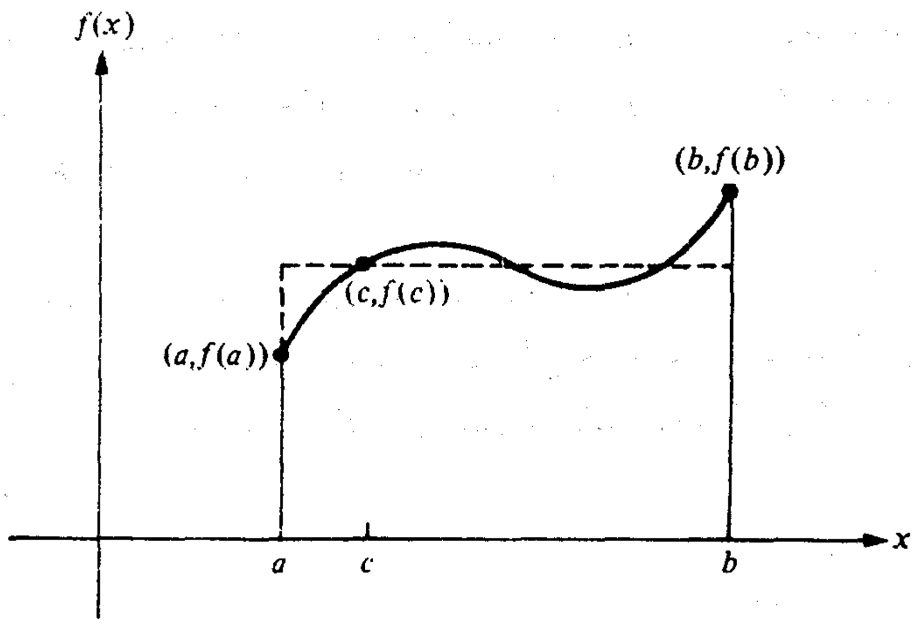
بعلاوه، هرگاه f در هرنقطه از $[a, b]$ مشتقپذیر باشد، آنگاه، به‌ازای هر $i = 1, 2$ ، $c_i = a$ ، یا $c_i = b$ ، یا $f'(c_i) = 0$.

در بررسی روشهای عددی به دونتیجهٔ دیگر نیاز داریم . اولی تعمیمی از قضیهٔ معمولی مقدار میانگین برای انتگرالها است . این قضیه، وقتی $g(x) \equiv 1$ ، **مقدار متوسط** تابع را روی بازهٔ $[a, b]$ بدست می‌دهد (ر.ک. شکل ۱۰۴).



شکل ۱۰۴

برهان قضیهٔ ۱۰۰۱ عموماً" در یک درس حساب دیفرانسیل و انتگرال پایه نمی‌آید، اما می‌توان آن را در هر کتاب حساب دیفرانسیل و انتگرال پیشرفتهٔ متعارف یافت (مثلاً"، ر.ک. فولکس[1] [۲۹] ، صفحهٔ ۱۲۵).

قضیهٔ ۱۰۰۱ (قضیهٔ مقدار میانگین و زندار برای انتگرالها). هرگاه $f \in C[a, b]$ ، g بر $[a, b]$ **انتگرالپذیر باشد، و** $g(x) \geq 0$ ، **آنگاه عددی مانند** c ، **که** $a < c < b$ ، **وجود دارد بطوری که**

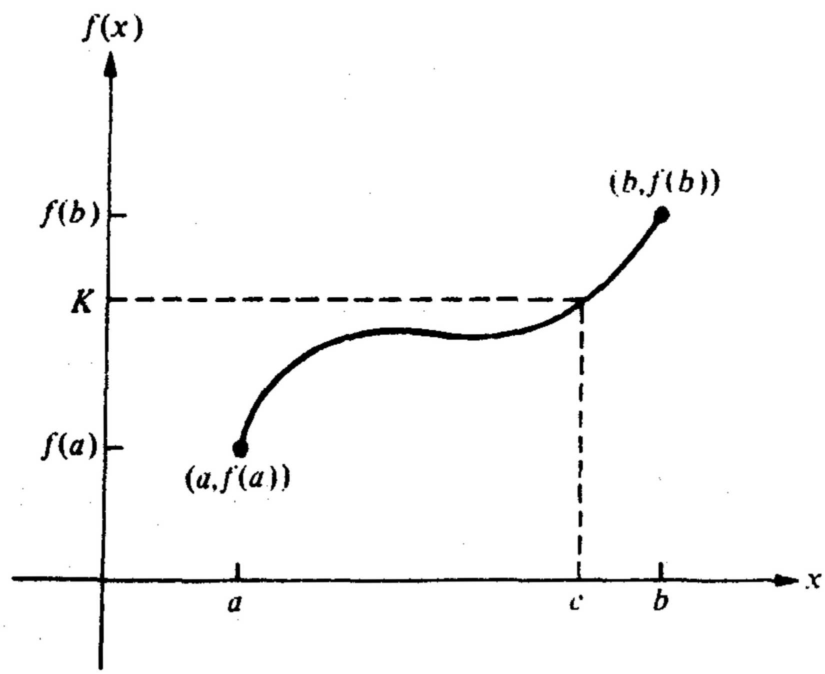
1. Fulks

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

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

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

قضیهٔ بعدی قضیهٔ مقدار میانی است. باآنکه شهودا" قضیهٔ واضحی است، اثباتش از حوصلهٔ درس حساب دیفرانسیل و انتگرال معمولی خارج است. برهان آن را می‌توان در اغلب کتب حساب دیفرانسیل و انتگرال پیشرفته یافت (مثلا"، ر.ک. فولکس[۲۹]، صفحهٔ ۵۹).



شکل ۵۰۱

قضیهٔ ۱۲۰۱ (قضیهٔ مقدار میانی) . هرگاه $f \in C[a, b]$ و K عددی بین $f(a)$ و $f(b)$ باشد ، آنگاه نقطه‌ای مانند c در (a, b) وجود دارد که به‌ازای آن $f(c) = K$ (ر.ک. شکل ۵۰۱) .

مثال ۱ . برای اثبات اینکه $x^5 - 2x^3 + 3x^2 - 1 = 0$ بر بازهٔ $[0, 1]$ جواب دارد ، تابع $f(x) = x^5 - 2x^3 + 3x^2 - 1$ را درنظر می‌گیریم . واضح است که f بر $[0, 1]$ پیوسته است و $f(0) = -1$ درحالی که $f(1) = 1$. چون $f(0) < 0 < f(1)$ ، قضیهٔ مقدارمیانی ایجاب می‌کند که عددی مانند x ، که $0 < x < 1$ ، وجود داشته باشد که به‌ازای آن $x^5 - 2x^3 + 3x^2 - 1 = 0$.

همانطور که در این مثال ملاحظه شد ، قضیهٔ مقدار میانی از جهت کمک در تعیین اینکه چه‌وقت مسائل معینی جواب دارند مهم است . لکن ، برای یافتن این جوابها راهی ارائه نمی‌دهد . این مطلب در فصل ۲ کاملتر بحث خواهد شد .

آخرین قضیهٔ این مرور از حساب دیفرانسیل و انتگرال بسط تیلور[1] را توصیف می‌کند . از اهمیت بسط تیلور در آنالیز عددی هرچه بگوییم کم است ، و نتیجهٔ زیر بارها مورد استفادهٔ ما قرار می‌گیرد . به‌اثبات این قضیه ، در بخش ۲۰۳ ، تمرین ۱۱ ، اشارهٔ مختصری خواهد شد .

قضیهٔ ۱۳۰۱ (قضیهٔ تیلور) . فرض‌کنیم $f \in C^n[a, b]$ و $f^{(n+1)}$ بر $[a, b]$ موجود باشد . همچنین ، $x_0 \in [a, b]$. دراین صورت ، به‌ازای هر $x \in [a, b]$ ، نقطه‌ای مانند $\xi(x) \in (a, b)$ وجود دارد بطوری که

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

که درآن

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

---

1. Taylor

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

و

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

دراینجا $P_n(x)$ **چندجمله‌ای تیلور درجهٔ** n م f حول x_0 ، و $R_n(x)$ **جملهٔ باقیمانده (یا خطای برشی)** وابسته به $P_n(x)$ نامیده می‌شود .

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

مثال ۲ . فرض‌کنیم $f(x) = \cos x$. چون $f \in C^{\infty}(R)$ ، می‌توان قضیهٔ ۱۳۰۱ را به ازای هر $n > 0$ بکار برد . به‌ازای $n = 2$ و $x_0 = 0$ ، قضیهٔ ۱۳۰۱ نتیجه می‌دهد که

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

که درآن $\xi(x)$ عددی بین 0 و x است .

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

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

که درآن $0 < \xi(x) < .001$. (تیره روی آخرین‌رقم در .166 نشان آن است که این رقم بی‌نهایت بار تکرار می‌شود .)

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

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

و ، درواقع ، دقتی تا 13 رقم اعشار وجود دارد .

دراین مثال ، اگر چندجمله‌ای تیلور درجهٔ سوم به‌ازای $x_0 = 0$ بکار رود ، چون $f'''(0) = 0$ ،

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

که درآن $.001 > \xi(x) > 0$. چندجمله‌ای تقریب ساز همان چندجمله‌ای قبلی است، و تقریب همان $.9999995$ است، لیکن دراینجا نیز انتظار دقت تا 13 رقم اعشار می‌رود، چرا که

$$|\tfrac{1}{24}x^4 \cos \xi(x)| \leq \tfrac{1}{24}(.001)^4(1) \approx 4.2 \times 10^{-14}.$$

این مقدار به مقدار واقعی نزدیکتر است.

مجموعه تمرینات ۱۰۱

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

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

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

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

۵ . تعیین کنید کـه قضیهٔ مقدار میانگین (قضیهٔ ۸۰۱) را درحالات زیر می‌توان بکار برد یا نه. اگرمی‌توانید، نقطهٔ c را که دراین قضیه صدق می‌کند بیابید. درغیر این‌صورت، نشان دهید چرا چنین نقطه‌ای وجود ندارد:

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

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

۶ . با استفاده‌ازقضیهٔ تیلور (قضیهٔ ۱۳۰۱)، به‌ازای $n = 2$ و $x_0 = 0$ ، تقریبی برای $\sin .001$ بدست آورید. آیا انتظار دارید دقت تقریب دراین مسئله همان دقت حاصل برای $\cos .001$ در مثال ۲ باشد؟

۷ گوییم تابع $f:[a,b] \to R$ بر $[a,b]$ در شرط لیپ شیتس[1]، و با ثابت لیپ شیتس L ، صدق می‌کند اگر، به‌ازای هر $x, y \in [a,b]$ ،

---

1. Lipschitz

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

(آ) نشان دهید کـه اگر f در شرط لیپ شیتس، و با ثابت لیپ شیتس L ، بر بازهٔ $[a, b]$ صدق کند ، $f \in C[a, b]$.

(ب) نشان دهید که اگر مشتق f بر $[a, b]$ بهوسیلهٔ L کراندار باشد ، f بر $[a, b]$ در شرط لیپ شیتس، و با ثابت لیپ شیتس L ، صدق می‌کند .

(پ) تابعی مثال بزنید که بریک بازهٔ بسته پیوسته باشد ولی در شرط لیپ شیتس برآن بازه صدق نکند .

۲.۱ خطاهای گردکردن وحساب کامپیوتر

وقتی در محــاسبات عــددی از ماشین حساب و یا کامپیوتر رقمی استفاده می‌شود ، باید خطایی غیــرقابل اجتناب ، به‌نام **خطای گردکردن**، درنظر گرفته شود . این خطا به این دلیل‌رخ می‌دهد که حساب ماشین با اعدادی صورت می‌گیرد که فقط تعدادی متناهی رقم دارند ؛ و درنتیجه ، خیلی‌از محاسبات با نمایشهای تقریبی از این اعداد انجام می‌شود . در یک کامپیوتر نوعی ، برای نمایش همهٔ اعداد حقیقی فقط زیرمجموعهٔ نسبتا" کوچکی از دستگاه اعــداد حقیقی بکار می‌رود . این زیرمجموعه فقط شامل اعداد گویا ، هم مثبت و هم منفی ، اســت ، و یک جــزء کسری ، به‌نام **مانتیس**، همراه با یک جزء نمایی ، به نام **مشخصه** ، را ذخیره می‌کند . مثلا" ، یک عدد با دقت معمولی و با ممیز سیار که در IBM سری 360 یا 370 بکار می‌رود از یک مانتیس 24 رقمی دودویی و یک نمای 7 رقمی دودویی با پایهٔ 16 تشکیل شده است . چون 24 رقم دودویی متناظر 7 تا 8 رقم اعشاری است ، می‌توان فرض کرد که IBM سری 360 یا 370 برای دستگاه اعداد با ممیز سیار دقتی دست کم تا هفــت رقم اعشار دارد . نمای هفت رقم دودویی برد 0 تا 127 را بدست می‌دهـد ، اما ، بخاطر پیشقدر نمایی ، برد 64– تا 63+ است ؛ یعنی ، 64 خودبخود از نمای ثبت شده کم می‌شود . عدد ماشینی

0	1000010	101100110000010000000000

دقیقا" نمایش عدد اعشاری

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

است، زیرا اولین رقم دودویی نمایش علامت است، 0 برای بعلاوه و 1 برای منها، هفت رقم دودویی بعدی نمایش نما است، و بیست و چهار رقم دودویی آخر نمایش مانتیس است و در واقع برای نمایش هر عدد در بازهٔ

$$[179.0156173706054687 5, 179.0156326293945312 5]$$

بکار می‌رود.

در این نمایش، کوچکترین عدد مثبت قابل بیان $16^{-64} \approx 10^{-77}$، و بزرگترین عدد $16^{63} \approx 10^{75}$ است. چون دست‌کم یکی از سمت چپ‌ترین چهار رقم دودویی هر عدد ناصفر بزرگتر از 16^{-64} باید یک باشد، $15 \cdot 2^{28}$ عدد به شکل

$$\pm .d_1 d_2 \cdots d_{24} \times 16^{e_1 e_2 \cdots e_7} \qquad (۱ \cdot ۱)$$

وجود دارد، که در این دستگاه برای نمایش جمیع اعداد حقیقی بکار می‌روند. در محاسبات، اعداد کوچکتر از 16^{-64} منتج به **پاریز** شده و اغلب آنها را صفر می‌گیرند، درحالی‌که اعداد بزرگتر از 16^{63} منتج به **سرریز** شده و باعث توقف محاسبات می‌گردند.

دستگاه نمایش عددی فوق در همهٔ ماشینهای کامپیوتر مرسوم نیست، لیکن مشکلات احتمالی را تا حدودی نشان می‌دهد. تا آخر این بحث، بخاطر سادگی، فرض می‌کنیم نمایش اعداد ماشینی به‌شکل اعشاری نرمال شدهٔ

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

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

نمایش یک عدد حقیقی دلخواه با ممیز سیار و به‌شکل (۱ ۰ ۱) سودمند است. هر عدد حقیقی و مثبت y را، درصورتی که در برد عددی ماشین باشد، می‌توان به شکل نرمال شدهٔ

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

درآورد. شکل با ممیز سیار (۱ ۰ ۱)، که با $fl(y)$ نموده می‌شود، از مختوم کردن مانتیس y به k رقم اعشار بدست می‌آید. برای انجام این مختومیت دوراه وجود دارد. یک روش جدا کردن ارقام $d_{k+1} d_{k+2} \cdots$ است که بدست می‌دهد

$$fl(y) = .d_1 d_2 \cdots d_k \times 10^n,$$

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

$$fl(y) = .\delta_1 \delta_2 \cdots \delta_k \times 10^n.$$

روش دوم را اغلب گردکردن عدد می‌نامند، چراکه، اگر $d_{k+1} \geq 5$ ، به d_k یک می‌افزاییم تا $fl(y)$ بدست آید و، اگر $d_{k+1} < 5$ ، همهٔ ارقام جز k رقم اول را جدا می‌کنیم مثلاً، اگر $k = 5$ و روش گرد کردن بکار رود، π و e به ترتیب با $.31416 \times 10^1$ و $.27183 \times 10^1$ نمایش داده می‌شوند. و اگر $k = 4$ و از روش جدا کردن استفاده‌شود، $\frac{1}{3}$ و $\frac{1}{7}$ به ترتیب با $.3333$ و $.1428$ نموده می‌شوند.

چون اعداد حقیقی همیشه‌دقیقاً در ماشین قابل نمایش نیستند، لازم است‌خطای ناشی از این تقریب متناهی رقم درنظر گرفته شود. تعریف زیر طرز اندازه‌گیری این خطا را مشخص می‌کند.

تعریف ۱۴۰۱ . هرگاه p^* تقریبی از p باشد، **خطای مطلق** مساوی $|p - p^*|$ و **خطای نسبی** برابر $(|p - p^*|/|p|)$ ، به‌شرط $p \neq 0$ ، است.

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

مثال ۱

(آ) هرگاه $p = .3000 \times 10$ و $p^* = .3100 \times 10$ ، خطای مطلق $.1$ و خطای نسبی $.333\bar{3} \times 10^{-1}$ است.

(ب) هرگاه $p = .3000 \times 10^{-3}$ و $p^* = .3100 \times 10^{-3}$ ، خطای مطلق $.1 \times 10^{-4}$ و خطای نسبی $.333\bar{3} \times 10^{-1}$ است.

(پ) هرگاه $p = .3000 \times 10^4$ و $p^* = .3100 \times 10^4$ ، خطای مطلق $.1 \times 10^3$ و خطای نسبی $.333\bar{3} \times 10^{-1}$ می‌باشد.

این مثال نشان می‌دهد که یک خطای نسبی، یعنی $.333\bar{3} \times 10^{-1}$ ، برای چند خطای مطلق کاملاً متفاوت رخ می‌دهد. درنتیجه، خطای مطلق، به‌عنوان شاخصی از دقت، ممکن‌است گمراه کننده و خطای نسبی مناسبتر باشد همانطور که تعریف ذیل‌نشان می‌دهد، با استفاده از خطای نسبی، می‌توان درباب عدهٔ ارقام درست یک تقریب با نمایش سخن گفت.

تعریف ۱۵۰۱ . گوییم p^* ، p را تا t **رقم** (یا پیکر) با معنی تقریب می‌کند اگر t بزرگترین عدد صحیح نامنفیی باشد که به‌ازای آن

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

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

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

صدق کند . درنتیجه ، باید $999.5 < p^* < 1000.5$ ، که با تعریف شهودی ارقام با‌معنی مطابقت‌دارد . اگر همین مسئله را برای 5000 و 9990 درنظر بگیریم ، p^* باید به‌ترتیب در $5002.5 > p^* > 4997.5$ و $9985.005 < p^* < 9994.995$ صدق کنـد تا اینکه تا چهار رقم با‌معنی دقیق باشد . این ممکن است با مفهوم شهودی ارقام با معنی مطابقت نداشته باشد . بهـرحال، توجه کنید که برای آنکه p^* یک تقریب 10,000 با چهار رقم با معنی باشد ، p^* باید در $9995 < p^* < 10005$ صدق‌کند ، که‌مجدداً " با وضع شهودی‌مطابقت دارد . جدول زیر سرشت پیوستهٔ این‌مفهوم را با ذکر کوچکترین کران بالایی $|p-p^*|$ ، به‌ازای مقادیر مختلف p ، یعنی $\max|p-p^*|$ ، وقتی p^* با p تا چهار رقم با‌معنی مطابقت دارد ، نشان می‌دهد .

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

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

می‌تواند به کتب علوم کامپیوتر، نظیر چو [1] [۲۰] *، مراجعه کند.)

فرض کنیم نمایش با ممیز سیار $fl(y)$، به‌ازای عدد حقیقی

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

چنین باشد:

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

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

فرض کنیم حساب k رقمی عبارت باشد از

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

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

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

$$x \oplus\!\!\!\!\div\, y = fl(fl(x) \div fl(y)).$$

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

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

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

عمل	نتیجه	مقدار واقعی	خطای نسبی
$x \oplus y$	$.10476 \times 10^1$	$\frac{22}{21}$	$.182 \times 10^{-4}$
$y \ominus x$	$.38095$	$\frac{8}{21}$	$.625 \times 10^{-5}$
$x \otimes y$	$.23809$	$\frac{5}{21}$	$.220 \times 10^{-4}$
$y \oplus\!\!\!\!\div\, x$	$.21428 \times 10^1$	$\frac{15}{7}$	$.267 \times 10^{-4}$

جدول ۱۰۱

1. Chu

* *Introduction to Computer Organization.*

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

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

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

اما خطای نسبی محاسبات برای این حساب درحد معقول و تقریباً 3×10^{-4} است.

از دست دادن ارقام با معنی در بعضی از اعمال حسابی نیز رخ می‌دهد، زیرا، همانطورکه مثال زیر نشان می‌دهد، اعمال با تعداد ثابتی رقم انجام می‌شوند. (همچنین، ر.ک. تمرین ۸.)

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

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

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

اگر در بدست آوردن این چهار رقم جدا کردن ساده‌ای انجام گیرد، تقریبهای چهاررقمی به p، q، و r به‌ترتیب $p^* = .5461$، $q^* = .5460$، و $r^* = p^* - q^* = .0001$ می‌باشند. درنتیجه،

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

که هنوز یک رقم با معنی دقت دارد .

ازدست دادن ارقام با معنی اغلب با تنظیم مجدد مسئله قابل‌اجتناب است .

مثال ۳. بااستفاده‌ازحساب سه‌رقمی، $f(x) = x^3 - 6x^2 + 3x - .149$ را در $x = 4.71$ محاسبه کنید .

	x	x^2	x^3	$6x^2$	$3x$
دقیق	4.71	22.1841	104.487111	133.1046	14.13
سه‌رقمی (جداکردن)	4.71	22.1	104.	132.	14.1
سه‌رقمی (گردکردن)	4.71	22.2	105.	133.	14.1

توجه کنید که اولین مقادیر که با حساب 3 رقمی بدست آمده‌اند فقط سه رقم پیشرو را حفظ می‌کنند، بدون هیچ‌گردکردن، و با مقادیرگرد شدهٔ سه‌رقمی تفاوت آشکاردارند .

دقیق
$$f(4.71) = 104.487111 - 133.1046 + 14.13 - .149 = -14.636489;$$

سه‌رقمی (جداکردن)
$$f(4.71) = 104. - 132. + 14.1 - .149 = -14.0;$$

سه‌رقمی (گردکردن)
$$f(4.71) = 105. - 133. + 14.1 - .149 = -14.0.$$

خطای نسبی در هر دو روش سه رقمی برابر است با

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

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

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

که نتیجه می‌دهد

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

و جواب گرد شدهٔ 3 رقمی 14.6– است . خطاهای نسبی جدید عبارت خواهند بود از

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

و

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

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

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

برای ملاحظهٔ اثرخطای‌گردکردن بر محاسبات متوالی ، شامل محاسبهٔ تابعی ، فرض کنیم p^* تقریب p باشد . همچنین ، فرض‌کنیم محاسبه با $g(p)$ نموده شده باشد ، که ، بخاطر این‌بررسی ، g را بربازه‌ای شامل p و p^* دوبار مشتقپذیر می‌گیریم . بنابرقضیهٔ تیلور (قضیهٔ ۱۳۰۱) ،

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

که‌درآن ξ بین p و p^* قراردارد . اگر $(p - p^*)^2$ در مقایسه با $|p - p^*|$ کوچک‌باشد ،

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

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

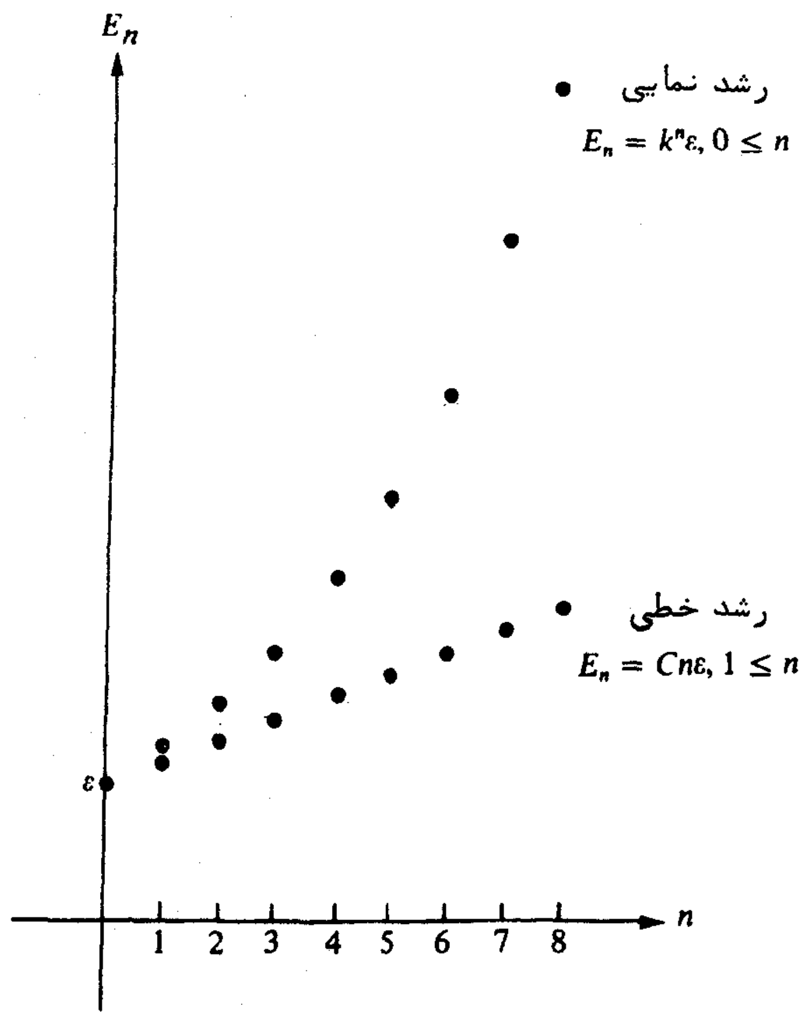
مثال ۴ . فرض‌کنیم p^* یک تقریب p باشد و در محاسبهٔ $g(p) = p^{3.2}$ بکار رفته باشد . طبق تذکرات فوق ،

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

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

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

تعریف ۱۶۰۱ . هرگاه $|E_n| \approx Cn\varepsilon$ ، که درآن C ثابتی مستقل از n است ، رشد خطا را **خطی** می‌نـامیم . هرگـاه ، به‌ازای $k > 1$ ای ، $|E_n| \approx k^n\varepsilon$ ، رشد خطا **نمایی** خوانده می‌شود .(ر .ک . شکل ۶۰۱ .)



شکل ۶۰۱

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

مثال ۵. دنبالهٔ $1, \frac{1}{3}, \frac{1}{9}, \frac{1}{27}, \frac{1}{81}, \ldots$ از اعداد را درنظر می‌گیریم. این دنباله را می‌توان با تعریف $p_0 = 1$ ، $p_1 = \frac{1}{3}$ ، و

$$p_n = \frac{10}{3}p_{n-1} - p_{n-2} \qquad (۲\cdot۱)$$

تولید کرد. اگر $\{p_n\}$ با حساب پنج رقمی و گرد کردن تولید شود، مقادیر جدول ۲۰۱ بدست می‌آیند.

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

جدول ۲۰۱

$p_n = C_1(\frac{1}{3})^n + C_2(3)^n$ ، به ازای هر دو عدد حقیقی C_1 و C_2 ، در معادلهٔ (۲۰۱) صدق می‌کند. برای آنکه $p_0 = 1$ و $p_1 = \frac{1}{3}$ ، باید $C_1 = 1$ و $C_2 = 0$ اختیار شوند. چون نمایش پنج رقمی عبارت است از $\hat{p}_0 = 1.0000$ و $\hat{p}_1 = .33333$ ، جواب تولید شده به وسیلهٔ معادلهٔ (۲۰۱) عملا" دارای $C_1 = 1.0000$ و $C_2 = -.12500 \times 10^{-5}$ است. بنابراین، خطای گرد کردن در نمایش $\hat{p}_1 = \frac{1}{3}$ مساوی

$$(-.125 \times 10^{-5})(3)^n$$