طرفی $f'(x) = ۲x + ۵(۱ - x)^{۴}$ و اگر $x \geq ۰$ آن‌گاه $f'(x) > ۰$ و بنابراین f در $[۰, \infty)$ صعودی است. هم‌چنین $f(x) = x^{۵} - ۵x^{۴} + ۱۰x^{۳} - ۹x^{۲} + ۵x - ۱$ و از این که ضرایب توان‌های زوج منفی و ضرایب توان‌های فرد مثبت است نتیجه می‌گیریم که اگر $x < ۰$ آن‌گاه $f(x) < ۰$ و این یعنی f ریشه‌ای در $(-\infty, ۰)$ ندارد.* $\triangle$

۲.۲ دنباله‌های همگرا

بیشتر روش‌های عددی، ساختاری تکراری دارند و به همین دلیل به آن‌ها روش‌های تکراری نیز گفته می‌شود. یک روش تکراری دنباله $\{x_n\}$ را تولید می‌کند که امیدواریم به α (صفر f) همگرا باشد. یک روش همگرا روشی است که دنباله‌ای همگرا به صفر تابع تولید می‌کند. در ادامه این بخش چند اصطلاح مربوط به همگرایی را مرور می‌کنیم.

تعریف ۱.۲ *فرض کنید $\{x_n\}$ و $\{\alpha_n\}$ دو دنباله متفاوت باشند. می‌نویسیم*

$$x_n = O(\alpha_n)$$

و می‌خوانیم x_n اُی بزرگ α_n است، هرگاه ثابت‌های $C > ۰$ و $N \in \mathbb{N}$ چنان وجود داشته باشند که

$$|x_n| \leq C\,|\alpha_n|, \qquad \forall n \geq N.$$

اگر برای هر n داشته باشیم $\alpha_n \neq ۰$، آن‌گاه می‌توان گفت نسبت $|x_n/\alpha_n|$ کران‌دار (با کران C) باقی می‌ماند هرگاه $n \to \infty$. هر چند که $\{\alpha_n\}$ یک دنباله دلخواه است ولی در عمل، بیشتر مواقع $\alpha_n = \frac{۱}{n^p}$ اختیار می‌شود که در آن $p > ۰$ و در صدد یافتن بزرگ‌ترین مقدار p هستیم به طوری که

$$x_n = O\left(\frac{۱}{n^p}\right).$$

گاهی مواقع می‌نویسیم

$$x_n = o(\alpha_n),$$

و می‌خوانیم x_n اُی کوچک α_n است هرگاه داشته باشیم $\lim_{n\to\infty} \frac{x_n}{\alpha_n} = ۰$.

تذکر ۲.۲ *بیشتر مواقع دو دنباله بیان‌شده در تعریف قبل همگرا به صفر هستند و اگر در این صورت داشته باشیم $x_n = O(\alpha_n)$ آن‌گاه می‌توان گفت همگرایی x_n به صفر متناسب با همگرایی α_n به صفر است، در حالی که اگر داشته باشیم $x_n = o(\alpha_n)$، بلافاصله نتیجه می‌گیریم همگرایی x_n به صفر سریعتر از همگرایی α_n به صفر است.*

مثال ۴.۲ برای $n \in \mathbb{N}_0$ فرض کنید

$$x_n = \frac{n+1}{n^2}, \qquad y_n = \frac{n+1}{n^3}, \qquad z_n = \frac{n^2+1}{n^3}.$$

واضح است که هر سه دنباله به صفر همگرا هستند. به جدول زیر نگاه کنید.

n	۱	۲	۳	۴	۵	۱۰۰
x_n	۲/۰۰۰۰۰	۰/۷۵۰۰۰	۰/۴۴۴۴۴	۰/۳۱۲۵۰	۰/۲۴۰۰۰	۰/۰۱۰۱۰
y_n	۲/۰۰۰۰۰	۰/۳۷۵۰۰	۰/۱۴۸۱۵	۰/۰۷۸۱۳	۰/۰۴۸۰۰	۰/۰۰۰۱۰
z_n	۲/۰۰۰۰۰	۰/۶۲۵۰۰	۰/۳۷۰۳۷	۰/۲۶۵۶۳	۰/۲۰۸۰۰	۰/۰۱۰۰۰

اگر قرار دهیم $\alpha_n = 1/n$ و $\beta_n = 1/n^2$ آن‌گاه

$$\lim_{n\to\infty} \frac{x_n}{\alpha_n} = \lim_{n\to\infty} \frac{n+1}{n} = 1, \qquad \lim_{n\to\infty} \frac{y_n}{\beta_n} = \lim_{n\to\infty} \frac{n+1}{n} = 1, \qquad \lim_{n\to\infty} \frac{z_n}{\alpha_n} = \lim_{n\to\infty} \frac{n^2+1}{n^2} = 1.$$

بنابراین

$$x_n = O\left(\frac{1}{n}\right), \qquad y_n = O\left(\frac{1}{n^2}\right), \qquad z_n = O\left(\frac{1}{n}\right).$$

هم‌چنین می‌توان نوشت $x_n = O\left(z_n\right)$، $z_n = O\left(x_n\right)$، $y_n = o(x_n)$ و $y_n = o(z_n)$. △

تذکر ۳.۲ گاهی با فرض $h = 1/n$ روابط را بر حسب h بررسی می‌کنیم زمانی که $h \to 0$.

مثال ۵.۲ به کمک بسط مکلورن خواهیم داشت

$$\cos h = 1 - \frac{h^2}{2!} + \frac{h^4}{4!} - \frac{h^6}{6!} + - \cdots.$$

بنابراین

$$\cos h - 1 + \frac{h^2}{2} = h^4 \left(\frac{1}{4!} - \frac{h^2}{6!} + \cdots \right),$$

و در نتیجه

$$\lim_{h\to 0} \frac{|\cos h - 1 + \frac{1}{2}h^2|}{h^4} = \frac{1}{24}.$$

پس $\cos h - 1 + \frac{1}{2}h^2 = O(h^4)$. △

تعریف ۲.۲ فرض کنید $\{x_n\}$ دنباله‌ای همگرا به x باشد و برای هر n داشته باشیم $x_n \neq x$. گوییم این دنباله به طور خطی به x همگرا است هرگاه ثابت $C \in (0, 1)$ چنان وجود داشته باشد که

$$\lim_{n\to\infty} \frac{|x_{n+1} - x|}{|x_n - x|} = C.$$

در این حالت C به نرخ همگرایی[۱] معروف است. برای $C = ۰$ ($C = ۱$) همگرایی زبرخطی[۲] (زیرخطی[۳]) نامیده می‌شود. اگر اعداد مثبت و ثابت C و $p > ۱$ وجود داشته باشند به طوری که

$$\lim_{n\to\infty} \frac{|x_{n+۱} - x|}{|x_n - x|^p} = C$$

آن‌گاه گوییم مرتبه همگرایی[۴] این دنباله p است. C به ثابت خطای مجانبی[۵] معروف است. اگر $p = ۲$ ($p = ۳$) همگرایی مربعی (مکعبی) نامیده می‌شود ($C < ۱$ لزومی ندارد).

مثال ۶.۲ فرض کنید $\{x_n\}$ دنباله‌ای با همگرایی خطی به صفر با

$$\lim_{n\to\infty} \frac{|x_{n+۱}|}{|x_n|} = ۰/۵,$$

و $\{y_n\}$ دنباله‌ای با همگرایی مربعی به صفر با همان ثابت خطای مجانبی باشد یعنی

$$\lim_{n\to\infty} \frac{|y_{n+۱}|}{|y_n|^۲} = ۰/۵.$$

برای سادگی فرض می‌کنیم برای $n \geq N$ داریم

$$\frac{|x_{n+۱}|}{|x_n|} \simeq ۰/۵, \qquad \frac{|y_{n+۱}|}{|y_n|^۲} \simeq ۰/۵.$$

بنابراین می‌توان نوشت

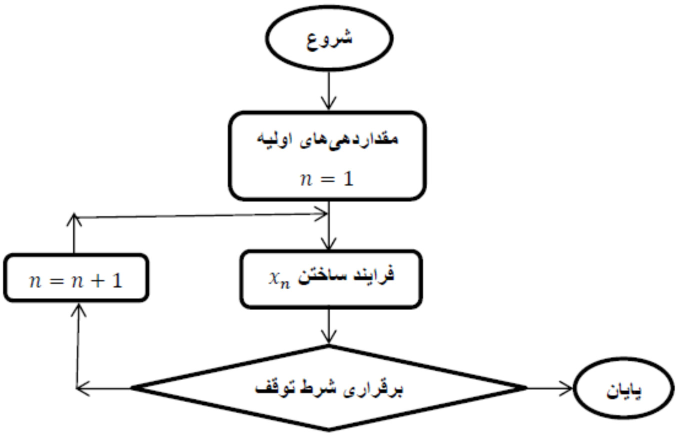
$$\begin{aligned}
|x_n - ۰| = |x_n| &\simeq ۰/۵\,|x_{n-۱}| \simeq (۰/۵)^۲\,|x_{n-۲}| \simeq \cdots \simeq (۰/۵)^n\,|x_۰|\,,\\
|y_n - ۰| = |y_n| &\simeq ۰/۵\,|y_{n-۱}|^۲ \simeq (۰/۵)(۰/۵\,|y_{n-۲}|^۲)^۲ = (۰/۵)^۳\,|y_{n-۲}|^۴\\
&\simeq (۰/۵)^۳(۰/۵\,|y_{n-۳}|^۲)^۴ = (۰/۵)^۷\,|y_{n-۳}|^۸\\
&\simeq \cdots \simeq (۰/۵)^{۲^n-۱}\,|y_۰|^{۲^n}\,.
\end{aligned}$$

△

مثال ۷.۲ دنباله $a_n = \frac{۱}{۲^n}$ با نرخ همگرایی $\frac{۱}{۲}$، همگرایی خطی به صفر دارد. همگرایی دنباله $b_n = \frac{۱}{۲^{۲^n}}$ به صفر زبرخطی است. در اصل همگرایی این دنباله مربعی است. دنباله $c_n = \frac{۱}{n+۱}$ همگرایی زیرخطی به صفر دارد. △

---

[۱] rate of convergence
[۲] superlinear
[۳] sublinear
[۴] order of convergence
[۵] asymptotic error constant



شکل ۲.۲: طرح کلی روش‌های تکراری

۳.۲ روش‌های عددی

برای برخورد با دومین مشکل اساسی مطرح‌شده در ابتدای فصل، به بررسی روش‌های عددی می‌پردازیم. همان‌طور که در طرح کلی روش‌های تکراری آمده در شکل ۲.۲ مشاهده می‌شود یک گام اساسی در این روش‌ها شرط توقف است که برای مسئله ریشه‌یابی انواع آن عبارتند از

۱) $|x_n - \alpha| < \epsilon$؛

۲) $|f(x_n)| < \epsilon$؛

۳) $|x_n - x_{n-۱}| < \epsilon$؛

۴) $\frac{|x_n - x_{n-۱}|}{|x_n|} < \epsilon$ (زمانی که ریشه خیلی کوچک یا بزرگ باشد)؛

۵) تعیین بیش‌ترین تعداد تکرار؛

۶) تلفیق موارد قبل؛

۷) استفاده از کران بالای خطا[۶].

تذکر ۴.۲ *در عمل چون به α دسترسی نداریم کمتر از اولین مورد استفاده می‌کنیم و اگر کران بالای خطا در دسترس باشد، مطمئن‌ترین شرط توقف است.*

تذکر ۵.۲ *برقراری $|f(x_n)| < \epsilon$ برقراری $|x_n - \alpha| < \epsilon$ را تضمین نمی‌کند. به عنوان مثال، صفر تابع $f(x) = \tan^{-۱} x - ۱/۵۵$ با دقت $۳D$ عبارت است از ۴۸/۰۷۹ (رادیان). از طرفی داریم $f(۵۰) \simeq ۸ \times ۱۰^{-۴}$ در حالی که $۵۰ - ۴۸/۰۷۹ = ۱/۹۲۱$.*

[۶] Upper error bound

تذکر ۶.۲ *اگر* $|x_n - x_{n-۱}| < \epsilon$ *برقرار باشد لزومی ندارد* $|x_n - \alpha| < \epsilon$ *برقرار باشد زیرا اگر* $\lim x_n = \alpha$ *آن‌گاه* $\lim |x_n - x_{n-۱}| = ۰$ *ولی عکس این مطلب درست نیست. به عنوان نمونه دنباله* $x_n = \sum_{k=۱}^{n} \frac{۱}{k}$ *واگرا است حال آن که داریم* $|x_n - x_{n-۱}| = \frac{۱}{n}$.

۱.۳.۲ روش دوبخشی

روش دوبخشی[۷] به روش نصف کردن، تنصیف و نیم‌سازی نیز شهرت دارد. ایده اصلی این روش آن است که، پس از حصول اطمینان از وجود ریشه منحصر به فرد در بازه $[a,b]$، بازه را نصف کرده و (به کمک قضیه بولزانو) نیم‌بازه‌ای که ریشه در آن وجود ندارد را کنار گذاشته و این فرایند را تکرار می‌کنیم. به شکل ۳.۲ نگاه کنید.

شکل ۳.۲: بررسی روش دوبخشی به کمک رسم

الگوریتم روش دوبخشی

- **ورودی.** a, b, f (با این فرض که تابع f در بازه $[a,b]$ صفر یکتا داشته باشد)
- **خروجی.** تقریبی از صفر تابع f

۱) قرار دهید $x = \frac{a+b}{۲}$

۲) اگر شرط توقف (یک یا چند شرط توقف بیان‌شده) برقرار باشد a، b و x را چاپ کرده و متوقف شوید

۳) اگر $f(a)f(x) < ۰$ قرار دهید $b = x$ و به گام ۱ برگردید

۴) قرار دهید $a = x$ و به گام ۱ برگردید

مثال ۸.۲ *معادله* $f(x) = x^۳ + ۴x^۲ - ۱۰ = ۰$ *ریشه‌ای یکتا در بازه* $[۱,۲]$ *دارد. زیرا* $f(۱) = -۵$، $f(۲) = ۱۴$ *و برای* $x \in [۱,۲]$ *داریم* $f'(x) = ۳x^۲ + ۸x > ۰$. *با اِعمال روش دوبخشی، جدول* ۲.۲ *به دست می‌آید. پس از* ۱۳ *تکرار* $x_{۱۳} = ۱/۳۶۵۱۱۲۳۰۵$، α *را با خطایی به اندازه*

$$|\alpha - x_{۱۳}| < |b_{۱۴} - a_{۱۴}| = |۱/۳۶۵۲۳۴۳۷۵ - ۱/۳۶۵۱۱۲۳۰۵| = ۰/۰۰۰۱۲۲۰۷۰,$$

تقریب می‌زند. چون $|a_{۱۴}| < |\alpha|$، *از*

$$\delta(x_{۱۳}) = \frac{|\alpha - x_{۱۳}|}{|\alpha|} < \frac{|b_{۱۴} - a_{۱۴}|}{|a_{۱۴}|} \le ۰/۹ \times ۱۰^{-۴} < ۰/۵ \times ۱۰^{-۳},$$

[۷] Bisection method

نتیجه می‌شود $x_{۱۳}$ دست کم سه رقم بامعنای درست دارد. ممکن است از رابطه $|f(x_{۱۳})| < |f(x_۹)|$ گمان کنیم $x_۹$ بهتر از $x_{۱۳}$، α را تقریب می‌زند ولی چون از α اطلاعی نداریم معلوم نیست آیا رابطه $|\alpha - x_۹| < |\alpha - x_{۱۳}|$ برقرار است یا نه! $\triangle$

n	a_n	b_n	x_n	$f(x_n)$
۱	۱/۰	۲/۰	۱/۵	۲/۳۷۵
۲	۱/۰	۱/۵	۱/۲۵	−۱/۷۹۶۸۷
۳	۱/۲۵	۱/۵	۱/۳۷۵	۰/۱۶۲۱۱
۴	۱/۲۵	۱/۳۷۵	۱/۳۱۲۵	−۰/۸۴۸۳۹
۵	۱/۳۱۲۵	۱/۳۷۵	۱/۳۴۳۷۵	−۰/۳۵۰۹۸
۶	۱/۳۴۳۷۵	۱/۳۷۵	۱/۳۵۹۳۷۵	−۰/۰۹۶۴۱
۷	۱/۳۵۹۳۷۵	۱/۳۷۵	۱/۳۶۷۱۸۷۵	۰/۰۳۲۳۶
۸	۱/۳۵۹۳۷۵	۱/۳۶۷۱۸۷۵	۱/۳۶۳۲۸۱۲۵	−۰/۰۳۲۱۵
۹	۱/۳۶۳۲۸۱۲۵	۱/۳۶۷۱۸۷۵	۱/۳۶۵۲۳۴۳۷۵	۰/۰۰۰۰۷۲
۱۰	۱/۳۶۳۲۸۱۲۵	۱/۳۶۵۲۳۴۳۷۵	۱/۳۶۴۲۵۷۸۱۳	−۰/۰۱۶۰۵
۱۱	۱/۳۶۴۲۵۷۸۱۳	۱/۳۶۵۲۳۴۳۷۵	۱/۳۶۴۷۴۶۰۹۴	−۰/۰۰۷۹۹
۱۲	۱/۳۶۴۷۴۶۰۹۴	۱/۳۶۵۲۳۴۳۷۵	۱/۳۶۴۹۹۰۲۳۵	−۰/۰۰۳۹۶
۱۳	۱/۳۶۴۹۹۰۲۳۵	۱/۳۶۵۲۳۴۳۷۵	۱/۳۶۵۱۱۲۳۰۵	−۰/۰۰۱۹۴

جدول ۲.۲: روش دوبخشی برای معادله $x^۳ + ۴x^۲ - ۱۰ = ۰$

تذکر ۷.۲ *روش دوبخشی برای یافتن ریشه چندگانه (با چندگانگی زوج) کارایی ندارد.*

قضیه ۲.۲ *اگر $f \in C[a,b]$ و $f(a)f(b) < ۰$ آن‌گاه روش دوبخشی دنباله $\{x_n\}_{n\geq ۱}$ را تولید می‌کند که به α (صفر f) همگرا است*[۸] *و داریم*

$$|x_n - \alpha| \leq \frac{b-a}{۲^n}, \qquad n \geq ۱.$$

تذکر ۸.۲ می‌توان نوشت $x_n = \alpha + O\left(\frac{۱}{۲^n}\right)$ *یعنی همگرایی دنباله $\{x_n\}_{n\geq ۱}$ به α متناسب با همگرایی $\{\frac{۱}{۲^n}\}_{n\geq ۱}$ به صفر خواهد بود و چون $۰/۰۰۱ < ۲^{-۱۰} = \frac{۱}{۱۰۲۴}$ می‌توان گفت پس از ۱۰ تکرار، بیشتر از سه رقم به ارقام بامعنای درست اضافه نمی‌شود و این نشان می‌دهد که همگرایی روش دوبخشی کند است.*

تمرین ۱.۲ *چند تکرار روش دوبخشی لازم است تا ریشه ساده و یکتای $f(x) = ۰$ در بازه $[a,b]$ با دقت rD به دست آید؟*

[۸] $f \in C[a,b]$ یعنی تابع f بر بازه $[a,b]$ پیوسته است.

مثال ۹.۲ *در ریاضیات مالی ثابت می‌شود اگر یک وام A تومانی با نرخ بهره ماهانه x در n قسط ماهانه به مبلغ P تومان مستهلک شود، رابطه $A = \frac{P}{x}(1-(1+x)^{-n})$ برقرار خواهد بود. اگر شخصی یک وام مسکن سی ساله به مبلغ ۱۲۱۵۰۰۰۰۰ تومان نیاز داشته باشد و قادر به پرداخت اقساط ماهانه ۹۰۰۰۰۰ تومان باشد حداکثر نرخ بهره‌ای که می‌تواند بپذیرد را با به کار بردن روش دوبخشی و با دقت $4D$ تعیین کنید.*

باید x را به قسمی تعیین کنیم که

$$121500000 = \frac{900000}{x}(1-(1+x)^{-30\times 12}),$$

به عبارت دیگر باید صفر تابع $f(x) = 135 + \frac{(1+x)^{-360}-1}{x}$ مشخص شود. برای این منظور اگر $a = 0/00600$ و $b = 0/00800$ انتخاب شود از $f(a) = -12/32136$ و $f(b) = 17/09771$ نتیجه می‌گیریم دست کم یک ریشه در بازه $[0/00600, 0/00800]$ وجود دارد. با یک بررسی ساده مشخص می‌شود این ریشه منحصر به فرد است!؟ از طرفی از نابرابری $\frac{0/00800-0/00600}{2^n} < 0/5 \times 10^{-4}$ در می‌یابیم کمترین تعداد تکرار لازم برای رسیدن به دقت مطلوب، $n = 6$ خواهد بود. از سطر آخر جدول ۳.۲ نتیجه می‌گیریم $0/00672 < \alpha < 0/00675$ و بنابراین با دقت $4D$ خواهیم داشت $x \simeq 0/0067$. تقریب بهتر، $0/00674992$ است و علت تفاوت محسوس دو مقدار $f(0/0067) \simeq -0/76719$ و $f(0/00674992) \simeq 4 \times 10^{-5}$ تنها کم بودن تعداد ارقام بامعنای $0/0067$ است. $\triangle$

n	a	b	$x = \frac{a+b}{2}$	$f(x)$
۱	۰/۰۰۶۰۰	۰/۰۰۸۰۰	۰/۰۰۷۰۰	۳/۷۳۸۴۴
۲	۰/۰۰۶۰۰	۰/۰۰۷۰۰	۰/۰۰۶۵۰	−۳/۹۱۳۸۷
۳	۰/۰۰۶۵۰	۰/۰۰۷۰۰	۰/۰۰۶۷۵	۰/۰۰۱۲۷
۴	۰/۰۰۶۵۰	۰/۰۰۶۷۵	۰/۰۰۶۶۳	−۱/۸۵۵۱۵
۵	۰/۰۰۶۶۳	۰/۰۰۶۷۵	۰/۰۰۶۶۹	−۰/۹۲۱۷۵
۶	۰/۰۰۶۶۹	۰/۰۰۶۷۵	۰/۰۰۶۷۲	−۰/۴۵۸۹۵

جدول ۳.۲: نتایج روش دوبخشی برای مثال ۹.۲

۲.۳.۲ روش نابجایی

در روش نابجایی[۹] مانند روش دوبخشی پس از حصول اطمینان از وجود ریشه منحصر به فرد در بازه $[a,b]$، بازه را به دو زیربازه تقسیم کرده و زیربازه‌ای که ریشه در آن وجود ندارد را کنار گذاشته و این فرایند را تکرار می‌کنیم. به شکل ۴.۲ نگاه کنید.

الگوریتم روش نابجایی

- **ورودی.** a, b, f (با این فرض که تابع f در بازه $[a,b]$ صفر یکتا داشته باشد)
- **خروجی.** تقریبی از صفر تابع f

[۹] Method of false position (regula falsi)

۱) قرار دهید $x = a - f(a)\frac{b-a}{f(b)-f(a)} = \frac{af(b)-bf(a)}{f(b)-f(a)}$

۲) اگر شرط توقف (یک یا چند شرط توقف بیان‌شده) برقرار باشد a، b و x را چاپ کرده و متوقف شوید

۳) اگر $f(a)f(x) < ۰$ قرار دهید $b = x$ و به گام ۱ برگردید

۴) قرار دهید $a = x$ و به گام ۱ برگردید

شکل ۴.۲: بررسی روش نابجایی به کمک رسم

مثال ۱۰.۲ *روش نابجایی در حل معادله* $f(x) = x^۳ + ۴x^۲ - ۱۰ = ۰$ *بر بازه* $[۱,۲]$ *نتایج جدول ۴.۲ را تولید می‌کند که با توجه به اختلاف* $x_۸$ *و* $x_۹$ *نتیجه می‌گیریم با دقت* $۴D$ *صفر تابع عبارت است از* $x \simeq ۱/۳۶۵۲$. $\triangle$

n	a_n	b_n	x_n	$f(x_n)$
۱	۱/۰	۲/۰	۱/۲۶۳۱۶	−۱/۶۰۲۲۷
۲	۱/۲۶۳۱۶	۲/۰	۱/۳۳۸۸۳	−۰/۴۳۰۳۷
۳	۱/۳۳۸۸۳	۲/۰	۱/۳۵۸۵۵	−۰/۱۱۰۰۱
۴	۱/۳۵۸۵۵	۲/۰	۱/۳۶۳۵۵	−۰/۰۲۷۷۶
۵	۱/۳۶۳۵۵	۲/۰	۱/۳۶۴۸۱	−۰/۰۰۶۹۸
۶	۱/۳۶۴۸۱	۲/۰	۱/۳۶۵۱۲	−۰/۰۰۱۷۶
۷	۱/۳۶۵۱۲	۲/۰	۱/۳۶۵۲۰	−۰/۰۰۰۴۴
۸	۱/۳۶۵۲۰	۲/۰	۱/۳۶۵۲۲	−۰/۰۰۰۱۱
۹	۱/۳۶۵۲۲	۲/۰	۱/۳۶۵۲۳	−۰/۰۰۰۰۳

جدول ۴.۲: روش نابجایی برای معادله $x^۳ + ۴x^۲ - ۱۰ = ۰$

تذکر ۹.۲ *روش نابجایی مانند روش دوبخشی همگرایی تضمین شده‌ای دارد و با آن که حجم عملیات آن حدود دو برابر روش دوبخشی است، بعضی مواقع سریع‌تر است. ولی ممکن است اکثر یا تمام عناصر دنباله* $\{x_n\}_{n\geq ۱}$ *در یک طرف ریشه* (α) *قرار بگیرند (مانند مثال اخیر) که در این حالت همگرایی روش نابجایی ممکن است از روش دوبخشی نیز کندتر باشد. در این صورت یا از روش‌های دیگر استفاده می‌شود و یا اگر بعد از دو تکرار* a *یا* b *تغییر مکان نداد باید در محاسبات* $\frac{f(a)}{۲}$ *یا* $\frac{f(b)}{۲}$ *با* $f(a)$ *یا* $f(b)$ *جایگزین شده و تکرارها دنبال شوند*[۱۰].

[۱۰] روش نابجایی اصلاح‌شده یا تغییریافته

۳.۳.۲ روش تکرار ساده

روش تکرار ساده[۱۱] که به روش تکرار نقطه ثابت[۱۲] و هم‌چنین روش تکرار تابعی[۱۳] نیز مشهور است، ایده ساده‌ای دارد که در ادامه آورده می‌شود. ابتدا معادله $f(x)=0$ را به صورت $x=g(x)$ به گونه‌ای تغییر می‌دهیم که اگر $f(\alpha)=0$ آن‌گاه $\alpha=g(\alpha)$. سپس به جای یافتن صفر تابع f، نقطه ثابت تابع g را جستجو می‌کنیم.

تعریف ۳.۲ *برای تابع $y=f(x)$ نقطه $z\in D_f$ را نقطه ثابت نامیم هرگاه $z=f(z)$ (محل برخورد نمودار f با نیمساز ناحیه اول و سوم).*

تذکر ۱۰.۲ *لزومی ندارد همه ریشه‌های $f(x)=0$، نقطه ثابت g باشند و بر عکس ممکن است همه نقاط ثابت g، ریشه $f(x)=0$ نباشند. به عنوان مثال $f_1(x)=\sqrt{x}-2\cos x$ یک صفر در $[1,2]$ و $f_2(x)=\sqrt{x}+2\cos x$ یک صفر در $[2,3]$ و یک صفر در $[3,4]$ دارد و اگر $x=g(x):=4\cos^2 x$ آن‌گاه g سه نقطه ثابت در $[1,2]$، $[2,3]$ و $[3,4]$ دارد. هم‌چنین $f(x)=x^2+x-1$ دو صفر در $[0,1]$ و $[-2,-1]$ دارد و $g_1(x)=\sqrt{1-x}$ یک نقطه ثابت در $[0,1]$ و $g_2(x)=-\sqrt{1-x}$ یک نقطه ثابت در $[-2,-1]$ دارد.*

قضیه ۳.۲ **(نقطه ثابت)** *آ) اگر $g:[a,b]\longrightarrow[a,b]$ و $g\in C[a,b]$ آن‌گاه g در $[a,b]$ حداقل یک نقطه ثابت دارد؛*

ب) هم‌چنین اگر g' بر (a,b) موجود باشد و عدد ثابت و مثبت L چنان وجود داشته باشد که

$$\forall\, x\in(a,b),\qquad |g'(x)|\le L<1,$$

آن‌گاه نقطه ثابت g در $[a,b]$ یکتا است؛

پ) به علاوه دنباله $\{x_n\}_{n\ge 1}$ تولیدشده توسط فرایند $x_n=g(x_{n-1})$ به ازای هر x_0 در $[a,b]$ به نقطه ثابت g همگرا است و داریم

$$|x_n-\alpha|\le\frac{L^n}{1-L}|x_1-x_0|.$$

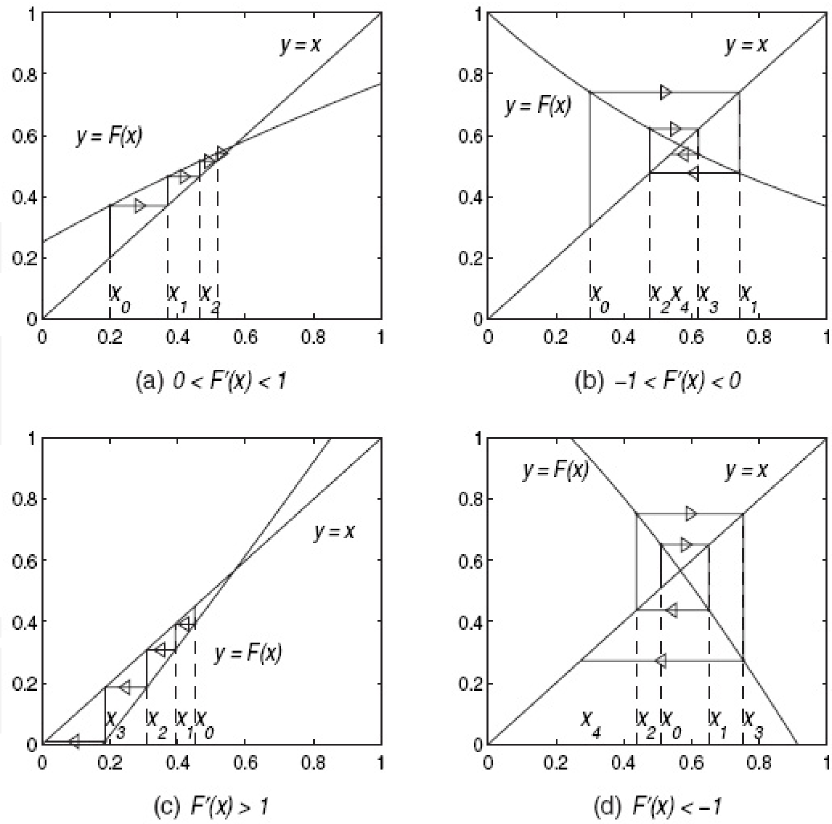
اگر شرایط قضیه نقطه ثابت برقرار باشد بلافاصله نتایج زیر به دست می‌آیند.

نتیجه ۱.۳.۲ $|x_n-\alpha|\le L^n\max\{x_0-a,b-x_0\}$.

نتیجه ۲.۳.۲ *اگر به ازای هر x در (a,b) داشته باشیم $0<g'(x)\le L<1$، آن‌گاه دنباله تولیدشده توسط روش تکرار ساده یعنی $\{x_n\}_{n\ge1}$ یکنوا است و اگر به ازای هر x در (a,b) داشته باشیم $-1<-L\le g'(x)<0$، آن‌گاه عناصر متوالی دنباله تولیدشده توسط روش تکرار ساده در دو طرف α قرار می‌گیرند.*

با توجه به کران خطا، هر چه L به صفر نزدیک‌تر باشد همگرایی سریع‌تر و هر چه L به یک نزدیک‌تر باشد همگرایی کندتر خواهد بود. در شکل ۵.۲ نمونه‌هایی از همگرایی یا واگرایی روش تکرار ساده آورده شده است.

---
[۱۱] Simple iteration
[۱۲] Fixed point iteration
[۱۳] Functional iteration



شکل ۵.۲: تعبیر هندسی روش تکرار ساده

الگوریتم روش تکرار ساده

- **ورودی.** $x_\circ, \epsilon, g$ (با این فرض که تابع g در فرضیات قضیه نقطه ثابت صدق کند)
- **خروجی.** مقدار x_n که به ازای آن $|x_n - x_{n-۱}| < \epsilon$

۱) قرار دهید $n = ۱$

۲) قرار دهید $x_n = g(x_{n-۱})$

۳) اگر $|x_n - x_{n-۱}| < \epsilon$ آن‌گاه x_n را چاپ کرده و متوقف شوید

۴) قرار دهید $n = n + ۱$ و به گام ۲ برگردید

مثال ۱۱.۲ *تابع $f(x) = x^۳ - x^۲ + ۱$ روی بازه $I = [-۱, -۰/۵]$ مفروض است. نشان دهید روش تکراری $x_n = x_{n-۱} - \frac{۱}{۶} f(x_{n-۱}),\ n = ۱, ۲, \dots$ برای هر $x_\circ$ متعلق به بازه I به صفر تابع f در این بازه همگرا است و این ریشه را با استفاده از این روش تکرار ساده با نقطه شروع $x_\circ = -۰/۷$ و با دقت $۳D$ به دست آورید.*

از این‌که $g(x) = x - \frac{۱}{۶} f(x) = -\frac{۱}{۶}x^۳ + \frac{۱}{۶}x^۲ + x - \frac{۱}{۶}$ پس $g'(x) = -\frac{۱}{۲}x^۲ + \frac{۱}{۳}x + ۱$ و از $g'(x) = ۰$ نتیجه می‌شود $x_۱ = \frac{۱-\sqrt{۱۹}}{۳} \simeq -۱/۱۲۰$ و $x_۲ = \frac{۱+\sqrt{۱۹}}{۳} \simeq ۱/۷۸۶$ و چون $I \subset [x_۱, x_۲]$ بنابراین برای هر $x \in I$ داریم $g'(x) > ۰$، یعنی تابع g بر I صعودی اکید است. از طرفی $g(-۱) = -\frac{۵}{۶}$ و $g(-۰/۵) = -\frac{۲۹}{۴۸}$ پس

$$g : [-۱, -۰/۵] \to [-\frac{۵}{۶}, -\frac{۲۹}{۴۸}] \subset [-۱, -۰/۵].$$

هم‌چنین داریم

$$-1 < x < -\frac{1}{2} \Rightarrow -\frac{4}{3} < x - \frac{1}{3} < -\frac{5}{6} \Rightarrow \frac{25}{36} < (x-\frac{1}{3})^2 < \frac{16}{9} \Rightarrow \frac{1}{6} < -\frac{1}{2}(x-\frac{1}{3})^2 + \frac{19}{18} < \frac{17}{24}$$

و چون $g'(x) = -\frac{1}{2}(x-\frac{1}{3})^2 + \frac{19}{18}$ *پس* $\forall x \in (-1, -0/5),\ |g'(x)| < \frac{17}{24}$. *بنابراین شرایط قضیه نقطه ثابت برقرار است. با توجه به مقادیر* x_n *گزارش‌شده در جدول ۵.۲ در می‌یابیم که* $\alpha = -0/755(3D)$.

n	۱	۲	۳	۴	۵
x_n	$-0/7278$	$-0/7419$	$-0/7488$	$-0/7520$	$-0/7536$
n	۶	۷	۸	۹	۱۰
x_n	$-0/7543$	$-0/7546$	$-0/7547$	$-0/7548$	$-0/7548$

جدول ۵.۲: نتایج روش تکرار ساده برای مثال ۱۱.۲

باید توجه داشت که با $L = \frac{17}{24}$ *و* $x_0 = -0/7$*، از کران خطای*

$$\frac{L^n}{1-L}|x_1 - x_0| \leq 0/5 \times 10^{-3},$$

خواهیم داشت $n \geq 16$*. چرا در عمل با تعداد تکرار کمتری جواب مطلوب به دست آمده است؟* △

تذکر ۱۱.۲ *در حالت کلی بهتر است کران‌های* g *و* g' *را با بررسی رفتار آنها (محاسبه اکسترمم‌ها) به دست آوردیم زیرا ممکن است در استفاده از نامساوی‌ها مرتکب اشتباه شویم و کران‌های بدبینانه‌ای به دست آوریم.*

مثال ۱۲.۲ *تقریبی از بزرگترین ریشه معادله* $2x - \ln x - 4 = 0$ *را با دقت* $4D$ *به دست آورید.*
همان طور که در شکل ۶.۲ مشاهده می‌شود این معادله دو ریشه به ترتیب در بازه‌های $[0, 1]$ *و* $[2, 3]$ *دارد.*

شکل ۶.۲: نمودار توابع $\ln x$ و $2x - 4$

با انتخاب $g(x) = 2 + \frac{1}{2}\ln x$ *داریم* $g'(x) = \frac{1}{2x}$*. چون* g' *در بازه* $[2, 3]$ *مثبت است پس* g *در این بازه صعودی است و بنابراین*

$$2 < g(2) = 2 + \frac{1}{2}\ln 2 \leq g(x) \leq g(3) = 2 + \frac{1}{2}\ln 3 < 3, \qquad \forall x \in [2, 3],$$

یعنی تابع g *به وضوح پیوسته بوده و بازه* $[2, 3]$ *را به توی بازه* $[2, 3]$ *می‌برد. از طرف دیگر چون در این بازه داریم*

$g''(x) = \frac{-1}{2x^2} < 0$، بنابراین g' در این بازه نزولی بوده و

$$g'(3) = \frac{1}{6} < g'(x) < g'(2) = \frac{1}{4}, \qquad \forall x \in (2,3),$$

و در نتیجه $L = \frac{1}{4}$. بنابراین شرایط قضیه نقطه ثابت برقرار است و از نابرابری

$$\frac{L^n}{1-L}|g(x_0) - x_0| \leq 0/5 \times 10^{-4},$$

به ازای $x_0 = 2/5$ داریم $n \geq 6$.

n	۱	۲	۳	۴	۵	۶	۷	۸
x_n	۲/۴۵۸۱۵	۲/۴۴۹۷۰	۲/۴۴۷۹۸	۲/۴۴۷۶۳	۲/۴۴۷۵۶	۲/۴۴۷۵۵	۲/۴۴۷۵۴	۲/۴۴۷۵۴

جدول ۶.۲: نتایج روش تکرار ساده برای مثال ۱۲.۲

نتایج جدول ۶.۲ حاکی از آن است که $\alpha = 2/4475(4D)$. چرا در تکرار هفتم به نقطه ثابت رسیده‌ایم؟ △

تذکر ۱۲.۲ در پاسخ به پرسش‌های پایانی در دو مثال ۱۱.۲ و ۱۲.۲ باید گفت این اتفاق‌ها ناشی از رشد خطای گرد کردن است. بنابراین توصیه می‌شود در عمل تکرارها را تا جایی ادامه دهیم که ثابت شوند.

قضیه ۴.۲ اگر $\{x_n\}_{n\geq 1}$ دنباله‌ای باشد که توسط روش تکرار ساده به دست آمده باشد و به α (نقطه ثابت g) همگرا باشد و

$$g'(\alpha) = g''(\alpha) = \cdots = g^{(p-1)}(\alpha) = 0 \neq g^{(p)}(\alpha),$$

آن‌گاه مرتبه همگرایی دنباله $\{x_n\}_{n\geq 1}$ برابر p است.

مثال ۱۳.۲ دنباله $\{x_n\}_{n\geq 1}$ در رابطه بازگشتی زیر صدق می‌کند

$$x_{n+1} = \frac{x_n(x_n^2 + 3a)}{3x_n^2 + a}, \qquad n \geq 0,$$

که در آن $a > 0$. حد دنباله و مرتبه همگرایی دنباله به حد آن را به دست آورید.
فرض کنید $\lim_{n\to\infty} x_n = \alpha$. با حدگیری داریم

$$\alpha = \frac{\alpha(\alpha^2 + 3a)}{3\alpha^2 + a},$$

و از آنجا $2\alpha(\alpha^2 - a) = 0$ و بنابراین $\alpha = 0, \pm\sqrt{a}$. با معرفی

$$g(x) = \frac{x(x^2 + 3a)}{3x^2 + a},$$

خواهیم داشت

$$g'(x)=\frac{3x^4-6ax^2+3a^2}{(3x^2+a)^2},\qquad g''(x)=\frac{48ax(x^2-a)}{(3x^2+a)^3},\qquad g'''(x)=\frac{48a(-9x^4+18ax^2-a^2)}{(3x^2+a)^4}.$$

چون $g'(0)\neq 0$ مرتبه همگرایی دنباله به صفر، یک است. از طرفی بنابر

$$g'(\pm\sqrt{a})=g''(\pm\sqrt{a})=0\neq 24a=g'''(\pm\sqrt{a}),$$

نتیجه می‌گیریم مرتبه همگرایی دنباله به $\pm\sqrt{a}$، سه است. △

تذکر ۱۳.۲ *اگر بتوان تابع g را به گونه‌ای انتخاب نمود که $g'(\alpha)$ صفر شود، بنابر قضیه ۴.۲، می‌توان انتظار همگرایی سریع‌تری داشت. به همین منظور فرض کنید g چنان انتخاب شده است که $\alpha=g(\alpha)$. با فرض $\lambda\neq 0$ قرار دهید $x+\lambda x=g(x)+\lambda x$ و از آن جا $x=\frac{g(x)+\lambda x}{1+\lambda}$. قرار دهید $g_\lambda(x)=\frac{g(x)+\lambda x}{1+\lambda}$. حال λ را به گونه‌ای انتخاب می‌کنیم که $g'_\lambda(\alpha)=0$. اما $g'_\lambda(x)=\frac{g'(x)+\lambda}{1+\lambda}$ پس باید داشته باشیم $\lambda=-g'(\alpha)$. ولی از α همین قدر اطلاع داریم که $\alpha\in[a,b]$. پس با این فرض که $b-a$ کوچک باشد آن‌گاه $\alpha\simeq\frac{a+b}{2}$ (یک تکرار از روش دوبخشی) و قرار می‌دهیم $\lambda=-g'(\frac{a+b}{2})$.*

مثال ۱۴.۲ *تقریبی از ریشه $x-\tan(x)=0$ به دست آورید که به $\frac{3\pi}{2}$ نزدیک باشد.*
با انتخاب $x=\tan(x)=g(x)$ واضح است که $g'(x)=1+\tan^2(x)>1$، یعنی g شرایط قضیه نقطه ثابت را ندارد. اگر روش تکراری $x_n=g(x_{n-1})$ را با $x_0=4/5$ دنبال کنیم خواهیم داشت

$$x_1=4/64,\qquad x_2=13/79,\qquad x_3=2/76,\qquad x_4=-0/4.$$

حال با فرض $a=4/4$ و $b=4/6$ خواهیم داشت $\alpha\simeq 4/5$ و قرار می‌دهیم $\lambda=-g'(4/5)=-22/5$ بنابراین $g_\lambda(x)=\frac{\tan(x)-22/5x}{-21/5}$. روش تکراری $x_n=g_\lambda(x_{n-1})$ با $x_0=4/5$، نتایج $x_1=4/4936$ و $x_i=4/4934$ برای $i=2,3,\ldots$ را تولید می‌کند. △

تمرین ۲.۲ *در مثال قبل $g(x)=\tan^{-1}(x)$ را بررسی کنید.*

تذکر ۱۴.۲ *فرض کنید تابع g با ویژگی‌های زیر در دسترس باشد.*

- g بر $[a,b]$ پیوسته باشد؛
- $g:[a,b]\to[a,b]$ یک به یک و پوشا باشد و یا $g:[a,b]\subset[c,d]\to[c,d]$؛
- g بر (a,b) مشتق‌پذیر بوده و $K>1$ چنان موجود باشد که برای هر $x\in(a,b)$ داشته باشیم $|g'(x)|\geq K$.

اگر بتوان ضابطه g^{-1} را به دست آورد و قرار دهیم $G(x)=g^{-1}(x)$ آن‌گاه واضح است که از $\alpha=g(\alpha)$ خواهیم داشت $\alpha=G(\alpha)$ و

- G بر $[a,b]$ (یا (c,d)) پیوسته است و این بازه را به توی خود می‌نگارد؛
- G بر (a,b) (یا (c,d)) مشتق‌پذیر بوده و داریم

$$\forall x\ :\ |G'(x)| = \frac{۱}{|g'(G(x))|} \leq \frac{۱}{K} < ۱.$$

یعنی G در شرایط قضیه نقطه ثابت صادق است و دنباله تولیدشده با $x_{n+۱} = G(x_n)$ به ازای هر $x_۰$ در $[a,b]$ به α همگرا می‌شود.

مثال ۱۵.۲ معادله $x^۳ + x - ۹۹۵ = ۰$ را با دقت $۹D$ حل کنید.
تنها ریشه معادله در فاصله $[۹/۹, ۱۰]$ قرار دارد. اگر معادله را به صورت هم‌ارز آن $x = ۹۹۵ - x^۳$ بنویسیم آن‌گاه تابع $g(x) = ۹۹۵ - x^۳$ در شرایط قضیه نقطه ثابت صدق نمی‌کند. داریم $g([۹/۹, ۱۰]) = [-۵, ۲۴/۷۰۱] \subseteq [۹/۹, ۱۰]$ و چون $|g'(x)| = ۳x^۲$ پس روی $[۹/۹, ۱۰]$ داریم $|g'(x)| \geq ۳ \times (۹/۹)^۲ = ۲۹۴/۰۳$. از طرفی واضح است که $g^{-۱}(x) = \sqrt[۳]{۹۹۵ - x}$، $g^{-۱}([۹/۹, ۱۰]) \subseteq [۹/۹, ۱۰]$ و $|(g^{-۱})'(x)| \leq \frac{۱}{۲۹۴/۰۳} = L$. بنابراین شرایط قضیه نقطه ثابت برای $g^{-۱}$ برقرار است. به ازای $x_۰ = ۱۰$ با دقت $۱۰D$ داریم $x_۱ = ۹/۹۴۹۷۴۷۸۹۵۶$. تعداد گام‌های لازم n از رابطه $\frac{L^n}{۱ - L}|x_۱ - x_۰| < ۰/۵ \times ۱۰^{-۹}$ به دست می‌آید. از حل این نابرابری نتیجه می‌شود $n \geq ۴$. در جدول ۷.۲ چند جمله از دنباله تولیدشده به کمک $g^{-۱}$ با مقدار آغازی $x_۰ = ۱۰$ و خطای تقریب داده شده است. بنابراین جواب معادله $f(x) = ۰$ با دقت $۹D$ عدد $x = ۹/۹۴۹۹۱۶۵۲۸$ است. $\triangle$

n	۱	۲	۳	۴
$x_n = g^{-۱}(x_{n-۱})$	۹/۹۴۹۷۴۷۸۹۵۶	۹/۹۴۹۹۱۷۰۹۶۰	۹/۹۴۹۹۱۶۵۲۶۳	۹/۹۴۹۹۱۶۵۲۸۳
$\left\lvert\frac{x_n - x_{n-۱}}{x_n}\right\rvert$	۰/۰۵۰۲۵۲۱۰۴۴	۰/۰۰۰۱۶۸۶۳۲۶	۰/۰۰۰۰۰۰۵۶۷۸	۰/۰۰۰۰۰۰۰۰۰۲

جدول ۷.۲: نتایج روش تکرار ساده برای مثال ۱۵.۲

۴.۳.۲ روش نیوتن

این روش به روش نیوتن-رفسون[۱۴] و روش مماسی نیز معروف است. برای به دست آوردن طرح تکراری این روش، فرض کنید[۱۵] $f \in C^۲[a,b]$ و $x_۰$ تقریبی از α باشد به طوری که $f'(x_۰) \neq ۰$. با انتخاب $h = \alpha - x_۰$، بنا بر بسط تیلور خواهیم داشت

$$f(\alpha) = f(x_۰ + h) = f(x_۰) + hf'(x_۰) + \frac{h^۲}{۲!}f''(\xi),$$

که در آن ξ بین α و $x_۰$ است. اگر h کوچک باشد، $h^۲ = (\alpha - x_۰)^۲$ قابل چشم‌پوشی بوده و داریم

$$۰ = f(\alpha) \simeq f(x_۰) + (\alpha - x_۰)f'(x_۰),$$

[۱۴] Newton-Raphson
[۱۵] $f \in C^۲[a,b]$ یعنی تابع f و مشتق اول و دوم آن بر بازه $[a,b]$ پیوسته هستند.

که نتیجه می‌دهد

$$\alpha \simeq x_0 - \frac{f(x_0)}{f'(x_0)} := x_1,$$

و با تکرار این فرایند خواهیم داشت

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \qquad n = 0, 1, \ldots.$$

یک راه دیگر برای به دست آوردن طرح تکراری روش نیوتن استفاده از تعبیر هندسی روش است که در شکل ۷.۲ آورده شده است.

شکل ۷.۲: تعبیر هندسی روش نیوتن

الگوریتم روش نیوتن

- **ورودی.** x_0, ϵ, f با این فرض که x_0 به α نزدیک باشد
- **خروجی.** مقدار x_n که به ازای آن $|x_n - x_{n-1}| < \epsilon$

۱) قرار دهید $n = 1$

۲) قرار دهید $x_n = x_{n-1} - \dfrac{f(x_{n-1})}{f'(x_{n-1})}$

۳) اگر $|x_n - x_{n-1}| < \epsilon$، x_n را چاپ کرده و متوقف شوید

۴) قرار دهید $n = n + 1$ و به گام ۲ برگردید

تذکر ۱۵.۲ *با انتخاب $g(x) = x - \frac{f(x)}{f'(x)}$ ملاحظه می‌شود که، روش نیوتن حالت خاصی از روش تکرار ساده است.*

مثال ۱۶.۲ *می‌خواهیم نقطه ثابت تابع کسینوس را تعیین کنیم. به کمک رسم، متوجه می‌شویم، نقطه ثابت تابع در بازه $[0, \frac{\pi}{2}]$ قرار دارد. با انتخاب $x_0 = \frac{\pi}{4}$ و استفاده از طرح تکرار ساده $x_{n+1} = \cos x_n$، نتایج جدول ۸.۲ (سمت چپ) به دست می‌آید که دلالت بر مقدار تقریبی ۰/۷۴ دارد. اما با انتخاب $f(x) = \cos x - x$ و $x_0 = \frac{\pi}{4}$ و با دنبال کردن طرح تکراری نیوتن یعنی*

$$x_n = x_{n-1} - \frac{\cos x_{n-1} - x_{n-1}}{-\sin x_{n-1} - 1}, \qquad n = 1, 2, \ldots,$$

نتایج جدول ۸.۲ (سمت راست) تولید می‌شود که دلالت بر آن دارد که نقطه ثابت تابع کسینوس با دقت $9D$ عبارت است از ۰/۷۳۹۰۸۵۱۳۳. $\triangle$

n	x_n
۰	۰/۷۸۵۳۹۸۱۶۳۵
۱	۰/۷۰۷۱۰۶۷۸۱۰
۲	۰/۷۶۰۲۴۴۵۹۷۲
۳	۰/۷۲۴۶۶۷۴۸۰۸
۴	۰/۷۴۸۷۱۹۸۸۵۸
۵	۰/۷۳۲۵۶۰۸۴۴۶
۶	۰/۷۴۳۴۶۴۲۱۱۳
۷	۰/۷۳۶۱۲۸۲۵۶۵

n	x_n
۰	۰/۷۸۵۳۹۸۱۶۳۵
۱	۰/۷۳۹۵۳۶۱۳۳۷
۲	۰/۷۳۹۰۸۵۱۷۸۱
۳	۰/۷۳۹۰۸۵۱۳۳۲
۴	۰/۷۳۹۰۸۵۱۳۳۲

جدول ۸.۲: تقریب‌هایی برای یافتن نقطه ثابت تابع کسینوس

مثال ۱۷.۲ *ثابت می‌شود یکی از مدل‌های رشد جمعیت یک جامعه، تابعی بر حسب زمان t به صورت زیر است*

$$N(t) = N_۰ e^{\lambda t} + \frac{\nu}{\lambda}(e^{\lambda t} - ۱),$$

که در آن $N_۰$ جمعیت آغازی، ν میزان مهاجرت (که ثابت فرض می‌شود) و λ بیانگر ثابت نرخ تولد است. فرض کنید جمعیت آغازی یک جامعه ۱۰۰۰۰۰۰ نفر بوده و در سال اول ۴۳۵۰۰۰ نفر مهاجر داشته باشد. اگر در انتهای سال اول جمعیت جامعه ۱۵۶۴۰۰۰ نفر شده باشد، ثابت نرخ تولد در این جامعه را مشخص کنید. سپس با استفاده از آن و با این فرض که میزان مهاجرت در سال دوم تغییر نکند، جمعیت جامعه را در پایان سال دوم پیش‌بینی کنید.

در واقع باید جواب معادله $۱۵۶۴۰۰۰ = ۱۰۰۰۰۰۰e^{\lambda} + \frac{۴۳۵۰۰۰}{\lambda}(e^{\lambda} - ۱)$ را به دست آوریم. به عبارت دیگر باید صفر تابع $f(\lambda) = ۱۰۰۰e^{\lambda} + \frac{۴۳۵}{\lambda}(e^{\lambda} - ۱) - ۱۵۶۴$ معین شود. برای این منظور از روش نیوتن استفاده می‌کنیم و جوابی به دست می‌آوریم که دارای دقت $۶D$ باشد. بنابراین با منظور نمودن دقت $۷D$ و به ازای $\lambda_۰ = ۱$، به کمک رابطه بازگشتی

$$\lambda_{n+۱} = \lambda_n - \frac{f(\lambda_n)}{f'(\lambda_n)}, \quad n = ۰, ۱, ۲, \ldots,$$

به نتایج جدول ۹.۲ دست می‌یابیم. دلیل توقف در تکرار پنجم آن است که

$$|\lambda_۵ - \lambda_۴| = ۰/۰۰۰۰۰۰۰۲ < ۰/۵ \times ۱۰^{-۶}.$$

در این صورت $\lambda = ۰/۱۰۰۹۹۸$ و جمعیت در انتهای سال دوم عبارت است از

$$N(۲) = ۱۰۰۰۰۰۰e^{۲\times ۰/۱۰۰۹۹۸} + \frac{۴۳۵۰۰۰}{۰/۱۰۰۹۹۸}(e^{۲\times ۰/۱۰۰۹۹۸} - ۱) = ۲۱۸۷۹۳۹.$$

اگر محاسبات میانی با دقت $۳D$ انجام شود، $\lambda = ۰/۰۷$ خواهد بود و جمعیت در انتهای سال دوم $N(۲) = ۲۰۸۴۱۲۰$ به دست می‌آید که با مقدار قبلی بیش از ۱۰۰۰۰۰ نفر (حدود ۵ درصد) تفاوت دارد. توجه داشته باشید اگر بخواهیم بر مبنای جمعیت در انتهای سال دوم تخصیص بودجه داشته باشیم این اختلاف می‌تواند خیلی از موازنه‌ها را بر هم زند.

توجه داشته باشید که به ازای $\lambda = ۰/۰۷$ خواهیم داشت $N(۱) = ۱۵۲۳۰۹۳$، در حالی که به ازای $\lambda = ۰/۱۰۰۹۹۸$ مقدار واقعی $N(۱) = ۱۵۶۴۰۰۰$ به دست می‌آید. △

n	۰	۱	۲	۳	۴	۵
λ_n	۱/۰۰۰۰۰۰۰	۰/۳۹۶۹۰۳۱	۰/۱۳۸۸۶۹۰	۰/۱۰۱۶۶۶۶	۰/۱۰۰۹۹۸۱	۰/۱۰۰۹۹۷۹

جدول ۹.۲: روش نیوتن در یافتن ثابت نرخ تولد

قضیه ۵.۲ *فرض کنید $f \in C^۲[a,b]$ و $\alpha \in [a,b]$ صفر ساده f باشد یعنی $f(\alpha) = ۰ \neq f'(\alpha)$. در این صورت δیی مثبت موجود است که دنباله حاصل از روش نیوتن به ازای هر $x_۰$ در $[\alpha - \delta, \alpha + \delta]$ به α همگرا است.*

تذکر ۱۶.۲ *از قضیه ۵.۲ چنین نتیجه‌گیری می‌شود که برای تضمین همگرایی روش نیوتن باید $x_۰$ به α نزدیک باشد. به همین منظور در عمل ابتدا چند تکرار یک روش ساده (با حجم عملیات کم) مانند روش دوبخشی را دنبال کرده و نتیجه را به عنوان $x_۰$ روش نیوتن اختیار می‌کنیم.*

قضیه ۶.۲ *تحت شرایط قضیه ۵.۲ مرتبه همگرایی روش نیوتن دست کم ۲ است.*

برهان. روش نیوتن را به عنوان یک روش تکرار ساده در نظر گرفته، با فرض $g(x) = x - \frac{f(x)}{f'(x)}$ خواهیم داشت $g'(x) = \frac{f(x)f''(x)}{f'^۲(x)}$ و در نتیجه $g'(\alpha) = \frac{f(\alpha)f''(\alpha)}{f'^۲(\alpha)} = ۰$. هم‌چنین $g''(\alpha) = \frac{f''(\alpha)}{f'(\alpha)}$ و بنابر قضیه ۴.۲ مرتبه همگرایی روش نیوتن در این حالت کمتر از دو نیست. □

تمرین ۳.۲ *نشان دهید مرتبه همگرایی روش نیوتن در یافتن ریشه $\alpha = ۰$ معادله $\sin x = ۰$ دقیقا ۳ است.*

مثال ۱۸.۲ *می‌خواهیم $\sqrt[m]{a}$ را با روش نیوتن تعیین کنیم که در آن a و m اعدادی مثبت هستند. به همین منظور تابع $f(x) = x^m - a$ را تعریف می‌کنیم که $\alpha = \sqrt[m]{a}$ صفر آن است. طرح تکراری نیوتن به صورت زیر در می‌آید*

$$x_{n+۱} = x_n - \frac{x_n^m - a}{mx_n^{m-۱}} = \frac{(m-۱)x_n^m + a}{mx_n^{m-۱}}, \qquad n = ۰, ۱, \ldots.$$

اگر $m = ۲$ و $a = ۲$ اختیار شود، این طرح به صورت $x_{n+۱} = \frac{x_n}{۲} + \frac{۱}{x_n}$ خلاصه می‌شود و نتایج

$$x_۱ = ۱/۵,\ x_۲ = ۱/۴۱۶۶۶۶۶۶۷,\ x_۳ = ۱/۴۱۴۲۱۵۶۸۶,\ x_۴ = ۱/۴۱۴۲۱۳۵۶۲,\ x_۵ = ۱/۴۱۴۲۱۳۵۶۲$$

به ازای $x_۰ = ۱$ تولید می‌شود. △

مثال ۱۹.۲ *با استفاده از روش نیوتن یک طرح تکراری برای محاسبه وارون عدد ناصفر a به دست آورده و به کمک آن تقریبی برای $\frac{۱}{۳}$ تعیین کنید.*

اگر $f(x) = \frac{۱}{x} - a$ آن‌گاه $f'(x) = \frac{-۱}{x^۲}$ و از آنجا داریم

$$x_{n+۱} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{\frac{۱}{x_n} - a}{\frac{-۱}{x_n^۲}},$$

و در نتیجه $x_{n+1} = x_n(2 - ax_n)$. *با فرض* $a = 3$ و $x_0 = 0.3$ *خواهیم داشت* $x_1 = 0.33$، $x_2 = 0.3333$ و $x_3 = 0.33333333$. △

تعریف ۴.۲ *α صفر چندگانه f از مرتبه تکرار $m > 1$ نامیده می‌شود هرگاه*

$$f(\alpha) = f'(\alpha) = \cdots = f^{(m-1)}(\alpha) = 0 \neq f^{(m)}(\alpha),$$

و یا بتوان نوشت $f(x) = (x-\alpha)^m h(x)$ که در آن $h(\alpha) \neq 0$.

قضیه ۷.۲ *اگر α صفر m-گانه f باشد، آن‌گاه مرتبه همگرایی روش نیوتن یک و نرخ همگرایی آن $1 - \frac{1}{m}$ است.*

قضیه ۸.۲ *روش $x_{n+1} = x_n - m\frac{f(x_n)}{f'(x_n)}$ که به روش نیوتن تعمیم‌یافته (ترمیم‌یافته، تغییریافته، اصلاح‌شده) معروف است، برای یافتن صفر m-گانه f، مرتبه همگرایی حداقل دو دارد.*

مثال ۲۰.۲ *$\alpha = 0$ صفر مرتبه چهارم $f(x) = 2\cos x - 2 + x^2$ است زیرا*

$$f'(x) = -2\sin x + 2x, \qquad f''(x) = -2\cos x + 2, \qquad f'''(x) = 2\sin x, \qquad f^{(4)}(x) = 2\cos x,$$

و به وضوح $f(0) = f'(0) = f''(0) = f'''(0) = 0 \neq f^{(4)}(0)$. هم‌چنین به کمک بسط تیلور می‌توان نوشت

$$f(x) = 2\cos x - 2 + x^2 = x^4\left(\frac{2}{4!} - \frac{2x^2}{6!} + \cdots\right).$$

با به کاربردن روش نیوتن به ازای $x_0 = 0.01$ خواهیم داشت $x_{12} = 0.000424341$ که نشان‌دهنده همگرایی کند روش است. اگر روش

$$x_{n+1} = x_n - 4\frac{2\cos x_n - 2 + x_n^2}{-2\sin x_n + 2x_n},$$

را به کار بریم، به ازای $x_0 = 0.01$ خواهیم داشت $x_2 = -0.163206 \times 10^{-7}$، اما برای محاسبه x_3 صورت و مخرج کسر صفر می‌شود (چون ریشه چندگانه است). △

قضیه ۹.۲ *اگر α صفر m-گانه f باشد، آن‌گاه α صفر ساده $f^{(m-1)}$ بوده و روش $x_{n+1} = x_n - \frac{f^{(m-1)}(x_n)}{f^{(m)}(x_n)}$، دنباله‌ای تولید می‌کند که با مرتبه همگرایی دست کم دو به α همگرا است.*

برهان. روش نیوتن برای یافتن ریشه ساده معادله $f^{(m-1)}(x) = 0$، دارای مرتبه همگرایی دست کم دو است. □

مثال ۲۱.۲ *$\alpha = 0$ صفر مرتبه چهارم $f(x) = 2\cos x - 2 + x^2$ و صفر ساده $f'''(x) = 2\sin x$ است. بنابراین طرح تکراری*

$$x_{n+1} = x_n - \frac{f'''(x_n)}{f^{(4)}(x_n)} = x_n - \frac{2\sin x_n}{2\cos x_n} = x_n - \tan x_n,$$

به ازای $x_0 = 0.01$ نتیجه می‌دهد $x_1 = -0.333347 \times 10^{-6}$، $x_2 = 0.123349 \times 10^{-19}$ و $x_3 = 0.0$. △

تذکر ۱۷.۲ *در به کار بردن روش نیوتن دو مشکل اساسی وجود دارد که عبارتند از*

- $x_۰$ *باید نزدیک* α *اختیار شود؛*
- *وجود، محاسبه و مخالف صفر بودن* $f'(x_n)$.

برای درک بهتر به شکل ۸.۲ نگاه کنید.

شکل ۸.۲: تعبیر هندسی مشکلات اساسی روش نیوتن

۵.۳.۲ روش وتری

روش وتری به روش خط قاطع[16] نیز معروف است و برای رفع مشکل دوم روش نیوتن مطرح شده است. با توجه به رابطه

$$f'(x_n) = \lim_{x \to x_n} \frac{f(x_n) - f(x)}{x_n - x},$$

اگر x به اندازه کافی به x_n نزدیک باشد به عنوان مثال $x = x_{n-۱}$، می‌توان نوشت

$$f'(x_n) \simeq \frac{f(x_n) - f(x_{n-۱})}{x_n - x_{n-۱}}.$$

با جای‌گذاری در طرح تکراری نیوتن، برای $n \geq ۱$ داریم

$$x_{n+۱} = x_n - \frac{f(x_n)}{\frac{f(x_n) - f(x_{n-۱})}{x_n - x_{n-۱}}} = x_n - \frac{f(x_n)(x_n - x_{n-۱})}{f(x_n) - f(x_{n-۱})} = \frac{x_{n-۱} f(x_n) - x_n f(x_{n-۱})}{f(x_n) - f(x_{n-۱})}.$$

یک راه دیگر به دست آوردن طرح تکراری روش وتری، استفاده از تعبیر هندسی روش است که در شکل ۹.۲ آورده شده است.

شکل ۹.۲: تعبیر هندسی روش وتری

---

[16] Secant method

الگوریتم روش وتری

- **ورودی.** x_0, x_1, ϵ, f
- **خروجی.** مقدار x_{n+1} که به ازای آن $|x_{n+1} - x_n| < \epsilon$

۱) قرار دهید $n = 1$

۲) قرار دهید $x_{n+1} = x_n - \frac{f(x_n)(x_n - x_{n-1})}{f(x_n) - f(x_{n-1})} = \frac{x_{n-1}f(x_n) - x_n f(x_{n-1})}{f(x_n) - f(x_{n-1})}$

۳) اگر $|x_{n+1} - x_n| < \epsilon$ آن‌گاه x_{n+1} را چاپ کرده و متوقف شوید

۴) قرار دهید $n = n + 1$ و به گام ۲ برگردید

مثال ۲۲.۲ *همان طور که بررسی شد $\alpha = 0$ صفر مرتبه چهار $f(x) = 2\cos x - 2 + x^2$ است. با به کاربردن روش وتری به ازای $x_0 = 0/01$ و $x_1 = 0/002$ خواهیم داشت $x_{10} = 0/000381888$.* $\triangle$

تذکر ۱۸.۲ *ثابت می‌شود مرتبه همگرایی روش وتری $1/618 \simeq \frac{1+\sqrt{5}}{2}$ است.*

تذکر ۱۹.۲ *روش وتری برای شروع به دو مقدار آغازی نیاز دارد که لزومی ندارد دو طرف α واقع باشند؛ ولی روش نابجایی برای شروع به a و bیی نیاز دارد که $f(a)f(b) < 0$. هم‌چنین در روش وتری همیشه نقطه‌ای با کوچک‌ترین اندیس کنار گذاشته می‌شود ولی در روش نابجایی ممکن است یک نقطه در چند تکرار متوالی ثابت بماند.*

تمرین

---

۱. درستی روابط زیر را بررسی کنید

$$\frac{1}{n\ln n} = o\left(\frac{1}{n}\right), \qquad \sin h = h - \frac{h^3}{6} + O(h^5), \qquad e^{-n} = o(1/n^2).$$

۲. ابتدا نشان دهید معادله $x^2 - 4\sin x = 0$ فقط یک ریشه در بازه $[1, 2]$ دارد. سپس برای آن که تقریبی از این ریشه با دقت $3D$ به دست آید به چند بار نصف کردن نیاز است؟ در پایان ریشه را با این دقت تعیین کنید.

۳. با شروع از بازه $[0, 5]$، روش دوبخشی به کدام صفر تابع $f(x) = (x-1)(x-2)(x-3)^2(x-4)$ همگرا می‌شود؟

۴. برای محاسبه ریشه مثبت معادله $x^2 + x - 1 = 0$، می‌توان توابع زیر را در نظر گرفت

$$x = g_1(x) := 1 - x^2, \quad x = g_2(x) := \sqrt{1-x}, \quad x = g_3(x) := \frac{1}{1+x}, \quad x = g_4(x) := \frac{x^2+1}{2x+1}.$$

به کمک قضیه نقطه ثابت نشان دهید g_1 مناسب نیست و به ترتیب g_4، g_3 و g_2 مناسب‌تر هستند.

۵. برای مثال ۱۲.۲، آیا تابع $g(x) = e^{۲x-۴}$ در شرایط قضیه نقطه ثابت صدق می‌کند؟

۶. یک طرح تکراری برای یافتن صفر مرتبه دو تابع $f(x) = ۱ - e^{x^۲}$ با مرتبه همگرایی دو به دست آورید.

۷. برای یافتن ریشه $x = ۰$ معادله $x + x^۳ - \sin(x) = ۰$ کدام دنباله مرتبه همگرایی بزرگتری دارد؟

الف) $x_{n+۱} = x_n - \frac{f(x_n)}{f'(x_n)}$ ب) $x_{n+۱} = x_n - ۲\frac{f(x_n)}{f'(x_n)}$ ج) $x_{n+۱} = x_n - ۳\frac{f(x_n)}{f'(x_n)}$ د) $x_{n+۱} = \sin(x_n) - x_n^۳$

۸. یک طرح تکراری برای تعیین $\sqrt[۳]{۲}$ به کمک روش نیوتن بسازید.

۹. کدام گزینه نادرست است؟

الف) روش نیوتن حالت خاصی از روش تکرار ساده است که همگرایی تضمین شده‌ای دارد.

ب) در یافتن ریشه‌های مضاعف یک تابع مرتبه همگرایی روش نیوتن دو نیست.

ج) اگر $g'(\alpha) = ۰$ آنگاه مرتبه همگرایی روش تکرار ساده حداقل دو است.

د) روش وتری لزوما همگرا نیست.

۱۰. کدام بازه $[a,b]$ این خاصیت را دارد که "دنباله حاصل از روش تکرار ساده $x_{n+۱} = ۰/۳(x_n - ۰/۱)^۳ - ۰/۲$ برای هر $x_۰ \in [a,b]$ به $\alpha \in [a,b]$ همگرا است"؟

الف) $[۱, ۱/۵]$ ب) $[-۲, -۱]$ ج) $[-۰/۵, ۰]$ د) $[۲, ۲/۵]$

۱۱. اگر $x_{n+۱} = \frac{۴}{۵}\cos(x_n)$ با $x_۰ = ۱$ به $\alpha \in [۰, \frac{\pi}{۲}]$ همگرا باشد، حداکثر مقدار $|x_{۱۰} - \alpha|$ کدام است؟

الف) ۰/۱۹۷۴ ب) ۰/۱۰۷۴ ج) ۰/۲۱۴۸ د) ۰/۳۰۴۸

۱۲. اگر بخواهیم جواب معادله $\sin(x) - \cos(x) = ۰/۱$ را با $x_۰ = ۰/۸$ و $x_۱ = ۱$، به روش وتری حل کنیم $x_۳$ کدام گزینه است؟ الف) ۰/۸۵۶۱۶۶۱۱۲ ب) ۰/۸۵۶۱۴۳۹۹۹ ج) ۰/۸۵۶۶۲۹۴۵۶ د) ۰/۸۵۶۲۳۶۹۹۰

۱۳. کدام یک از عبارات زیر صحیح است؟

الف) روش نیوتن برای حل معادلات غیرخطی همیشه همگرا است.

ب) روش دوبخشی برای یافتن تمام ریشه‌ها همگرا است.

ج) مرتبه همگرایی روش نیوتن به مرتبه ریشه بستگی دارد.

د) مرتبه همگرایی روش وتری بیشتر از مرتبه همگرایی روش نیوتن است.

۱۴. اگر از روش دوبخشی برای محاسبه ریشه مثبت معادله $۳xe^x = ۱$ در بازه $[۰, ۱]$ با دقت $۰/۵ \times ۱۰^{-۴}$ استفاده شود، حداقل چند تکرار از روش دوبخشی لازم است؟

الف) ۱۴ ب) ۱۵ ج) ۱۹ د) ۱۷

۱۵. مرتبه همگرایی روش نیوتن برای پیدا کردن ریشه‌های $f(x) = x^۴ - ۲x^۲ = ۰$ در صورتی که به ریشه همگرا شود، چند است؟

الف) برای همه ریشه‌ها خطی است.

ب) برای همه ریشه‌ها از مرتبه دو است.

ج) برای ریشه‌ی صفر از مرتبه دو و برای ریشه‌های ناصفر خطی است.

د) برای ریشه‌ی صفر خطی است و برای ریشه‌های ناصفر از مرتبه دو است.

۱۶. طرح تکراری $x_{n+1} = \frac{2}{3}\left(x_n + \frac{1}{x_n^2}\right)$ با

الف) مرتبه همگرایی دو به $\sqrt[3]{2}$ همگرا است.
ب) مرتبه همگرایی حداقل دو به $\sqrt[3]{2}$ همگرا است.
ج) مرتبه همگرایی دو به $\sqrt{2}$ همگرا است.
د) مرتبه همگرایی حداقل دو به $\sqrt{2}$ همگرا است.

۱۷. فرض کنید $f(x) = x + \cos x$ و $x_0 = 0$ و $x_1 = 0/5$ باشد. مقدار x_2 در روش وتری کدام است؟

الف) $-1/4232$
ب) $1/3242$
ج) $1/4232$
د) $-1/3242$

۱۸. کدام یک از گزینه‌های زیر در مورد ریشه $\alpha = 0$ برای $f(x) = \sin x - \sinh x$ صحیح است؟

الف) روش نیوتن همگرای مرتبه دو است.
ب) روش دوبخشی همگرا نیست.
ج) روش نیوتن همگرای مرتبه یک است.
د) روش نیوتن همگرای حداقل مرتبه دو است.

۱۹. کدام گزینه درست است؟

الف) معادله $x = 1 + \tan^{-1} x$ نقطه ثابت یکتا در بازه $[2, 3]$ دارد.

ب) تابع $g(x) = 1 + \tan^{-1} x$ ریشه یکتا در بازه $[2, 3]$ دارد.

ج) تابع $g(x) = 1 + \tan^{-1} x$ نقطه ثابت یکتا در بازه $[2, 3]$ دارد.

د) معادله $x = 1 + \tan x$ ریشه یکتا در بازه $[2, 3]$ دارد.

۲۰. با این فرض که $f(x) = \cos(e^x)$، $x_0 = 0$، x_1 با روش نیوتن و x_2 با روش وتری به دست آید، x_1 و x_2 به ترتیب کدامند؟

الف) $0/6420$ و $0/4015$
ب) $0/6421$ و $0/4015$
ج) $0/6420$ و $0/4016$
د) $0/6421$ و $0/4016$

۲۱. چند تکرار روش دوبخشی در بازه $[-0/5, 1/5]$ لازم است تا ریشه $f(x) = 1 - e^x$ با حداکثر خطای 10^{-3} محاسبه شود؟

الف) ۹
ب) ۱۰
ج) ۱۱
د) ۸

۲۲. کدامیک از عبارات زیر صحیح است؟

الف) روش دوبخشی برای یافتن تمامی ریشه‌ها همگرا است.
ب) مرتبه همگرایی روش نیوتن به مرتبه ریشه بستگی دارد.
ج) مرتبه همگرایی روش وتری بیشتر از مرتبه همگرایی روش نیوتن است.
د) روش نیوتن برای حل معادلات غیرخطی همیشه همگرا است.

فصل ۳

درون‌یابی

مقدمه

در جمهوری اسلامی ایران هر ده سال یک بار سرشماری جمعیت انجام می‌شد و قرار شده است از این به بعد هر پنج سال یک بار انجام پذیرد. جدول ۱.۳ جمعیت کشور را در سال‌های ۱۳۳۵ تا ۱۳۹۵ نشان می‌دهد. در ارتباط با این جدول سوال‌های زیر مطرح می‌شوند

- در سال ۱۳۵۹ (آغاز جنگ تحمیلی ایران با عراق) جمعیت ایران چقدر بوده است؟
- در سال ۱۴۰۰ جمعیت ایران چقدر خواهد بود؟
- در سال ۱۳۳۰ جمعیت ایران چقدر بوده است؟
- در چه سالی جمعیت ایران حدود ۴۰ میلیون نفر بوده است؟

سال	۱۳۳۵	۱۳۴۵	۱۳۵۵	۱۳۶۵	۱۳۷۵	۱۳۸۵	۱۳۹۰	۱۳۹۵
میلیون نفر	۱۸/۹۵	۲۵/۷۹	۳۳/۷۱	۴۹/۴۵	۶۰/۰۶	۷۰/۴۷	۷۵/۱۵	۷۹/۹۳

جدول ۱.۳: جمعیت ایران در سال‌های ۱۳۳۵ تا ۱۳۹۵

تعریف ۱.۳ *فرض کنید تابع $y = f(x)$ در دسترس نباشد (یا ضابطه‌ای پیچیده داشته باشد) ولی مقدار آن در $n + 1$ نقطه $x_0 < x_1 < \cdots < x_n$ معلوم باشد. مسئله یافتن مقدار تابع f در نقطه x متعلق به بازه $[x_0, x_n]$ به مسئله درون‌یابی (سوال اول)، مسئله یافتن مقدار تابع f در نقطه x که به بازه $[x_0, x_n]$ تعلق ندارد به مسئله برون‌یابی (سوال دوم و سوم) و مسئله تعیین x زمانی که $f(x)$ معلوم باشد به مسئله درون‌یابی وارون (سوال چهارم) معروف است.*

۱.۳ درون‌یابی

قضیه ۱.۳ (تقریب وایرشتراس) *فرض کنید $f \in C[a,b]$. به ازای هر $\epsilon > 0$ چندجمله‌ای p چنان وجود دارد که*

$$|f(x) - p(x)| < \epsilon, \quad \forall x \in [a,b].$$

با توجه به این قضیه و با توجه به همواری چندجمله‌ای‌ها (مشتق و انتگرال یک چندجمله‌ای، چندجمله‌ای است)، چنین به نظر می‌رسد که ساده‌ترین روش برای حل مسئله درون‌یابی، ساختن چندجمله‌ای درون‌یاب تابع f است.

تعریف ۲.۳ *فرض کنید مقدار تابع f در $n+1$ نقطه $x_0 < x_1 < \cdots < x_n$ معلوم باشد. چندجمله‌ای p را چندجمله‌ای درون‌یاب تابع f نامند اگر*

$$i = 0, 1, \ldots, n \quad : \quad p(x_i) = f(x_i).$$

این رابطه به ویژگی چندجمله‌ای درون‌یاب معروف است.

قضیه ۲.۳ *اگر مقدار تابع f در $n+1$ نقطه متمایز $x_0, x_1, \ldots, x_n$ معلوم باشد، آن‌گاه یک و فقط یک چندجمله‌ای با حداکثر درجه n با ویژگی درون‌یابی وجود دارد.*

برهان. فرض کنید $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ چندجمله‌ای درون‌یاب (با ضرایب نامعین) تابع f در نقاط $x_0, x_1, \ldots, x_n$ باشد. بنابر ویژگی درون‌یابی داریم

$$\begin{aligned}
p(x_0) &= a_0 + a_1 x_0 + \cdots + a_n x_0^n = f_0,\\
p(x_1) &= a_0 + a_1 x_1 + \cdots + a_n x_1^n = f_1,\\
&\vdots\\
p(x_n) &= a_0 + a_1 x_n + \cdots + a_n x_n^n = f_n,
\end{aligned}$$

که در آن $f_i = f(x_i)$. این معادله‌ها را می‌توان به شکل فشرده $A\alpha = F$ نوشت که در آن

$$A = \begin{bmatrix} 1 & x_0 & \cdots & x_0^n \\ 1 & x_1 & \cdots & x_1^n \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \cdots & x_n^n \end{bmatrix}, \quad F = \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \end{bmatrix}, \quad \alpha = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

ماتریس A به ماتریس واندرموند معروف است و ثابت می‌شود $\det(A) = \prod_{0 \le i < j \le n}(x_j - x_i)$. چون نقاط x_i و x_j متمایز هستند واضح است که $\det(A) \neq 0$ و بنابراین دستگاه $A\alpha = F$ جواب یکتا دارد. □

تذکر ۱.۳ *اثبات این قضیه دلالت دارد بر روش ضرایب نامعین برای تعیین چندجمله‌ای درون‌یاب که به تولید یک دستگاه پُر و بدحالت منجر می‌شود (در عمل جواب‌های عددی رضایت‌بخشی به دست نمی‌آید) و برای nهای بزرگ کارایی ندارد. در ادامه روش‌های کاراتر بررسی می‌شوند.*

۱.۱.۳ روش لاگرانژ

فرض کنید L_j به ازای $j = ۰, ۱, \dots, n$ یک چندجمله‌ای درجه n باشد و قرار دهید

$$p(x) = f_۰ L_۰(x) + f_۱ L_۱(x) + \cdots + f_n L_n(x) = \sum_{j=۰}^{n} f_j L_j(x).$$

برای آن که p در ویژگی درون‌یابی صدق کند باید داشته باشیم

$$L_j(x_i) = \delta_{ij} = \begin{cases} ۱, & i = j, \\ ۰, & i \neq j. \end{cases}$$

یعنی L_j باید در n نقطه $x_۰, \dots, x_{j-۱}, x_{j+۱}, \dots, x_n$ صفر شود. بنابراین

$$L_j(x) = c(x - x_۰) \cdots (x - x_{j-۱})(x - x_{j+۱}) \cdots (x - x_n),$$

و چون $L_j(x_j) = ۱$ پس

$$c = \frac{۱}{(x_j - x_۰) \cdots (x_j - x_{j-۱})(x_j - x_{j+۱}) \cdots (x_j - x_n)},$$

و در نتیجه

$$L_j(x) = \frac{(x - x_۰) \cdots (x - x_{j-۱})(x - x_{j+۱}) \cdots (x - x_n)}{(x_j - x_۰) \cdots (x_j - x_{j-۱})(x_j - x_{j+۱}) \cdots (x_j - x_n)} = \prod_{j \neq k = ۰}^{n} \frac{x - x_k}{x_j - x_k}.$$

تذکر ۲.۳ *چندجمله‌ای‌های L_j از درجه n هستند و به چندجمله‌ای‌های لاگرانژ معروف هستند و به کمک چندجمله‌ای*

$$\psi_n(x) = (x - x_۰) \cdots (x - x_n) = \prod_{i=۱}^{n} (x - x_i),$$

می‌توان نوشت

$$p(x) = \psi_n(x) \sum_{j=۰}^{n} \frac{f_j}{(x - x_j)\psi'_n(x_j)}.$$

تمرین ۱.۳ *نشان دهید چندجمله‌ای‌های لاگرانژ مستقل خطی بوده و $\sum_{j=۰}^{n} L_j(x) = ۱$.*

الگوریتم روش لاگرانژ

- **ورودی.** نقاط $(x_۰, f_۰), (x_۱, f_۱), \dots, (x_n, f_n)$
- **خروجی.** چندجمله‌ای درون‌یاب f

۱) برای $j = ۰, ۱, \dots, n$ قرار دهید $L_j(x) = \prod_{j \neq k = ۰}^{n} \frac{x - x_k}{x_j - x_k}$

۲) قرار دهید $p(x) = \sum_{j=۰}^{n} f_j L_j(x)$

مثال ۱.۳ *چندجمله‌ای درون‌یاب مربوط به داده‌های جدول زیر را با روش لاگرانژ تعیین کنید.*

x_i	-۱	۰	۲
f_i	۱	۱	۷

به وضوح $n = ۲$ *و داریم*

$$L_۰(x) = \frac{(x-x_۱)(x-x_۲)}{(x_۰-x_۱)(x_۰-x_۲)} = \frac{(x-۰)(x-۲)}{(-۱-۰)(-۱-۲)} = \frac{x^۲-۲x}{۳},$$
$$L_۱(x) = \frac{(x-(-۱))(x-۲)}{(۰-(-۱))(۰-۲)} = \frac{-x^۲+x+۲}{۲}, \qquad L_۲(x) = \frac{(x-(-۱))(x-۰)}{(۲-(-۱))(۲-۰)} = \frac{x^۲+x}{۶}$$

و بنابراین

$$p(x) = f_۰L_۰(x) + f_۱L_۱(x) + f_۲L_۲(x) = ۱ \times (\frac{x^۲-x}{۲}) + ۱ \times \frac{-x^۲+x+۲}{۲} + ۷ \times (\frac{x^۲+x}{۶}) = x^۲+x+۱$$

پس می‌توان به عنوان مثال $p(\frac{۱}{۲}) = (\frac{۱}{۲})^۲ + \frac{۱}{۲} + ۱ = \frac{۷}{۴}$ *را به عنوان تقریبی از* $f(\frac{۱}{۲})$ *پذیرفت.* △

مثال ۲.۳ *چندجمله‌ای درون‌یاب مربوط به داده‌های جدول زیر را با روش لاگرانژ تعیین کنید.*

x_i	-۱	۰	۱	۲
f_i	۱	۱	۳	۷

به وضوح $n = ۳$ *و داریم*

$$L_۰(x) = \frac{(x-۰)(x-۱)(x-۲)}{(-۱-۰)(-۱-۱)(-۱-۲)} = \frac{x^۳-۳x^۲+۲x}{-۶},$$
$$L_۱(x) = \frac{x^۳-۲x^۲-x+۲}{۲}, \quad L_۲(x) = \frac{x^۳-x^۲-۲x}{-۲}, \qquad L_۳(x) = \frac{x^۳-x}{۶}$$

و بنابراین

$$p(x) = ۱ \times L_۰(x) + ۱ \times L_۱(x) + ۳ \times L_۲(x) + ۷ \times L_۳(x) = x^۲+x+۱$$

با آن که L_j*ها درجه ۳ هستند ولی* p *درجه ۲ است.* △

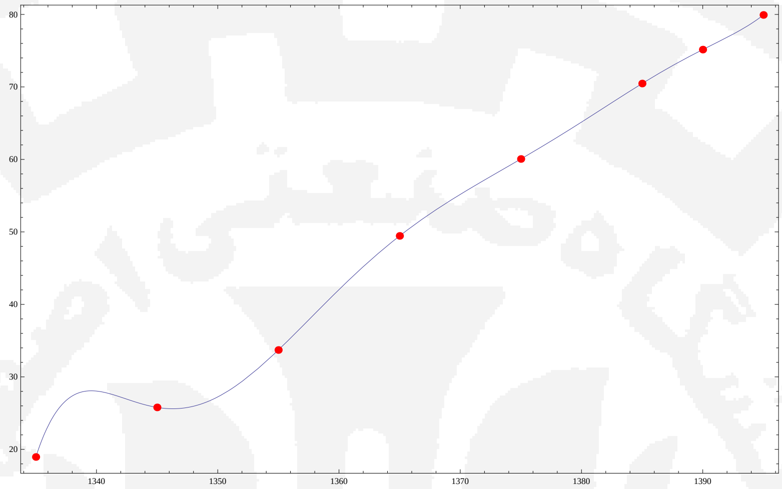
اشکالات روش لاگرانژ

- محاسبات حتی زمانی که n کوچک باشد زیاد است و برای nهای بزرگ روش کارایی چندانی ندارد؛
- با اضافه شدن یک نقطه به داده‌های قبلی، باید تمام عملیات را از سر گرفت (از محاسبات قبلی استفاده نمی‌شود)؛
- قبل از اتمام عملیات، درجه چندجمله‌ای درون‌یاب معلوم نیست.

مثال ۳.۳ *چندجمله‌ای درون‌یاب داده‌های جمعیت ایران را ساخته و نمودار آن را در شکل ۱.۳ رسم کرده‌ایم. نتایج*

$$p(۱۳۳۰)=-۴۴٫۹۵,\ p(۱۳۴۰)=۲۸٫۰۴,\ p(۱۳۵۹)=۴۰٫۳۹,\ p(۱۳۶۸)=۵۱٫۹۲,\ p(۱۴۰۰)=۹۳٫۲۹$$

به دست می‌آیند که باید به دقت تفسیر شوند. △



شکل ۱.۳: نمودار جمعیت ایران در سال‌های ۱۳۹۵–۱۳۳۵

۲.۱.۳ روش تفاضلات تقسیم‌شده نیوتن

در این بخش روشی ساخته می‌شود که اشکالات روش لاگرانژ را برطرف کند. به همین منظور، فرض کنید فضای تمام چندجمله‌ای‌های از درجه حداکثر n را با π_n نمایش دهیم.

تمرین ۲.۳ *نشان دهید*

$$\pi_n=\text{span}\{۱,(x-x_۰),(x-x_۰)(x-x_۱),\dots,(x-x_۰)\cdots(x-x_{n-۱})\}.$$

یعنی نشان دهید چندجمله‌ای‌های

$$۱,(x-x_۰),(x-x_۰)(x-x_۱),\dots,(x-x_۰)\cdots(x-x_{n-۱})$$

مستقل خطی هستند.

بنابراین، می‌توان هر چندجمله‌ای از درجه حداکثر n را به صورت ترکیب خطی از چندجمله‌ای‌های اخیر نوشت. به ویژه برای چندجمله‌ای درون‌یاب p خواهیم داشت

$$p(x)=a_۰+a_۱(x-x_۰)+a_۲(x-x_۰)(x-x_۱)+\cdots+a_n(x-x_۰)\cdots(x-x_{n-۱}).$$

ضرایب $a_0, a_1, \ldots, a_n$ را به گونه‌ای پیدا می‌کنیم که p در ویژگی درون‌یابی صدق کند. پس باید داشته باشیم

$$p(x_0) = a_0 = f_0 \longrightarrow a_0 = f_0,$$
$$p(x_1) = a_0 + a_1(x_1 - x_0) = f_1 \longrightarrow a_1 = \tfrac{f_1 - f_0}{x_1 - x_0}.$$

پیش از آن که سایر ضرایب را به همین صورت به دست آوریم، تفاضلات تقسیم‌شده نیوتن را تعریف می‌کنیم.

تعریف ۳.۳ *فرض کنید $x_0, x_1, \ldots, x_n$ نقاطی متمایز باشند. تفاضل تقسیم‌شده مرتبه اول f در نقاط x_i و x_{i+1} که با نماد $f[x_i, x_{i+1}]$ نمایش داده می‌شود، به صورت زیر است*

$$f[x_i, x_{i+1}] = \frac{f_{i+1} - f_i}{x_{i+1} - x_i},$$

و تفاضل تقسیم‌شده مرتبه j تابع f در نقاط $x_i, x_{i+1}, \ldots, x_{i+j}$ به صورت زیر تعریف می‌شود

$$f[x_i, \ldots, x_{i+j}] = \frac{f[x_{i+1}, \ldots, x_{i+j}] - f[x_i, \ldots, x_{i+j-1}]}{x_{i+j} - x_i}.$$

بنابراین با توجه به نماد تعریف‌شده می‌توان نوشت $a_1 = f[x_0, x_1]$ و داریم

$$p(x_2) = a_0 + a_1(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1) = f_2,$$

و از آن‌جا خواهیم داشت

$$a_2 = \frac{f_2 - f_0 - \frac{f_1 - f_0}{x_1 - x_0}(x_2 - x_0)}{(x_2 - x_0)(x_2 - x_1)} = \frac{\frac{f_2 - f_0}{x_2 - x_1} - \frac{f_1 - f_0}{x_1 - x_0}\frac{x_2 - x_0}{x_2 - x_1}}{(x_2 - x_0)} = \frac{\frac{f_2 - f_1 + f_1 - f_0}{x_2 - x_1} - \frac{f_1 - f_0}{x_1 - x_0}\frac{x_2 - x_1 + x_1 - x_0}{x_2 - x_1}}{x_2 - x_0},$$

و یا

$$a_2 = \frac{\frac{f_2 - f_1}{x_2 - x_1} - \frac{f_1 - f_0}{x_1 - x_0}}{x_2 - x_0} = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}.$$

در نتیجه

$$a_2 = f[x_0, x_1, x_2].$$

با یک روند استقرایی می‌توان نشان داد

$$a_j = f[x_0, x_1, \ldots, x_j], \qquad j = 1, \ldots, n,$$

و بنابراین

$$p(x) = f_0 + \sum_{j=1}^{n} f[x_0, x_1, \ldots, x_j](x - x_0)\cdots(x - x_{j-1}).$$

الگوریتم روش تفاضلات تقسیم‌شده نیوتن

- **ورودی.** نقاط $(x_۰, f_۰), (x_۱, f_۱), \dots, (x_n, f_n)$
- **خروجی.** اعداد $F_{۰,۰}, F_{۱,۱}, \dots, F_{n,n}$ به طوری که $p(x) = \sum_{i=۰}^{n} F_{i,i} \prod_{k=۰}^{i-۱}(x - x_k)$

۱) برای $i = ۰, ۱, \dots, n$ قرار بده $F_{i,۰} = f_i$

۲) برای $i = ۱, \dots, n$ و $j = ۱, \dots, i$ قرار بده $F_{i,j} = \frac{F_{i,j-۱} - F_{i-۱,j-۱}}{x_i - x_{i-j}}$

مثال ۴.۳ *چندجمله‌ای درون‌یاب مربوط به جدول زیر را با روش تفاضلات تقسیم‌شده نیوتن تعیین کنید.*

x_i	-۱	۰	۲
f_i	۱	۱	۷

برای ساختن چندجمله‌ای درون‌یاب، جدولی معروف به جدول تفاضلات تقسیم‌شده به صورت زیر می‌سازیم

x_i	f_i	$f[x_i, x_{i+۱}]$	$f[x_۰, x_۱, x_۲]$
-۱	۱		
		$\frac{۱-۱}{۰-(-۱)} = ۰$	
۰	۱		$\frac{۳-۰}{۲-(-۱)} = ۱$
		$\frac{۷-۱}{۲-۰} = ۳$	
۲	۷		

و به کمک آن داریم $p(x) = ۱ + (۰)(x-(-۱)) + (۱)(x-(-۱))(x-(۰)) = x^۲ + x + ۱$. △

مثال ۵.۳ *اگر داده $(۱, ۳)$ به جدول مثال قبل اضافه شود، آن را دوباره حل کنید.*
به راحتی می‌توان جدول مثال قبل را به صورت زیر اصلاح کرد.

x_i	f_i	$f[x_i, x_{i+۱}]$	$f[x_i, x_{i+۱}, x_{i+۲}]$	$f[x_۰, x_۱, x_۲, x_۳]$
-۱	۱			
		۰		
			۱	
۰	۱			
		۳		$\frac{۱-۱}{۱-(-۱)} = ۰$
۲	۷			
			$\frac{۴-۳}{۱-۰} = ۱$	
		$\frac{۳-۷}{۱-۲} = ۴$		
۱	۳			

و سپس $q(x) = p(x) + 0(x+1)x(x-2) = p(x) = x^2 + x + 1$. △

مزایای روش تفاضلات تقسیم‌شده نیوتن

- حجم عملیات چندان زیاد نیست؛
- با اضافه شدن یک نقطه (نقاطی) به جدول، از محاسبات قبلی استفاده می‌شود؛
- چندجمله‌ای درون‌یاب به تدریج ساخته می‌شود و درجه آن، پس از ساختن جدول مشخص می‌شود.

تذکر ۳.۳ *چندجمله‌ای درون‌یاب به ترتیب نقاط بستگی ندارد، به بیان دیگر اگر p چندجمله‌ای درون‌یاب تابع f در مجموعه نقاط $\{x_0, x_1, \dots, x_n\}$ و q چندجمله‌ای درون‌یاب تابع f در مجموعه نقاط $\{y_0, y_1, \dots, y_n\}$ باشد و داشته باشیم $\{x_0, x_1, \dots, x_n\} = \{y_0, y_1, \dots, y_n\}$ آن‌گاه بنابر یکتایی چندجمله‌ای درون‌یاب داریم $p = q$. به علاوه با توجه به ضریب x^n در p و q داریم $f[x_0, x_1, \dots, x_n] = f[y_0, y_1, \dots, y_n]$ یعنی تفاضل تقسیم‌شده به ترتیب نقاط بستگی ندارد.*

تذکر ۴.۳ *اگر p_n چندجمله‌ای درون‌یاب در نقاط $x_0, \dots, x_n$ و p_{n+1} چندجمله‌ای درون‌یاب در نقاط x_0، ...، x_{n+1} باشد، آنگاه داریم*

$$p_{n+1}(x) = p_n(x) + f[x_0, \dots, x_{n+1}](x - x_0)\cdots(x - x_n).$$

۳.۱.۳ روش‌های پیشرو/پسرو نیوتن

روش‌های لاگرانژ و تفاضلات تقسیم‌شده نیوتن برای نقاط $x_0, \dots, x_n$ چه هم‌فاصله باشند و چه نباشند به کار برده می‌شود، اما اگر نقاط هم‌فاصله باشند چندجمله‌ای درون‌یاب به شکل ساده‌تری قابل بیان است که در ادامه با نحوه نمایش آن آشنا می‌شویم. فرض کنید

$$x_{i+1} - x_i = h, \qquad i = 0, 1, \dots, n-1,$$

و یا

$$x_i = x_0 + ih, \qquad i = 0, 1, \dots, n.$$

تعریف ۴.۳ *عملگر انتقال که با E نمایش داده می‌شود به صورت $Ef_i = f_{i+1} = f(x_{i+1})$ تعریف می‌شود و برای هر k طبیعی داریم $E^k f_i = f_{i+k}$. با این فرض که به ازای هر α حقیقی داشته باشیم $x_{i+\alpha} = x_i + \alpha h$ و $f_{i+\alpha} = f(x_{i+\alpha})$ می‌توان تعریف کرد $E^\alpha f_i = f_{i+\alpha}$، به ویژه $E^{-1} f_i = f_{i-1}$.*

تعریف ۵.۳ *عملگر تفاضل پیشرو که با Δ نمایش داده می‌شود به صورت $\Delta = E - 1$ بیان می‌شود. بنابراین $\Delta f_i = (E-1)f_i = f_{i+1} - f_i$ و برای هر k طبیعی داریم*

$$\Delta^{k+1} f_i = \Delta(\Delta^k f_i) = \Delta^k(\Delta f_i) = \Delta^k(f_{i+1} - f_i) = \Delta^k f_{i+1} - \Delta^k f_i.$$

به عنوان مثال $\Delta^{2} f_i = \Delta(\Delta f_i) = \Delta(f_{i+1} - f_i) = f_{i+2} - 2f_{i+1} + f_i$. به طور مشابه عملگر تفاضل پسرو که با ∇ نمایش داده می‌شود، به صورت $\nabla = 1 - E^{-1}$ بیان شده و در نتیجه $\nabla f_i = (1 - E^{-1}) f_i = f_i - f_{i-1}$ و برای هر k طبیعی داریم

$$\nabla^{k+1} f_i = \nabla(\nabla^k f_i) = \nabla^k(\nabla f_i) = \nabla^k f_i - \nabla^k f_{i-1}.$$

به عنوان مثال $\nabla^{2} f_i = \nabla(\nabla f_i) = \nabla(f_i - f_{i-1}) = f_i - 2f_{i-1} + f_{i-2}$.

بررسی روابط زیر به سادگی امکان‌پذیر است.

$$1) E\Delta = \Delta E, \qquad 2) E\nabla = \nabla E, \qquad 3) \Delta\nabla = \nabla\Delta,$$

$$4) \Delta f_i = \nabla f_{i+1}, \quad 5) \nabla f_i = \Delta f_{i-1}, \quad 6) \Delta^k f_i = \nabla^k f_{i+k} \quad (k \in \mathbb{N}).$$

تمرین ۳.۳ *با فرض $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ با استقرا نشان دهید*

$$\begin{cases} \Delta^n p(x_i) = n! h^n a_n, \\ \Delta^m p(x_i) = 0 \qquad\qquad m > n. \end{cases}$$

لم ۱.۳ *اگر k عددی طبیعی باشد، آن‌گاه به ازای هر عدد صحیح نامنفی i داریم*

$$f[x_i, x_{i+1}, \ldots, x_{i+k}] = \frac{\Delta^k f_i}{k! h^k} = \frac{\nabla^k f_{i+k}}{k! h^k}.$$

قضیه ۳.۳ **(چندجمله‌ای درون‌یاب پیشروی نیوتن)** *چندجمله‌ای درون‌یاب f در نقاط هم‌فاصله $x_0, x_1, \ldots, x_n$ به صورت زیر است*

$$p(x) = f_0 + \theta \Delta f_0 + \frac{\theta(\theta - 1)}{2!} \Delta^2 f_0 + \cdots + \frac{\theta(\theta - 1)\cdots(\theta - n + 1)}{n!} \Delta^n f_0 = f_0 + \sum_{l=1}^{n} \binom{\theta}{l} \Delta^l f_0,$$

که در آن $\theta = \frac{x - x_0}{h}$ و برای هر l در $\mathbb{N}$ و هر θ در $\mathbb{R}$ داریم

$$\binom{\theta}{l} = \frac{\theta(\theta - 1)\cdots(\theta - l + 1)}{l!}.$$

نتیجه ۱.۳.۳ *چندجمله‌ای درون‌یاب در نقاط هم‌فاصله $x_i, x_{i+1}, \ldots, x_{i+k}$ به صورت زیر بیان می‌شود*

$$p(x) = f_i + \theta \Delta f_i + \frac{\theta(\theta - 1)}{2!} \Delta^2 f_i + \cdots + \frac{\theta(\theta - 1)\cdots(\theta - k + 1)}{k!} \Delta^k f_i, \qquad \theta = \frac{x - x_i}{h}.$$

قضیه ۴.۳ **(چندجمله‌ای درون‌یاب پسرو نیوتن)** *چندجمله‌ای درون‌یاب f در نقاط هم‌فاصله $x_0, x_1, \ldots, x_n$ به صورت زیر است (با این فرض که $\theta = \frac{x-x_i}{h}$)*

$$p(x) = f_n + \theta \nabla f_n + \frac{\theta(\theta+1)}{2!}\nabla^2 f_n + \cdots + \frac{\theta(\theta+1)\cdots(\theta+n-1)}{n!}\nabla^n f_n = f_n + \sum_{l=1}^{n} \binom{\theta+l-1}{l} \nabla^l f_n.$$

نتیجه ۱.۴.۳ *چندجمله‌ای درون‌یاب در نقاط هم‌فاصله $x_{i-k}, x_{i-k+1}, \ldots, x_i$ به صورت زیر قابل بیان است*

$$p(x) = f_i + \theta \nabla f_i + \frac{\theta(\theta+1)}{2!}\nabla^2 f_i + \cdots + \frac{\theta(\theta+1)\cdots(\theta+k-1)}{k!}\nabla^k f_i, \qquad \theta = \frac{x-x_i}{h}.$$

تذکر ۵.۳ *در عمل هنگام استفاده از نتایج ۱.۳.۳ و ۱.۴.۳، این سوال مطرح می‌شود که کدام انتخاب (پیشرو یا پسرو، انتخاب x_i و درجه چندجمله‌ای (k)) مناسب‌تر است؟ در پاسخ باید توجه داشت که x_i را چنان انتخاب می‌کنیم که $\theta = \frac{x-x_i}{h}$ (که در آن x نقطه‌ای است که می‌خواهیم درون‌یابی کنیم) کوچک باشد. اگر x به ابتدای جدول نزدیک باشد به صورت پیشرو و اگر x به انتهای جدول نزدیک باشد به صورت پسرو عمل می‌کنیم. برای پرهیز از افزایش حجم محاسبات، درجه چندجمله‌ای درون‌یاب را بی‌جهت اضافه نمی‌کنیم؛ البته این مطلب بستگی به h دارد (برای h کوچک درون‌یابی خطی (درجه یک) نیز جواب خوبی می‌دهد).*

مثال ۶.۳ *با توجه به جدول داده‌شده مطلوب است مقدار* $\sin(5^\circ)$، $\sin(25^\circ)$ و $\sin(45^\circ)$.

x_i	۰°	۱۰°	۲۰°	۳۰°	۴۰°	۵۰°
$\sin(x_i)$	۰	۰٫۱۷۳۶	۰٫۳۴۲۰	۰٫۵	۰٫۶۴۲۸	۰٫۷۶۶۰

ابتدا جدولی به صورت زیر می‌سازیم

x_i	f_i	Δf_i	$\Delta^2 f_i$	$\Delta^3 f_i$	$\Delta^4 f_i$	$\Delta^5 f_i$
۰	۰					
		۰٫۱۷۳۶				
۱۰	۰٫۱۷۳۶		-۰٫۰۰۵۲			
		۰٫۱۶۸۴		-۰٫۰۰۵۲		
۲۰	۰٫۳۴۲۰		-۰٫۰۱۰۴		۰٫۰۰۰۴	
		۰٫۱۵۸۰		-۰٫۰۰۴۸		۰
۳۰	۰٫۵		-۰٫۰۱۵۲		۰٫۰۰۰۴	
		۰٫۱۴۲۸		-۰٫۰۰۴۴		
۴۰	۰٫۶۴۲۸		-۰٫۰۱۹۶			
		۰٫۱۲۳۲				
۵۰	۰٫۷۶۶۰					
x_i	f_i	∇f_i	$\nabla^2 f_i$	$\nabla^3 f_i$	$\nabla^4 f_i$	$\nabla^5 f_i$

برای درون‌یابی در $x = ۵^\circ$، با انتخاب $x_۰ = ۰^\circ$ داریم $\theta = \frac{x-x_۰}{h} = \frac{۵^\circ-۰^\circ}{۱۰^\circ} = \frac{۱}{۲}$. به کمک نتیجه۱.۳.۳ خواهیم داشت

$$\sin(۵^\circ) \simeq f_۰ + \theta\Delta f_۰ + \frac{\theta(\theta-۱)}{۲!}\Delta^۲ f_۰ + \cdots + \frac{\theta(\theta-۱)\cdots(\theta-۵+۱)}{۵!}\Delta^۵ f_۰,$$

که پس از جایگذاری، مقادیر تقریبی $\sin(۵^\circ)$ در ادامه در جدول آورده شده است. برای درون‌یابی در $x = ۴۵^\circ$، با انتخاب $x_۴ = ۴۰^\circ$ داریم $\theta = \frac{x-x_۴}{h} = \frac{۴۵^\circ-۴۰^\circ}{۱۰^\circ} = \frac{۱}{۲}$. به کمک لم۱.۴.۳ خواهیم داشت

$$\sin(۴۵^\circ) \simeq f_۴ + \theta\nabla f_۴ + \frac{\theta(\theta+۱)}{۲!}\nabla^۲ f_۴ + \cdots + \frac{\theta(\theta+۱)\cdots(\theta+۴-۱)}{۴!}\nabla^۴ f_۴,$$

که پس از جایگذاری، مقادیر تقریبی $\sin(۴۵^\circ)$ در ادامه در جدول آورده شده است. برای درون‌یابی در $x = ۲۵^\circ$ اگر از $x_۲ = ۲۰^\circ$ به صورت پیشرو استفاده کنیم آن‌گاه

$$\theta = \frac{x-x_۲}{h} = \frac{۲۵^\circ - ۲۰^\circ}{۱۰^\circ} = \frac{۱}{۲}.$$

به کمک نتیجه۱.۳.۳ داریم

$$\sin(۲۵^\circ) \simeq f_۲ + \theta\Delta f_۲ + \frac{\theta(\theta-۱)}{۲!}\Delta^۲ f_۲ + \frac{\theta(\theta-۱)(\theta-۲)}{۳!}\Delta^۳ f_۲,$$

که پس از جایگذاری، مقادیر تقریبی $\sin(۲۵^\circ)$ در ادامه در جدول آورده شده است و اگر برای درون‌یابی در $x = ۲۵^\circ$ از $x_۳ = ۳۰^\circ$ به صورت پسرو استفاده کنیم آن‌گاه

$$\theta = \frac{x-x_۳}{h} = \frac{۲۵^\circ - ۳۰^\circ}{۱۰^\circ} = -\frac{۱}{۲}.$$

به کمک نتیجه۱.۴.۳ داریم

$$\sin(۲۵^\circ) \simeq f_۳ + \theta\nabla f_۳ + \frac{\theta(\theta+۱)}{۲!}\nabla^۲ f_۳ + \frac{\theta(\theta+۱)(\theta+۲)}{۳!}\nabla^۳ f_۳,$$

که پس از جایگذاری، مقادیر تقریبی $\sin(۲۵^\circ)$ در ادامه در جدول آورده شده است.

مقدار تقریبی $\sin(۵^\circ)$	k*(درجه چندجمله‌ای درون‌یاب)*
۰	۰
۰ + ۰٫۰۸۶۸ = ۰٫۰۸۶۸	۱
۰ + ۰٫۰۸۶۸ + ۰٫۰۰۰۶ = ۰٫۰۸۷۴	۲
۰ + ۰٫۰۸۶۸ + ۰٫۰۰۰۶ − ۰٫۰۰۰۳ = ۰٫۰۸۷۱	۳
۰ + ۰٫۰۸۶۸ + ۰٫۰۰۰۶ − ۰٫۰۰۰۳ + ۰٫۰۰۰۰ = ۰٫۰۸۷۱	۴
۰ + ۰٫۰۸۶۸ + ۰٫۰۰۰۶ − ۰٫۰۰۰۳ + ۰٫۰۰۰۰ − ۰ = ۰٫۰۸۷۱	۵

مقدار تقریبی $\sin(45^\circ)$	k(درجه چندجمله‌ای درون‌یاب)
۰/۶۴۲۸	۰
۰/۶۴۲۸ + ۰/۰۷۱۴ = ۰/۷۱۴۲	۱
۰/۶۴۲۸ + ۰/۰۷۱۴ − ۰/۰۰۵۷ = ۰/۷۰۸۵	۲
۰/۶۴۲۸ + ۰/۰۷۱۴ − ۰/۰۰۵۷ − ۰/۰۰۱۵ = ۰/۷۰۷۰	۳
۰/۶۴۲۸ + ۰/۰۷۱۴ − ۰/۰۰۵۷ − ۰/۰۰۱۵ + ۰/۰۰۰۱ = ۰/۷۰۷۱	۴

مقدار تقریبی $\sin(25^\circ)$ (پیشرو)	k(درجه چندجمله‌ای درون‌یاب)
۰/۳۴۲۰	۰
۰/۳۴۲۰ + ۰/۰۷۹۰ = ۰/۴۲۱۰	۱
۰/۳۴۲۰ + ۰/۰۷۹۰ + ۰/۰۰۱۹ = ۰/۴۲۲۹	۲
۰/۳۴۲۰ + ۰/۰۷۹۰ + ۰/۰۰۱۹ − ۰/۰۰۰۳ = ۰/۴۲۲۶	۳

مقدار تقریبی $\sin(25^\circ)$ (پسرو)	k(درجه چندجمله‌ای درون‌یاب)
۰/۵۰۰۰	۰
۰/۵۰۰۰ − ۰/۰۷۹۰ = ۰/۴۲۱۰	۱
۰/۵۰۰۰ − ۰/۰۷۹۰ + ۰/۰۰۱۳ = ۰/۴۲۲۳	۲
۰/۵۰۰۰ − ۰/۰۷۹۰ + ۰/۰۰۱۳ + ۰/۰۰۰۳ = ۰/۴۲۲۶	۳

$\triangle$

تذکر ۶.۳ *ممکن است چندجمله‌ای درون‌یاب پیشرو (پسرو) نیوتن برای درون‌یابی f زمانی که x در اواسط جدول قرار دارد مناسب نباشد، زیرا از تمام اطلاعات جدول استفاده نمی‌شود. در این صورت بهتر است درون‌یاب مرکزی مورد استفاده قرار گیرد.*

۲.۳ خطای چندجمله‌ای درون‌یاب

در این بخش قصد داریم به بررسی خطای چندجمله‌ای درون‌یاب بپردازیم.

قضیه ۵.۳ *فرض کنید p چندجمله‌ای درون‌یاب تابع f در نقاط متمایز $a = x_0 \le x_1 \le \cdots \le x_n = b$ باشد. اگر $f \in C^{n+1}[a,b]$ آن‌گاه به ازای هر x در $[a,b]$ عدد $\xi(x)$ در (a,b) چنان وجود دارد که*

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

تذکر ۷.۳ *بنابر شرایط قضیه ۵.۳، به ازای هر x در $[a,b]$ داریم*

$$|f(x)-p(x)| \leq \frac{M_۱ M_۲}{(n+۱)!},$$

که در آن

$$M_۱ = \max_{a\leq x\leq b} |f^{(n+۱)}(x)|,$$

و

$$M_۲ = \max_{a\leq x\leq b} |(x-x_۰)(x-x_۱)\cdots(x-x_n)|.$$

واضح است که برای nهای بزرگ، یافتن $M_۲$ مقدور نیست و در چنین حالتی از کران بدبینانه $M_۲ \leq (b-a)^{n+۱}$ استفاده می‌کنیم.

مثال ۷.۳ *چندجمله‌ای درون‌یاب تابع $y=f(x)=\cos(\frac{\pi x}{۸})$ را در نقاط $x_۰=۰, x_۱=۲, x_۲=۳$ به دست آورده و سپس کران بالای خطای درون‌یابی را تعیین نموده و آن را با خطای واقعی در نقطه $x=۱$ مقایسه کنید (محاسبات را با سه رقم اعشار دنبال کنید).* $\triangle$

x_i	f_i	$f[x_i,x_{i+1}]$	$f[x_۰,x_۱,x_۲]$
۰	۱		
		$\frac{۰/۷۰۷-۱}{۲-۰}=-۰/۱۴۷$	
۲	۰/۷۰۷		$\frac{-۰/۳۲۴+۰/۱۴۷}{۳-۰}=-۰/۰۵۹$
		$\frac{۰/۳۸۳-۰/۷۰۷}{۳-۲}=-۰/۳۲۴$	
۳	۰/۳۸۳		

پس چندجمله‌ای درون‌یاب به صورت زیر است

$$p(x) = ۱ - ۰/۱۴۷x - ۰/۰۵۹x(x-۲) = ۱ - ۰/۰۲۹x - ۰/۰۵۹x^۲.$$

چون $n=۲$ و $f(x)=\cos(\frac{\pi x}{۸})$ داریم

$$f'(x) = -\frac{\pi}{۸}\sin(\frac{\pi x}{۸}), \qquad f''(x) = -\frac{\pi^۲}{۶۴}\cos(\frac{\pi x}{۸}), \qquad f'''(x) = \frac{\pi^۳}{۵۱۲}\sin(\frac{\pi x}{۸}).$$

بنابراین

$$M_۱ = \max_{۰\leq x\leq ۳} |f'''(x)| = \frac{\pi^۳}{۵۱۲}|\sin(\frac{\pi x}{۸})| \leq \frac{\pi^۳}{۵۱۲} \simeq ۰/۰۶۱,$$

و

$$M_۲ = \max_{۰\leq x\leq ۳} |x(x-۲)(x-۳)| \leq ۳^۳ = ۲۷.$$

در نتیجه

$$|f(x)-p(x)| \le \frac{M_۱ M_۲}{۶} \le \frac{۰/۰۶۱ \times ۲۷}{۶} = ۰/۲۷۳.$$

اما این کران بالای خطا بدبینانه است و در ادامه کران بالای واقع‌بینانه‌تری به دست می‌آوریم. چون تابع $y=\sin(\frac{\pi x}{۸})$ در بازه $[۰,۳]$ صعودی است!؟ پس داریم

$$M_۱ = \max_{۰\le x\le ۳} |f'''(x)| = \frac{\pi^۳}{۵۱۲} \max_{۰\le x\le ۳} |\sin(\frac{\pi x}{۸})| \le \frac{\pi^۳}{۵۱۲} |\sin(\frac{۳\pi}{۸})| \simeq ۰/۰۵۶.$$

هم‌چنین با فرض $g(t)=t(t-۲)(t-۳)$ داریم $g'(t)=۳t^۲-۱۰t+۶$ و بنابراین از $g'(t)=۰$ نتیجه می‌شود $t_۱=\frac{۵+\sqrt{۷}}{۳}$ و $t_۲=\frac{۵-\sqrt{۷}}{۳}$. حال چون $t_۱$ و $t_۲$ به بازه $[۰,۳]$ تعلق دارند، خواهیم داشت

$$M_۲ = \max\{|g(۰)|, |g(t_۱)|, |g(t_۲)|, |g(۳)|\} = \{۰, ۲/۱۱۳, ۰/۶۳۱, ۰\} = ۲/۱۱۳.$$

پس

$$|f(x)-p(x)| \le \frac{M_۱ M_۲}{۶} \le \frac{۰/۰۵۶ \times ۲/۱۱۳}{۶} = ۰/۰۲۰.$$

از طرفی داریم

$$|f(۱)-p(۱)| = |۰/۹۲۴ - ۰/۹۱۲| = ۰/۰۱۲ < ۰/۰۲۰ < ۰/۲۷۳.$$

تذکر ۸.۳ *اگر p چندجمله‌ای درون‌یاب f در نقاط $x_۰,\ldots,x_n$ با روش تفاضلات تقسیم‌شده نیوتن باشد، چندجمله‌ای درون‌یاب f در نقاط $x_۰,\ldots,x_n,t$ با همان روش به صورت زیر به دست می‌آید*

$$q(x) = p(x) + f[x_۰, x_۱, \ldots, x_n, t](x-x_۰)\cdots(x-x_n),$$

و چون $f(t)=q(t)$ بنابراین

$$f(t) = p(t) + f[x_۰, x_۱, \ldots, x_n, t](t-x_۰)\cdots(t-x_n),$$

که با مقایسه با قضیه خطای چندجمله‌ای درون‌یاب خواهیم داشت $f[x_۰,\ldots,x_n,t] = \frac{f^{(n+۱)}(\xi_t)}{(n+۱)!}$ و می‌توان نوشت $f[x_۰,\ldots,x_n] = \frac{f^{(n)}(\xi)}{n!}$.

۳.۳ برون‌یابی و درون‌یابی وارون

در حالت کلی برای برون‌یابی ابزارهای پیشرفته‌تری نیاز است ولی برای برون‌یابی در نقاطی که نزدیک دو انتهای بازه داده شده $[a,b]$ باشند می‌توان از همان چندجمله‌ای درون‌یاب استفاده کرد و با توجه به مشکلات درون‌یابی باید توجه داشت که هرچه از دو انتها دور شویم اعتبار نتایج کمتر می‌شود. اما برای درون‌یابی وارون می‌توان از ابزارهای درون‌یابی

به خوبی سود برد. در ادامه دو ایده برای حل این مسئله مطرح می‌گردد. ایده اول آن است که فرض کنید p چندجمله‌ای درون‌یاب تابع f در نقاط $x_۰ < x_۱ < \cdots < x_n$ باشد. با به کار بردن یکی از روش‌های فصل ریشه‌یابی مانند نیوتن یا تکرار ساده در حل معادله $p(\bar{x}) = f(\bar{x})$، تقریبی از $\bar{x}$ به دست می‌آید.

مثال ۸.۳ *با توجه به جدول داده‌شده مطلوب است مقدار $\bar{x}$ به طوری که* $\sinh(\bar{x}) = ۵$.

x_i	۱	۲	۳	۴
$\sinh(x_i)$	۱/۱۷۵۲	۳/۶۲۶۹	۱۰/۰۱۷۹	۲۷/۲۸۹۹

ابتدا جدولی به صورت زیر می‌سازیم

x_i	f_i	Δf_i	$\Delta^۲ f_i$	$\Delta^۳ f_i$
۱	۱/۱۷۵۲			
		۲/۴۵۱۷		
۲	۳/۶۲۶۹		۳/۹۳۹۳	
		۶/۳۹۱۰		۶/۹۴۱۷
۳	۱۰/۰۱۷۹		۱۰/۸۸۱۰	
		۱۷/۲۷۲۰		
۴	۲۷/۲۸۹۹			

به کمک نتیجه۱.۳.۳ می‌توان نوشت

$$p(\bar{x}) = f_۰ + \theta \Delta f_۰ + \frac{\theta(\theta-۱)}{۲!}\Delta^۲ f_۰ + \frac{\theta(\theta-۱)(\theta-۲)}{۳!}\Delta^۳ f_۰.$$

که در آن با انتخاب $x_۰ = ۱$ *داریم* $\theta = \frac{\bar{x}-x_۰}{h} = \bar{x} - ۱$ *و در نتیجه*

$$p(\bar{x}) = ۱/۱۷۵۲ + ۲/۴۵۱۷(\bar{x}-۱) + \frac{۳/۹۳۹۳}{۲}(\bar{x}-۱)(\bar{x}-۲) + \frac{۶/۹۴۱۷}{۶}(\bar{x}-۱)(\bar{x}-۲)(\bar{x}-۳),$$

و یا

$$p(\bar{x}) = ۱/۱۵۷۰\bar{x}^۳ - ۴/۹۷۲۱\bar{x}^۲ + ۹/۲۶۹۲\bar{x} - ۴/۲۷۸۹.$$

بنابراین $\bar{x}$ از حل معادله $p(\bar{x}) = ۵$ به دست می‌آید. جدول زیر تکرارهای روش نیوتن را با انتخاب $t_۰ = ۲/۳$ برای تابع $g(t) = p(t) - ۵ = ۱/۱۵۷۰t^۳ - ۴/۹۷۲۱t^۲ + ۹/۲۶۹۲t - ۹/۲۷۸۹$ نشان می‌دهد

n	۱	۲	۳
t_n	۲/۳۳۸۹	۲/۳۳۸۰	۲/۳۳۸۰

پس با دقت $۳D$ داریم $\bar{x} = ۲/۳۳۸$ و $g(۲/۳۳۸) = ۰/۰۰۰۰$ اما $\sinh(۲/۳۳۸) - ۵ = ۰/۱۳۲۰$!؟ $\triangle$

اما ایده دوم آن است که فرض کنید تابع $y = f(x)$ در بازه‌ای شامل x_iها وارون‌پذیر است و جدول زیر را در نظر بگیرید.

y_i	$y_۰$	$y_۱$	$\cdots$	y_n
x_i	$x_۰$	$x_۱$	$\cdots$	x_n

اگر $x = q(y)$ چندجمله‌ای درون‌یاب صادق در جدول باشد که با یکی از روش‌های درون‌یابی لاگرانژ یا تفاضلات تقسیم‌شده نیوتن به دست آمده باشد، آن‌گاه داریم $\bar{x} \simeq q(f(\bar{x}))$. یعنی $x = q(y)$ را به عنوان تقریبی از تابع وارون $y = f(x)$ می‌پذیریم.

مثال ۹.۳ *یک کاربرد جالب از درون‌یابی وارون در ریشه‌یابی است. تقریبی از ریشه تابعی که از آن تابع فقط اطلاعات زیر در دسترس است بیابید. سپس جواب خود را آزمایش کنید.*

$$f(۰) = -۱, \qquad f(۰/۵) = -۰/۳۷۷۶, \qquad f(۱) = ۰/۴۵۹۷, \qquad f(۱/۵) = ۱/۴۲۹۳.$$

ابتدا جدول تفاضلات تقسیم‌شده را به صورت زیر ساخته

f_i	x_i	مرتبه اول	مرتبه دوم	مرتبه سوم
-۱	۰			
		$\frac{۰/۵-۰}{-۰/۳۷۷۶-(-۱)} = ۰/۸۰۳۳$		
$-۰/۳۷۷۶$	$۰/۵$		$-۰/۱۴۱۲$	
		$\frac{۱-۰/۵}{۰/۴۵۹۷-(-۰/۳۷۷۶)} = ۰/۵۹۷۲$		$۰/۰۳۹۶$
$۰/۴۵۹۷$	۱		$-۰/۰۴۵۱$	
		$\frac{۱/۵-۱}{۱/۴۲۹۳-۰/۴۵۹۷} = ۰/۵۱۵۷$		
$۱/۴۲۹۳$	$۱/۵$			

و از $f(\alpha) = ۰$ *داریم* $\alpha \simeq ۰ + ۰/۸۰۳۳(۱) - ۰/۱۴۱۲(۱)(۰/۳۷۷۶) + ۰/۰۳۹۶(۱)(۰/۳۷۷۶)(-۰/۴۵۹۷)$.
پس $\alpha \simeq ۰/۷۴۳۱$. *برای آزمایش جواب به جدول تفاضلاتی زیر نیاز داریم.*

x_i	f_i	Δf_i	$\Delta^۲ f_i$	$\Delta^۳ f_i$
۰	-۱			
		$۰/۶۲۲۴$		
$۰/۵$	$-۰/۳۷۷۶$		$۰/۲۱۴۹$	
		$۰/۸۳۷۳$		$-۰/۰۸۲۶$
۱	$۰/۴۵۹۷$		$۰/۱۳۲۳$	
		$۰/۹۶۹۶$		
$۱/۵$	$۱/۴۲۹۳$			

سپس به کمک نتیجه۱.۳.۳ می‌توان نوشت

$$p(x) = f_۰ + \theta \Delta f_۰ + \frac{\theta(\theta-۱)}{۲!}\Delta^۲ f_۰ + \frac{\theta(\theta-۱)(\theta-۲)}{۳!}\Delta^۳ f_۰.$$

که در آن با انتخاب $x_۰ = ۰$ *داریم* $\theta = \frac{x-x_۰}{h} = \frac{۰/۷۴۳۱-۰}{۰/۵} = ۱/۴۸۶۲$ *و در نتیجه*

$$\begin{aligned} p(۰/۷۴۳۱) &= -۱ + ۰/۶۲۲۴(۱/۴۸۶۲) + \tfrac{۰/۲۱۴۹}{۲}(۱/۴۸۶۲)(۰/۴۸۶۲) + \\ &\tfrac{-۰/۰۸۲۶}{۶}(۱/۴۸۶۲)(۰/۴۸۶۲)(-۰/۵۱۳۸), \end{aligned}$$

و بنابراین $p(۰/۷۴۳۱) \simeq ۰/۰۰۷۸$ *که بیانگر آن است که* $۰/۷۴۳۱$ *تقریبی از* α *است.* $\triangle$

۴.۳ تقریب کم‌ترین مربعات گسسته

در فصل ۱ با چندجمله‌ای تیلور درجه n حول نقطه $x_۰$ آشنا شدیم که تقریب خوبی برای یک تابع $n+۱$ بار مشتق‌پذیر در همسایگی کوچکی از $x_۰$ است. از چندجمله‌ای درون‌یاب نیز می‌توان به عنوان تقریبی از یک تابع استفاده نمود ولی این چندجمله‌ای فقط در نقاط معلوم دقیق است (صرف نظر از خطای گرد کردن) و در سایر نقاط ممکن است حتی جوابی دور از انتظار تولید کند. در این فصل، قصد داریم یک چندجمله‌ای بسازیم که تقریب مناسبی! برای یک تابع معلوم (مجهول) باشد. در اینجا با یکی از دو مسئله کلی زیر مواجه هستیم

- در جستجوی تابعی (چندجمله‌ای) هستیم که برای داده‌های یک جدول مناسب! باشد؛

- تابعی با ضابطه پیچیده در دسترس است و می‌خواهیم به جای کار کردن با آن، از نوع ساده‌تری از توابع مانند چندجمله‌ای‌ها استفاده کنیم که تقریب مناسبی! برای تابع باشد.

به منظور بررسی مسئله اول، فرض کنید از تابع f فقط داده‌های جدولی

x_i	$x_۱$	$\cdots$	x_m
f_i	$f_۱$	$\cdots$	f_m

در دسترس باشد و بخواهیم چندجمله‌ای $p_n(x) = a_n x^n + \cdots + a_۱ x + a_۰$ را چنان بیابیم که تقریب مناسبی برای تابع نامعلوم f باشد. برای مفهوم دادن به واژه‌ی «تقریب مناسب»، ابتدا به نظر می‌رسد باید ضرایب $a_۰, \ldots, a_n$ را به گونه‌ای یافت که عبارت

$$E_\infty(a_۰, \ldots, a_n) = \max_{۱\le k \le m} |f_k - p_n(x_k)|$$

کمینه شود. این مسئله از نوع اقل‌اکثر بوده و در اینجا قادر به حل آن نخواهیم بود. ایده دیگری که به ذهن می‌رسد آن است که برای تعیین ضرایب $a_0,\dots,a_n$ تابع

$$E_1(a_0,\dots,a_n)=\sum_{k=1}^{m}|f_k-p_n(x_k)|$$

کمینه شود و به همین منظور بنابر آن چه که از حساب دیفرانسیل و انتگرال آموختیم باید عبارت $\frac{\partial E_1}{\partial a_i}$ را یافته و برابر صفر قرار دهیم که عدم مشتق‌پذیری تابع قدرمطلق مانع از ادامه کار می‌گردد. اما می‌توان برای تعیین ضرایب $a_0,\dots,a_n$ تابع

$$E_2(a_0,\dots,a_n)=\sum_{k=1}^{m}(f_k-p_n(x_k))^2$$

را کمینه کرد. این مسئله به مسئله کم‌ترین مربعات[1] گسسته معروف است و برای حل آن برای $i=0,1,\dots,n$ باید داشته باشیم $\frac{\partial E_2}{\partial a_i}=0$ و در نتیجه

$$0=\frac{\partial E_2}{\partial a_i}=\frac{\partial}{\partial a_i}\sum_{k=1}^{m}\left(f_k-\sum_{j=0}^{n}a_jx_k^j\right)^2=\sum_{k=1}^{m}\frac{\partial}{\partial a_i}\left(f_k-\sum_{j=0}^{n}a_jx_k^j\right)^2$$

و یا

$$0=-2\sum_{k=1}^{m}x_k^i\left(f_k-\sum_{j=0}^{n}a_jx_k^j\right)$$

و از آن جا

$$\sum_{j=0}^{n}\left(\sum_{k=1}^{m}x_k^{i+j}\right)a_j=\sum_{k=1}^{m}f_kx_k^i,\qquad i=0,1,\dots,n.$$

این دستگاه $(n+1)\times(n+1)$ به دستگاه معادلات نرمال معروف است و از حل آن ضرایب $a_0,\dots,a_n$ به دست می‌آیند. با قرار دادن

$$\begin{aligned}
&\alpha=[a_i]_{(n+1)\times 1},\\
&\beta=[\beta_i]_{(n+1)\times 1},\qquad \beta_i=\sum_{k=1}^{m}f_kx_k^i,\qquad i=0,1,\dots,n,\\
&S=[s_{ij}]_{(n+1)\times(n+1)},\qquad s_{ij}=\sum_{k=1}^{m}x_k^{i+j},\qquad i,j=0,1,\dots,n
\end{aligned}$$

می‌توان دستگاه معادلات نرمال را به صورت فشرده $S\alpha=\beta$ و یا به شکل گسترده زیر نیز بیان نمود

$$\begin{bmatrix} s_{00} & s_{01} & \cdots & s_{0n}\\ \vdots & \vdots & \cdots & \vdots\\ s_{n0} & s_{n1} & \cdots & s_{nn}\end{bmatrix}\cdot\begin{bmatrix} a_0\\ \vdots\\ a_n\end{bmatrix}=\begin{bmatrix}\beta_0\\ \vdots\\ \beta_n\end{bmatrix}$$

[1] least squares

مثال ۱۰.۳ یک چندجمله‌ای درجه دو (سهمی) مناسب داده‌های جدولی زیر بسازید.

x_i	۰	۰/۲۵	۰/۵۰	۰/۷۵	۱/۰۰
f_i	۱/۰۰۰۰	۱/۲۸۴۰	۱/۶۴۸۷	۲/۱۱۷۰	۲/۷۱۸۳

در اینجا $m = ۵$ و $n = ۲$ و دستگاه معادلات نرمال به صورت زیر است

$$\begin{aligned} ۵a_۰ &+ ۲/۵a_۱ + ۱/۸۷۵a_۲ = ۸/۷۶۸۰ \\ ۲/۵a_۰ &+ ۱/۸۷۵a_۱ + ۱/۵۶۲۵a_۲ = ۵/۴۵۱۴ \\ ۱/۸۷۵a_۰ &+ ۱/۵۶۲۵a_۱ + ۱/۳۸۲۸a_۲ = ۴/۴۰۱۵ \end{aligned}$$

و از حل آن داریم $a_۰ = ۱/۰۰۵۱$، $a_۱ = ۰/۸۶۴۶۸$ و $a_۲ = ۰/۸۴۳۱۶$. پس

$$p_۲(x) = ۰/۸۴۳۱۶x^۲ + ۰/۸۶۴۶۸x + ۱/۰۰۵۱$$

و بنابراین

x_i	۰	۰/۲۵	۰/۵۰	۰/۷۵	۱/۰۰
f_i	۱/۰۰۰۰	۱/۲۸۴۰	۱/۶۴۸۷	۲/۱۱۷۰	۲/۷۱۸۳
$p_۲(x_i)$	۱/۰۰۵۱	۱/۲۷۴۰	۱/۶۴۸۲	۲/۱۲۷۹	۲/۷۱۲۹
$f_i - p_۲(x_i)$	−۰/۰۰۵۱	۰/۰۱۰۰	۰/۰۰۰۴	−۰/۰۱۰۹	۰/۰۰۵۴

هم‌چنین

$$E_۲ = \sum_{k=۱}^{۵} (f_k - p_۲(x_k))^۲ = ۲/۷۴ \times ۱۰^{-۴}.$$

△

مثال ۱۱.۳ برای یافتن تابعی به یکی از شکل‌های زیر، مناسب داده‌های (x_i, y_i) راه‌کار ارایه دهید.

$$y = ae^{bx}, \qquad y = ax^۴ + bx^۲.$$

برای مورد $y = ae^{bx}$ داریم $\ln y = \ln(ae^{bx})$ و یا $\ln y = \ln a + bx$ و با تغییر متغیرهای $Y = \ln y$ و $A = \ln a$ کافی است خط $Y = A + bx$ را چنان یافت که مناسب داده‌های $(x_i, \ln y_i)$ باشد و در آخر قرار دهیم $a = e^A$. برای مورد $y = ax^۴ + bx^۲$ داریم $\frac{y}{x^۲} = ax^۲ + b$ و با تغییر متغیرهای $Y = \frac{y}{x^۲}$ و $X = x^۲$ کافی است خط $Y = aX + b$ را چنان یافت که مناسب داده‌های $\left(x_i^۲, \frac{y_i}{x_i^۲}\right)$ باشد. △

تمرین

۱. الف) برای نقاط جدول زیر چندجمله‌ای درونیاب را به روش مناسب بیابید.

x	-۱	۰	۱	۳
$f(x)$	-۱	۱	-۱	۰

ب) اگر بدانیم برای هر $x \in [-۱, ۳]$ داریم $|f^{(۴)}(x)| \leq ۱$ خطای درونیابی در $x = ۲$ حداکثر چقدر است؟

ج) آیا با اضافه کردن نقطه $(۴, \frac{۲۳}{۲})$ درجه چندجمله‌ای درونیاب بیشتر می‌شود؟ چرا؟

۲. یک تابع به شکل $y = ax + bx^{۳}$ به داده‌های جدول زیر برازش دهید.

x_i	-۲	-۱	۱	۲
y_i	۴	۱	-۱	۴

۳. جدول زیر را برای تابع $y = f(x)$ در بازه $[۰, ۱]$ در نظر بگیرید. به روش درونیابی مناسب برای نقاط هم‌فاصله مقدار $f(۰/۱)$ را تقریب بزنید.

x	۰	۰/۲۵	۰/۵	۰/۷۵	۱
$f(x)$	۳/۵	۱/۷۵	۱	۱/۲۵	۲

۴. با نقاط $x_k = \frac{۱}{۲}$ ، $k = ۰, ۱, ۲$، کران خطای درونیابی $\ln(۱ + x)$ را برای $۰ \leq x \leq ۱$ به دست آورید.

۵. الف) فرض کنید $P_۱(x)$ چندجمله‌ای درونیاب درجه یک تابع $f(x) = e^{-x^{۲}}$ در نقاط $x_۰ = ۰$ و $x_۱ = ۲$ باشد. نشان دهید برای هر $x \in [۰, ۲]$ داریم

$$|f(x) - P_۱(x)| \leq ۱$$

ب) a را طوری تعیین کنید که با اضافه کردن نقطه $(۵, a)$ به جدول زیر، درجه چندجمله‌ای درونیاب تغییر نکند.

x	۱	۲	۳	۴
$g(x)$	۲	۱۱	۳۲	۷۱

۶. ضرایب a و b را چنان بیابید که تابع $y = \frac{۱}{ax^{۲} + b}$ بهترین تقریب (به مفهوم کمترین مربعات) برای داده‌های جدول زیر باشد.

x_i	۰/۱	۰/۲	۰/۳	۰/۴
y_i	۱	۰/۵	۰/۲	۰/۱

۷. چندجمله‌ای درون‌یاب تابع $f(x) = x^{۳} + x^{۲} - x - ۱$ در نقاط $-۱, ۰, ۱, ۲$ کدام است؟

الف) $p(x) = x^{۳} + x^{۲} - x - ۱$ ب) $p(x) = x^{۳}$ ج) $p(x) = x(x^{۲} - ۱)$ د) $p(x) = x^{۳} - ۱$

۸. صرف نظر از خطای گرد کردن، کدام گزینه در مورد چندجمله‌ای درون‌یاب درجه $n - ۱$ تابع f درست است؟

الف) روش‌های لاگرانژ و تفاضلات تقسیم‌شده نیوتن همواره جواب یکسانی می‌دهند.

ب) این دو روش فقط وقتی f چندجمله‌ای درجه n باشد، جواب یکسانی می‌دهند.

ج) این دو روش هیچ‌گاه جواب یکسانی نمی‌دهند.

د) این دو روش فقط وقتی f چندجمله‌ای درجه $n-۱$ باشد، جواب یکسانی می‌دهند.

۹. برای آن که منحنی $y=\frac{۱}{(ax+b)^۲}$، داده‌های جدول

x_i	۰	۰٫۵	۱
y_i	۱	۰٫۲۵	۰٫۱۶

را به روش کمترین مربعات برازش کند، مقادیر a,b کدامند؟

الف) $a=۲,\ b=۱$ ب) $a=۱,\ b=۲$ ج) $a=۱٫۵,\ b=۱٫۰۸$ د) $a=۱٫۰۸,\ b=۱٫۵$

۱۰. درجه چندجمله‌ای درون‌یاب نظیر داده‌های جدول

x	۱	۲	-۱	۰	۳
y	۳	۸	-۱	۰	۱۵

چند است؟

الف) ۱ ب) ۲ ج) ۴ د) ۳

۱۱. فرض کنید $p(x)=x^۲+x+۱$ چندجمله‌ای درون‌یاب نظیر نقاط $(۰,۱)$، $(۱,۳)$ و $(-۱,۱)$ باشد. با اضافه شدن نقطه $(x_۰,f(x_۰))$، اگر بدانیم $f[-۱,۱,۰,x_۰]=۲$ درون‌یاب جدید کدام خواهد بود؟

الف) $۲x^۳-۲x^۲-۲x+۱$ ب) $۲x^۳-x^۲+x+۱$ ج) $۲x^۳+x^۲-x+۱$ د) $۲x^۳-x^۲-۳x+۱$

۱۲. اگر $L_۱(x),\ L_۲(x),\ L_۳(x)$ چندجمله‌ای‌های لاگرانژ متناظر با نقاط $x_۱=۱,\ x_۲=۲,\ x_۳=۳$ باشند و به ازای هر x داشته باشیم $c_۱L_۱(x)+c_۲L_۲(x)+c_۳L_۳(x)=۱$ آن‌گاه $c_۱+c_۲+c_۳$ کدام است؟

الف) صفر ب) ۱ ج) ۳ د) ۳−

۱۳. مقدار تقریبی $\sqrt{۰٫۲۵}$ را به کمک روش درونیابی چندجمله‌ای مناسب در نقاط $۰٫۱, ۰٫۱۵, ۰٫۲, ۰٫۳$ تقریب بزنید. سپس کرانی برای خطای تقریب محاسبه کنید.

۱۴. الف) چندجمله‌ای درون‌یاب درجه دو تابع $f(x)=e^x$ را در نقاط $x=۰,۰٫۵,۱٫۵$ با روش درون‌یابی مناسب به دست آورید.
ب) کوچکترین کران خطای حاصل از درون‌یابی قسمت قبل را تعیین کنید.

۱۵. ضرایب a و b را به گونه‌ای بیابید که منحنی $y=\frac{۲}{ax+b}$ بهترین تقریب (به مفهوم کمترین مربعات) داده‌های جدول زیر باشد.

x_i	۰	۱	۳	۵	۷
y_i	-۱	۱	۲	۲	۳

۱۶. برای یافتن تابعی به یکی از شکل‌های زیر، مناسب داده‌های (x_i,y_i) راه‌کار ارایه دهید.

$$y=ax^۳+b,\qquad y=ax^۴+bx^۳+cx^۲,\qquad y=\frac{a}{bx+c},\qquad y=\frac{ax^۲}{bx^۲+c},\qquad y=a\cos x+b.$$

۱۷. الف) در جدول تفاضلات تقسیم‌شده زیر، جاهای خالی را مشخص کنید.

x_i	y_i	تفاضل تقسیم‌شده مرتبه ۱	تفاضل تقسیم‌شده مرتبه ۲	تفاضل تقسیم‌شده مرتبه ۳
$\square$	-۱			
		۲		
۰	$\square$		-۲	
		$\square$		$\frac{۵}{۶}$
$\square$	-۱		$\frac{۴}{۳}$	
		۲		
۳	$\square$			

ب) چندجمله‌ای درون‌یاب را در نقاط $(-۱,-۱)$، $(۰,۱)$، $(۱,-۱)$ و $(۳,۳)$ به دست آورید. سپس نشان دهید اگر نقطه $(۲,-۲)$ به داده‌های قبلی اضافه شود، درجه چندجمله‌ای افزایش نمی‌یابد.

۱۸. چندجمله‌ای بهترین تقریب درجه دوم مناسب جدول

x_i	-۲	-۱	۱	۲
y_i	۰	-۳	۳	۴

را به کمک روش کمترین مربعات بیابید.

۱۹. اگر $x_0 = -۱$، $x_1 = ۰$، $x_2 = -۲$، $x_3 = ۴$ و $x_4 = -۳$، مقدار $(L_4(x_3))^2 - ۴L_0(x_1) + ۳L_2(x_2)$ کدام است؟

الف) ۱ ب) ۰ ج) ۴ د) ۳

۲۰. حداکثر خطای درون‌یابی تابع $f(x) = \ln(x)$ در نقاط $x_0 = ۱$ و $x_1 = ۱/۵$ چند است؟

الف) ۰/۰۳۱۲ ب) ۰/۰۲۱۳ ج) ۰/۰۱۲۳ د) ۰/۵

۲۱. مقدار درونیاب مبتنی بر نقاط $(-۰/۵,۱)$، $(۰,۰)$، $(۰/۵,-۱)$ و $(۱,۰)$ در نقطه $x = ۰/۲۵$ کدام است؟

الف) ۰/۶۵۲ ب) ۰/۶۵۲− ج) ۰/۶۲۵ د) ۰/۶۲۵−

۲۲. درجه چندجمله‌ای درونیاب مبتنی بر نقاط $(-۱,-۶)$ ، $(۰,-۳)$ ، $(۱,۲)$ ، $(۳,۱۸)$ و $(-۲,-۷)$ کدام است؟

الف) ۵ ب) ۴ ج) ۳ د) ۲

۲۳. کدامیک از گزینه‌های زیر نادرست است؟

الف) خطای درون‌یابی در $n+۱$ نقطه واقع بر یک چندجمله‌ای از درجه n، صفر است.

ب) درون‌یاب در $n+۱$ نقطه، یک چندجمله‌ای دقیقاً از درجه n است.

ج) در محاسبه تفاضلات تقسیم‌شده نیوتن، امکان انتشار خطا وجود دارد.

د) در محاسبه تفاضلات پیشرو، امکان انتشار خطا وجود دارد.

۲۴. بهترین تابع به شکل $y = \frac{۱}{ax^2 + b}$ به مفهوم کمترین مربعات مناسب داده‌های زیر را مشخص کنید؟

x_i	-۱	۰	۱	۲
y_i	۱	۳	۲	۱

فصل ۴

مشتق‌گیری و انتگرال‌گیری عددی

در این فصل قصد داریم روش‌های عددی مشتق‌گیری و انتگرال‌گیری را بررسی کنیم. در مشتق‌گیری عددی می‌خواهیم تقریبی از $f'(x)$ را به دست آوریم به طوری که x یک نقطه داده‌شده و معلوم است. در انتگرال‌گیری عددی سعی بر آن داریم که مقدار عددی $\int_a^b f(x)dx$ را تقریب بزنیم.

۱.۴ مشتق‌گیری عددی

در این بخش یا با یک تابع مواجه هستیم که ترجیح می‌دهیم به طور مستقیم از ضابطه آن مشتق نگیریم و یا ممکن است با جدولی از مقدارهای یک تابع مشتق‌پذیر روبرو باشیم و قصد داشته باشیم مشتق‌گیری را به صورت عددی (تقریبی) انجام دهیم و برای این منظور از بین روش‌های مبتنی بر چندجمله‌ای درون‌یاب، مبتنی بر بسط تیلور و روش گاوس، روش بسط تیلور را بررسی خواهیم کرد.

۱.۱.۴ روش مبتنی بر بسط تیلور

برای به دست آوردن رابطه‌های تقریبی مشتق، می‌توان از بسط تیلور توابع

$$\ldots, f(x-۲h), f(x-h), f(x+h), f(x+۲h), \ldots$$

استفاده کرده و یک ترکیب خطی از آن‌ها به گونه‌ای ساخت که خطا از مرتبه $O(h^p)$ و یا درجه دقت روش n باشد.

تعریف ۱.۴ *گفته می‌شود مرتبه دقت (صحت) یک عبارت تقریبی n است اگر آن رابطه برای چندجمله‌ای‌های از درجه حداکثر n دقیق باشد. به عنوان نمونه اگر ضریب h^p در یک رابطه با خطای $O(h^p)$ برابر $f^{(p+۱)}(\xi)$ باشد آن‌گاه مرتبه دقت آن روش به وضوح p خواهد بود.*

به کمک بسط‌های تیلور

$$f(x+h) = f(x) + hf'(x) + \frac{h^۲}{۲!}f''(x) + \frac{h^۳}{۳!}f'''(x) + \cdots,$$

و

$$f(x-h)=f(x)-hf'(x)+\frac{h^{۲}}{۲!}f''(x)-\frac{h^{۳}}{۳!}f'''(x)+\cdots$$

داریم

$$f'(x)=\frac{f(x+h)-f(x)}{h}-\frac{h}{۲!}f''(x)-\frac{h^{۲}}{۳!}f'''(x)+\cdots,$$
$$f'(x)=\frac{f(x)-f(x-h)}{h}+\frac{h}{۲!}f''(x)-\frac{h^{۲}}{۳!}f'''(x)+\cdots$$

که از آنجا برای نقاط هم‌فاصله $\ldots,x_{i-۲},x_{i-۱},x_i,x_{i+۱},x_{i+۲},\ldots$ (با فاصله $x_{i+۱}-x_i=h$) و با فرض $x=x_i$، خواهیم داشت

$$f_i'=f'(x_i)=\frac{f_{i+۱}-f_i}{h}+O(h),\qquad f_i'=f'(x_i)=\frac{f_i-f_{i-۱}}{h}+O(h)$$

که به ترتیب به فرمول‌های پیشرو و پسرو دو نقطه‌ای برای مشتق اول معروف بوده و درجه دقت آنها یک است یعنی این فرمول‌ها برای چندجمله‌ای‌های با درجه حداکثر یک دقیق هستند. با کم کردن بسط‌های تیلور $f(x-h)$ و $f(x+h)$ از هم نتیجه می‌گیریم

$$f'(x)=\frac{f(x+h)-f(x-h)}{۲h}-\frac{h^{۲}}{۳!}f^{(۳)}(x)-\frac{h^{۴}}{۵!}f^{(۵)}(x)-\cdots$$

و از آنجا یک فرمول مرکزی سه نقطه‌ای به صورت

$$f_i'=\frac{f_{i+۱}-f_{i-۱}}{۲h}+O(h^{۲})$$

برای مشتق اول به دست می‌آید که دارای درجه دقت دو است. حال می‌توان به کمک بسط‌های تیلور $f_{i+۱}$ و $f_{i+۲}$ یک فرمول پیشرو سه نقطه‌ای به صورت

$$f_i'=f'(x_i)\simeq\frac{-۳f_i+۴f_{i+۱}-f_{i+۲}}{۲h}$$

ساخت که دارای درجه دقت دو است و خطایی از مرتبه $O(h^{۲})$ دارد و با تغییر h به $-h$ در آن، فرمول پسرو سه نقطه‌ای

$$f_i'\simeq\frac{f_{i-۲}-۴f_{i-۱}+۳f_i}{۲h}$$

به دست می‌آید که دارای درجه دقت دو است و خطایی از مرتبه $O(h^{۲})$ دارد. هم‌چنین به کمک بسط‌های تیلور

$$f(x\pm ۲h)=f(x)\pm ۲hf'(x)+\frac{(۲h)^{۲}}{۲!}f''(x)\pm\frac{(۲h)^{۳}}{۳!}f'''(x)+\cdots$$

و ساختن ترکیب خطی $f(x-۲h)-۸f(x-h)+۸f(x+h)-f(x+۲h)$ به رابطه

$$f'(x)=\frac{۱}{۱۲h}(f(x-۲h)-۸f(x-h)+۸f(x+h)-f(x+۲h))+\frac{h^{۴}}{۳۰}f^{(۵)}(\xi)$$

خواهیم رسید که یک رابطه مرکزی پنج نقطه‌ای با خطای برشی $O(h^{۴})$ است و در آن $x-۲h\le\xi\le x+۲h$.

همچنین با جمع بسط‌های تیلور f_{i+1} و f_{i-1} داریم

$$f''(x_i) = f_i'' = \frac{f_{i-1} - 2f_i + f_{i+1}}{h^2} - \frac{2h^2}{4!}f^{(4)}(x_i) - \frac{2h^4}{6!}f^{(6)}(x_i) - \cdots$$

که یک فرمول مرکزی سه نقطه‌ای برای مشتق مرتبه دوم با درجه دقت سه است و خطایی از مرتبه $O(h^2)$ دارد و به طور مشابه می‌توان فرمول‌های پیشرو یا پسرو نه تنها برای مشتق مرتبه دو بلکه برای سایر مراتب مشتق به دست آورد.

تذکر ۱.۴ *در حالت کلی رابطه‌های مشتق دارای خطایی از مرتبه $O(h^p)$ هستند که بستگی به تعداد نقاط درون‌یابی دارد. در ظاهر یا باید p بزرگ باشد که منجر به افزایش محاسبات (فراخوانی بیشتر تابع f) و رشد خطای گرد کردن می‌شود و یا باید h کوچک باشد که این نیز ممکن است مشکل‌ساز باشد، به عنوان نمونه در محاسبه $\frac{f_{i+1}-f_i}{h}$ اگر h کوچک باشد f_i و f_{i+1} به هم نزدیک خواهند بود و تفاضل دو عدد نزدیک به هم باعث از بین رفتن ارقام بامعنا خواهد شد که ممکن است به تولید خطا منجر شود.به علاوه این خطا به عدد کوچک h تقسیم می‌شود که خطای تولیدشده را افزایش خواهد داد. بنابر این سعی می‌کنیم مشتق‌گیری عددی را با احتیاط انجام دهیم.*

مثال ۱.۴ *با استفاده از جدول زیر برای مقادیر تابع $y = f(x)$، مقدار تقریبی $f'(۲)$ را با استفاده از فرمول‌های سه و پنج نقطه‌ای و همچنین تقریبی برای $f''(۲)$ به دست آورید.*

x_i	۱/۸	۱/۹	۲/۰	۲/۱	۲/۲
f_i	۱۰/۸۸۹۳۶۵	۱۲/۷۰۳۱۹۹	۱۴/۷۷۸۱۱۲	۱۷/۱۴۸۹۵۷	۱۹/۸۵۵۰۳۰

به ازای $h = ۰/۱$ و فرمول سه نقطه‌ای پیشرو داریم

$$f'(۲) = \frac{۱}{۰/۲} \times [-۳f(۲) + ۴f(۲/۱) - f(۲/۲)] = ۲۲/۰۳۲۳۱۰$$

به همین ترتیب به ازای $h = ۰/۱$ و فرمول سه نقطه‌ای پسرو خواهیم داشت

$$f'(۲) = \frac{۱}{۰/۲} \times [۳f(۲) - ۴f(۱/۹) + f(۱/۸)] = ۲۲/۰۵۴۵۲۵$$

اما به ازای $h = ۰/۱$ و فرمول سه نقطه‌ای مرکزی داریم

$$f'(۲) = \frac{۱}{۰/۲} \times [f(۲/۱) - f(۱/۹)] = ۲۲/۲۲۸۷۹۰$$

و به ازای $h = ۰/۲$ و فرمول سه نقطه‌ای مرکزی می‌توان نوشت

$$f'(۲) = \frac{۱}{۰/۴} \times [f(۲/۲) - f(۱/۸)] = ۲۲/۴۱۴۱۶۳$$

سرانجام به ازای $h = ۰/۱$ و فرمول پنج نقطه‌ای مرکزی داریم

$$f'(۲) = \frac{۱}{۱/۲} \times [f(۱/۸) - ۸f(۱/۹) + ۸f(۲/۱) - f(۲/۲)] = ۲۲/۱۶۶۹۹۹$$

روشن است که برای استفاده از سایر فرمول‌های پنج نقطه‌ای داده‌ها کافی نیستند. در این مثال $f(x)=xe^x$ و مقدار واقعی $f'(۲)=۲۲٫۱۶۷۱۶۸$ نشان می‌دهد که خطای این روش‌ها به ترتیب عبارت است از $۱٫۱۳\times ۱۰^{-۱}$، $۱٫۳۵\times ۱۰^{-۱}$، $-۶٫۱۶\times ۱۰^{-۲}$، $-۲٫۴۷\times ۱۰^{-۱}$ و $۱٫۶۹\times ۱۰^{-۴}$. حال به ازای $h=۰٫۱$ و $h=۰٫۲$ به ترتیب نتایج زیر از فرمول مرکزی به دست می‌آیند

$$f''(۲)\simeq \frac{۱}{۰٫۰۱}[f(۱٫۹)-۲f(۲)+f(۲٫۱)]=۲۹٫۵۹۳۲۰۰,$$

$$f''(۲)\simeq \frac{۱}{۰٫۰۴}[f(۱٫۸)-۲f(۲)+f(۲٫۲)]=۲۹٫۷۰۴۲۷۵$$

که خطای آنها به ترتیب عبارت است از $-۳٫۷۰\times ۱۰^{-۲}$ و $-۱٫۴۸\times ۱۰^{-۱}$. $\triangle$

۲.۴ انتگرال‌گیری عددی

محاسبه $\int_a^b f(x)dx$ زمانی که تابع اولیه f در دسترس نیست یا محاسبه تابع اولیه به سادگی امکان‌پذیر نیست و یا وقتی از f فقط داده‌های جدولی در دسترس است از مسایل اساسی انتگرال‌گیری است. برای حل این مسئله، در این بخش قصد داریم انتگرال‌گیری را به صورت عددی (تقریبی) انجام دهیم. به عنوان مثال می‌توان به انتگرال‌های $\int_۰^b e^{-x^۲}dx$، $\int_{-a}^a \sqrt[n]{۱+x^n}dx$ و $\int_۰^۱ \frac{۱}{۱+x^n}dx$ اشاره کرد. ایده اصلی انتگرال‌گیری عددی همان ایده یافتن مساحت ناحیه محصور به منحنی $y=f(x)$، محور xها و خطوط $x=a$ و $x=b$ است. یعنی ابتدا افرازی منظم به صورت $a=x_۰<x_۱<\cdots<x_n=b$ برای بازه $[a,b]$ در نظر می‌گیریم که در آن $x_{i+۱}-x_i=h$ و $h=\frac{b-a}{n}$، سپس چندجمله‌ای درون‌یاب f در نقاط $x_i,\ldots,x_{i+m}$ یعنی $p_m(x)$ را ساخته و $\int_{x_i}^{x_{i+m}} p_m(x)dx$ را محاسبه نموده و با جمع نمودن این انتگرال‌ها، تقریبی برای $\int_a^b f(x)dx$ به دست می‌آوریم. به شکل ۱.۴ توجه کنید.

شکل ۱.۴: ایده انتگرال‌گیری عددی

۱.۲.۴ قاعده ذوزنقه

چندجمله‌ای درون‌یاب درجه اول ($m=1$) تابع f در نقاط x_i و x_{i+1} به صورت $p_1(x)=f_i+\theta\Delta f_i$ است که در آن $\theta=\frac{x-x_i}{h}$. سپس با انتگرال‌گیری خواهیم داشت

$$\begin{aligned}\int_{x_i}^{x_{i+1}} f(x)dx \simeq \int_{x_i}^{x_{i+1}} p_1(x)dx &= \int_{x_i}^{x_{i+1}} (f_i+\theta\Delta f_i)dx \\ &= \int_0^1 (f_i+\theta\Delta f_i)hd\theta = h\left(\theta f_i+\frac{\theta^2}{2}\Delta f_i\right)_0^1\end{aligned}$$

در نتیجه قاعده ذوزنقه ساده[1] به صورت

$$\int_{x_i}^{x_{i+1}} f(x)dx \simeq \frac{h}{2}(f_i+f_{i+1})$$

ساخته می‌شود. هم‌چنین می‌توان نوشت

$$\begin{aligned}\int_a^b f(x)dx = \int_{x_0}^{x_n} f(x)dx &= \sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}} f(x)dx \\ &\simeq \sum_{i=0}^{n-1}\frac{h}{2}(f_i+f_{i+1}) = \frac{h}{2}\left(f_0+2\sum_{i=1}^{n-1} f_i+f_n\right)\end{aligned}$$

و از آن‌جا قاعده ذوزنقه مرکب[2] به صورت

$$\int_a^b f(x)dx \simeq T(h) = \frac{h}{2}\left(f_0+2\sum_{i=1}^{n-1} f_i+f_n\right)$$

به دست می‌آید. به شکل ۲.۴ توجه کنید.

شکل ۲.۴: قاعده ذوزنقه ساده و مرکب

قضیه ۱.۴ (خطای قاعده ذوزنقه) *اگر* $f\in C^2[x_i,x_{i+1}]$ *آن‌گاه*

$$E_i=\int_{x_i}^{x_{i+1}} f(x)dx-\frac{h}{2}(f_i+f_{i+1})=\frac{-h^3}{12}f''(\xi_i),\qquad \xi_i\in[x_i,x_{i+1}]$$

[1] simple Trapezoidal rule
[2] composite Trapezoidal rule

و اگر $f\in C^{2}[a,b]$ آن‌گاه

$$E_T(h)=\int_a^b f(x)dx-T(h)=\frac{-nh^3}{12}f''(\xi)=\frac{-(b-a)h^2}{12}f''(\xi),\qquad \xi\in[a,b].$$

تذکر ۲.۴ *در عمل اگر به ازای هر x در بازه $[a,b]$ داشته باشیم $|f''(x)|\leq M_2$، آن‌گاه خواهیم داشت*

$$|E_T(h)|\leq\frac{(b-a)M_2}{12}h^2$$

و این یعنی خطا از مرتبه $O(h^2)$ است. هم‌چنین با توجه به ظاهر شدن f'' در عبارت خطای قاعده ذوزنقه، می‌توان نتیجه گرفت که این روش برای چندجمله‌ای‌های حداکثر از درجه یک دقیق است.

مثال ۲.۴ *به کمک قاعده ذوزنقه، تقریبی از $\int_0^1 x\sin x dx$ به دست آورید که خطای آن حداکثر 10^{-2} باشد.*
چون $a=0$، $b=1$ و $f(x)=x\sin x$ داریم

$$f'(x)=\sin x+x\cos x,\qquad f''(x)=2\cos x-x\sin x$$

و در نتیجه

$$|f''(x)|=|2\cos x-x\sin x|\leq 2|\cos x|+|x||\sin x|\leq 2+1=3=M_2$$

حال باید داشته باشیم

$$\frac{(b-a)M_2}{12}h^2\leq 10^{-2}$$

و یا $\frac{(1-0)\times 3}{12}h^2\leq 10^{-2}$ یعنی $h^2\leq 0/04$ و از آن‌جا $h\leq 0/2$. با انتخاب $h=0/2$ داریم $n=\frac{b-a}{h}=\frac{1}{0/2}=5$ و به ازای $x_0=0$ و $x_i=x_0+ih=\frac{i}{5}$ خواهیم داشت

$$T(0/2)=\frac{0/2}{2}\left(f_0+2\sum_{i=1}^{4}f_i+f_5\right).$$

در نتیجه

$$T(0/2)=\frac{0/2}{2}(0+2(0/2\sin(0/2)+0/4\sin(0/4)+0/6\sin(0/6)+0/8\sin(0/8))+sin(1))$$

و یا $T(0/2)=0/3058$. از طرفی

$$\int_0^1 x\sin xdx=(-x\cos x)_0^1+\int_0^1\cos xdx=-\cos 1+\sin 1=0/3012$$

در نتیجه

$$\left|\int_0^1 x\sin xdx-T(0/2)\right|=|0/3012-0/3058|=0/0046<10^{-2}.$$

اگر قاعده ذوزنقه ساده را به کار ببریم، یعنی با انتخاب $n = ۱$ یا $h = b - a = ۱$ خواهیم داشت

$$T(۱) = \frac{۱}{۲}(f_۰ + f_۱) = \frac{۱}{۲}(\sin ۰ + \sin ۱) = ۰/۴۲۰۷.$$

$\triangle$

۲.۲.۴ قاعده سیمسون

چندجمله‌ای درون‌یاب درجه دوم $(m = ۲)$ در نقاط x_i،$x_{i+۱}$ و $x_{i+۲}$ به صورت زیر است

$$p_۲(x) = f_i + \theta \Delta f_i + \frac{\theta(\theta - ۱)}{۲!} \Delta^۲ f_i, \qquad \theta = \frac{x - x_i}{h}.$$

با قرار دادن

$$\int_{x_i}^{x_{i+۲}} f(x)dx \simeq \int_{x_i}^{x_{i+۲}} p_۲(x)dx$$

و اِعمال تغییر متغیر $\theta = \frac{x-x_i}{h}$ داریم

$$\begin{aligned}
\int_{x_i}^{x_{i+۲}} f(x)dx &\simeq \int_۰^۲ (f_i + \theta \Delta f_i + \frac{\theta(\theta - ۱)}{۲!} \Delta^۲ f_i) h d\theta \\
&= h\left(\theta f_i + \tfrac{\theta^۲}{۲} \Delta f_i + (\tfrac{\theta^۳}{۶} - \tfrac{\theta^۲}{۴}) \Delta^۲ f_i\right)_۰^۲ \\
&= h(۲ f_i + ۲ \Delta f_i + \tfrac{۱}{۳} \Delta^۲ f_i)
\end{aligned}$$

و با جایگزینی Δf_i و $\Delta^۲ f_i$، قاعده سیمسون ساده[۳] به صورت زیر به دست می آید

$$\int_{x_i}^{x_{i+۲}} f(x)dx \simeq \frac{h}{۳}(f_i + ۴ f_{i+۱} + f_{i+۲}).$$

حال با فرض زوج بودن n می‌توان نوشت

$$\begin{aligned}
\int_a^b f(x)dx &= \int_{x_۰}^{x_n} f(x)dx = \int_{x_۰}^{x_۲} f(x)dx + \int_{x_۲}^{x_۴} f(x)dx + \cdots + \int_{x_{n-۲}}^{x_n} f(x)dx \\
&\simeq \tfrac{h}{۳}(f_۰ + ۴ f_۱ + f_۲) + \tfrac{h}{۳}(f_۲ + ۴ f_۳ + f_۴) + \cdots + \tfrac{h}{۳}(f_{n-۲} + ۴ f_{n-۱} + f_n)
\end{aligned}$$

بنابراین قاعده‌ی سیمسون مرکب[۴] به صورت زیر ساخته می‌شود

$$\int_a^b f(x)dx \simeq S(h) = \frac{h}{۳}(f_۰ + ۴(f_۱ + f_۳ + \cdots + f_{n-۱}) + ۲(f_۲ + f_۴ + \cdots + f_{n-۲}) + f_n).$$

---

[۳] simple Simpson's rule

[۴] composite Simpson's rule

قضیه ۲.۴ (خطای قاعده سیمسون) اگر $f \in C^{4}[x_i, x_{i+2}]$ آن‌گاه

$$E_i = \int_{x_i}^{x_{i+2}} f(x)dx - \frac{h}{3}(f_i + 4f_{i+1} + f_{i+2}) = \frac{-h^5}{90} f^{(4)}(\xi_i), \qquad x_i \le \xi_i \le x_{i+2}$$

و اگر $f \in C^{4}[a,b]$ آن‌گاه

$$E_S(h) = \int_a^b f(x)dx - S(h) = \frac{-nh^5}{180} f^{(4)}(\xi) = \frac{-(b-a)h^4}{180} f^{(4)}(\xi), \qquad \xi \in [a,b].$$

تذکر ۳.۴ در عمل اگر به ازای هر x در $[a,b]$ داشته باشیم $|f^{(4)}(x)| \le M_4$، آن‌گاه خواهیم داشت

$$|E_S(h)| \le \frac{(b-a)M_4}{180} h^4$$

و این یعنی خطای قاعده سیمسون از مرتبه $O(h^4)$ است و با توجه به ظاهر شدن جمله $f^{(4)}$ در عبارت خطای قاعده سیمسون، ملاحظه می‌شود این قاعده برای چندجمله‌ای‌های حداکثر از درجه ۳ دقیق است.

مثال ۳.۴ در قاعده سیمسون، بازه $[0,2]$ را به چند قسمت تقسیم کنیم تا مقدار تقریبی $\int_0^2 e^{-x^2} dx$ با خطایی کمتر از 10^{-4} به دست آید.
ابتدا داریم

$$f(x) = e^{-x^2}, \quad f'(x) = -2xe^{-x^2}, \quad f''(x) = (-2 + 4x^2)e^{-x^2}, \quad f'''(x) = (12x - 8x^3)e^{-x^2}$$

و $f^{(4)}(x) = (12 - 48x^2 + 16x^4)e^{-x^2}$. برای یافتن M_4 با متحد صفر قرار دادن

$$f^{(5)}(x) = (-120x + 160x^3 - 32x^5)e^{-x^2}$$

خواهیم داشت $x = 0,\ x \simeq \pm 0{,}495,\ x \simeq \pm 1{,}043$ و از آنجا داریم

$$M_4 = \max\{f^{(4)}(0), f^{(4)}(0{,}495), f^{(4)}(1{,}043), f^{(4)}(2)\} = 12.$$

بنابراین باید داشته باشیم

$$\frac{(2-0) \times 12}{180} h^4 < 10^{-4}$$

در نتیجه $h < 0{,}165000$ که نتیجه می‌دهد $n > 12{,}12000$. پس $n = 14$. △

شکل ۳.۴: قاعده نقطه میانی ساده و مرکب

۳.۲.۴ قاعده نقطه میانی

در قاعده‌های ذوزنقه و سیمسون به مقدار $f(a)$ و $f(b)$ نیاز است و بنابراین چنین روش‌هایی برای محاسبه انتگرال‌هایی به شکل $\int_{-1}^{1}\frac{dx}{\sqrt{1-x^2}}$ یا $\int_0^1 \ln x dx$ مناسب نیستند. راه چاره استفاده از روش‌هایی است که به محاسبه $f(a)$ و $f(b)$ نیاز نداشته باشند. یکی از این روش‌ها قاعده نقطه میانی ساده[۵] است که به صورت زیر معرفی می‌شود

$$\int_{x_i}^{x_{i+1}} f(x)dx \simeq hf(x_i+\frac{h}{2}).$$

بنابراین می‌توان نوشت

$$\int_a^b f(x)dx = \int_{x_0}^{x_n} f(x)dx = \sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}} f(x)dx \simeq \sum_{i=0}^{n-1} hf(x_i+\frac{h}{2})$$

و در نتیجه قاعده نقطه میانی مرکب[۶] به صورت زیر ساخته می‌شود

$$\int_a^b f(x)dx \simeq M(h) = h\sum_{i=0}^{n-1} f(x_i+\frac{h}{2}).$$

تعبیر هندسی این روش در شکل ۳.۴ نشان داده شده است.

قضیه ۳.۴ **(خطای قاعده نقطه میانی)** *اگر $f\in C^2[x_i,x_{i+1}]$ آن‌گاه*

$$E_i = \int_{x_i}^{x_{i+1}} f(x)dx - hf(x_i+\frac{h}{2}) = \frac{h^3}{24}f''(\xi_i), \qquad \xi_i\in[x_i,x_{i+1}]$$

و در نتیجه اگر $f\in C^2[a,b]$ آن‌گاه

$$E_M(h) = \int_a^b f(x)dx - M(h) = \frac{nh^3}{24}f''(\xi) = \frac{(b-a)h^2}{24}f''(\xi), \qquad \xi\in[a,b].$$

تذکر ۴.۴ *در عمل اگر به ازای هر x در $[a,b]$ داشته باشیم $|f''(x)|\le M_2$ آن‌گاه خواهیم داشت*

[۵] simple midpoint rule

[۶] composite midpoint rule

$$|E_M(h)| \leq \frac{(b-a)M_2}{24}h^2.$$

به وضوح مشاهده می‌شود در بیشتر مواقع خطای روش نقطه میانی حدود نصف خطای روش ذوزنقه‌ای است و هم‌چنین این روش برای چندجمله‌ای‌های حداکثر از درجه یک دقیق است.

مثال ۴.۴ *مقدار تقریبی $\int_0^1 \ln x dx$ را به روش نقطه میانی با $n = 4$ به دست آورده با مقدار واقعی مقایسه کنید.*

$$-1 = \int_0^1 \ln x dx \simeq 0/25(\ln\frac{1}{8} + \ln\frac{3}{8} + \ln\frac{5}{8} + \ln\frac{7}{8}) = -0/915\cdots.$$

△

۴.۲.۴ قاعده‌های نیوتن-کاتس

با نگاهی به قاعده‌های انتگرال‌گیری قبلی دیده می‌شود که شکل کلی یک قاعده انتگرال‌گیری یا کوادراتور[7]، به صورت زیر است

$$\int_a^b f(x)dx = \sum_{i=0}^{n} a_i f(x_i) + E$$

که در آن $x_0, x_1, \ldots, x_n$ نقاطی در بازه $[a,b]$ و E بیان‌گر خطای روش است. در کوادراتورهای نیوتن-کاتس[8]، نقاط $x_0, x_1, \ldots, x_n$ معلوم و هم‌فاصله فرض می‌شوند و ضرایب $a_0, a_1, \ldots, a_n$ به گونه‌ای تعیین می‌شوند که قاعده انتگرال‌گیری دارای درجه دقت (صحت) n باشد، یعنی برای $f(x) = 1, x, \ldots, x^n$ دقیق باشد $(E = 0)$ و برای تعیین جمله خطا قرار می‌دهیم $f(x) = x^{n+1}$ و $E \neq 0$. کوادراتورهای نیوتن-کاتس به دو دسته بسته و باز تقسیم می‌شوند. در نوع بسته از دو مقدار $f(a)$ و $f(b)$ استفاده می‌شود حال آن‌که در نوع باز نیازی به آنها نیست.

مثال ۵.۴ *یک قاعده انتگرال‌گیری به صورت*

$$\int_0^{3h} f(x)dx = \sum_{i=0}^{3} a_i f(x_i) + E$$

با نقاط $x_0 = 0, x_1 = h, x_2 = 2h, x_3 = 3h$ در نظر بگیرید. برای تعیین ضرایب $a_0, \ldots, a_3$، با قرار دادن $E = 0$ به ازای $f(x) = 1, x, x^2, x^3$، یک دستگاه خطی به صورت

$$\begin{aligned}
f(x) &= 1, \quad \int_0^{3h} 1dx = a_0 + a_1 + a_2 + a_3 = 3h\\
f(x) &= x, \quad \int_0^{3h} xdx = ha_1 + 2ha_2 + 3ha_3 = \tfrac{9h^2}{2}\\
f(x) &= x^2, \quad \int_0^{3h} x^2dx = h^2a_1 + 4h^2a_2 + 9h^2a_3 = 9h^3\\
f(x) &= x^3, \quad \int_0^{3h} x^3dx = h^3a_1 + 8h^3a_2 + 27h^3a_3 = \tfrac{81h^4}{4}
\end{aligned}$$

---

[7] quadrature

[8] Newton-Cotes

به دست می‌آید که از حل آن خواهیم داشت

$$a_۰ = \frac{۳h}{۸}, \qquad a_۱ = \frac{۹h}{۸}, \qquad a_۲ = \frac{۹h}{۸}, \qquad a_۳ = \frac{۳h}{۸}.$$

بنابراین

$$\int_۰^{۳h} f(x)dx \simeq \frac{۳h}{۸}(f(۰) + ۳f(h) + ۳f(۲h) + f(۳h)).$$

حال با تغییر متغیر $t = x + x_۰$ یک قاعده چهار نقطه‌ای به نام قاعده $\frac{۳}{۸}$ سیمسون به صورت زیر به دست می‌آید

$$\int_{x_۰}^{x_۳} f(t)dt \simeq \frac{۳h}{۸}(f_۰ + ۳f_۱ + ۳f_۲ + f_۳).$$

با توجه به تذکر بعد، خطای این قاعده $\frac{۳h^۵}{۸۰} f^{(۴)}(\xi)$ است و برای چندجمله‌ای‌های حداکثر از درجه سه دقیق است. به عنوان تمرین، قاعده مرکب نظیر را به دست آورید. $\triangle$

تذکر ۵.۴ در مثال ۵.۴، از روش ضرایب نامعین استفاده کرده و یک قاعده به دست آوردیم. به صورت مشابه می‌توان قاعده‌های نیوتن_کاتس $(m + ۱)$_نقطه‌ای از نوع بسته را به دست آورد. شکل کلی این قاعده‌ها به صورت زیر است

$$\int_{x_۰}^{x_m} f(x)dx = A_۰ h \sum_{i=۰}^{m} a_i f_i + A_۱ h^{l+۱} f^{(l)}(\xi), \qquad \xi \in [x_۰, x_m]$$

که در آن

$$l = \begin{cases} m + ۱ & \text{اگر } m \text{ فرد باشد} \\ m + ۲ & \text{اگر } m \text{ زوج باشد} \end{cases}$$

و برای تعیین سایر مجهولات می‌توان از جدول ۱.۴ کمک گرفت. باید توجه داشت که ضرایب در جدول داده‌شده متقارن هستند. چون به ازای $m = ۸$ به بعد، ضرایب منفی آشکار می‌شوند، جهت جلوگیری از انجام عمل تفاضل، بهتر است از mهای کوچک استفاده کرد.

تذکر ۶.۴ افراز منظم $a = t_۰ < t_۱ < \cdots < t_n = b$ از بازه $[a, b]$ را در نظر بگیرید به طوری که

$$t_i = t_۰ + ih, \quad i = ۰, ۱, \ldots, n$$

که در آن $h = \frac{b-a}{n}$. سپس تعریف کنید $x_۱ = t_۰ + h/۲$ و $x_i = x_۱ + (i - ۱)h,\ i = ۱, \ldots, n$. قاعده‌های نیوتن_کاتس باز با استفاده از نقاط $x_۱, \ldots, x_n$ ساخته می‌شوند. به عنوان مثال قاعده نقطه میانی یک کوادراتور نیوتن_کاتس دو نقطه‌ای از نوع باز است.

m	A_0	a_0	a_1	a_2	a_3	a_4	A_1
۱	$\frac{1}{2}$	۱	۱				$\frac{-1}{12}$
۲	$\frac{1}{3}$	۱	۴	۱			$\frac{-1}{90}$
۳	$\frac{3}{8}$	۱	۳	۳	۱		$\frac{-3}{80}$
۴	$\frac{4}{45}$	۷	۳۲	۱۲	۳۲	۷	$\frac{-8}{905}$
۵	$\frac{5}{288}$	۱۹	۷۵	۵۰	۵۰	۷۵	$\frac{-275}{2096}$
۶	$\frac{1}{140}$	۴۱	۲۱۶	۲۷	۲۷۲	۲۷	$\frac{-9}{1400}$
۷	$\frac{7}{17280}$	۷۵۱	۳۵۷۷	۱۳۲۳	۲۹۸۹	۲۹۸۹	$\frac{-8183}{518400}$
۸	$\frac{4}{14175}$	۹۸۹	۵۸۸۸	−۹۲۸	۱۰۹۴۸	−۴۵۰۴۰	$\frac{-2368}{467775}$

جدول ۱.۴: قاعده‌های نیوتن-کاتس

۵.۲.۴ کوادراتور گاوس

در این روش یک کوادراتور به صورت زیر در نظر گرفته می‌شود

$$\int_{-1}^{1} f(x)dx = \sum_{i=0}^{n} \omega_i f(x_i) + E$$

که در آن نه تنها ضرایب بلکه نقاط نیز مجهول فرض می‌شوند، بنابراین $2n+2$ مجهول داریم و $2n+2$ معادله به این صورت ساخته می‌شوند که فرض می‌شود درجه دقت (صحت) کوادراتور $2n+1$ باشد، یعنی برای چندجمله‌ای‌های تا درجه $2n+1$ دقیق باشد به عبارت دیگر برای $f(x) = 1, x, x^2, \ldots, x^{2n+1}$ قرار می‌دهیم $E = 0$ و برای تعیین E قرار می‌دهیم $f(x) = x^{2n+2}$.

مثال ۶.۴ *برای به دست آوردن کوادراتور دو نقطه‌ای گاوس با فرض*

$$\int_{-1}^{1} f(x)dx = \omega_0 f(x_0) + \omega_1 f(x_1) + E$$

و با قرار دادن $E = 0$ به ازای $f(x) = 1, x, x^2, x^3$ خواهیم داشت

$$\begin{cases} 2 = \int_{-1}^{1} dx = \omega_0 + \omega_1 \\ 0 = \int_{-1}^{1} x dx = \omega_0 x_0 + \omega_1 x_1 \\ \frac{2}{3} = \int_{-1}^{1} x^2 dx = \omega_0 x_0^2 + \omega_1 x_1^2 \\ 0 = \int_{-1}^{1} x^3 dx = \omega_0 x_0^3 + \omega_1 x_1^3 \end{cases}$$

و بنابراین با حل این دستگاه غیرخطی (که چندان هم راحت نیست) داریم

$$\omega_0 = \omega_1 = 1 \qquad x_1 = -x_0 = \frac{\sqrt{3}}{3}.$$

در نتیجه

$$\int_{-1}^{1} f(x)dx = f(\frac{-\sqrt{3}}{3}) + f(\frac{\sqrt{3}}{3}) + E.$$

برای یافتن E به ازای $f(x) = x^4$ قرار می‌دهیم $E = \tilde{E}$ و خواهیم داشت

$$\int_{-1}^{1} x^4 dx = (\frac{-\sqrt{3}}{3})^4 + (\frac{\sqrt{3}}{3})^4 + \tilde{E}$$

و از آن‌جا $\tilde{E} = \frac{8}{45}$. در نتیجه

$$E = \tilde{E} \times \frac{f^{(4)}(\xi)}{4!} = \frac{f^{(4)}(\xi)}{135}, \qquad \xi \in [-1, 1].$$

پس کوادراتور دو نقطه‌ای گاوس به صورت زیر به دست می‌آید

$$\int_{-1}^{1} f(x)dx = f(\frac{-\sqrt{3}}{3}) + f(\frac{\sqrt{3}}{3}) + \frac{f^{(4)}(\xi)}{135}$$

البته به شرط آن که $f \in C^4[-1, 1]$. △

در حالت کلی برای تعیین کوادراتورهای گاوس، باید یک دستگاه $2n + 2$ معادله غیرخطی حل شود که کار چندان ساده‌ای نیست. در عمل از قضیه زیر کمک می‌گیریم.

قضیه ۴.۴ **(کوادراتور گاوس-لژاندر)** *کوادراتور n نقطه‌ای گاوس-لژاندر در بازه $[-1, 1]$ به صورت زیر است*

$$\int_{-1}^{1} f(x)dx = \sum_{i=0}^{n-1} \omega_i f(x_i) + E_n$$

که در آن $x_0, x_1, \ldots, x_{n-1}$ ریشه‌های p_n یعنی چندجمله‌ای درجه n لژاندر هستند و ضرایب (وزن‌ها) از رابطه

$$\omega_i = \frac{2(1 - {x_i}^2)}{n^2 (p_{n-1}(x_i))^2}, \qquad i = 0, 1, \cdots, n-1$$

به دست می‌آیند و هم‌چنین

$$E_n = \frac{2^{2n+1}(n!)^4}{(2n+1)((2n)!)^3} f^{(2n)}(\xi), \qquad \xi \in [-1, 1].$$

مثال ۷.۴ *کوادراتور سه نقطه‌ای گاوس-لژاندر را به دست آورید.*
می‌دانیم چندجمله‌ای‌های لژاندر به کمک رابطه بازگشتی

$$p_{n+1}(x) = \frac{2n+1}{n+1} x p_n(x) - \frac{n}{n+1} p_{n-1}(x), \qquad n = 1, 2, \ldots$$

با دو جمله ۱ $= p_0(x)$ و $p_1(x) = x$ ساخته می‌شوند و یا می‌توان از رابطه زیر کمک گرفت

$$p_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}\left((x^2-1)^n\right).$$

در اینجا ۳ $= n$ و داریم

$$p_2(x) = \frac{3}{2}x^2 - \frac{1}{2}, \qquad p_3(x) = \frac{5}{2}x^3 - \frac{3}{2}x.$$

از حل ۰ $= p_3(x)$ خواهیم داشت

$$x_0 = -\sqrt{\frac{3}{5}}, \qquad x_1 = 0, \qquad x_2 = \sqrt{\frac{3}{5}}.$$

بنابراین

$$\omega_0 = \frac{2(1-x_0^2)}{9(p_2(x_0))^2} = \frac{5}{9} = \omega_2, \qquad \omega_1 = \frac{2(1-x_1^2)}{9(p_2(x_1))^2} = \frac{8}{9}$$

و هم‌چنین

$$E_3 = \frac{2^7 \times (3!)^4}{7 \times (6!)^3} f^{(6)}(\xi) = \frac{1}{15750} f^{(6)}(\xi), \qquad \xi \in [-1, 1].$$

در نتیجه

$$\int_{-1}^{1} f(x)dx = \frac{1}{9}\left(5f(-\sqrt{\frac{3}{5}}) + 8f(0) + 5f(\sqrt{\frac{3}{5}})\right) + \frac{f^{(6)}(\xi)}{15750}.$$

این قاعده برای چندجمله‌ای‌های تا درجه ۵ دقیق است. به عنوان مثال اگر $f(x) = \frac{1}{1+x^2}$ آن‌گاه

$$\int_{-1}^{1} \frac{dx}{1+x^2} \simeq \frac{1}{9}\left(\frac{5}{1+\frac{3}{5}} + \frac{8}{1+0} + \frac{5}{1+\frac{3}{5}}\right) = \frac{19}{12} = 1/58\bar{3}$$

از طرف دیگر

$$\int_{-1}^{1} \frac{dx}{1+x^2} = \tan^{-1}(x)]_{-1}^{1} = \frac{\pi}{4} + \frac{\pi}{4} = \frac{\pi}{2} = 1/57079.$$

△

کوادراتورهای گاوس-لژاندر را مانند کوادراتورهای نیوتن-کاتس یک بار برای همیشه به دست آورده و در این راستا به نکات زیر توجه داریم

- اگر n زوج باشد

$$\begin{cases} x_{n-i} = -x_i \\ & i = 0, 1, \ldots, \frac{n-2}{2} \\ \omega_{n-i} = \omega_i \end{cases}$$

و اگر n فرد باشد

$$\begin{cases} x_{n-i} = -x_i \\ x_{\frac{n-1}{2}} = 0 \qquad i = 0, 1, \ldots, \frac{n-1}{2} - 1 \\ \omega_{n-i} = \omega_i \end{cases}$$

- ω_i ها همه مثبت هستند و $0 < \omega_i \leq 1$ (اگر $f(x_i)$ دارای خطا باشد ضریب آن کوچک است)

- با تغییر متغیر $x = \frac{1}{2}((b-a)t + (b+a))$ $(dx = \frac{1}{2}(b-a)dt)$ خواهیم داشت

$$\int_a^b f(x)dx = \frac{b-a}{2} \int_{-1}^{1} f(\frac{1}{2}((b-a)t + (b+a)))dt.$$

مثال ۸.۴ *از میان قاعده‌های انتگرال‌گیری دو و سه نقطه‌ای گاوس و سیمسون با $h = 0/1, 0/01$، کدام یک تقریب بهتری برای انتگرال $I = \int_0^4 x^5 dx$ نتیجه می‌دهد؟ می‌دانیم در عبارت خطای روش‌های دو نقطه‌ای گاوس، سیمسون و سه نقطه‌ای گاوس به ترتیب $f^{(4)}$، $h^4 f^{(4)}$ و $f^{(6)}$ ظاهر می‌شود و با توجه به تابع $f(x) = x^5$، انتظار می‌رود به ترتیب قاعده سه نقطه‌ای گاوس، روش سیمسون با $h = 0/01$، روش سیمسون با $h = 0/1$ و قاعده دو نقطه‌ای گاوس تقریب‌های بهتری تولید کنند. البته بدون تردید قاعده سه نقطه‌ای گاوس جواب دقیق را به دست می‌دهد و داریم*

$$I = 2\int_{-1}^{1}(2t+2)^5 dt = \frac{2}{9}(5(-2\sqrt{\frac{3}{5}}+2)^5 + 8(2\times 0+2)^5 + 5(2\sqrt{\frac{3}{5}}+2)^5) = \frac{46}{6} = 682/66\cdots.$$

$\triangle$

مثال ۹.۴ *ضرایب ω_i در قاعده زیر را به گونه‌ای به دست آورید که این قاعده دارای درجه دقت ۲ باشد.*

$$\int_0^h f(\sqrt{x})dx \simeq \omega_1 f(0) + \omega_2 f'(0) + \omega_3 f(h)$$

به ازای $f(x) = 1, x, x^2$ خواهیم داشت

$$\begin{aligned} h &= \textstyle\int_0^h dx = \omega_1 + 0 + \omega_3 \\ \tfrac{2}{3}h^{\frac{3}{2}} &= \textstyle\int_0^h \sqrt{x}dx = 0 + \omega_2 + h\omega_3 \\ \tfrac{h^2}{2} &= \textstyle\int_0^h xdx = 0 + 0 + h^2\omega_3 \end{aligned}$$

و از حل آن داریم

$$\omega_1 = h - \frac{1}{2}, \qquad \omega_2 = \frac{2}{3}h^{\frac{3}{2}} - \frac{h}{2}, \qquad \omega_3 = \frac{1}{2}.$$

$\triangle$

۶.۲.۴ روش رامبرگ

روش رامبرگ از قاعده انتگرال‌گیری ذوزنقه (سیمسون) استفاده کرده و به کمک برون‌یابی ریچاردسون تقریب‌های بهتری به دست می‌آورد. ثابت می‌شود برای توابع به اندازه کافی مشتق‌پذیر داریم

$$\int_a^b f(x)dx = T(h) + a_۲ h^۲ + a_۴ h^۴ + a_۶ h^۶ + \cdots$$

که در آن h اندازه گام نقاط هم فاصله، $T(h)$ قاعده ذوزنقه مرکب و $a_۲$، $a_۴$،... ضرایبی ثابت و مستقل از h هستند. با تبدیل h به $\frac{h}{۲}$ در این رابطه و حذف $a_۲$ خواهیم داشت

$$\int_a^b f(x)dx = \frac{۴T(\frac{h}{۲}) - T(h)}{۳} - \frac{a_۴}{۴}h^۴ - \frac{۵a_۶}{۱۶}h^۶ - \cdots$$

و این یعنی $\frac{۴T(\frac{h}{۲})-T(h)}{۳}$ تقریبی برای مقدار انتگرال با خطای $O(h^۴)$ است، در حالی که $T(h)$ تقریبی با خطای $O(h^۲)$ است. برای ساختن یک روند تکراری، به ازای $i = ۰, ۱, ۲, \ldots$ قرار می‌دهیم

$$h_i = \frac{b-a}{۲^i}, \quad x_j = a + jh_i, \ j = ۰, ۱, \ldots, ۲^i, \quad T_{۰i} = T(h_i) = \frac{h_i}{۲}\left(f(a) + ۲\sum_{j=۱}^{۲^i-۱} f(x_j) + f(b)\right)$$

و برای $p = ۱, ۲, \ldots$ قرار می‌دهیم

$$T_{pi} = \frac{۴^p T_{(p-۱)(i+۱)} - T_{(p-۱)i}}{۴^p - ۱}$$

در عمل جدولی به صورت زیر تهیه کرده و روند را تا جایی ادامه می‌دهیم که برای نمونه شرط توقف $|T_{p۰} - T_{(p-۱)۰}| < \varepsilon$ برقرار شود.

$T_{۰۰}$				
$T_{۰۱}$	$T_{۱۰}$			
$T_{۰۲}$	$T_{۱۱}$	$T_{۲۰}$		
$T_{۰۳}$	$T_{۱۲}$	$T_{۲۱}$	$T_{۳۰}$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

تذکر ۷.۴ *خطای T_{pi} از مرتبه $O(h^{۲p+۲})$ است و برای چندجمله‌ای‌های تا درجه $۲p+۱$ دقیق است و به علاوه ثابت می‌شود*

$$\lim_{p \longrightarrow \infty} T_{p۰} = \int_a^b f(x)dx.$$

تذکر ۸.۴ *بعضی از قاعده‌های نیوتن-کاتس در حین محاسبه جملات قاعده رامبرگ به دست می‌آیند، به عنوان مثال می‌دانیم*

$$T(h) = \tfrac{h}{۲}(f(a) + f(b))$$

$$T(\tfrac{h}{۲}) = \tfrac{h}{۴}\left(f(a) + ۲f(\tfrac{a+b}{۲}) + f(b)\right)$$

و در نتیجه

$$\frac{۴T(\frac{h}{۲}) - T(h)}{۳} = \frac{h}{۶}\left(f(a) + ۴f(\frac{a+b}{۲}) + f(b)\right) = S(\frac{h}{۲}).$$

مثال ۱۰.۴ *مقدار تقریبی $\int_{۰}^{\frac{\pi}{۴}} \sec x dx$ را با روش رامبرگ و با دقت $۴D$ به دست آورید.*

h	$T_{۰i}$	$T_{۱i}$	$T_{۲i}$	$T_{۳i}$
$\frac{\pi}{۴}$	۰/۹۴۸۰۶			
$\frac{\pi}{۸}$	۰/۸۹۹۰۸	۰/۸۸۲۷۶		
$\frac{\pi}{۱۶}$	۰/۸۸۵۸۹	۰/۸۸۱۴۹	۰/۸۸۱۴۰	
$\frac{\pi}{۳۲}$	۰/۸۸۲۵۱	۰/۸۸۱۳۸	۰/۸۸۱۳۷	۰/۸۸۱۳۷
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

داریم

$$|T_{۳۰} - T_{۲۰}| = |۰/۸۸۱۳۷ - ۰/۸۸۱۴۰| = ۰/۳ \times ۱۰^{-۴} < ۰/۵ \times ۱۰^{-۴}$$

و با مقایسه با مقدار واقعی

$$\int_{۰}^{\frac{\pi}{۴}} \sec x dx = (\ln(\sec x + \tan x))_{۰}^{\frac{\pi}{۴}} = \ln(\sqrt{۲}+۱) = ۰/۸۸۱۳۷۳۵۸۷$$

در می‌یابیم

$$\int_{۰}^{\frac{\pi}{۴}} \sec x dx - ۰/۸۸۱۳۷ < ۰/۴ \times ۱۰^{-۵}.$$

$\triangle$

تمرین

۱. از تابع $y = f(x)$ در بازه $[۰, ۱]$، داده‌های جدول زیر در دسترس است.

x	۰	۰/۲۵	۰/۳۷۵	۰/۵	۰/۶۲۵	۰/۷۵	۱
$f(x)$	۰	۰/۱۳۵۰۶	۰/۱۶۰۶۱	۰/۱۶۸۸۷	۰/۱۶۵۵۲	۰/۱۵۴۹۰	۰/۱۲۳۸۵

تقریب‌هایی برای $f'(۰)$، $f'(۰/۵)$ و $f'(۱)$ و همچنین $f''(۰)$، $f''(۰/۵)$ و $f''(۱)$ به چند روش حساب کنید.

۲. به کمک روش بسط تیلور، رابطه پنج نقطه‌ای پیشرو یعنی

$$f'(x) = \frac{۱}{۱۲h}(-۲۵f(x) + ۴۸f(x+h) - ۳۶f(x+۲h) + ۱۶f(x+۳h) - ۳f(x+۴h)) + \frac{h^۴}{۵}f^{(۵)}(\xi)$$

را استخراج کنید که در آن $x \le \xi \le x + ۴h$. سپس با تبدیل h به $-h$ شکل پسروی آن را نیز به دست آورید.

۳. با استفاده از بسط تیلور $f(x_i+h)$ و $f(x_i+\frac{h}{۲})$ نشان دهید $\frac{\Delta f_i}{h}$ برای تقریب $f'_{i+\frac{۱}{۲}}$ دقیق‌تر است تا برای تقریب f'_i، به عبارت دیگر داریم

$$f'_i = \frac{\Delta f_i}{h} + O(h), \qquad f'_{i+\frac{۱}{۲}} = \frac{\Delta f_i}{h} + O(h^۲).$$

۴. قاعده مشتق‌گیری زیر برای چندجمله‌ای‌های حداکثر از درجه چند دقیق است؟

$$f'(x) \simeq \frac{f(x+h/۲) - f(x-h/۲)}{h}$$

۵. به کمک قاعده سیمسون، تقریبی از $\int_{۰}^{۱} x\sin x dx$ به دست آورید که خطای آن حداکثر $۱۰^{-۵}$ باشد.

۶. مقدار $\int_{۰}^{۰/۰۹} \frac{dx}{\sqrt{x}}$ را به کمک روش نقطه میانی چنان به دست آورید که خطای آن حداکثر $۱۰^{-۲}$ باشد.

۷. مقدار $\int_{۰}^{۱}(x^۳+۱)dx$ را به کمک روش دو نقطه‌ای گاوس و روش ذوزنقه با طول گام $h = ۰/۲$ تقریب بزنید. جواب کدام یک از دو روش قسمت قبل دقیق‌تر است؟ (برای ادعای خود بدون محاسبه مقدار دقیق انتگرال دلیل بیاورید).

۸. مقدار $\int_{۱}^{۲} \ln x\, dx$ را به کمک روش دو نقطه‌ای گاوس و روش سیمپسون با طول گام $h = ۰/۲۵$ تقریب بزنید. سپس کرانی برای خطای روش سیمپسون به دست آورید. اگر در قسمت قبل طول گام را در روش سیمپسون $h = ۰/۱۲۵$ بگیریم، بدون محاسبه مقدار تقریبی و فقط به کمک فرمول خطا توضیح دهید که خطا چه تغییری می‌کند؟

۹. فرض کنید h یک عدد حقیقی معلوم باشد. ضرایب مجهول $w_۱$، $w_۲$، $w_۳$ و $w_۴$ را چنان بیابید که فرمول

$$\int_{۰}^{h} f(x)dx = w_۱ f(۰) + w_۲ f(h) + w_۳ f(h/۲) + w_۴ f'(h/۲)$$

برای چندجمله‌ای‌های تا درجه ۳ دقیق باشد.

۱۰. حداقل تعداد تقسیمات n را چنان بیابید که خطای محاسبه $\int_{۱}^{۲} e^{-x^۲}\, dx$ به روش ذوزنقه کمتر از $۱۰^{-۴}$ باشد.

۱۱. حداقل تعداد تقسیمات n را چنان بیابید که خطای محاسبه $\int_{۰}^{۱} \sin(۱/۲x)dx$ به کمک روش‌های ذوزنقه و سیمپسون کمتر از $۱۰^{-۴}$ باشد. با n به دست آمده از قسمت قبل مقدار تقریبی انتگرال را با هر دو روش (محاسبات با دقت $۴D$) محاسبه و با مقدار واقعی مقایسه کنید.

۱۲. فرض کنید h یک عدد حقیقی معلوم باشد. ضرایب مجهول $w_۰$، $w_۱$، $w_۲$ و $w_۳$ را چنان بیابید که فرمول

$$\frac{۱}{۳h}\int_{۰}^{۳h} f(x)dx = w_۰ f(۰) + w_۱ f(h) + w_۲ f(۲h) + w_۳ f(۳h)$$

برای چندجمله‌ای‌های تا درجه ۳ دقیق باشد.

۱۳. فرمول تقریبی $f'''(x_i) \simeq \frac{f_{i+۲} - ۲f_{i+۱} + ۲f_{i-۱} - f_{i-۲}}{۲h^۳}$ برای چندجمله‌ای‌های حداکثر تا درجه چند دقیق است؟ چرا؟

۱۴. الف) انتگرال $\int_{۰}^{۱} f(x)dx$ را در نظر بگیرید و فرض کنید $T(۱) = ۰/۷۵۰۰۰$، $T(\frac{۱}{۲}) = ۰/۸۱۹۴۴$، $T(\frac{۱}{۴}) = ۰/۸۳۱۷۰$. با استفاده از قاعده انتگرال‌گیری رامبرگ تقریبی از $\int_{۰}^{۱} f(x)dx$ با دقت $O(h^۶)$ بیابید.

ب) مقدار تقریبی انتگرال $\int_{0}^{0/4} \frac{\cos x}{1+x^{2}} dx$ را به کمک روش گاوس دو نقطه‌ای به دست آورید.

۱۵. ضرایب a_0, a_1, a_2 را چنان بیابید که فرمول تقریبی $f'(x_i) \simeq a_0 f(x_i - h) + a_1 f(x_i) + a_2 f(x_i + h)$ برای چندجمله‌ای‌های با درجه حداکثر ۲ دقیق باشد.

۱۶. الف) بازه $[1, 2]$ را به چند قسمت مساوی تقسیم کنیم تا مقدار انتگرال $I = \int_{1}^{2} \ln \frac{x^{2}}{2} dx$ به کمک روش ذوزنقه با خطایی کمتر از 10^{-2} به دست آید؟

ب) با استفاده از قاعده گاوسی مناسبی که برای چندجمله‌ای‌های از درجه ۴ دقیق باشد، تقریبی از انتگرال داده شده بیابید.

فصل ۵

دستگاه معادلات خطی

در این فصل قصد داریم از دستگاه معادلات هم‌زمان

$$\begin{cases} E_1 & : & a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ E_2 & : & a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \vdots & & \cdots & & \vdots & & \vdots \\ E_n & : & a_{n1}x_1 & + & a_{n2}x_2 & + & \cdots & + & a_{nn}x_n & = & b_n \end{cases}$$

که در آن b_iها و a_{ij}ها معلوم هستند، $x_1, \dots, x_n$ را به دست آوریم. چنین دسته معادلات را یک دستگاه n معادله n مجهولی خطی نامند. با معرفی بردارهای x و b و ماتریس A به صورت

$$x = [x_i]_{n\times 1}, \qquad b = [b_i]_{n\times 1}, \qquad A = [a_{ij}]_{n\times n}$$

و یا

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}, \qquad A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

می‌توان این دستگاه معادلات را به شکل فشرده $Ax = b$ نوشت.

تذکر ۱.۵ *حل این مسئله معادل با یافتن A^{-1} است، به این معنی که اگر A^{-1} موجود باشد $x = A^{-1}b$ جواب مسئله بیان‌شده است و برعکس اگر solve فرآیندی باشد که A و b را به عنوان ورودی گرفته و x را به عنوان خروجی بدهد یعنی $x = \mathrm{solve}(A, b)$ آن‌گاه برای $i = 1, \dots, n$ قرار می‌دهیم $\hat{A}_i = \mathrm{solve}(A, e_i)$ و سپس تعریف می‌کنیم $\hat{A} = [\hat{A}_1 \ \hat{A}_2 \ \cdots \ \hat{A}_n]$ و خواهیم داشت*

$$A\hat{A} = A[\hat{A}_1 \ \hat{A}_2 \ \cdots \ \hat{A}_n] = [A\hat{A}_1 \ A\hat{A}_2 \ \cdots \ A\hat{A}_n] = [e_1 \ e_2 \ \cdots \ e_n] = I$$

یعنی $\hat{A} = A^{-1}$. چون محاسبه A^{-1} با روش‌های سنتی (کلاسیک) حجم عملیات بالایی دارد، روش‌هایی مطلوب هستند که x را بدون محاسبه A^{-1} به دست آورند.

روش‌های حل مسئله بیان‌شده به دو دسته روش‌های مستقیم[۱] و روش‌های تکراری[۲] دسته‌بندی می‌شوند که در ادامه به بررسی آنها می‌پردازیم.

۱.۵ روش‌های مستقیم

روش‌های مستقیم به روش‌هایی گفته می‌شود که در تعدادی متناهی (از قبل مشخص) تکرار خاتمه می‌یابند و اگر خطای گرد کردن وجود نداشته باشد، جواب واقعی را به دست می‌آورند. روش‌های مستقیمی که از نظر عددی پایدار باشند از اهمیت بالایی برخوردار هستند و وجه تمایز آن‌ها در تعداد اعمال (حجم عملیات)[۳] است.

۱.۱.۵ روش حذف گاوسی

قبل از آن که به معرفی این روش بپردازیم، توجه داریم که اگر $A = [a_{ij}]_{n\times n}$، ماتریسی پایین مثلثی باشد یعنی داشته باشیم

$$A = \begin{bmatrix} a_{11} & & \bigcirc \\ \vdots & \ddots & \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

آن‌گاه از $Ax = b$ نتیجه می‌شود

$$\begin{cases} E_1 : a_{11}x_1 & = b_1 \\ E_2 : a_{21}x_1 + a_{22}x_2 & = b_2 \\ \vdots \quad\quad \vdots \quad\quad \vdots \quad\quad \ddots & \vdots \\ E_n : a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n & = b_n \end{cases}$$

و از آن‌جا به کمک جای‌گذاری پیشرو (از اول)[۴] خواهیم داشت

$$\begin{cases} x_1 = \frac{b_1}{a_{11}} \\ x_i = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j\right), \qquad i = 2, \dots, n \end{cases}$$

---

[۱] direct methods
[۲] iterative methods
[۳] operations count
[۴] forward substitution

و اگر A ماتریسی بالامثلثی باشد یعنی داشته باشیم

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ \bigcirc & & a_{nn} \end{bmatrix}$$

آن‌گاه از $Ax = b$ نتیجه می‌شود

$$\begin{cases} E_1 & : & a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ E_2 & : & & & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & & & \ddots & & \vdots & & \vdots \\ E_n & : & & & & & & & a_{nn}x_n & = & b_n \end{cases}$$

و از آن‌جا به کمک جای‌گذاری پسرو (از آخر)[۵] خواهیم داشت

$$\begin{cases} x_n = \frac{b_n}{a_{nn}} \\ x_i = \frac{1}{a_{ii}}\left(b_i - \sum_{j=i+1}^{n} a_{ij}x_j\right), \qquad i = n-1, n-2, \dots, 1. \end{cases}$$

به وضوح شرط لازم و کافی برای وجود جواب در این حالت آن است که درایه‌های قطری A مخالف صفر باشند. اما ایده روش حذف گاوسی[۶] آن است که به کمک اَعمال سطری مقدماتی[۷] دستگاه $Ax = b$ را به دستگاه $Rx = c$ تبدیل کنیم که دو دستگاه هم‌ارز باشند (جواب یکسانی داشته باشند) و R یک ماتریس مثلثی باشد. واضح است که اگر R ماتریسی پایین/بالامثلثی باشد با جای‌گذاری پیشرو/پسرو به جواب می‌رسیم. **قرارداد** : *ماتریس افزوده*[۸] *متناظر با دستگاه* $Ax = b$*، ماتریسی است به صورت زیر*

$$S = [A\ b] = \begin{bmatrix} a_{11} & \cdots & a_{1n} & a_{1,n+1} \\ \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & a_{n,n+1} \end{bmatrix}$$

که در آن برای $i = 1, \dots, n$ *داریم* $a_{i,n+1} = b_i$ *و منظور از* R_i *سطر* i*اُم ماتریس* S *است.*

تعریف ۱.۵ *عملیات زیر به اَعمال سطری مقدماتی معروف هستند.*

- *ضرب یک سطر در عددی مخالف صفر* $(R_i \leftarrow \lambda R_i)$

[۵] backward substitution
[۶] Gaussian elimination method
[۷] elementary row operations
[۸] augmented matrix

- *تعویض دو سطر با هم* ($R_i \leftrightarrow R_j$)
- *افزودن مضربی از یک سطر به سطر دیگر* ($R_i \leftarrow R_i + \lambda R_j$)

تذکر ۲.۵ *انجام این اَعمال روی سطرهای ماتریس S متناظر با اِعمال آن روی معادلات دستگاه $Ax = b$ است. با اِعمال تعداد متناهی اَعمال سطری مقدماتی، ماتریس S به ماتریس $\tilde{S}$ تبدیل می‌شود که هم‌ارز هستند (دستگاه معادلات نظیر آن‌ها جواب یکسان دارند).*

حال می‌توان روش حذف گاوسی را در الگوریتم زیر خلاصه کرد. **الگوریتم روش حذف گاوسی**

- **ورودی.** ماتریس افزوده‌ی $S = [A\ b]_{n\times n+1}$
- **خروجی.** ماتریس افزوده‌ی $S = [U\ c]_{n\times n+1}$ که در آن U ماتریسی بالا مثلثی است

۱) قرار دهید $j = 1$

۲) rای $(j \le r \le n)$ پیدا کنید که $a_{rj} \neq 0$. اگر چنین rای یافت نشد متوقف شوید مسئله جواب یکتا ندارد

۳) اگر $r \neq j$ آن‌گاه انجام دهید $R_r \leftrightarrow R_j$

۴) برای $i = j+1, \dots, n$ قرار دهید $l_{ij} = \frac{a_{ij}}{a_{jj}}$ و انجام دهید $R_i \leftarrow R_i - l_{ij}R_j$

۵) قرار دهید $j = j+1$ و اگر $j < n$ به گام ۲ بروید

تعریف ۲.۵ *در این الگوریتم، گام ۲ به محورگیری*[۹] *و عنصر a_{rj} به عنصر محوری*[۱۰] *معروف است.*

تذکر ۳.۵ *چون در هر تکرار از روش حذف گاوسی اَعمال سطری مقدماتی انجام می‌شود بنابراین داریم*

$$\begin{aligned}\det A &= \det A^{(1)} = (-1)^{m_2} \det A^{(2)} = (-1)^{m_2}(-1)^{m_3} \det A^{(3)} = \cdots \\ &= (-1)^{m_2} \cdots (-1)^{m_n} \det A^{(n)}\end{aligned}$$

که در آن برای $j = 2, \dots, n$ خواهیم داشت

$$m_j = \begin{cases} 1 & \text{اگر در گام ۳ تکرار } (j-1)\text{ام جابجایی سطر داشته باشیم} \\ 0 & \text{در غیر این صورت} \end{cases}$$

به ویژه

$$\det A = (-1)^{m_2} \cdots (-1)^{m_n} \det A^{(n)} = (-1)^{m_2} \cdots (-1)^{m_n} \det U = (-1)^{m_2} \cdots (-1)^{m_n} u_{11}u_{22} \cdots u_{nn}.$$

---

[۹] pivoting

[۱۰] pivot element

مثال ۱.۵ *دستگاه داده‌شده را با روش حذف گاوسی و جای‌گذاری پسرو حل کنید.*

$$\begin{cases} x_1 + x_2 - x_3 + x_4 = 1 \\ x_1 + x_2 + x_3 - x_4 = 1 \\ x_2 + x_3 - x_4 = -1 \\ x_1 - x_2 - x_3 - x_4 = -1 \end{cases}$$

ماتریس افزوده عبارت است از

$$S = \begin{bmatrix} 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 \\ 0 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

و برای $j = 1$ داریم $r = 1$ زیرا $a_{11} = 1 \neq 0$. بنابراین

$$l_{21} = \frac{a_{21}}{a_{11}} = \frac{1}{1} = 1, \qquad R_2 \leftarrow R_2 - R_1$$

$$l_{41} = \frac{a_{41}}{a_{11}} = \frac{1}{1} = 1, \qquad R_4 \leftarrow R_4 - R_1$$

پس داریم

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

و برای $j = 2$ داریم $r = 3$ زیرا $a_{22} = 0 \neq 1 = a_{32}$ بنابراین با انجام $R_2 \leftrightarrow R_3$ داریم

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

و $l_{42} = \frac{a_{42}}{a_{22}} = \frac{-2}{1} = -2$ و با انجام $R_4 \leftarrow R_4 + 2R_2$ خواهیم داشت

$$S = \begin{bmatrix} 1 & 1 & -1 & 1 & 1 \\ 0 & 1 & 1 & -1 & -1 \\ 0 & 0 & 2 & -2 & 0 \\ 0 & 0 & 2 & -4 & -4 \end{bmatrix}.$$

در آخر برای $j=۳$ داریم $r=۳$ زیرا $a_{۳۳}=۲\neq ۰$. حال $l_{۴۳}=\frac{a_{۴۳}}{a_{۳۳}}=\frac{۲}{۲}=۱$ و با انجام عمل $R_۴\leftarrow R_۴-R_۳$ داریم

$$S=\begin{bmatrix} ۱ & ۱ & -۱ & ۱ & ۱ \\ ۰ & ۱ & ۱ & -۱ & -۱ \\ ۰ & ۰ & ۲ & -۲ & ۰ \\ ۰ & ۰ & ۰ & -۲ & -۴ \end{bmatrix}.$$

به کمک جای‌گذاری پسرو داریم

$$\begin{cases} x_۴=\frac{-۴}{-۲}=۲ \\ x_۳=\frac{۱}{۲}(۰+۲x_۴)=۲ \\ x_۲=\frac{۱}{۱}(-۱-x_۳+x_۴)=-۱ \\ x_۱=\frac{۱}{۱}(۱-x_۲+x_۳-x_۴)=۲ \end{cases}$$

$\triangle$

تذکر ۴.۵ *با شمارش اعمال ممیز شناور، ثابت می‌شود حجم عملیات الگوریتم حذف گاوسی و جای‌گذاری پسرو/پیشرو به ترتیب از مرتبه $O(n^۳)$ و $O(n^۲)$ است (n بعد ماتریس A است).*

با توجه به مثال‌هایی که در ادامه خواهند آمد، باید روش حذف گاوسی با جای‌گذاری پسرو با حساب ممیز شناور را با احتیاط کامل به کار برد.

مثال ۲.۵ *با این فرض که ماشین حسابی با دقت $۳S$ در اختیار است، می‌خواهیم به کمک روش حذف گاوسی با جای‌گذاری پسرو جواب دستگاه داده‌شده را به دست آوریم.*

$$\begin{cases} E_۱ : & ۰/۰۰۰۱۰۰x_۱ & + & ۱/۰۰x_۲ & = & ۱/۰۰ \\ E_۲ : & ۱/۰۰x_۱ & + & ۱/۰۰x_۲ & = & ۲/۰۰ \end{cases}$$

با عمل سطری مقدماتی $E_۲\leftarrow E_۲-\frac{۱/۰۰}{۰/۰۰۰۱۰۰}E_۱$ خواهیم داشت

$$\begin{cases} E_۱ : & ۰/۰۰۰۱۰۰x_۱ & + & ۱/۰۰x_۲ & = & ۱/۰۰ \\ E_۲ : & & & -۱۰۰۰۰x_۲ & = & -۱۰۰۰۰ \end{cases}$$

و به کمک جای‌گذاری پسرو داریم

$$\begin{cases} x_۲\simeq ۱/۰۰ \\ x_۱\simeq \frac{۱/۰۰-۱/۰۰\times ۱/۰۰}{۰/۰۰۰۱۰۰}=۰/۰۰ \end{cases}$$

که با مقایسه با جواب واقعی دستگاه با دقت $۴D$ یعنی $x_۱=۱/۰۰۰۱$ و $x_۲=۰/۹۹۹۹$، متوجه می‌شویم درصد خطای نسبی $x_۱$ زیاد است. علت آن است که در محاسبه $x_۲$ خطایی به اندازه $۰/۰۰۰۱$ مرتکب شده‌ایم و در محاسبه

x_1 این عدد در ضریب $\frac{1/00}{0/000100} = 10000$ ضرب می‌شود (تقسیم به عدد کوچک) و باعث می‌شود خطایی نزدیک به ۱ واحد در محاسبه x_1 تولید شود. اما چاره کار آن است که از محورگیری جزیی[11] استفاده کنیم. در این نوع محورگیری باید گام ۲ الگوریتم روش حذف گاوسی به صورت زیر تغییر یابد.

۲) rای $(j \le r \le n)$ پیدا کنید که $|a_{rj}| = \max_{j \le k \le n} |a_{kj}|$. اگر $a_{rj} = 0$ متوقف شوید مسئله جواب یکتا ندارد

پس در این مثال برای $j = 1$ خواهیم داشت $r = 2$ و با انجام $R_2 \leftrightarrow R_1$ داریم

$$\begin{cases} E_1 : & 1/00x_1 + 1/00x_2 = 2/00 \\ E_2 : & 0/000100x_1 + 1/00x_2 = 1/00 \end{cases}$$

سپس به کمک عمل سطری مقدماتی $E_2 \leftarrow E_2 - \frac{0/000100}{1/00}E_1$ خواهیم داشت

$$\begin{cases} E_1 : & 1/00x_1 + 1/00x_2 = 2/00 \\ E_2 : & 1/00x_2 = 1/00 \end{cases}$$

با جای‌گذاری پسرو به دست می‌آوریم $x_1 \simeq 1/00$ و $x_2 \simeq 1/00$. $\triangle$

مثال‌هایی وجود دارند که نشان می‌دهند محورگیری جزیی مشکل ناپایداری روش حذف گاوسی را برطرف نمی‌کند و باید از سایر روش‌ها مانند محورگیری جزیی مقیاس‌شده (محورگیری جزیی وزنی)[12] استفاده کرد. اما چون در عمل بیشتر مواقع مشکل ناپایداری روش حذف گاوسی با همان محورگیری جزیی برطرف می‌شود می‌توان روش حذف گاوسی با محورگیری جزیی را یک روش پایدار دانست.

۲.۱.۵ روش حذفی گاوس-جردن

قبل از آن که به معرفی این روش بپردازیم، توجه داریم که اگر $A = [a_{ij}]_{n\times n}$ ماتریسی قطری باشد یعنی داشته باشیم

$$A = \mathrm{diag}[a_{11}, \ldots, a_{nn}] = \begin{bmatrix} a_{11} & & \bigcirc \\ & \ddots & \\ \bigcirc & & a_{nn} \end{bmatrix}$$

[11] partial pivoting
[12] scaled partial pivoting

آن‌گاه از $Ax = b$ نتیجه می‌شود

$$\begin{cases} E_1 & : & a_{11}x_1 & & & = & b_1 \\ E_2 & : & & a_{22}x_2 & & = & b_2 \\ \vdots & & & \ddots & & & \vdots \\ E_n & : & & & a_{nn}x_n & = & b_n \end{cases}$$

و از آن‌جا خواهیم داشت

$$x_i = \frac{b_i}{a_{ii}}, \qquad i = 1, \dots, n.$$

به وضوح شرط لازم و کافی برای وجود جواب در این حالت آن است که درایه‌های قطری A مخالف صفر باشند. اما ایده روش حذفی گاوس–جردن[13] آن است که به کمک اَعمال سطری مقدماتی دستگاه $Ax = b$ را به دستگاه $Dx = c$ تبدیل کنیم که دو دستگاه هم‌ارز باشند (جواب یکسانی داشته باشند) و D یک ماتریس قطری باشد. مراحل این روش در الگوریتم زیر خلاصه شده است. **الگوریتم روش حذفی گاوس–جردن**

- **ورودی.** ماتریس افروده $S = [A\ b]_{n\times n+1}$
- **خروجی.** ماتریس افروده $S = [D\ c]_{n\times n+1}$ که در آن D ماتریسی قطری است

۱) قرار دهید $j = 1$

۲) rای $(j \le r \le n)$ پیدا کنید که $a_{rj} \neq 0$. اگر چنین rای یافت نشد متوقف شوید مسئله جواب یکتا ندارد

۳) اگر $r \neq j$ آن‌گاه انجام دهید $R_r \leftrightarrow R_j$

۴) برای $i = j+1, \dots, n$ قرار دهید $l_{ij} = \frac{a_{ij}}{a_{jj}}$ و انجام دهید $R_i \leftarrow R_i - l_{ij}R_j$

۵) قرار دهید $j = j + 1$ و اگر $j < n$ به گام ۲ بروید

۶) اگر $a_{jj} = 0$ متوقف شوید مسئله جواب یکتا ندارد

۷) برای $i = 1, \dots, j-1$ قرار دهید $l_{ij} = \frac{a_{ij}}{a_{jj}}$ و انجام دهید $R_i \leftarrow R_i - l_{ij}R_j$

۸) قرار دهید $j = j - 1$ و اگر $j > 1$ به گام ۶ بروید

مثال ۳.۵ *دستگاه داده‌شده را با روش حذفی گاوس–جردن حل کنید.*

$$\begin{cases} x_1 + x_2 - x_3 + x_4 = 1 \\ x_1 + x_2 + x_3 - x_4 = 1 \\ \phantom{x_1 +{}} x_2 + x_3 - x_4 = -1 \\ x_1 - x_2 - x_3 - x_4 = -1 \end{cases}$$

[13] Gauss-Jordan elimination method

ماتریس افزوده عبارت است از

$$S=\begin{bmatrix} 1 & 1 & -1 & 1 & 1\\ 1 & 1 & 1 & -1 & 1\\ 0 & 1 & 1 & -1 & -1\\ 1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

و برای $j=1$ داریم $r=1$ زیرا $a_{11}=1\neq 0$. بنابراین

$$l_{21}=\frac{a_{21}}{a_{11}}=\frac{1}{1}=1,\qquad R_2\leftarrow R_2-R_1$$
$$l_{31}=\frac{a_{31}}{a_{11}}=\frac{0}{1}=0,\qquad R_3\leftarrow R_3$$
$$l_{41}=\frac{a_{41}}{a_{11}}=\frac{1}{1}=1,\qquad R_4\leftarrow R_4-R_1$$

پس داریم

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

و برای $j=2$ داریم $r=3$ زیرا $a_{22}=0\neq 1=a_{32}$ و با انجام $R_2\leftrightarrow R_3$ داریم

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

و در نتیجه

$$l_{32}=\frac{a_{32}}{a_{22}}=\frac{0}{1}=0,\qquad R_3\leftarrow R_3$$
$$l_{42}=\frac{a_{42}}{a_{22}}=\frac{-2}{1}=-2,\qquad R_4\leftarrow R_4+2R_2$$

بنابراین

$$S=\begin{bmatrix} 1 & 1 & -1 & 1 & 1\\ 0 & 1 & 1 & -1 & -1\\ 0 & 0 & 2 & -2 & 0\\ 0 & 0 & 2 & -4 & -4 \end{bmatrix}.$$

در آخر برای $j=3$ داریم $r=3$ زیرا $a_{33}=2\neq 0$. پس

$$l_{43}=\frac{a_{43}}{a_{33}}=\frac{2}{2}=1,\qquad R_4\leftarrow R_4-R_3$$

و در نتیجه

$$S = \begin{bmatrix} ۱ & ۱ & -۱ & ۱ & ۱ \\ ۰ & ۱ & ۱ & -۱ & -۱ \\ ۰ & ۰ & ۲ & -۲ & ۰ \\ ۰ & ۰ & ۰ & -۲ & -۴ \end{bmatrix}.$$

حال چون $a_{۴۴} = -۲ \neq ۰$ بنابراین

$$l_{۳۴} = \frac{a_{۳۴}}{a_{۴۴}} = \frac{-۲}{-۲} = ۱, \qquad R_۳ \leftarrow R_۳ - R_۴$$

$$l_{۲۴} = \frac{a_{۲۴}}{a_{۴۴}} = \frac{-۱}{-۲} = \frac{۱}{۲}, \qquad R_۲ \leftarrow R_۲ - \frac{۱}{۲}R_۴$$

$$l_{۱۴} = \frac{a_{۱۴}}{a_{۴۴}} = \frac{۱}{-۲} = -\frac{۱}{۲}, \qquad R_۱ \leftarrow R_۱ + \frac{۱}{۲}R_۴$$

و داریم

$$S = \begin{bmatrix} ۱ & ۱ & -۱ & ۰ & -۱ \\ ۰ & ۱ & ۱ & ۰ & ۱ \\ ۰ & ۰ & ۲ & ۰ & ۴ \\ ۰ & ۰ & ۰ & -۲ & -۴ \end{bmatrix}.$$

سپس چون $a_{۳۳} = ۲ \neq ۰$ با انجام اعمال

$$l_{۲۳} = \frac{a_{۲۳}}{a_{۳۳}} = \frac{۱}{۲}, \qquad R_۲ \leftarrow R_۲ - \frac{۱}{۲}R_۳$$

$$l_{۱۳} = \frac{a_{۱۳}}{a_{۳۳}} = \frac{-۱}{۲} = -\frac{۱}{۲}, \qquad R_۱ \leftarrow R_۱ + \frac{۱}{۲}R_۳$$

خواهیم داشت

$$S = \begin{bmatrix} ۱ & ۱ & ۰ & ۰ & ۱ \\ ۰ & ۱ & ۰ & ۰ & -۱ \\ ۰ & ۰ & ۲ & ۰ & ۴ \\ ۰ & ۰ & ۰ & -۲ & -۴ \end{bmatrix}.$$

در آخر چون $a_{۲۲} = ۱ \neq ۰$ با انجام عمل

$$l_{۱۲} = \frac{a_{۱۲}}{a_{۲۲}} = \frac{۱}{۱} = ۱, \qquad R_۱ \leftarrow R_۱ - R_۲$$

داریم

$$S = \begin{bmatrix} ۱ & ۰ & ۰ & ۰ & ۲ \\ ۰ & ۱ & ۰ & ۰ & -۱ \\ ۰ & ۰ & ۲ & ۰ & ۴ \\ ۰ & ۰ & ۰ & -۲ & -۴ \end{bmatrix}$$

در پایان خواهیم داشت

$$x_۱ = \frac{۲}{۱} = ۲, \qquad x_۲ = \frac{-۱}{۱} = -۱, \qquad x_۳ = \frac{۴}{۲} = ۲, \qquad x_۴ = \frac{-۴}{-۲} = ۲.$$

$\triangle$

تذکر ۵.۵ *همانند روش حذف گاوسی، هنگام کار با حساب ممیز شناور باید در روش حذفی گاوس-جردن نیز از محورگیری مناسبی استفاده کرد.*

تذکر ۶.۵ *روش حذفی گاوس-جردن به جای‌گذاری پسرو/پیشرو نیازی ندارد و حجم عملیات آن دو برابر روش حذف گاوسی است و به همین دلیل در عمل برای حل دستگاه، کمتر از آن استفاده می‌شود و با توجه به فرآیند زیر بیشتر برای یافتن وارون ماتریس از آن استفاده می‌شود. اگر B وارون ماتریس $A = [a_{ij}]_{n\times n}$ باشد آن‌گاه از $AB = I$ خواهیم داشت*

$$A[B_۱ \; B_۲ \cdots B_n] = [e_۱ \; e_۲ \cdots e_n]$$

و از آن‌جا

$$AB_i = e_i, \qquad i = ۱, \ldots, n$$

و اگر این n دستگاه را با روش حذفی گاوس-جردن حل کنیم ستون‌های B مشخص می‌شوند. اما چون ماتریس ضرایب این n دستگاه یکسان است، در عمل به کمک اعمال سطری مقدماتی ماتریس افزوده $[A \; I]$ را در صورت امکان به ماتریس افزوده $[I \; B]$ تبدیل می‌کنیم.

مثال ۴.۵ *آیا ماتریس داده‌شده وارون‌پذیر است؟*

$$A = \begin{bmatrix} ۱ & ۲ & ۳ \\ ۲ & ۳ & ۴ \\ ۳ & ۴ & ۵ \end{bmatrix}$$

ابتدا ماتریس افزوده‌ای به صورت

$$\left[\begin{array}{ccc|ccc} ۱ & ۲ & ۳ & ۱ & ۰ & ۰ \\ ۲ & ۳ & ۴ & ۰ & ۱ & ۰ \\ ۳ & ۴ & ۵ & ۰ & ۰ & ۱ \end{array}\right]$$

ساخته و روش گاوس-جردن را دنبال می‌کنیم. با صفرسازی درایه‌های زیر قطر ستون اول داریم

$$\left[\begin{array}{ccc|ccc} ۱ & ۲ & ۳ & ۱ & ۰ & ۰ \\ ۰ & -۱ & -۲ & -۲ & ۱ & ۰ \\ ۰ & -۲ & -۴ & -۳ & ۰ & ۱ \end{array}\right]$$

سپس داریه‌های زیر قطر ستون دوم را صفر کرده، خواهیم داشت

$$\left[\begin{array}{ccc|ccc} ۱ & ۲ & ۳ & ۱ & ۰ & ۰ \\ ۰ & -۱ & -۲ & -۲ & ۱ & ۰ \\ ۰ & ۰ & ۰ & ۱ & -۲ & ۱ \end{array}\right]$$

چون ادامه کار امکان‌پذیر نیست، ماتریس داده شده وارون‌پذیر نیست. △

مثال ۵.۵ *وارون ماتریس داده‌شده را بیابید.*

$$A = \begin{bmatrix} ۱ & ۲ & -۱ \\ ۲ & ۱ & ۰ \\ -۱ & ۱ & ۲ \end{bmatrix}$$

ابتدا ماتریس افزوده‌ای به صورت

$$\left[\begin{array}{ccc|ccc} ۱ & ۲ & -۱ & ۱ & ۰ & ۰ \\ ۲ & ۱ & ۰ & ۰ & ۱ & ۰ \\ -۱ & ۱ & ۲ & ۰ & ۰ & ۱ \end{array}\right]$$

ساخته و داریه‌های زیر قطر ستون اول را صفر کرده، خواهیم داشت

$$\left[\begin{array}{ccc|ccc} ۱ & ۲ & -۱ & ۱ & ۰ & ۰ \\ ۰ & -۳ & ۲ & -۲ & ۱ & ۰ \\ ۰ & ۳ & ۱ & ۱ & ۰ & ۱ \end{array}\right]$$

سپس داریه‌های زیر قطر ستون دوم را صفر کرده، داریم

$$\left[\begin{array}{ccc|ccc} ۱ & ۲ & -۱ & ۱ & ۰ & ۰ \\ ۰ & -۳ & ۲ & -۲ & ۱ & ۰ \\ ۰ & ۰ & ۳ & -۱ & ۱ & ۱ \end{array}\right]$$

سطر سوم را به ۳ تقسیم کرده سپس داریه‌های بالای قطر ستون سوم را صفر کرده، خواهیم داشت

$$\left[\begin{array}{ccc|ccc} ۱ & ۲ & ۰ & \frac{۲}{۳} & \frac{۱}{۳} & \frac{۱}{۳} \\ ۰ & -۳ & ۰ & -\frac{۴}{۳} & \frac{۱}{۳} & -\frac{۲}{۳} \\ ۰ & ۰ & ۱ & -\frac{۱}{۳} & \frac{۱}{۳} & \frac{۱}{۳} \end{array}\right]$$

حال سطر دوم را به ۳− تقسیم کرده سپس درایه بالای قطر ستون دوم را صفر کرده، خواهیم داشت

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ 0 & 1 & 0 & \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ 0 & 0 & 1 & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right]$$

$\triangle$

۳.۱.۵ تجزیه مثلثی

تعریف ۳.۵ *فرض کنید ماتریس A را به صورت $A = LU$ تجزیه کرده باشیم. اگر L ماتریسی پایین‌مثلثی و U ماتریسی بالامثلثی باشد چنین تجزیه‌ای به تجزیه مثلثی[۱۴] معروف است.*

قضیه ۱.۵ *اگر در اِعمال روش حذف گاوسی بر روی دستگاه $Ax = b$ نیازی به جابجایی سطر نباشد آن‌گاه برای ماتریس A یک تجزیه مثلثی به صورت $A = LU$ موجود است که در آن*

$$L = \begin{bmatrix} 1 & & & \bigcirc \\ l_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ l_{n1} & \cdots & l_{n,n-1} & 1 \end{bmatrix}, \qquad U = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & \ddots & \vdots \\ \bigcirc & & & a_{nn}^{(n)} \end{bmatrix}.$$

تذکر ۷.۵ *اگر تجزیه مثلثی ماتریس ضرایب دستگاه $Ax = b$ به صورت $A = LU$ در دسترس باشد آن‌گاه داریم $LUx = b$. از حل دستگاه $Ly = b$ با جای‌گذاری پیشرو y را به دست آورده سپس x از حل دستگاه $Ux = y$ با جای‌گذاری پسرو تعیین می‌شود. همچنین* $\det A = \det L \det U = \det U$.

مثال ۶.۵ *با انجام اعمال* $E_2 \leftarrow E_2 - 2E_1$، $E_3 \leftarrow E_3 - 3E_1$، $E_4 \leftarrow E_4 + E_1$، $E_3 \leftarrow E_3 - 4E_2$ و $E_4 \leftarrow E_4 + 3E_2$ بر روی دستگاه

$$\begin{cases} E_1 : & x_1 + x_2 + 3x_4 = 4 \\ E_2 : & 2x_1 + x_2 - x_3 + x_4 = 1 \\ E_3 : & 3x_1 - x_2 - x_3 + 2x_4 = -3 \\ E_4 : & -x_1 + 2x_2 + 3x_3 - x_4 = 4 \end{cases}$$

[۱۴] triangular decomposition(factorization)

خواهیم داشت

$$\begin{cases} E_۱ : & x_۱ + x_۲ + ۳x_۴ = ۴ \\ E_۲ : & -x_۲ - x_۳ - ۵x_۴ = -۷ \\ E_۳ : & ۳x_۳ + ۱۳x_۴ = ۱۳ \\ E_۴ : & -۱۳x_۴ = -۱۳ \end{cases}$$

بنابراین

$$A = \begin{bmatrix} ۱ & ۱ & ۰ & ۳ \\ ۲ & ۱ & -۱ & ۱ \\ ۳ & -۱ & -۱ & ۲ \\ -۱ & ۲ & ۳ & -۱ \end{bmatrix} = \begin{bmatrix} ۱ & ۰ & ۰ & ۰ \\ ۲ & ۱ & ۰ & ۰ \\ ۳ & ۴ & ۱ & ۰ \\ -۱ & -۳ & ۰ & ۱ \end{bmatrix} \begin{bmatrix} ۱ & ۱ & ۰ & ۳ \\ ۰ & -۱ & -۱ & -۵ \\ ۰ & ۰ & ۳ & ۱۳ \\ ۰ & ۰ & ۰ & -۱۳ \end{bmatrix} = LU$$

بنابراین $\det(A) = \det(U) = ۱ \times -۱ \times ۳ \times -۱۳ = ۳۹$ و برای حل دستگاه

$$Ax = LUx = \begin{bmatrix} ۱ & ۰ & ۰ & ۰ \\ ۲ & ۱ & ۰ & ۰ \\ ۳ & ۴ & ۱ & ۰ \\ -۱ & -۳ & ۰ & ۱ \end{bmatrix} \begin{bmatrix} ۱ & ۱ & ۰ & ۳ \\ ۰ & -۱ & -۱ & -۵ \\ ۰ & ۰ & ۳ & ۱۳ \\ ۰ & ۰ & ۰ & -۱۳ \end{bmatrix} \begin{bmatrix} x_۱ \\ x_۲ \\ x_۳ \\ x_۴ \end{bmatrix} = \begin{bmatrix} ۴ \\ ۱ \\ -۳ \\ ۴ \end{bmatrix}$$

ابتدا دستگاه

$$Ly = \begin{bmatrix} ۱ & ۰ & ۰ & ۰ \\ ۲ & ۱ & ۰ & ۰ \\ ۳ & ۴ & ۱ & ۰ \\ -۱ & -۳ & ۰ & ۱ \end{bmatrix} \begin{bmatrix} y_۱ \\ y_۲ \\ y_۳ \\ y_۴ \end{bmatrix} = \begin{bmatrix} ۴ \\ ۱ \\ -۳ \\ ۴ \end{bmatrix}$$

را باجای‌گذاری پیشرو حل کرده داریم

$$y_۱ = ۴, \qquad y_۲ = ۱ - ۲y_۱ = -۷,$$
$$y_۳ = -۳ - ۳y_۱ - ۴y_۲ = ۱۳, \qquad y_۴ = ۴ + y_۱ + ۳y_۲ = -۱۳.$$

سپس از حل دستگاه

$$Ux = \begin{bmatrix} ۱ & ۱ & ۰ & ۳ \\ ۰ & -۱ & -۱ & -۵ \\ ۰ & ۰ & ۳ & ۱۳ \\ ۰ & ۰ & ۰ & -۱۳ \end{bmatrix} \begin{bmatrix} x_۱ \\ x_۲ \\ x_۳ \\ x_۴ \end{bmatrix} = \begin{bmatrix} ۴ \\ -۷ \\ ۱۳ \\ -۱۳ \end{bmatrix}$$

باجای‌گذاری پسرو خواهیم داشت

$$x_4 = 1, \qquad x_3 = 13 - 13x_4 = 0,$$
$$x_2 = 7 - x_3 - 5x_4 = 2, \qquad x_1 = 4 - x_2 - 3x_4 = -1.$$

$\triangle$

تعریف ۴.۵ *یک زیرماتریس اصلی[15] از ماتریس* $A = [a_{ij}]_{n\times n}$ *ماتریسی به صورت زیر است*

$$\begin{bmatrix} a_{i_1,i_1} & a_{i_1,i_2} & \cdots & a_{i_1,i_k} \\ a_{i_2,i_1} & a_{i_2,i_2} & \cdots & a_{i_2,i_k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_k,i_1} & a_{i_k,i_2} & \cdots & a_{i_k,i_k} \end{bmatrix}$$

که در آن $1 \le k \le n$ *و* $1 \le i_1 < i_2 < \cdots < i_k \le n$. *هم‌چنین یک زیرماتریس اصلی پیشرو[16] مرتبه* k $(1 \le k \le n)$ *از ماتریس* $A = [a_{ij}]_{n\times n}$ *ماتریسی به صورت زیر است*

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix}.$$

قضیه ۲.۵ (تجزیه مثلثی) *فرض کنید* $A = [a_{ij}]_{n\times n}$ *ماتریسی وارون‌پذیر باشد.* A *تجزیه‌ای یکتا به صورت* $A = LU$ *دارد که در آن* L *ماتریسی پایین مثلثی با درایه‌های قطری ۱ و* U *ماتریسی بالا مثلثی است اگر و فقط اگر تمام زیرماتریس‌های اصلی پیشرو* A *وارون‌پذیر باشند.*

نتیجه ۱.۲.۵ تحت شرایط قضیه ۲.۵ ماتریس A تجزیه‌ای یکتا به صورت

$$A = \begin{bmatrix} 1 & & & \bigcirc \\ l_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ l_{n1} & \cdots & l_{n,n-1} & 1 \end{bmatrix} \begin{bmatrix} d_{11} & & & \bigcirc \\ & d_{22} & & \\ & & \ddots & \\ \bigcirc & & & d_{nn} \end{bmatrix} \begin{bmatrix} 1 & u_{12} & \cdots & u_{1n} \\ & \ddots & \ddots & \vdots \\ & & \ddots & u_{n-1,n} \\ \bigcirc & & & 1 \end{bmatrix} = LDU$$

[15] principal submatrix
[16] leading principal submatrix

دارد و به علاوه اگر A ماتریسی متقارن باشد آن‌گاه

$$A=\begin{bmatrix} 1 & & & \bigcirc \\ l_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ l_{n1} & \cdots & l_{n,n-1} & 1 \end{bmatrix}\begin{bmatrix} d_{11} & & & \bigcirc \\ & d_{22} & & \\ & & \ddots & \\ \bigcirc & & & d_{nn} \end{bmatrix}\begin{bmatrix} 1 & l_{21} & \cdots & l_{n1} \\ & \ddots & \ddots & \vdots \\ & & \ddots & l_{n,n-1} \\ \bigcirc & & & 1 \end{bmatrix}=LDL^T.$$

در ادامه قصد داریم مسئله تجزیه مثلثی را به طور مستقیم مورد بررسی قرار دهیم.

فرض کنید ماتریس $A=[a_{ij}]_{n\times n}$ داده شده باشد و بخواهیم ماتریس پایین‌مثلثی $L=[l_{ij}]_{n\times n}$ و ماتریس بالامثلثی $U=[u_{ij}]_{n\times n}$ را چنان بیابیم که $A=LU$. بنابراین باید داشته باشیم

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}=\begin{bmatrix} l_{11} & & \bigcirc \\ \vdots & \ddots & \\ l_{n1} & \cdots & l_{nn} \end{bmatrix}\begin{bmatrix} u_{11} & \cdots & u_{1n} \\ & \ddots & \vdots \\ \bigcirc & & u_{nn} \end{bmatrix}$$

که از آن‌جا n^2 معادله به شکل

$$a_{ij}=\sum_{p=1}^{\min\{i,j\}} l_{ip}u_{pj}, \qquad i,j=1,\ldots,n$$

به دست می‌آید که در آن n^2+n مجهول (درایه‌های L و U) ظاهر شده است. با قرار دادن $l_{ii}=1$ برای $i=1,\ldots,n$، تعداد مجهولات به n^2 مجهول کاهش یافته، خواهیم داشت

$$\begin{cases} u_{ij}=a_{ij}-\sum_{p=1}^{i-1} l_{ip}u_{pj}, & 1\le i\le j\le n \\ l_{ij}=\frac{1}{u_{jj}}\left(a_{ij}-\sum_{p=1}^{j-1} l_{ip}u_{pj}\right), & 1\le j<i\le n \end{cases} \tag{1.5}$$

دولیتل[17]، طرز استفاده از (۱.۵) را به طور ماهرانه‌ای در الگوریتم زیر خلاصه کرده است. **الگوریتم تجزیه مثلثی دولیتل**

- **ورودی.** ماتریس $A=[a_{ij}]_{n\times n}$
- **خروجی.** ماتریس بالامثلثی $U=[u_{ij}]_{n\times n}$ و ماتریس پایین‌مثلثی $L=[l_{ij}]_{n\times n}$ با درایه‌های قطری ۱ به طوری که $A=LU$

۱) برای $i=1,\ldots,n$ قرار دهید $l_{ii}=1$

۲) برای $k=1,\ldots,n$ گام‌های ۳ و ۴ را تکرار کنید

[17] Doolittle

۳) برای $j=k,\ldots,n$ قرار دهید $u_{kj}=a_{kj}-\sum_{p=1}^{k-1} l_{kp}u_{pj}$

۴) برای $i=k+1,\ldots,n$ قرار دهید $l_{ik}=\frac{1}{u_{kk}}\left(a_{ik}-\sum_{p=1}^{k-1} l_{ip}u_{pk}\right)$

در این الگوریتم به ازای هر k ابتدا عناصر $u_{kk},\ldots,u_{kn}$ محاسبه شده و سپس عناصر $l_{k+1,k},\ldots,l_{nk}$ به دست می‌آیند. کروت[18] به طریقی مشابه اما متفاوت، نحوه استفاده از (۱.۵) را به این صورت تغییر داد که ابتدا عناصر $u_{1k},\ldots,u_{kk}$ محاسبه شده و سپس عناصر $l_{k+1,k},\ldots,l_{nk}$ به دست می‌آیند. این کار هنرمندانه در الگوریتم زیر خلاصه شده است.

الگوریتم تجزیه مثلثی کروت

- **ورودی.** ماتریس $A=[a_{ij}]_{n\times n}$
- **خروجی.** ماتریس بالامثلثی $U=[u_{ij}]_{n\times n}$ و ماتریس پایین‌مثلثی $L=[l_{ij}]_{n\times n}$ با درایه‌های قطری ۱ به طوری که $A=LU$

۱) برای $i=1,\ldots,n$ قرار دهید $l_{ii}=1$

۲) برای $j=1,\ldots,n$ گام‌های ۳ و ۴ را تکرار کنید

۳) برای $i=1,\ldots,j$ قرار دهید $u_{ij}=a_{ij}-\sum_{p=1}^{i-1} l_{ip}u_{pj}$

۴) برای $i=j+1,\ldots,n$ قرار دهید $l_{ij}=\frac{1}{u_{jj}}\left(a_{ij}-\sum_{p=1}^{j-1} l_{ip}u_{pj}\right)$

مثال ۷.۵ *با دنبال کردن مراحل هر یک از دو الگوریتم، به راحتی تجزیه مثلثی ماتریس*

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

به صورت زیر به دست می‌آید.

$$A=\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & -2 & 4 & 0 \\ 1 & 2 & 5 & -1 \\ 1 & -1 & 0 & 1 \end{bmatrix}=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & 1 & -4 \\ 0 & 0 & 2 & -5 \\ 0 & 0 & 0 & -9 \end{bmatrix}=LU$$

△

[18] Crout

تذکر ۸.۵ *هنگام کار با حساب ممیز شناور باید از محورگیری مناسب استفاده کرد. همچنین به سادگی می‌توان نشان داد حجم عملیات تجزیه مثلثی دولیتل (کروت) $O(n^3)$ است.*

تعریف ۵.۵ *ماتریس $A=[a_{ij}]_{n\times n}$ را غالب قطری سطری*[19] *گویند هرگاه*

$$|a_{ii}| \geq \sum_{i\neq j=1}^{n} |a_{ij}|, \qquad i=1,\dots,n$$

و اگر برابری برقرار نباشد به آن غالب قطری سطری اکید[20] *گفته می‌شود. از این به بعد منظور از غالب قطری (اکید) همان غالب قطری سطری (اکید) است.*

مثال ۸.۵ *A یک ماتریس غالب قطری اکید و B یک ماتریس غالب قطری است.*

$$A=\begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & -5 & 1 & -2 \\ -1 & 0 & 6 & 3 \\ 1 & -3 & -4 & 10 \end{bmatrix}, \qquad B=\begin{bmatrix} 2 & 1 & -1 & 0 \\ 3 & -5 & -1 & 0 \\ 0 & 0 & 2 & -1 \\ 1 & -2 & 0 & 4 \end{bmatrix}.$$

$\triangle$

قضیه ۳.۵ *فرض کنید $A=[a_{ij}]_{n\times n}$ یک ماتریس غالب قطری اکید باشد. A وارون‌پذیر است، روش حذف گاوسی را می‌توان روی هر دستگاه خطی با ماتریس ضرایب A بدون نیاز به جابجایی سطر اِعمال نمود و محاسبات نسبت به رشد خطای گرد کردن پایدار است.*

۲.۵ روش‌های تکراری

روش‌های تکراری بر خلاف روش‌های مستقیم روندی نامتناهی دارند و حتی اگر خطای گرد کردن هم وجود نداشته باشد، جواب تقریبی تولید می‌کنند. هم‌گرایی و حجم عملیات دو معیار برای مقایسه این روش‌ها است. اما برای تعریف هم‌گرایی به مفاهیمی چون نرم[21]، فاصله[22] و غیره نیاز است.

۱.۲.۵ نرم برداری و ماتریسی

تعریف ۶.۵ *یک نرم برداری روی $\mathbb{R}^n$ تابعی مانند $\|\cdot\|$ از $\mathbb{R}^n$ به توی $\mathbb{R}$ با خواص زیر است*

[19] row diagonally dominant
[20] strictly row diagonally dominant
[21] norm
[22] distance

آ ـ *به ازای هر x در $\mathbb{R}^n$ داشته باشیم* $\|x\| \geq 0$

ب ـ $\|x\| = 0$ *اگر و فقط اگر* $x = 0$

پ ـ *به ازای هر x در $\mathbb{R}^n$ و به ازای هر α در $\mathbb{R}$ داشته باشیم* $\|\alpha x\| = |\alpha| \|x\|$

ت ـ *(نابرابری مثلثی) به ازای هر x و y در $\mathbb{R}^n$ داشته باشیم* $\|x+y\| \leq \|x\| + \|y\|$

و اگر از $\|x\| = 0$ نتیجه نشود $x = 0$ آن‌گاه $\|\cdot\|$ را نیم‌ـنرم[۲۳] *نامند. پس از تعریف نرم، بلافاصله می‌توان فاصله بین دو بردار x و y در $\mathbb{R}^n$ را به صورت $\|x-y\|$ تعریف کرد.*

مثال ۹.۵ *فرض کنید $x = [x_i]_{n\times 1}$ برداری در $\mathbb{R}^n$ باشد. توابع*

$$\begin{cases} \|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}, & 1 \leq p < \infty \\ \|x\|_\infty = \max_{1\leq i\leq n} |x_i| \end{cases}$$

نرم‌های برداری متداولی روی $\mathbb{R}^n$ هستند. $\|\cdot\|_p$ به نرم p، $\|\cdot\|_2$ به نرم اقلیدسی[۲۴] *و $\|\cdot\|_\infty$ به نرم ماکزیمم (نرم یکنواخت)*[۲۵] *معروف هستند.* △

تعریف ۷.۵ *فرض کنید $\{x^{(k)}\}_{k=1}^{\infty}$ دنباله‌ای از بردارهای $\mathbb{R}^n$ و x برداری در $\mathbb{R}^n$ باشد. x را حد دنباله $\{x^{(k)}\}_{k=1}^{\infty}$ نامیده و می‌نویسیم $\lim_{k\to\infty} x^{(k)} = x$ هرگاه*

$$\lim_{k\to\infty} x_i^{(k)} = x_i, \qquad i = 1, \dots, n$$

که در آن $x = [x_i]_{n\times 1}$ و $x^{(k)} = [x_i^{(k)}]_{n\times 1}$. با توجه به ارتباط تناتنگ حد برداری و حد اسکالری، بررسی برقراری خواصی نظیر منحصر به فرد بودن حد، خاصیت خطی داشتن حد و غیره چندان سخت نیست.

قضیه ۴.۵ *با این فرض که $\lim_{k\to\infty} x^{(k)} = x$ و $\lim_{k\to\infty} y^{(k)} = y$ به ازای هر دو اسکالر دل‌خواه α و β داریم*

$$\lim_{k\to\infty} (\alpha x^{(k)} + \beta y^{(k)}) = \alpha x + \beta y.$$

شایان توجه است که بسیاری از خواص حد در $\mathbb{R}$ را می‌توان به $\mathbb{R}^n$ تعمیم داد (مشابه قضیه اخیر) ولی بعضی از احکام مانند قضیه فشار (ساندویچ) قابل تعمیم نیستند و در این راستا تعریف زیر راه‌گشا است.

تعریف ۸.۵ *دنباله $\{x^{(k)}\}_{k=1}^{\infty}$ از بردارهای در $\mathbb{R}^n$ هم‌گرا به بردار x در $\mathbb{R}^n$ نسبت به نرم برداری $\|\cdot\|$ گفته می‌شود و می‌نویسیم*

$$\lim_{k\to\infty} \left\| x^{(k)} - x \right\| = 0$$

[۲۳] semi-norm

[۲۴] Euclidean norm

[۲۵] maximum norm (uniform norm)

هرگاه به ازای هر $\epsilon > 0$ داده‌شده عدد صحیح N_ϵ چنان موجود باشد که

$$\left\|x^{(k)} - x\right\| < \epsilon, \qquad \forall k \geq N_\epsilon.$$

قضیه ۵.۵ فرض کنید $\{x^{(k)}\}_{k=1}^{\infty}$ دنباله‌ای از بردارهای $\mathbb{R}^n$ و x برداری در $\mathbb{R}^n$ باشد.

$$\lim_{k\to\infty} \left\|x^{(k)} - x\right\| = 0 \quad \Leftrightarrow \quad \lim_{k\to\infty} x^{(k)} = x$$

تعریف ۹.۵ یک نرم ماتریسی تابعی مانند $\|\cdot\|$ از $\mathbb{R}^{n\times n}$ به توی $\mathbb{R}$ با خواص زیر است

آ- به ازای هر A در $\mathbb{R}^{n\times n}$ داشته باشیم $\|A\| \geq 0$

ب- $\|A\| = 0$ اگر و فقط اگر $A = 0$

پ- به ازای هر A در $\mathbb{R}^{n\times n}$ و به ازای هر α در $\mathbb{R}$ داشته باشیم $\|\alpha A\| = |\alpha|\,\|A\|$

ت- (نابرابری مثلثی) به ازای هر A و B در $\mathbb{R}^{n\times n}$ داشته باشیم $\|A+B\| \leq \|A\| + \|B\|$

مثال ۱۰.۵ فرض کنید $A = [a_{ij}]$ ماتریس دل‌خواهی در $\mathbb{R}^{n\times n}$ باشد. توابع

$$\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|, \qquad \|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|, \qquad \|A\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2}$$

نرم‌های ماتریسی متداولی روی $\mathbb{R}^{n\times n}$ هستند. $\|\cdot\|_F$ به نرم فروبنیوس[۲۶] معروف است. $\triangle$

تعریف ۱۰.۵ عدد λ یک مقدار ویژه[۲۷] برای ماتریس A نامیده می‌شود هرگاه بردار مخالف صفر x چنان وجود داشته باشد که $Ax = \lambda x$. در این حالت x بردار ویژه[۲۸] نظیر λ نامیده می‌شود. مجموعه‌ای که شامل تمام مقدارهای ویژه ماتریس A باشد به طیف[۲۹] A معروف است و با $\sigma(A)$ نمایش داده می‌شود. شعاع طیفی[۳۰] ماتریس A که با $\rho(A)$ نمایش داده می‌شود عبارت است از $\rho(A) = \max_{\lambda\in\sigma(A)} |\lambda|$.

۲.۲.۵ روش‌های مبتنی بر تفکیک ماتریسی

در یک روش تکراری برای حل دستگاه معادلات خطی $Ax = b$، ابتدا دستگاهی معادل به صورت $x = Tx + c$ ساخته شده و سپس با انتخاب بردار آغازی $x^{(0)}$ دنباله‌ای از بردارها، از طرح تکراری $x^{(k)} = Tx^{(k-1)} + c$ تولید می‌شود. در واقع با یک ماتریس T و یک بردار c، می‌توان یک روش تکراری ساخت. بنابراین تفاوت روش‌های تکراری در ماتریس

[۲۶] Frobenius norm
[۲۷] eigenvalue
[۲۸] eigenvector
[۲۹] spectrum
[۳۰] spectral radius

T و بردار c است. یک روش مهم برای تبدیل دستگاه $Ax=b$ به دستگاه $x=Tx+c$، استفاده از فن تفکیک (شکافت) ماتریس[۳۱] A است. برای این کار ماتریس

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

را به صورت $A=D-L-U$ تفکیک می‌کنیم که در آن $D=\operatorname{diag}[a_{11},\dots,a_{nn}]$ و

$$L=\begin{bmatrix} 0 & & & \bigcirc \\ -a_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ -a_{n1} & \cdots & -a_{n,n-1} & 0 \end{bmatrix}, \qquad U=\begin{bmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ & 0 & \ddots & \vdots \\ & & \ddots & -a_{n-1,n} \\ \bigcirc & & & 0 \end{bmatrix}.$$

بنابراین $Ax=b$ معادل است با $(D-L-U)x=b$ و از آن‌جا خواهیم داشت

$$\begin{cases} Dx=(L+U)x+b \\ (D-L)x=Ux+b \end{cases} \Longrightarrow \begin{cases} x=D^{-1}(L+U)x+D^{-1}b \\ x=(D-L)^{-1}Ux+(D-L)^{-1}b \end{cases}$$

بنابراین می‌توان دو طرح تکراری به صورت زیر ساخت

$$\begin{cases} x^{(k)}=T_j\,x^{(k-1)}+c_j \\ x^{(k)}=T_g\,x^{(k-1)}+c_g \end{cases} \qquad k=1,2,\dots$$

که در طرح تکراری اول که به روش تکراری ژاکوبی[۳۲] معروف است $T_j=D^{-1}(L+U)$ و $c_j=D^{-1}b$ و در طرح تکراری دوم که به روش تکراری گاوس–سیدل[۳۳] معروف است $T_g=(D-L)^{-1}U$ و $c_g=(D-L)^{-1}b$. شکل جبری این روش‌ها به صورت زیر است.

$$\begin{cases} x_i^{(k)}=\dfrac{b_i-\sum_{i\neq j=1}^{n}a_{ij}x_j^{(k-1)}}{a_{ii}} \\ x_i^{(k)}=\dfrac{b_i-\sum_{j=1}^{i-1}a_{ij}x_j^{(k)}-\sum_{j=i+1}^{n}a_{ij}x_j^{(k-1)}}{a_{ii}} \end{cases} \qquad i=1,2,\dots,n \quad k=1,2,\dots$$

تذکر ۹.۵ *برای آن‌که این روش‌ها خوش تعریف باشند باید برای $i=1,\dots,n$ داشته باشیم $a_{ii}\neq 0$.*

---

[۳۱] matrix spliting
[۳۲] Jacobi iterative method
[۳۳] Gauss-Seidel iterative method

k	۱	۲	۳	۴	۵
$x_{۱}^{(k)}$	۰٫۶۰۰۰	۱٫۰۴۷۳	۰٫۹۳۲۶	۱٫۰۱۵۲	۰٫۹۸۹۰
$x_{۲}^{(k)}$	۲٫۲۷۲۷	۱٫۷۱۵۹	۲٫۰۵۳۰	۱٫۹۵۳۷	۲٫۰۱۱۴
$x_{۳}^{(k)}$	−۱٫۱۰۰۰	−۰٫۸۰۵۲	−۱٫۰۴۹۳	−۰٫۹۶۸۱	−۱٫۰۱۰۳
$x_{۴}^{(k)}$	۱٫۸۷۵۰	۰٫۸۸۵۲	۱٫۱۳۰۹	۰٫۹۷۳۹	۱٫۰۲۱۴

جدول ۱.۵: تکرارهای روش ژاکوبی

تذکر ۱۰.۵ *یکی از شرایط زیر را می‌توان به عنوان شرط توقف روش‌های تکراری برگزید.*

$$\left\|x^{(k)} - x^{(k-۱)}\right\| < \epsilon, \quad \frac{\left\|x^{(k)} - x^{(k-۱)}\right\|}{\left\|x^{(k)}\right\|} < \epsilon, \quad \left\|r^{(k)} = Ax^{(k)} - b\right\| < \epsilon, \quad k \leq K$$

مثال ۱۱.۵ *ابتدا دستگاه*

$$\begin{cases} ۱۰x_{۱} - x_{۲} + ۲x_{۳} = ۶ \\ -x_{۱} + ۱۱x_{۲} - x_{۳} + ۳x_{۴} = ۲۵ \\ ۲x_{۱} - x_{۲} + ۱۰x_{۳} - x_{۴} = -۱۱ \\ ۳x_{۲} - x_{۳} + ۸x_{۴} = ۱۵ \end{cases}$$

را به صورت زیر بازنویسی می‌کنیم.

$$\begin{cases} x_{۱} = \frac{۳}{۵} + \frac{۱}{۱۰}x_{۲} - \frac{۱}{۵}x_{۳} \\ x_{۲} = \frac{۲۵}{۱۱} + \frac{۱}{۱۱}x_{۱} + \frac{۱}{۱۱}x_{۳} - \frac{۳}{۱۱}x_{۴} \\ x_{۳} = -\frac{۱۱}{۱۰} - \frac{۱}{۵}x_{۱} + \frac{۱}{۱۰}x_{۲} + \frac{۱}{۱۰}x_{۴} \\ x_{۴} = \frac{۱۵}{۸} - \frac{۳}{۸}x_{۲} + \frac{۱}{۸}x_{۳} \end{cases}$$

اگر بخواهیم روش تکراری ژاکوبی را به کار ببریم به صورت زیر اندیس‌گذاری می‌کنیم.

$$\begin{cases} x_{۱}^{(k)} = \frac{۳}{۵} + \frac{۱}{۱۰}x_{۲}^{(k-۱)} - \frac{۱}{۵}x_{۳}^{(k-۱)} \\ x_{۲}^{(k)} = \frac{۲۵}{۱۱} + \frac{۱}{۱۱}x_{۱}^{(k-۱)} + \frac{۱}{۱۱}x_{۳}^{(k-۱)} - \frac{۳}{۱۱}x_{۴}^{(k-۱)} \\ x_{۳}^{(k)} = -\frac{۱۱}{۱۰} - \frac{۱}{۵}x_{۱}^{(k-۱)} + \frac{۱}{۱۰}x_{۲}^{(k-۱)} + \frac{۱}{۱۰}x_{۴}^{(k-۱)} \\ x_{۴}^{(k)} = \frac{۱۵}{۸} - \frac{۳}{۸}x_{۲}^{(k-۱)} + \frac{۱}{۸}x_{۳}^{(k-۱)} \end{cases}$$

به ازای بردار آغازی $x^{(۰)} = [۰ \ ۰ \ ۰ \ ۰]^T$ پنج تکرار از این روش در جدول ۱.۵ آمده است. اگر بخواهیم روش تکراری گاوس-سیدل را به کار ببریم به صورت زیر اندیس‌گذاری می‌کنیم.

$$\begin{cases} x_1^{(k)} = \frac{3}{5} + \frac{1}{10}x_2^{(k-1)} - \frac{1}{5}x_3^{(k-1)} \\ x_2^{(k)} = \frac{25}{11} + \frac{1}{11}x_1^{(k)} + \frac{1}{11}x_3^{(k-1)} - \frac{3}{11}x_4^{(k-1)} \\ x_3^{(k)} = -\frac{11}{10} - \frac{1}{5}x_1^{(k)} + \frac{1}{10}x_2^{(k)} + \frac{1}{10}x_4^{(k-1)} \\ x_4^{(k)} = \frac{15}{8} - \frac{3}{8}x_2^{(k)} + \frac{1}{8}x_3^{(k)} \end{cases}$$

به ازای بردار آغازی $x^{(0)} = [0 \ 0 \ 0 \ 0]^T$ پنج تکرار از این روش در جدول ۲.۵ آمده است. △

k	۱	۲	۳	۴	۵
$x_1^{(k)}$	۰/۶۰۰۰	۱/۰۳۰۰	۱/۰۰۶۵	۱/۰۰۰۹	۱/۰۰۱۰
$x_2^{(k)}$	۲/۳۲۷۲	۲/۰۳۷۰	۲/۰۰۳۶	۲/۰۰۰۳	۲/۰۰۰۰
$x_3^{(k)}$	−۰/۹۸۷۳	−۱/۰۱۴۰	−۱/۰۰۲۵	−۱/۰۰۰۳	−۱/۰۰۰۰
$x_4^{(k)}$	۰/۸۷۸۹	۰/۹۸۴۴	۰/۹۹۸۳	۰/۹۹۹۹	۱/۰۰۰۰

جدول ۲.۵: تکرارهای روش گاوس-سیدل

مثال ۱۲.۵ *نتایج مثال قبل دلالت بر آن دارد که روش تکراری گاوس-سیدل بهتر از روش تکراری ژاکوبی است. گرچه این مطلب در بیشتر حالات درست است ولی روش گاوس-سیدل در حل دستگاه زیر شکست می‌خورد (با بردار آغازی $x^{(0)} = [0 \ 0 \ 0]^T$ تا ۲۵ تکرار هم تقریب خوبی تولید نمی‌کند) حال آن‌که روش ژاکوبی خیلی زود هم‌گرا می‌شود (بررسی کنید)*

$$\begin{cases} x_1 + 2x_2 - 2x_3 = 7 \\ x_1 + x_2 + x_3 = 2 \\ 2x_1 + 2x_2 + x_3 = 5 \end{cases}$$

△

قضیه ۶.۵ *برای هر بردار $x^{(0)}$ در $\mathbb{R}^n$ دنباله $\{x^{(k)}\}_{k=1}^{\infty}$ تولیدشده از طرح تکراری*

$$x^{(k)} = Tx^{(k-1)} + c, \qquad k = 1, 2, \ldots$$

به جواب یکتای معادله $x = Tx + c$ هم‌گرا است اگر و فقط اگر $\rho(T) < 1$.

قضیه ۷.۵ *اگر برای یک نرم ماتریسی طبیعی داشته باشیم $\|T\| < 1$ و c بردار دل‌خواهی در $\mathbb{R}^n$ باشد آن‌گاه دنباله $\{x^{(k)}\}_{k=1}^{\infty}$ تولیدشده توسط*

$$x^{(k)} = Tx^{(k-1)} + c, \qquad k = 1, 2, \ldots$$

به ازای هر $x^{(0)}$ در $\mathbb{R}^n$ به برداری مثل x در $\mathbb{R}^n$ هم‌گرا است و به علاوه کران‌های زیر برقرارند.

$$\begin{cases} \|x - x^{(k)}\| \le \|T\|^k \|x - x^{(0)}\| \\ \|x - x^{(k)}\| \le \frac{\|T\|^k}{1-\|T\|} \|x^{(1)} - x^{(0)}\| \end{cases}$$

قضیه ۸.۵ *اگر A ماتریسی غالب قطری اکید در $\mathbb{R}^{n\times n}$ باشد، آن‌گاه روش‌های ژاکوبی و گاوس-سیدل به ازای هر بردار آغازی $x^{(۰)}$ در $\mathbb{R}^n$ هم‌گرا هستند.*

تذکر ۱۱.۵ *با توجه به قضیه قبل به نظر می‌رسد بهتر است برای حل دستگاهی مانند*

$$\begin{cases} x_۱ + ۴x_۲ - ۲x_۳ = ۱ \\ x_۱ + x_۲ + ۴x_۳ = -۲ \\ -۵x_۱ + ۲x_۲ + x_۳ = ۳ \end{cases}$$

ابتدا آن را به صورت

$$\begin{cases} -۵x_۱ + ۲x_۲ + x_۳ = ۳ \\ x_۱ + ۴x_۲ - ۲x_۳ = ۱ \\ x_۱ + x_۲ + ۴x_۳ = -۲ \end{cases}$$

بازنویسی کرده و سپس به کمک روش‌های ژاکوبی یا گاوس-سیدل آن را حل کنیم.

۳.۵ آنالیز خطا

به نظر می‌رسد که اگر $\tilde{x}$ جواب تقریبی دستگاه $Ax=b$ باشد و نرم بردار باقی‌مانده نظیر یعنی $\|r\| = \|b - A\tilde{x}\|$ کوچک باشد آن‌گاه $\|x-\tilde{x}\|$ هم کوچک باشد ولی دستگاه‌هایی وجود دارند که این خاصیت برای آن‌ها برقرار نیست.

مثال ۱۳.۵ دستگاه

$$\begin{bmatrix} ۱ & ۲ \\ ۱/۰۰۰۱ & ۲ \end{bmatrix} \begin{bmatrix} x_۱ \\ x_۲ \end{bmatrix} = \begin{bmatrix} ۳ \\ ۳/۰۰۰۱ \end{bmatrix}$$

دارای جواب واقعی $x = [۱\ ۱]^T$ است و بردار باقی‌مانده نظیر تقریب ضعیف $\tilde{x} = [۳\ ۰]^T$ به قرار زیر است

$$r = b - A\tilde{x} = \begin{bmatrix} ۳ \\ ۳/۰۰۰۱ \end{bmatrix} - \begin{bmatrix} ۱ & ۲ \\ ۱/۰۰۰۱ & ۲ \end{bmatrix} \begin{bmatrix} ۳ \\ ۰ \end{bmatrix} = \begin{bmatrix} ۰ \\ -۰/۰۰۰۲ \end{bmatrix}.$$

کوچک بودن نرم بردار باقی‌مانده یعنی *$\|r\|_\infty = ۰/۰۰۰۲$ دلیلی بر خوب بودن تقریب نیست زیرا $\|x-\tilde{x}\|_\infty = ۲$.* $\triangle$

قضیه ۹.۵ *در نظر بگیرید که A ماتریسی وارون‌پذیر، $\tilde{x}$ جواب تقریبی دستگاه $Ax=b$ و r بردار باقی‌مانده نظیر $\tilde{x}$ باشد. آن‌گاه به ازای هر نرم طبیعی داریم*

$$\|x-\tilde{x}\| \le \|r\| \left\|A^{-۱}\right\|$$

و اگر $x \neq ۰ \neq b$ آن‌گاه

$$\frac{۱}{\|A\|\,\|A^{-۱}\|}\frac{\|r\|}{\|b\|} \le \frac{\|x-\tilde{x}\|}{\|x\|} \le \|A\| \left\|A^{-۱}\right\| \frac{\|r\|}{\|b\|}.$$

تعریف ۱۱.۵ عدد وضعیت[۳۴] *ماتریس وارون‌پذیر A نسبت به نرم $\|\cdot\|$ به صورت زیر تعریف می‌شود*

$$\kappa(A) = \|A\| \left\|A^{-1}\right\|$$

و اگر A ماتریسی وارون‌پذیر نباشد قرار می‌دهیم $\kappa(A) = \infty$.

تذکر ۱۲.۵ *برای هر ماتریس وارون‌پذیر دل‌خواه A و هر نرم ماتریسی طبیعی داریم $\kappa(A) \geq 1$ زیرا*

$$1 = \|I\| = \left\|AA^{-1}\right\| \leq \|A\| \left\|A^{-1}\right\| = \kappa(A).$$

تذکر ۱۳.۵ *با توجه به تعریف عدد وضعیت، نابرابری‌های قضیه قبل به صورت زیر بازنویسی می‌شوند*

$$\|x - \tilde{x}\| \leq \kappa(A)\frac{\|r\|}{\|A\|}$$

و اگر $x \neq 0 \neq b$ آن‌گاه

$$\frac{1}{\kappa(A)}\frac{\|r\|}{\|b\|} \leq \frac{\|x-\tilde{x}\|}{\|x\|} \leq \kappa(A)\frac{\|r\|}{\|b\|}.$$

بنابراین مسئله $Ax = b$ خوش‌وضع است هرگاه $\kappa(A)$ به ۱ نزدیک باشد و هر چه $\kappa(A)$ از ۱ دورتر باشد مسئله بدوضع‌تر خواهد شد.

مثال ۱۴.۵ *در مثال قبل داریم*

$$A = \begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix}$$

که برای آن $\|A\|_\infty = 3.0001$ چندان بزرگ نیست ولی برای

$$A^{-1} = \begin{bmatrix} -10000 & 10000 \\ 5000.5 & -5000 \end{bmatrix}$$

داریم $\|A^{-1}\|_\infty = 20000$ و از آن‌جا $\kappa(A) = 3.0001 \times 20000 = 60002$. $\triangle$

این بخش را با قضیه‌ای در رابطه با اثر اختلالات جزیی داده‌های ورودی، روی جواب دستگاه $Ax = b$ به پایان می‌رسانیم.

قضیه ۱۰.۵ *فرض کنید A ماتریسی وارون‌پذیر، $b \neq 0$ و $x \neq 0$ جواب دستگاه $Ax = b$ باشد. اگر بردار سمت راست و ماتریس ضرایب را به صورت $b + \delta b$ و $A + \delta A$ مختل کرده باشیم و $\tilde{x}$ جواب نظیر باشد یعنی $(A + \delta A)\tilde{x} = b + \delta b$ و*

$$\|\delta A\| < \frac{1}{\|A^{-1}\|}$$

[۳۴] condition number

آن‌گاه

$$\frac{\|x-\tilde{x}\|}{\|x\|} \leq \frac{\kappa(A)\,\|A\|}{\|A\|-\kappa(A)\,\|\delta A\|}\left(\frac{\|\delta b\|}{\|b\|}+\frac{\|\delta A\|}{\|A\|}\right).$$

تمرین

---

۱. فرض کنید ماشین حسابی با دقت $۴S$ در اختیار است. به کمک روش حذف گاوسی با جای‌گذاری پسرو دستگاه داده‌شده را یک بار بدون محورگیری جزیی و یک بار با محورگیری جزیی حل کنید.

$$\begin{cases} E_۱ : & ۰/۰۰۳۰۰۰x_۱ + ۵۹/۱۴x_۲ = ۵۹/۱۷ \\ E_۲ : & ۵/۲۹۱x_۱ - ۶/۱۳۰x_۲ = ۴۶/۷۸ \end{cases}$$

۲. با این فرض که یک ماشین حساب با دقت $۳S$ در اختیار داریم، دستگاه داده‌شده را یک بار با محورگیری جزیی و یک بار بدون آن حل کنید.

$$\begin{cases} E_۱ : & ۲/۱۱x_۱ - ۴/۲۱x_۲ + ۰/۹۲۱x_۳ = ۲/۰۱ \\ E_۲ : & ۴/۰۱x_۱ + ۱۰/۲x_۲ - ۱/۱۲x_۳ = -۳/۰۹ \\ E_۳ : & ۱/۰۹x_۱ + ۰/۹۸۷x_۲ + ۰/۸۳۲x_۳ = ۴/۲۱ \end{cases}$$

جواب واقعی عبارت است از $x_۱=-۰/۴۲۸,\quad x_۲=۰/۴۲۷,\quad x_۳=۵/۱۱$.

۳. در روش حذف گاوسی با محورگیری جزئی برای ماتریس $A=\begin{bmatrix} ۱ & ۲ & ۳ \\ ۲ & ۴ & ۸ \\ -۱ & ۰ & ۱ \end{bmatrix}$، $a_{۳۳}$ در پایان روش کدام است؟ الف) ۱ ب) ۰ ج) ۱− د) ۲

۴. جواب دستگاه $Ax=b$ کدام گزینه است اگر بدانیم ماتریس A پس از اینکه سطرهای دوم، سوم و چهارم را به ترتیب با $\frac{۱}{۲}$ برابر، ۲− برابر و ۱ برابر سطر اول جمع کرده و سپس سطر چهارم را با ۲− برابر سطر دوم و $-\frac{۱}{۲}$ برابر سطر سوم جمع کنیم به ماتریس $\begin{bmatrix} ۳ & ۰ & -۳ & -۱ \\ ۰ & ۱ & ۲ & ۱ \\ ۰ & ۰ & ۱/۵ & ۱ \\ ۰ & ۰ & ۰ & ۲/۵ \end{bmatrix}$ تبدیل شود و $b=\begin{bmatrix} ۶ \\ ۰ \\ ۱۵ \\ ۱ \end{bmatrix}$. الف) $\begin{bmatrix} \frac{۲}{۵} & \frac{۷۲}{۵} & -\frac{۱۴۶}{۵} & \frac{۲۴۸}{۱۵} \end{bmatrix}^t$ ب) $\begin{bmatrix} ۰ & ۲ & -۱ & ۴ \end{bmatrix}^t$ ج) $\begin{bmatrix} ۴ & -۱ & ۲ & ۰ \end{bmatrix}^t$ د) $\begin{bmatrix} \frac{۲۴۸}{۱۵} & -\frac{۱۴۶}{۵} & \frac{۷۲}{۵} & \frac{۲}{۵} \end{bmatrix}^t$

۵. در تجزیه LU ماتریس $\begin{bmatrix} ۱ & ۲ & ۳ \\ ۲ & ۴ & ۸ \\ -۱ & ۰ & ۱ \end{bmatrix}$ کدام گزینه درست است؟ الف) $L=\begin{bmatrix} ۱ & ۰ & ۰ \\ ۲ & ۱ & ۰ \\ -۱ & ۰ & ۱ \end{bmatrix}$ ب) $L=$

$$\begin{bmatrix} ۱ & ۰ & ۰ \\ -۲ & ۱ & ۰ \\ ۱ & ۰ & ۱ \end{bmatrix}$$
ج) $L = \begin{bmatrix} ۱ & ۰ & ۰ \\ ۲ & ۱ & ۰ \\ -۱ & ۲ & ۱ \end{bmatrix}$ د) وجود ندارد.

۶. دستگاه معادلات خطی زیر را در نظر بگیرید.

$$\begin{cases} ۶ax_۱ + ۳ax_۲ + ۲ax_۳ = ۱ \\ x_۱ - ۳ax_۲ + ۴x_۳ = ۲ \\ x_۱ + x_۲ - ۲ax_۳ = ۳ \end{cases}$$

الف) به ازای چه مقادیری از a در روش حذف گاوسی با محورگیری جزئی نیازی به جابجایی سطر نیست؟
ب) به ازای $a = ۲$ چند تکرار روش گاوس-سیدل لازم است تا با انتخاب بردار آغازی صفر جوابی با دقت $۲D$ به دست آوریم؟

۷. در تجزیه LU ماتریس $\begin{bmatrix} ۲ & -۴ & ۱ \\ ۱ & ۱ & ۴ \\ ۳ & ۱ & -۳ \end{bmatrix}$ مقدار $u_{۳۳}$ کدام است؟ الف) $\frac{۷}{۳}$ ب) $-\frac{۷}{۳}$ ج) $-\frac{۳۸}{۳}$ د) $\frac{۳۸}{۳}$

۸. فرض کنید $A = LU$ که در آن $L = \begin{bmatrix} ۱ & ۰ & ۰ \\ -۱ & ۱ & ۰ \\ ۲ & ۱ & ۱ \end{bmatrix}$ و $U = \begin{bmatrix} ۲ & ۱ & ۳ \\ ۰ & ۰/۵ & ۵ \\ ۰ & ۰ & ۱ \end{bmatrix}$. مقدار $\det A$ کدام است؟

الف) ۳ ب) -۳ ج) -۱ د) ۱

۹. کدام‌یک از گزینه‌های زیر در مورد روش حذف گاوسی برای حل دستگاه مربعی $Ax = b$ نادرست است؟

الف) اگر $a_{۱۱} = ۰$، این روش را نمی‌توان به کار برد. ب) بدون استفاده از محورگیری جزئی ممکن است جواب‌های حاصل از این روش دقیق نباشند. ج) اگر خطای گردکردن وجود نداشته باشد، این روش جواب دقیق را به دست می‌دهد. د) ماتریس بالامثلثی حاصل از این روش به بردار سمت راست b بستگی ندارد.

۱۰. دستگاه خطی زیر را در نظر بگیرید. دترمینان ماتریس ضرایب را به کمک تجزیه LU به دست آورید.

$$\begin{cases} ۲x_۱ - x_۲ - ۶x_۳ = -۲ \\ ۸x_۱ + x_۲ - x_۳ = ۱۳ \\ x_۱ - ۵x_۲ - x_۳ = -۸ \end{cases}$$

پس از اعمال شرایط کافی برای همگرایی، $X^{(۳)}$ را با استفاده از روش گاوس-سیدل، با نقطه شروع $X^{(۰)} = (۱ \ ۰ \ ۱)^T$ به دست آورید. (محاسبات با دو رقم اعشار دنبال شوند). $||T_G||_\infty$ را حساب کنید. (T_G ماتریس تکرار روش گاوس-سیدل در قسمت قبل است.)

۱۱. دستگاه معادلات خطی زیر را در نظر بگیرید.

$$\left\{\begin{array}{lllll} x_۱ & + & ۴x_۲ & -۲x_۳ & =-۲ \\ x_۱ & + & x_۲ & +۴x_۳ & =۱ \\ ۵x_۱ & - & x_۲ & +۲x_۳ & =۲ \end{array}\right.$$

با جابجایی معادلات، دستگاه را به دستگاهی تبدیل کنید که روش گاوس-سیدل برای تعیین جواب تقریبی آن همواره همگرا باشد و سپس نتایج سه گام از این روش (یعنی $X^{(۱)}$، $X^{(۲)}$ و $X^{(۳)}$) را با تقریب اولیه $X^{(۰)}=(۰,۰,۰)^t$ به دست آورید.

۱۲. دستگاه معادلات خطی زیر را در نظر بگیرید.

$$\left\{\begin{array}{lllll} ۷x_۱ & + & ۴x_۲ & -۲x_۳ & =۱ \\ x_۱ & - & ۶x_۲ & +۴x_۳ & =-۱ \\ ۵x_۱ & - & x_۲ & +۸x_۳ & =-۲ \end{array}\right.$$

چند تکرار روش ژاکوبی با بردار آغازی صفر لازم است تا جوابی با خطایی کمتر از $۱۰^{-۴}$ به دست آید؟
چند تکرار روش گاوس-سیدل با بردار آغازی صفر لازم است تا جوابی با خطایی کمتر از $۱۰^{-۴}$ به دست آید؟

۱۳. دستگاه معادلات زیر را در نظر بگیرید

$$\begin{cases} ۳x_۱-x_۲+x_۳=۱, \\ ۲x_۱+a\,x_۲+x_۳=۲, \\ x_۱-x_۲+۳x_۳=-۱. \end{cases}$$

الف) به ازای چه مقادیری از a در روش حذف گاوسی با محورگیری جزئی، نیازی به جابجایی سطر نیست.
ب) به ازای چه مقادیری از a، همگرایی روش گاوس-سایدل به ازای هر جواب اولیه $X^{(۰)}$ تضمین‌شده است.
ج) به ازای $a=۴$ دو تکرار روش ژاکوبی را با جواب اولیه $X^{(۰)}=[۰,۰,۰]^T$ به دست آورید (محاسبات با چهار رقم بامعنا).

۱۴. دستگاه معادلات زیر را در نظر بگیرید

$$\begin{cases} x_۱-۲x_۲+۴۰x_۳=۳۹, \\ ۱۰x_۱-x_۲+x_۳=۱۰, \\ ۲x_۱+۲۰x_۲-x_۳=۲۱. \end{cases}$$

الف) جواب دستگاه فوق را به کمک روش حذف گاوسی با محورگیری جزئی به دست آورید.
ب) دو تکرار روش ژاکوبی را با بردار اولیه $X^{(۰)}=[۰,۰,۰]^T$ به دست آورید.
ج) با چه تغییری در دستگاه داده شده می‌توان از همگرایی روش گاوس-سیدل با هر بردار اولیه مطمئن بود؟

فصل ۶

حل عددی معادلات دیفرانسیل عادی

معادله $y' = f(x, y)$ یک معادله دیفرانسیل عادی (معمولی) مرتبه اول نامیده می‌شود و در آن $f(x, y)$ یک تابع حقیقی دو متغیره است که به ازای هر x در $[a, b]$ و تمامی yهای حقیقی تعریف شده است (فرض می‌شود $y = y(x)$ تابعی حقیقی تعریف شده بر $[a, b]$ باشد). این معادله همراه با شرط $y(a) = \alpha$ ($y(x_۰) = y_۰$) که به شرط اولیه معروف است را یک مسئله مقدار اولیه[۱] می‌نامند.

مثال ۱.۶ *مسئله مقدار اولیه*

$$\begin{cases} y' = ۱ + x \sin(xy), \qquad x \in [۰, ۲], \\ y(۰) = ۰, \end{cases}$$

را در نظر بگیرید. با آن که ثابت می‌شود (به [۲] مراجعه شود) این مسئله مقدار اولیه جواب یکتا دارد، ولی به کمک روش‌های معمول نمی‌توان آن را به دست آورد. $\triangle$

به دلیل مشکلات موجود در حل تحلیلی چنین مسایلی (عدم وجود جواب تحلیلی یا پیچیدگی حل تحلیلی)، به بررسی روش‌های حل عددی آن‌ها می‌پردازیم. در این راستا قصد داریم در این فصل ابتدا روش بسط تیلور و روش‌های رانگ-کوتا را معرفی کرده و در آخر، این روش‌ها را برای حل دستگاه معادلات دیفرانسیل مرتبه اول و معادلات دیفرانسیل مرتبه بالاتر به کار گیریم.

در این روش‌ها پس از اختیار نمودن عدد کوچک h (اندازه گام) و با فرض $x_۰ = a$، به ازای $i = ۰, ۱, ۲, \ldots$ قرار می‌دهیم $x_i = x_۰ + ih$. سپس y_i به عنوان تقریبی از $y(x_i)$ در نظر گرفته می‌شود. واضح است که تفاوت این روش‌ها در نحوه ساختن y_i است. به شکل ۱.۶ توجه کنید.

۱.۶ روش بسط تیلور

به کمک بسط تیلور می‌توان نوشت

$$y(x_{i+۱}) = y(x_i + h) = y(x_i) + hy'(x_i) + \frac{h^۲}{۲!} y'(x_i) + \cdots + \frac{h^p}{p!} y^{(p)}(x_i) + \frac{h^{p+۱}}{(p+۱)!} y^{(p+۱)}(\xi) \qquad (۱.۶)$$

[۱] initial value problem (IVP)

شکل ۱.۶: تعبیر هندسی حل عددی معادلات دیفرانسیل

که در آن ξ بین x_i و x_{i+1} قرار دارد. اگر $y' = f(x,y)$ را به صورت دقیق‌تر $y' = f(x, y(x))$ بنویسیم آن‌گاه

$$y''(x) = \frac{d}{dx}(y'(x)) = \frac{d}{dx} f(x,y) = \frac{dx}{dx}\frac{\partial f}{\partial x} + \frac{dy}{dx}\frac{\partial f}{\partial y} = f_x + y' f_y = f_x + f f_y.$$

حال با فرض $f^{(0)}(x,y) = f(x,y)$، اگر برای $r \geq 1$ تعریف کنیم

$$f^{(r)}(x,y) = {f_x}^{(r-1)}(x,y) + f(x,y){f_y}^{(r-1)}(x,y)$$

آن‌گاه با استقرا روی r می‌توان ثابت نمود $y^{(r+1)}(x) = f^{(r)}(x, y(x))$. بنابراین (۱.۶) را می‌توان به صورت زیر بازنویسی کرد

$$y(x_{i+1}) = y(x_i) + hf(x_i, y(x_i)) + \frac{h^2}{2!} f^{(1)}(x_i, y(x_i)) + \cdots + \frac{h^p}{p!} f^{(p-1)}(x_i, y(x_i)) + \frac{h^{p+1}}{(p+1)!} f^{(p)}(\xi, y(\xi)).$$

صرف نظر از جمله باقیمانده، روش تیلور مرتبه p به صورت زیر خواهد بود

$$\begin{cases} y_{i+1} = y_i + hf(x_i, y_i) + \frac{h^2}{2!} f^{(1)}(x_i, y_i) + \cdots + \frac{h^p}{p!} f^{(p-1)}(x_i, y_i), \qquad i = 0, 1, 2, \ldots \\ y_0 = \alpha \end{cases}$$

عیب اصلی این روش محاسبه $f^{(r)}(x,y)$ است. به عنوان مثال برای $r = 2$ داریم

$$f^{(2)} = f_x^{(1)} + f f_y^{(1)} = f_{xx} + 2 f f_{xy} + f^2 f_{yy} + (f_x + f f_y) f_y.$$

مثال ۲.۶ *به کمک روش تیلور و با انتخاب $h = 0.1$ تقریبی با دقت $4D$ از $y(0.5)$ برای مسئله داده‌شده ارایه دهید.*

$$y' = x + y, \qquad y(0) = 1$$

داریم

$$\begin{aligned} &y' = x + y, && y'' = 1 + y' = 1 + x + y, \\ &y''' = 1 + y' = 1 + x + y, && y^{(4)} = 1 + y' = 1 + x + y = y^{(5)} = \cdots. \end{aligned}$$

بنابراین به ازای $i = ۰, ۱, ۲, ۳, \dots$ *خواهیم داشت*

$$y(x_{i+۱}) \simeq y_{i+۱} = y_i + h(x_i + y_i) + \frac{h^۲}{۲!}(۱ + x_i + y_i) + \frac{h^۳}{۳!}(۱ + x_i + y_i) + \frac{h^۴}{۴!}(۱ + x_i + y_i) + \cdots.$$

حال برای $i = ۰, ۱, ۲, \dots$ *قرار می‌دهیم* $x_i = \frac{i}{۱۰}$*. به عنوان نمونه با انتخاب پنج جمله (روش تیلور مرتبه چهار) از عبارت اخیر خواهیم داشت*

$$y_{i+۱} = ۰/۰۰۵۱۷ + ۰/۱۰۵۱۷x_i + ۱/۱۰۵۱۷y_i, \quad i = ۰, ۱, \dots$$

و از آن‌جا داریم

$$y_۱ = ۱/۱۱۰۳۴, \quad y_۲ = ۱/۲۴۲۸۰, \quad y_۳ = ۱/۳۹۹۷۱, \quad y_۴ = ۱/۵۷۳۶۴, \quad y_۵ = ۱/۷۹۷۴۳.$$

از طرفی جواب واقعی این مسئله $y(x) = ۲e^x - x - ۱$ *است و بنابراین* $y(۰/۵) = ۱/۷۹۷۴۴$ *و در نتیجه خطایی حدود* $۰/۰۰۰۰۱$ *مرتکب شده‌ایم.* $\triangle$

مثال ۳.۶ *برای معادله* $y' = \dfrac{۱}{۱ + y^۲}$ *با شرط اولیه* $y(۰) = ۱$ *داریم*

$$f(x,y) = \frac{۱}{۱ + y^۲}, \quad f^{(۱)}(x,y) = f_x + ff_y = \frac{-۲y}{(۱ + y^۲)^۳}, \quad f^{(۲)}(x,y) = \frac{۲(۵y^۲ - ۱)}{(۱ + y^۲)^۵}$$

بنابراین برای $p = ۳$ *خواهیم داشت*

$$\begin{cases} y_{i+۱} = y_i + \dfrac{h}{۱ + y_i^۲} + \dfrac{h^۲}{۲!}\dfrac{-۲y_i}{(۱ + y_i^۲)^۳} + \dfrac{h^۳}{۳!}\dfrac{۲(۵y_i^۲ - ۱)}{(۱ + y_i^۲)^۵}, \quad i = ۰, ۱, \dots, \\ y_۰ = ۱. \end{cases}$$

$\triangle$

تذکر ۱.۶ *در روش تیلور مرتبه* p *هر چه* p *بزرگ‌تر باشد دقت روش بیش‌تر است (خطا از مرتبه* $O(h^{p+۱})$ *است) ولی محاسبه جملات* $f^{(r)}(x,y)$ *برای* $r = ۱, ۲, ۳, \dots, p$ *ممکن است بسیار مشکل و زمان‌گیر باشد؛ برای پرهیز از این مشکل به روش‌های اویلر و رانگ-کوتا پناه می‌بریم.*

۲.۶ روش اویلر

روش اویلر همان روش تیلور مرتبه یک با خطایی از مرتبه $O(h)$ بوده و طرح تفاضلی آن به صورت زیر است

$$\begin{cases} y_{i+۱} = y_i + hf(x_i, y_i), \quad i = ۰, ۱, \dots, \\ y(x_۰) = \alpha. \end{cases}$$

x_i	y_i	$y(x_i)$	$\lvert y_i - y(x_i)\rvert$
۰/۰	۰/۵۰۰۰۰۰۰	۰/۵۰۰۰۰۰۰	۰/۰۰۰۰۰۰۰
۰/۲	۰/۸۰۰۰۰۰۰	۰/۸۲۹۲۹۸۶	۰/۰۲۹۲۹۸۶
۰/۴	۱/۱۵۲۰۰۰۰	۱/۲۱۴۰۸۷۷	۰/۰۶۲۰۸۷۷
۰/۶	۱/۵۵۰۴۰۰۰	۱/۶۴۸۹۴۰۶	۰/۰۹۸۵۴۰۶
۰/۸	۱/۹۸۸۴۸۰۰	۲/۱۲۷۲۲۹۵	۰/۱۳۸۷۴۹۵
۱/۰	۲/۴۵۸۱۷۶۰	۲/۶۴۰۸۵۹۱	۰/۱۸۲۶۸۳۱
۱/۲	۲/۹۴۹۸۱۱۲	۳/۱۷۹۹۴۱۵	۰/۲۳۰۱۳۰۳
۱/۴	۳/۴۵۱۷۷۳۴	۳/۷۳۲۴۰۰۰	۰/۲۸۰۶۲۶۶
۱/۶	۳/۹۵۰۱۲۸۱	۴/۲۸۳۴۸۳۸	۰/۳۳۳۳۵۵۷
۱/۸	۴/۴۲۸۱۵۳۸	۴/۸۱۵۱۷۶۳	۰/۳۸۷۰۲۲۵
۲/۰	۴/۸۶۵۷۸۴۵	۵/۳۰۵۴۷۲۰	۰/۴۳۹۶۸۷۴

جدول ۱.۶: مقایسه روش اویلر با جواب واقعی

مثال ۴.۶ *برای حل مسئله*

$$\begin{cases} y' = y - x^2 + 1, \quad x \in [0, 2], \\ y(0) = 0/5, \end{cases}$$

فاصله $[0, 2]$ را به ده قسمت مساوی تقسیم کرده و به کمک روش اویلر مقادیر تقریبی $y_1, \ldots, y_{10}$ را مشخص کنید.
با انتخاب $n = 10$ داریم $h = \frac{b-a}{10} = \frac{2-0}{10} = 0/2$ و بنابراین به ازای $i = 0, 1, \ldots, 10$ خواهیم داشت $x_i = x_0 + ih = 0/2i$. از طرفی چون $y_{i+1} = y_i + hf(x_i, y_i) = y_i + h(y_i - x_i^2 + 1)$، پس طرح تکراری به صورت زیر به دست می‌آید

$$\begin{cases} y_{i+1} = 1/2 y_i - 0/008 i^2 + 0/2, \quad i = 0, 1, \ldots, 9, \\ y_0 = y(0) = 0/5. \end{cases}$$

جواب‌های تقریبی به دست آمده با جواب واقعی $y(x) = (x+1)^2 - 0/5e^x$ در جدول ۱.۶ مقایسه شده‌اند. چون h چندان کوچک نیست خطاها رضایت‌بخش نیستند. $\triangle$

۳.۶ روش‌های رانگ-کوتا

حدود ۱۰۰ سال پیش دو ریاضیدان آلمانی به نام‌های رانگ[2] و کوتا[3] برای حل معادلات دیفرانسیل معمولی روشی ارایه دادند که نیازی به محاسبه مشتق‌های y نداشته و تنها با ارزیابی $f(x, y)$ جواب‌های تقریبی خوبی تولید می‌کند. روش

[2] C. Runge
[3] M. W. Kutta

x_i	$y(x_i)$	RK2	خطا
۰٫۰۰۰	۰٫۵۰۰	۰٫۵۰۰	۰٫۰۰۰
۰٫۲۰۰	۰٫۸۲۹	۰٫۸۲۶	۰٫۰۰۳
۰٫۴۰۰	۱٫۲۱۴	۱٫۲۰۷	۰٫۰۰۷
⋮	⋮	⋮	⋮
۱٫۰۰۰	۲٫۶۴۱	۲٫۶۱۸	۰٫۰۲۳
⋮	⋮	⋮	⋮
۲٫۰۰۰	۵٫۳۰۵	۵٫۲۳۳	۰٫۰۷۲

جدول ۲.۶: مقایسه نتایج روش اویلر اصلاح‌شده (RK2) با جواب دقیق

رانگ-کوتای r مرحله‌ای یک طرح تفاضلی است که برای محاسبه y_{i+1} باید f در r نقطه ارزیابی شود. جهت شناخت بهتر این روش‌ها، ابتدا طرح تفاضلی رانگ-کوتای دو مرحله‌ای (RK2) را بررسی می‌کنیم. RK2 همانند روش تیلور مرتبه دو است که خطایی از مرتبه $O(h^2)$ دارد و در آن برای $i = 0, 1, \ldots$ قرار می‌دهیم

$$\begin{cases} K_1 = hf(x_i, y_i), \\ K_2 = hf(x_i + h, y_i + K_1), \\ y_{i+1} = y_i + \frac{1}{2}(K_1 + K_2) \end{cases}$$

و یا

$$y_{i+1} = y_i + \frac{h}{2}(f(x_i, y_i) + f(x_i + h, y_i + hf(x_i, y_i))).$$

روش رانگ-کوتای دو مرحله‌ای به روش اویلر اصلاح‌شده (تعمیم‌یافته - تغییریافته - ترمیم‌یافته)[4] نیز معروف است.

مثال ۵.۶ *مسئله زیر را در نظر بگیرید.*

$$\begin{cases} y' = y - x^2 + 1, \qquad x \in [0, 2], \\ y(0) = 0.5. \end{cases}$$

با انتخاب $n = 10$ خواهیم داشت $h = 0.2$ و برای $i = 0, 1, 2, \ldots, 10$ داریم $x_i = 0.2i$. طرح تفاضلی اویلر اصلاح‌شده به صورت زیر ساده می‌شود

$$y_{i+1} = 1.22y_i - 0.0088i^2 - 0.008i + 0.216.$$

نتایج در جدول ۲.۶ آورده شده است. △

[4] modified Euler method

x_i	$y(x_i)$	RK4	خطا
۰٫۰۰۰۰۰۰۰	۰٫۵۰۰۰۰۰۰	۰٫۵۰۰۰۰۰۰	۰٫۰۰۰۰۰۰۰
۰٫۲۰۰۰۰۰۰	۰٫۸۲۹۲۹۸۶	۰٫۸۲۹۲۹۳۳	۰٫۰۰۰۰۰۵۳
۰٫۴۰۰۰۰۰۰	۱٫۲۱۴۰۸۷۷	۱٫۲۱۴۰۷۶۲	۰٫۰۰۰۰۱۱۴
⋮	⋮	⋮	⋮
۱٫۰۰۰۰۰۰۰	۲٫۶۴۰۸۵۹۱	۲٫۶۴۰۸۲۲۷	۰٫۰۰۰۰۳۶۴
⋮	⋮	⋮	⋮
۲٫۰۰۰۰۰۰۰	۵٫۳۰۵۴۷۲۰	۵٫۳۰۵۳۶۳۰	۰٫۰۰۰۱۰۸۹

جدول ۳.۶: روش رانگ–کوتای چهار مرحله‌ای

طرح تفاضلی رانگ–کوتای چهار مرحله‌ای (RK4) که خطای آن از مرتبه $O(h^4)$ است، از دیگر روش‌های پرکاربرد رانگ–کوتا بوده و در آن برای $i = 0, 1, 2, \ldots$ قرار می‌دهیم

$$\begin{cases} K_1 = hf(x_i, y_i), \\ K_2 = hf(x_i + \frac{h}{2}, y_i + \frac{K_1}{2}), \\ K_3 = hf(x_i + \frac{h}{2}, y_i + \frac{K_2}{2}), \\ K_4 = hf(x_i + h, y_i + K_3), \\ y_{i+1} = y_i + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4). \end{cases}$$

مثال ۶.۶ *با به کار بردن طرح تفاضلی رانگ–کوتای چهار مرحله‌ای بر روی مسئله*

$$\begin{cases} y' = y - x^2 + 1, \qquad x \in [0, 2], \\ y(0) = 0/5, \end{cases}$$

و با انتخاب $n = 10$، نتایج جدول ۳.۶ به دست می‌آید. △

۴.۶ دستگاه معادلات دیفرانسیل مرتبه اول

یک دستگاه معادلات دیفرانسیل مرتبه اول به صورت کلی

$$\begin{cases} x_1'(t) = f_1(t, x_1(t), \ldots, x_n(t)), \quad x_1(t_0) = x_{10} \\ \qquad \vdots \qquad\qquad\qquad\qquad \vdots \\ x_n'(t) = f_n(t, x_1(t), \ldots, x_n(t)), \quad x_n(t_0) = x_{n0} \end{cases}$$

است که در آن مشتق نسبت به متغیر مستقل t است و $x_1(t),...,x_n(t)$ توابع مجهول بر حسب متغیر مستقل t هستند و $x_{10},...,x_{n0}$ مقادیر اولیه هستند.

مثال ۷.۶ *به عنوان نمونه در زیر یک دستگاه معادلات دیفرانسیل مرتبه اول ۲ × ۲ غیرخطی (سمت چپ) و یک دستگاه معادلات دیفرانسیل مرتبه اول ۳ × ۳ خطی (سمت راست) ارایه شده است.*

$$\begin{cases} x'(t) &=& y^2(t), \quad x(0)=0 \\ y'(t) &=& -x(t), \quad y(0)=1 \end{cases} \qquad \begin{cases} x_1'(t) &=& x_2(t)-x_3(t)+t, & x_1(0)=1 \\ x_2'(t) &=& 3t^2, & x_2(0)=1 \\ x_3'(t) &=& x_2(t)+e^{-t}, & x_3(0)=-1 \end{cases}$$

△

در ادامه با استفاده از روش‌های عددی حل معادلات دیفرانسیل مرتبه اول، روش‌هایی برای حل عددی دستگاه‌های معادلات دیفرانسیل مرتبه اول ۲ × ۲

$$\begin{cases} x'(t) &=& f_1(t,x(t),y(t)), \quad x(t_0)=x_0 \\ y'(t) &=& f_2(t,x(t),y(t)), \quad y(t_0)=y_0 \end{cases} \tag{۲.۶}$$

و ۳ × ۳ زیر ارایه می‌کنیم.

$$\begin{cases} x'(t) &=& f_1(t,x(t),y(t),z(t)), \quad x(t_0)=x_0 \\ y'(t) &=& f_2(t,x(t),y(t),z(t)), \quad y(t_0)=y_0 \\ z'(t) &=& f_3(t,x(t),y(t),z(t)), \quad z(t_0)=z_0 \end{cases} \tag{۳.۶}$$

۱.۴.۶ روش اویلر

در این روش باید روش اویلر را به طور همزمان برای تمامی معادلات دستگاه به کار ببریم. به عنوان مثال برای دستگاه ۲ × ۲ (۲.۶)، اگر مقدار تقریبی x در t_n را با x_n و مقدار تقریبی y در t_n را با y_n نمایش دهیم، دستگاه زیر به دست می‌آید

$$\begin{cases} x_{n+1} &=& x_n + hf_1(t_n,x_n,y_n) \\ y_{n+1} &=& y_n + hf_2(t_n,x_n,y_n) \end{cases}$$

و برای دستگاه ۳ × ۳ (۳.۶)، اگر مقدار تقریبی x در t_n را با x_n، مقدار تقریبی y در t_n را با y_n و مقدار تقریبی z در t_n را با z_n نمایش دهیم دستگاه زیر به دست می‌آید

$$\begin{cases} x_{n+1} &=& x_n + hf_1(t_n,x_n,y_n,z_n) \\ y_{n+1} &=& y_n + hf_2(t_n,x_n,y_n,z_n) \\ z_{n+1} &=& z_n + hf_3(t_n,x_n,y_n,z_n) \end{cases}$$

مثال ۸.۶ *در دستگاه‌های زیر به ترتیب مقادیر $x(0/1)$ و $y(0/1)$ را برای دستگاه 2×2 و $x(0/2)$، $y(0/2)$ و $z(0/2)$ را برای دستگاه 3×3 به روش اویلر به دست آورید.*

$$\begin{cases} x'(t) &= y(t), \quad x(0) = 0 \\ y'(t) &= -x(t), \quad y(0) = 1 \end{cases} \qquad \begin{cases} x'(t) &= y(t) - z(t) + t, \quad x(0) = 1 \\ y'(t) &= -x(t), \quad y(0) = 1 \\ z'(t) &= y(t) + e^{-t}, \quad z(0) = -1 \end{cases}$$

برای دستگاه 2×2 داریم $f_1(t_n, x_n, y_n) = y_n$ و $f_2(t_n, x_n, y_n) = -x_n$. بنابراین برای $h = 0/1$ خواهیم داشت

$$x_{n+1} = x_n + 0/1 y_n, \qquad y_{n+1} = y_n - 0/1 x_n$$

و $f_1(t_n, x_n, y_n, z_n) = y_n - z_n + t_n$، $f_2(t_n, x_n, y_n, z_n) = -x_n$ و $f_3(t_n, x_n, y_n, z_n) = y_n + e^{-t_n}$ برای دستگاه 3×3 برقرار است و در نتیجه برای $h = 0/2$ داریم

$$x_{n+1} = x_n + 0/2(y_n - z_n + t_n), \qquad y_{n+1} = y_n - 0/2 x_n, \qquad z_{n+1} = z_n + 0/2(y_n + e^{-t_n}).$$

پس به ازای $n = 0$ برای دستگاه 2×2 نتایج زیر به دست می‌آیند

$$\begin{cases} x(0/1) = x_1 &= x_0 + 0/1 y_0 = 0 + 0/1 = 0/1 \\ y(0/1) = y_1 &= y_0 - 0/1 x_0 = 1 - 0 = 1 \end{cases}$$

و برای دستگاه 3×3 داریم

$$\begin{cases} x(0/2) &= x_0 + 0/2(y_0 - z_0 + t_0) = 1 + 0/2(1 + 1 + 0) = 1/4 \\ y(0/2) &= y_0 - 0/2 x_0 = 1 - 0/2 = 0/8 \\ z(0/2) &= z_0 + 0/2(y_0 + e^{-t_0}) = -1 + (0/2)(1 + 1) = -0/96 \end{cases}$$

$\triangle$

۲.۴.۶ روش اویلر اصلاح‌شده

در این‌جا باید روش اویلر اصلاح‌شده را به طور همزمان برای تمامی معادلات دستگاه به کار ببریم. به عنوان مثال برای دستگاه 2×2 (۲.۶)، خواهیم داشت

$$\begin{cases} x_{n+1} &= x_n + \frac{h}{2}[f_1(t_n, x_n, y_n) + f_1(t_{n+1}, x^*_{n+1}, y^*_{n+1})] \\ y_{n+1} &= y_n + \frac{h}{2}[f_2(t_n, x_n, y_n) + f_2(t_{n+1}, x^*_{n+1}, y^*_{n+1})] \end{cases}$$

به قسمی که $x^*_{n+1} = x_n + hf_1(t_n, x_n, y_n)$ و $y^*_{n+1} = y_n + hf_2(t_n, x_n, y_n)$. برای دستگاه (۳.۶)، با این فرض که $x^*_{n+1} = x_n + hf_1(t_n, x_n, y_n, z_n)$, $y^*_{n+1} = y_n + hf_2(t_n, x_n, y_n, z_n)$, $z^*_{n+1} = z_n + hf_3(t_n, x_n, y_n, z_n)$ داریم

$$\begin{cases} x_{n+1} = x_n + \frac{h}{2}[f_1(t_n, x_n, y_n, z_n) + f_1(t_{n+1}, x^*_{n+1}, y^*_{n+1}, z^*_{n+1})] \\ y_{n+1} = y_n + \frac{h}{2}[f_2(t_n, x_n, y_n, z_n) + f_2(t_{n+1}, x^*_{n+1}, y^*_{n+1}, z^*_{n+1})] \\ z_{n+1} = z_n + \frac{h}{2}[f_3(t_n, x_n, y_n, z_n) + f_3(t_{n+1}, x^*_{n+1}, y^*_{n+1}, z^*_{n+1})] \end{cases}$$

مثال ۹.۶ *مثال قبل را به کمک روش اویلر اصلاح‌شده حل کنید. برای دستگاه 2×2 داریم*

$$x_{n+1} = x_n + \frac{h}{2}(y_n + y^*_{n+1}), \qquad y_{n+1} = y_n + \frac{h}{2}(-x_n - x^*_{n+1}),$$

به قسمی که $x^*_{n+1} = x_n + hy_n$ و $y^*_{n+1} = y_n - hx_n$. بنابراین برای $h = 0/1$ داریم

$$\begin{cases} x_{n+1} &=& x_n + 0/05[y_n + y_n - 0/1x_n] = 0/995x_n + 0/1y_n \\ y_{n+1} &=& y_n + 0/05[-x_n - x_n - 0/1y_n] = -0/1x_n + 0/995y_n \end{cases}$$

پس به ازای $n = 0$ برای دستگاه 2×2 نتایج زیر به دست می‌آیند

$$x(0/1) = 0/995x_0 + 0/1y_0 = 0/1 \qquad y(0/1) = -0/1x_0 + 0/995y_0 = 0/995.$$

برای دستگاه 3×3 داریم

$$\begin{cases} x_{n+1} = x_n + \frac{h}{2}[y_n - z_n + t_n + y^*_{n+1} - z^*_{n+1} + t_{n+1}] \\ y_{n+1} = y_n + \frac{h}{2}[-x_n - x^*_{n+1}] \\ z_{n+1} = z_n + \frac{h}{2}[y_n + e^{-t_n} + y^*_{n+1} + e^{-t_{n+1}}] \end{cases}$$

که در آن $x^*_{n+1} = x_n + h(y_n - z_n + t_n),\ y^*_{n+1} = y_n - hx_n,\ z^*_{n+1} = z_n + h(y_n + e^{-t_n})$. برای $h = 0/2$ داریم

$$\begin{cases} x_{n+1} &=& x_n + 0/1[y_n - z_n + t_n + y_n - 0/2x_n - (z_n + 0/2(y_n + e^{-t_n})) + t_{n+1}] \\ y_{n+1} &=& y_n + 0/1[-x_n - x_n - 0/2(y_n - z_n + t_n)] \\ z_{n+1} &=& z_n + 0/1[y_n + e^{-t_n} + y_n - 0/2x_n + e^{-t_{n+1}}] \end{cases}$$

پس برای دستگاه 3×3 نتایج زیر به دست می‌آیند.

$$\begin{cases} x(0/2) &=& 1 + 0/1[1 + 1 + 1 - 0/2 - (-1 + 0/4) + 0/2] = 1/36 \\ y(0/2) &=& 1 + 0/1(-1 - 1 - 0/4) = 0/76 \\ z(0/2) &=& -1 + 0/1[1 + 1 + 1 - 0/2 + e^{-0/2}] = -0/6381269247 \end{cases}$$

$\triangle$

۳.۴.۶ روش تیلور

در این روش باید روش تیلور را به طور همزمان برای تمامی معادلات دستگاه به کار بریم. به عنوان مثال روش تیلور مرتبه سه برای دستگاه ۲ × ۲ (۲.۶)، به صورت زیر است

$$\begin{cases} x_{n+1} = x_n + hf_1(t_n, x_n, y_n) + \frac{h^2}{2!}f_1'(t_n, x_n, y_n) + \frac{h^3}{3!}f_1''(t_n, x_n, y_n), \\ y_{n+1} = y_n + hf_2(t_n, x_n, y_n) + \frac{h^2}{2!}f_2'(t_n, x_n, y_n) + \frac{h^3}{3!}f_2''(t_n, x_n, y_n) \end{cases}$$

که در آن $f_1'(t,x,y) = f_{1t} + f_1 f_{1x} + f_2 f_{1y}$، $f_2'(t,x,y) = f_{2t} + f_1 f_{2x} + f_2 f_{2y}$ و $f_1''(t,x,y)$ و $f_2''(t,x,y)$ به روشی مشابه به دست می‌آیند.

۴.۴.۶ روش رانگ_کوتای چهار مرحله‌ای

در این روش برای حل دستگاه ۲ × ۲ (۲.۶) داریم

$$\begin{aligned} k_1 &= hf_1(t_n, x_n, y_n), & l_1 &= hf_2(t_n, x_n, y_n), \\ k_2 &= hf_1(t_n + \tfrac{h}{2}, x_n + \tfrac{k_1}{2}, y_n + \tfrac{l_1}{2}), & l_2 &= hf_2(t_n + \tfrac{h}{2}, x_n + \tfrac{k_1}{2}, y_n + \tfrac{l_1}{2}), \\ k_3 &= hf_1(t_n + \tfrac{h}{2}, x_n + \tfrac{k_2}{2}, y_n + \tfrac{l_2}{2}), & l_3 &= hf_2(t_n + \tfrac{h}{2}, x_n + \tfrac{k_2}{2}, y_n + \tfrac{l_2}{2}), \\ k_4 &= hf_1(t_n + h, x_n + k_3, y_n + l_3), & l_4 &= hf_2(t_n + h, x_n + k_3, y_n + l_3), \\ x_{n+1} &= x_n + \tfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), & y_{n+1} &= y_n + \tfrac{1}{6}(l_1 + 2l_2 + 2l_3 + l_4) \end{aligned}$$

مثال ۱۰.۶ *در دستگاه زیر مقادیر $x(0.1)$ و $y(0.1)$ را به روش رانگ_کوتای چهار مرحله‌ای به دست آورید.*

$$\begin{cases} x'(t) = y(t), & x(0) = 0 \\ y'(t) = -x(t), & y(0) = 1 \end{cases}$$

بنابر روش رانگ_کوتای چهار مرحله‌ای به ازای $h = 0.1$، $n = 0$، $x_0 = 0$، $y_0 = 1$،

$$\begin{aligned}
& f_1(t_n, x_n, y_n) = y_n, \qquad f_2(t_n, x_n, y_n) = -x_n, \\
& k_1 = 0.1 y_n = 0.1, \qquad l_1 = -0.1 x_n = 0, \\
& k_2 = 0.1(y_n + \tfrac{l_1}{2}) = 0.1, \qquad l_2 = -0.1(x_n + \tfrac{k_1}{2}) = -0.005, \\
& k_3 = 0.1(y_n + \tfrac{l_2}{2}) = 0.09975, \qquad l_3 = -0.1(x_n + \tfrac{k_2}{2}) = -0.005, \\
& k_4 = 0.1(y_n + l_3) = 0.09950, \qquad l_2 = -0.1(x_n + k_3) = -0.009975, \\
& x(0.1) = x_0 + \tfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) = 0 + \tfrac{1}{6}(0.1 + 0.2 + 0.19950 + 0.09975) = 0.099875, \\
& y(0.1) = y_0 + \tfrac{1}{6}(l_1 + 2l_2 + 2l_3 + l_4) = 1 + \tfrac{1}{6}(0 - 0.01 - 0.01 - 0.009975) = 0.9950042.
\end{aligned}$$

$\triangle$

n	۰	۱
k_1	۰٫۳	۰٫۲۴۶۰۲۶۳۳۲۲
k_2	۰٫۲۶۹۸۶۶۶۹۳۷	۰٫۲۲۱۶۴۳۸۵۹۶
k_3	۰٫۲۷۳۳۹۲۷۳۰۰	۰٫۲۲۴۵۹۸۳۱۱۵
k_4	۰٫۲۴۵۷۲۹۳۷۸۰	۰٫۲۰۲۱۳۱۲۳۱۵
l_1	−۰٫۱	−۰٫۰۵۶۰۴۲۱۶۲۴۳
l_2	−۰٫۰۷۴۹۹۳۷۵۰۷۸	−۰٫۰۳۶۸۲۱۷۳۵۴۵
l_3	−۰٫۰۷۸۲۶۳۴۳۴۲۴	−۰٫۰۳۹۵۱۸۰۸۵۰۵
l_4	−۰٫۰۵۵۷۵۸۵۴۹۴۰	−۰٫۰۲۲۲۶۵۲۵۱۷۴
x_{n+1}	۰٫۲۷۲۰۴۱۳۷۰۹	۰٫۴۹۵۴۸۱۶۸۸۵
y_{n+1}	−۱٫۰۷۷۰۴۵۴۸۷	−۱٫۱۱۵۵۴۳۳۲۹

جدول ۴.۶: نتایج روش رانگ-کوتای چهار مرحله‌ای برای یک مثال ۱۱.۶

مثال ۱۱.۶ *دستگاه زیر را در بازه $[0, 0{,}2]$ با $h = 0{,}1$ به کمک روش رانگ-کوتای چهار مرحله‌ای حل کنید.*

$$\begin{cases} x'(t) = -4x(t) - 2y(t) + \cos(t) + 4\sin(t), & x(0) = 0 \\ y'(t) = 3x(t) + y(t) - 3\sin(t), & y(0) = -1 \end{cases}$$

بنابر روش رانگ-کوتای چهار مرحله‌ای به ازای $h = 0{,}1$، $x_0 = 0$، $y_0 = -1$ و

$$\begin{aligned}
&f_1(t_n, x_n, y_n) = -4x_n - 2y_n + \cos(t_n) + 4\sin(t_n),\\
&f_2(t_n, x_n, y_n) = x_n + y_n - 3\sin(t_n),\\
&k_1 = 0{,}1[-4x_n - 2y_n + \cos(t_n) + 4\sin(t_n)],\\
&l_1 = 0{,}1[x_n + y_n - 3\sin(t_n)],\\
&k_2 = 0{,}1[-4(x_n + \tfrac{k_1}{2}) - 2(y_n + \tfrac{l_1}{2}) + \cos(t_n + 0{,}05) + 4\sin(t_n + 0{,}05)],\\
&l_2 = 0{,}1[(x_n + \tfrac{k_1}{2}) + (y_n + \tfrac{l_1}{2}) - 3\sin(t_n + 0{,}05)],\\
&k_3 = 0{,}1[-4(x_n + \tfrac{k_2}{2}) - 2(y_n + \tfrac{l_2}{2}) + \cos(t_n + 0{,}05) + 4\sin(t_n + 0{,}05)],\\
&l_3 = 0{,}1[(x_n + \tfrac{k_2}{2}) + (y_n + \tfrac{l_2}{2}) - 3\sin(t_n + 0{,}05)],\\
&k_4 = 0{,}1[-4(x_n + k_3 - 2(y_n + l_3) + \cos(t_n + 0{,}1) + 4\sin(t_n + 0{,}1)],\\
&l_4 = 0{,}1[(x_n + k_3 + (y_n + l_3) - 3\sin(t_n + 0{,}1)],\\
&x_{n+1} = x_n + \tfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4),\\
&y_{n+1} = y_n + \tfrac{1}{6}(l_1 + 2l_2 + 2l_3 + l_4)
\end{aligned}$$

نتایج جدول ۴.۶ به دست می‌آیند. △

۵.۴.۶ معادلات دیفرانسیل عادی مرتبه بالاتر

معادله دیفرانسیل $y^{(n)}(t) = f(t, y(t), y'(t), ..., y^{(n-1)}(t))$ با شرایط اولیه $y(t_0) = y_{01}$،...،$y^{(n-1)}(t_0) = y_{0n}$ را می‌توان با تغییر متغیر $x_1(t) = y(t)$، $x_2(t) = y'(t)$،...، $x_n(t) = y^{(n-1)}(t)$ به دستگاه معادلات دیفرانسیل مرتبه اول زیر تبدیل کرد و به کمک یکی از روش‌های قبلی حل کرد

$$\begin{cases} x_1'(t) = x_2(t), & x_1(t_0) = y_{01}, \\ x_2'(t) = x_3(t), & x_2(t_0) = y_{02}, \\ \quad \vdots & \quad \vdots, \\ x_{n-1}'(t) = x_n(t), & x_{n-1}(t_0) = y_{0,n-1}, \\ x_n'(t) = f(t, x_1(t), x_2(t), ..., x_n(t)), & x_n(t_0) = y_{0n}. \end{cases}$$

مثال ۱۲.۶ *معادله دیفرانسیل مرتبه دوم $y''(t) = t + \cos(t) - y^2(t) + y'(t)$ با شرایط اولیه $y(0) = 0$ و $y'(0) = 1$ با تغییر متغیر $x_1(t) = y(t), x_2(t) = y'(t)$ به دستگاه زیر تبدیل می‌شود*

$$\begin{cases} x_1'(t) = x_2(t), & x_1(0) = 0, \\ x_2'(t) = t + \cos(t) - x_1^2(t) + x_2(t), & x_2(0) = 1. \end{cases}$$

برای این مثال، دو تکرار از روش رانگ_کوتای چهار مرحله‌ای را با $h = 0/01$ دنبال کنید. △

تمرین

۱. روش تیلور مرتبه چهار را برای مسئله زیر بنویسید.

$$y' = \sin(xy), \qquad y(\pi/2) = 1.$$

۲. روش تیلور مرتبه چهار را برای دستگاه (۳.۶) بنویسید.

۳. x و y توابعی از t و جواب‌های دستگاه زیر هستند. مطلوب است محاسبه $x(1/1)$ و $y(1/1)$ به کمک روش رانگ_کوتای دو مرحله‌ای. (e عدد نپر پایه لگاریتم طبیعی است. به عبارت دیگر $\ln e = 1$.)

$$\begin{cases} x' = \ln y + t & x(1) = 1 \\ y' = tx + y & y(1) = e \end{cases}$$

۴. الف) اگر $y = f(x)$ جواب معادله $y' = e^{x+y}$ با شرط اولیه $y(0) = 0$ باشد؛ مطلوب است محاسبه مقدار تقریبی $f(0/6)$ به کمک روش اویلر اصلاح‌شده و به ازای $h = 0/2$. جواب واقعی معادله را به دست آورده و با

جواب‌های تقریبی مقایسه کنید.

ب) x و y توابعی از t و جواب‌های دستگاه زیر هستند. مطلوب است $x(0/2)$ و $y(0/2)$ به کمک روش رانگ–کوتای چهار مرحله‌ای.

$$\begin{cases} x' = y^2 + t, & x(0) = 0, \\ y' = \cos(x), & y(0) = 1. \end{cases}$$

۵. فرض کنید y تابعی از متغیر x باشد. مطلوب است حل مسئله $y' = 1 + (x-y)^2, y(2) = 1$ به کمک روش‌های رانگ–کوتای چهار مرحله‌ای و تیلور مرتبه سه در بازه $[2,3]$ و به ازای $h = 0/5$.

۶. فرض کنید λ یک عدد حقیقی باشد و معادله دیفرانسیل $y' = \lambda y$ با شرط اولیه $y(t_0) = y_0$ را روی بازه $[t_0, t_N]$ در نظر بگیرید. نشان دهید روش‌های تیلور مرتبه ۴ و رانگ–کوتای ۴ مرحله‌ای در حل این مسئله به رابطه بازگشتی یکسانی برای y_{n+1} خواهند رسید (طول گام را h در نظر بگیرید).

۷. مسئله $\begin{cases} y' = 2t(1+y), & t \in [0,1], \\ y(0) = 0. \end{cases}$ را در نظر بگیرید. جدول زیر را با استفاده از روش رانگ–کوتای دومرحله‌ای کامل کنید. (محاسبات با دقت $5D$)

n	t_n	k_1	k_2	w_n
۰	…	…	…	…
۱	…	…	…	…
۲	…	…	…	…
۳	…	…	…	…
۴	…	…	…	…

۸. برای مسئله مقدار اولیه $\begin{cases} y' = e^x + y, & x \in [0, 0/2], \\ y(0) = 1 \end{cases}$ مقدار تقریبی $y(0/2)$ را با انتخاب $h = 0/1$ و به کمک روش رانگ–کوتای دو مرحله‌ای بیابید (محاسبات با چهار رقم بامعنا).

۹. معادله دیفرانسیل $y' = xy$ را با مقدار اولیه $y(1) = 1$ در نظر بگیرید. در یک گام مقدار $y(1/1)$ را به روش‌های تیلور مرتبه ۴ و رانگ – کوتای چهار مرحله‌ای تقریب بزنید. (محاسبات با چهار رقم بامعنا).

کتاب‌نامه

[۱] بهفروز غلامحسین و میرنیا میرکمال، نظریه و کاربرد آنالیز عددی، نوشته تیلور و دیگران (ترجمه)، مرکز نشر دانشگاهی، ۱۳۷۳.

[۲] عالم‌زاده علی اکبر، بابلیان اسماعیل و امیدوار محمدرضا، آنالیز عددی، نوشته بردن و دیگران (ترجمه)، انتشارات منصوری، ۱۳۶۸.

[3] American Heritage Dictionary, 1992.

[4] Chambers 20th Century Dictionary, 1983.

[5] Henrichi P., Elements of numerical analysis, 1964.

[6] Kincaid D. and Cheney E. W., Numerical analysis, Mathematics of scientific computing, 1991.

[7] Scarborough J.B., Numerical Mathematical Analysis, 1930.

[8] Traub J., Communications of the ACM, 1972.

[9] Trefethen L. N., SIAM News, November 1992.

[10] Webster's New Collegiate Dictionary, 1973.