Wstęp do metod numerycznych

8. Całkowanie numeryczne

P. F. Góra

http://th-www.if.uj.edu.pl/zfs/gora/

2019/20

Całka oznaczona

Obliczanie całek oznaczonych jest (lub też raczej, było) jednym z głównych zadań analizy numerycznej. Jak wiadomo,

$$I = \int_a^b f(x)\,dx = F(b) - F(a)\,, \tag{1}$$

gdzie $F(x)$ jest *funkcją pierwotną* $f(x)$, to znaczy taką, że $\frac{dF}{dx} = f(x)$. Problem w tym, że znajdywanie funkcji pierwotnych ("całek nieoznaczonych") jest zadaniem znacznie trudniejszym, niż różniczkowanie. Dla wielu funkcji funkcja pierwotna nie wyraża się poprzez skończoną kombinację funkcji elementarnych. Stąd bierze się potrzeba numerycznego obliczania całek.

Krzywa ROC

W uczeniu maszynowym, w ocenie klasyfikatorów binarnych — czyli zaliczających badane obiekty do jednej z dwu rozłącznych klas — dużą rolę odgrywa *Receiver Operating Characteristic* — wielkość zwana popularnie "krzywą ROC". Wyniki klasyfikatora binarnego mogą należeć do jednej z czterech grup

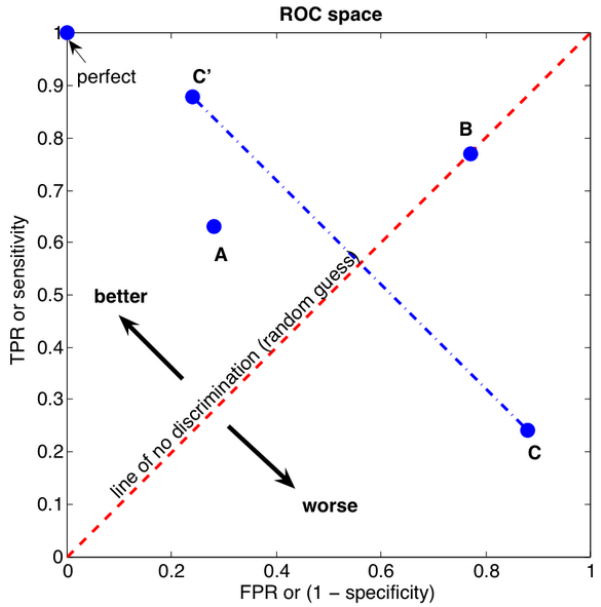
TP, *True positive* Badany obiekt posiada daną cechę i klasyfikator prawidłowo rozpoznaje, że obiekt ją posiada.

FP, *False positive* Badany obiekt nie posiada danej cechy, ale klasyfikator błędnie stwierdza, że obiekt ją posiada.

FN, *False negative* Badany obiekt posiada daną cechę, ale klasyfikator błędnie stwierdza, że obiekt jej nie poisada.

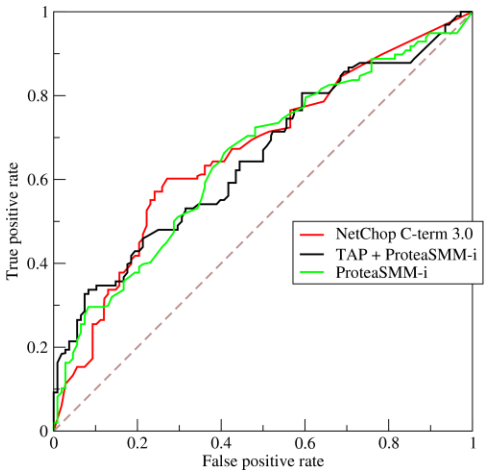
TN, *True negative* Badany obiekt nie poisada danej cechy i klasyfikator prawidłowo stwierdza, że obiekt jej nie posiada.

Z uwagi na zastosowania w diagnostyce medycznej, ale nie tylko, obecność błędów typu FN jest szczególnie niemile widziana. Wyniki działania klasyfikatora możemy przedstawić w przestrzeni ROC:



T — doskonałe działanie. A jest lepsze od B, C. B odpowiada losowemu działaniu klasyfikatora. C′, odbicie lustrzane C (odwrócenie działania klasyfikatora), jest jeszcze lepsze, niż A.

Jeżeli mamy do czynienia nie z pojedynczym klasyfikatorem, ale z rodziną klasyfikatorów, zależnych od pewnego parametru, mogącego wpływać na częstotliwość błędów typu FP, wyniki ich działania tworzą krzywe w przestrzeni ROC.



Pole pod krzywą (AUC — *Area Under Curve* ☺) jest miarą prawdopodobieństwa, że klasyfikator oceni wyżej losowo wybrany przypadek pozytywny (obiekt ma daną cechę) niż losowo wybrany przypadek negatywny

(obiekt nie ma danej cechy). Stąd pojawia się konieczność *numerycznego* obliczania AUC, czyli numerycznego całkowania krzywej ROC.

Przykład ten pokazuje, że choć całkowanie numeryczne ma główne zastosowania w obliczeniach inżynierskich i naukowych, występuje także w uczeniu maszynowym, a także w analizie sygnałów, analizie statystycznej i w innych zastosowaniach informatyki.

Caveat emptor!

Numerycznie wolno obliczać całki, o których wiemy, że istnieją. To, że jakaś procedura numeryczna daje skończony (a nawet pozornie sensowny) wynik, nie stanowi dowodu, że obliczana całka istnieje.

Przykład NEGATYWNY

Mamy numerycznie obliczyć całkę

$$I = \int_1^\infty \frac{dx}{x}\,. \tag{2}$$

Naiwnie przyjmujemy, że dzielimy przedział całkowania na podprzedziały o małej długości h każdy, bierzemy wartość w prawym krańcu przedziału i w duchu sum Riemanna przyjmujemy, że

$$I \simeq \sum_{n=1}^{M} \frac{1}{1+nh} \cdot h \quad \text{dla } M \gg 1. \tag{3}$$

Wartość M w (3) powiększamy tak długo, aż kolejne przyczynki nie staną się zaniedbywalnie małe. Tymczasem dla każdej skończonej dokładności obliczeń, przyczynki (kolejne wyrazy $1/(1+nh)$) dla odpowiednio dużych n staną się zaniedbywalnie małe, to znaczy porównywalne z błędem zaokrąglenia. Dostaniemy w ten sposób jakieś skończone "przybliżenie" wartości I, a tymczasem całka (2) jest rozbieżna!

Punkt wyjścia: interpolacja

Wzory na całkowanie przybliżone, tak zwane *kwadratury*, uzyskuje się przez całkowanie odpowiednich wielomianów interpolacyjnych. Ogólnie, jeśli

$$f(x) = y(x) + E(x)\,, \tag{4}$$

gdzie $y(x)$ jest wielomianem stopnia n postaci

$$y(x) = \sum_{i=0}^{n} h_i(x) f_i + (\text{ewentualne człony z pochodnymi}), \tag{5}$$

zaś $E(x)$ jest błędem interpolacji, jako przybliżenie całki otrzymujemy

$$\int_a^b f(x)\,dx = \sum_{i=0}^{n} H_i f_i + E\,, \tag{6}$$

gdzie

$$H_i = \int_a^b h_i(x)\,dx\,, \qquad E = \int_a^b E(x)\,dx\,. \tag{7}$$

Należy przy tym tak dobrać parametry interpolacji, aby całki z ewentualnych członów z pochodnymi znikały tożsamościowo. W ten sposób z interpolacji Hermite'a uzyskuje się tak zwane *kwadratury Gaussa*, których nie będziemy omawiać, natomiast z całkowania wielomianu interpolacyjnego Lagrange'a z węzłami równoodległymi uzyskuje się *kwadratury Newtona–Cotesa*. Jeżeli krańce przedziału całkowania są węzłami interpolacji — zwanymi w tym kontekscie węzłami kwadratury — dostajemy tak zwane zamknięte kwadratury Newtona–Cotesa, znajdujące najczęstsze zastosowania.

Błąd całkowania, E, powinien być proporcjonalny do pochodnej rzędu $(n+1)$ funkcji podcałkowej w pewnym (nieznanym) punkcie przedziału całkowania. Okazuje się jednak, iż ze względu na pewne warunki symetrii, wyrażenia na błąd zawierają tylko pochodne rzędu parzystego — kwadratury oparte na nieparzystej ilości węzłów, czyli wyprowadzone z całkowania wielomianu iterpolacyjnego parzystego stopnia ($n = 2, 4, \ldots$ — węzły numerujemy od zera), mają rząd wyższy, niżby to wynikało z rzędu interpolacji.

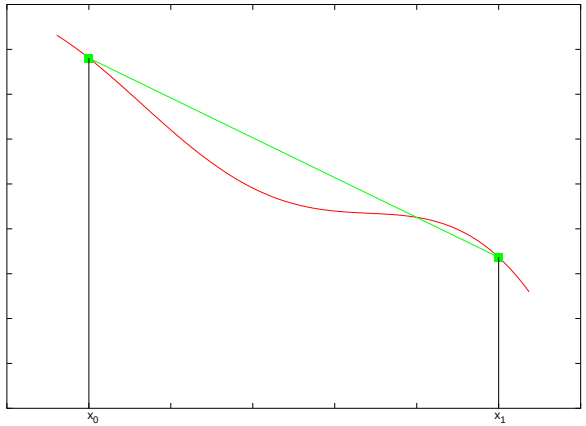
Kwadratury Newtona-Cotesa

W praktyce zamknięte kwadratury Newtona–Cotesa opiera się na interpolacji wielomianami niskiego stopnia. Jest to spowodowane czterema względami:

- prostostą obliczeniową,
- trudnością w szacowaniu pochodnych wysokiego rzędu,
- obawą przed oscylacjami Rungego, czyli „szalonym” zachowaniem wielomianów interpolacyjnych wysokiego rzędu,
- możliwoscią osiągnięcia większej dokładności przez zastosowanie kwadratur złożonych (patrz niżej).

Poniżej przedstawiam cztery najczęściej stosowane kwadratury: metodę trapezów, metodę Simpsona, metodę 3/8 i metodę Milne’a. Kwadratury wyższego rzędu występują w zastosowaniach praktycznych *niezmiernie* rzadko.

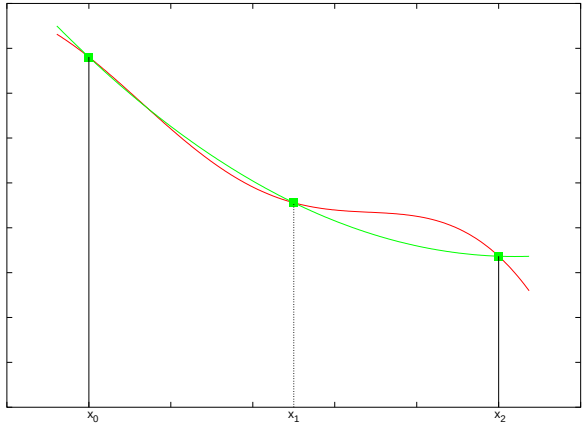
Metoda trapezów



$$\int_a^b f(x)\,dx \simeq \frac{b-a}{2}(f_0+f_1)$$

$$E = -\frac{1}{12}(b-a)^3 f''(\zeta)$$

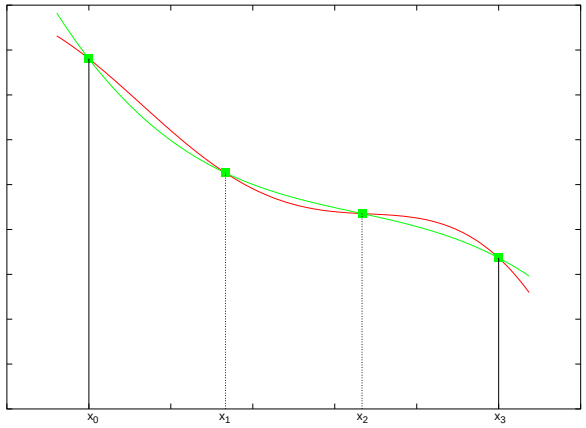
Metoda Simpsona



$$\int_a^b f(x)\,dx \simeq \frac{b-a}{6}(f_0+4f_1+f_2)$$

$$E = -\frac{1}{90}\left(\frac{b-a}{2}\right)^5 f^{(4)}(\zeta)$$

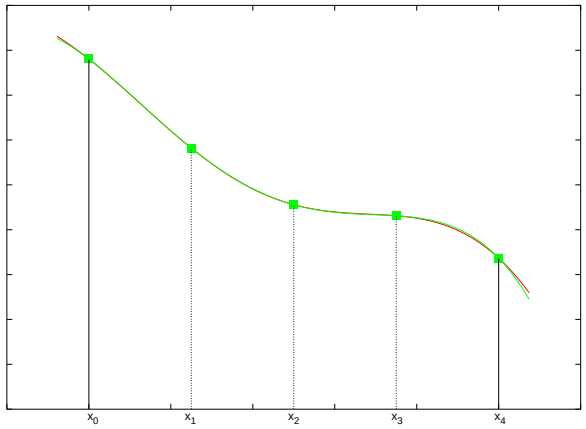
Metoda 3/8



$$\int_a^b f(x)\,dx \simeq \frac{b-a}{8}(f_0 + 3f_1 + 3f_2 + f_3)$$

$$E = -\frac{3}{80}\left(\frac{b-a}{3}\right)^5 f^{(4)}(\zeta)$$

Metoda Milne'a



$$\int_a^b f(x)\,dx \simeq \frac{b-a}{90}(7f_0 + 32f_1 + 12f_2 + 32f_3 + 7f_4)$$

$$E = -\frac{8}{945}\left(\frac{b-a}{4}\right)^7 f^{(6)}(\zeta)$$

Na powyższych rysunkach czerwone krzywe oznaczają “prawdziwą” funkcję podcałkową, natomiast krzywe zielone — jej kolejne wielomiany interpolacyjne. We wzorach na błąd kwadratury $\zeta \in [a, b]$.

Wzór na kwadraturę przybliżoną podaje pole pod wykresem odpowiedniego wielomianu interpolacyjnego, zgodnie z geometryczną interpretacją całki.

(Uwaga: W przypadku ogólnym wielomian interpolacyjny czwartego stopnia zastosowany we wzorze Milne’a niekoniecznie *aż tak dobrze* przybliża funkcję podcałkową.)

Kwadratury złożone

Zamiast stosować kwadratury wyższego rzędu, większą dokładność w numerycznym obliczaniu całek uzyskuje się stosując kwadratury złożone, to jest dzieląc przedział całkowania na podprzedziały i stosując do każdego z nich kwadraturę niższego rzędu. Procedurę tę można iterować, to znaczy sukcesywnie zagęszczać podział. Ponieważ najbardziej „kosztowną” częścią całkowania numerycznego jest obliczanie funkcji podcałkowej, podziały zagęszcza się w ten sposób, aby węzły podziału grubszego były też węzłami podziału gęstszego — w ten sposób bowiem można użyć *już obliczonych* wartości funcji. Szczególnie łatwo jest to osiągnąć, gdy krańce przedziału całkowania są węzłami kwadratury. Stąd właśnie bierze się popularność zamkniętych wzorów Newtona–Cotesa.

Redukcja błędu

Co ciekawe, stosowanie kwadratur złożonych prowadzi do dodatkowego zmniejszenia błędu. Rozpatrzmy to na przykładzie złożonego wzoru trapezów. Obliczamy całkę (1) i otrzymujemy błąd

$$E = -\frac{1}{12}(b-a)^3 f''(\zeta_0)\,, \tag{8}$$

gdzie $\zeta_0 \in [a,b]$. Teraz korzystamy z addytywności całki:

$$I = \int_a^b f(x)\,dx = \int_a^{(a+b)/2} f(x)\,dx + \int_{(a+b)/2}^b f(x)\,dx\,, \tag{9}$$

obliczamy przy pomocy wzoru trapezów całki po podprzedziałach $[a,(a+b)/2]$, $[(a+b)/2,b]$, zaś jako błąd bierzemy sumę błędów popełnionych

w obu podprzedziałach:

$$\begin{aligned}\widetilde{E} &= -\frac{1}{12}\left(\frac{a+b}{2}-a\right)^3 f''(\zeta_1) - \frac{1}{12}\left(b-\frac{a+b}{2}\right)^3 f''(\zeta_2) \\ &= -\frac{1}{4}\cdot\frac{1}{12}(b-a)^3\frac{f''(\zeta_1)+f''(\zeta_2)}{2}\,, \end{aligned} \tag{10a}$$

gdzie $\zeta_1 \in [a,(a+b)/2]$, $\zeta_2 = \in [(a+b)/2,b]$. Skorzystawszy z twierdzenia o wartości średniej zastosowanego do drugiej pochodnej funkcji podcałkowej, otrzymujemy ostatecznie

$$\widetilde{E} = -\frac{1}{4}\cdot\frac{1}{12}(b-a)^3 f''(\zeta_3)\,, \tag{10b}$$

gdzie $\zeta_3 \in [a,b]$. Porównując wzory (8) i (10b), widzimy, iż zagęszczenie podziału czterokrotnie zmniejszyło czynnik stały w wyrażeniu na błąd metody trapezów*.

*Nie można natomiast powiedzieć, iż wzory (8) i (10b) „różnią się o czynnik 1/4”, jako że pochodne występujące w tych wzorach obliczane są w innych punktach.

Całkowity błąd złożonej kwadratury trapezów jest (pesymistycznie) sumą błędów popełnianych na poszczególnych podprzedziałach. Dla n podprzedziałów dostajemy

$$E = \sum_{i=1}^{n} \left(-\frac{1}{12}\right) \left(\frac{b-a}{n}\right)^3 f''(\zeta_i) = -\frac{1}{12}\frac{(b-a)^3}{n^2} \cdot \underbrace{\frac{1}{n}\sum_{i=1}^{n} f''(\zeta_i)}$$

$$= -\frac{1}{n^2}\frac{(b-a)^3}{12} f''(\zeta)\,, \tag{11}$$

ponieważ długość każdego podprzedziału wynosi $(b-a)/n$, natomiast do podkreślonego fragmentu zastosowaliśmy twierdzenie o wartości średniej.

Złożony wzór trapezów

Zwróćmy uwagę, iż numeryczna wartość całki po jednokrotnym zagęszczeniu podziału i zastosowaniu wzoru trapezów dana jest przez

$$I \simeq \frac{1}{2}\cdot\frac{b-a}{2}(f_0+f_1)+\frac{1}{2}\cdot\frac{b-a}{2}(f_1+f_2) = \frac{b-a}{2}\left(\frac{1}{2}f_0+f_1+\frac{1}{2}f_2\right), \tag{12}$$

gdzie f_1 jest wartością funkcji podcałkowej w punkcie środkowym. W ogólności złożony wzór trapezów ma postać

$$I_N \simeq h\left(\frac{1}{2}f_0+f_1+f_2+\cdots+f_{N-1}+\frac{1}{2}f_N\right), \tag{13}$$

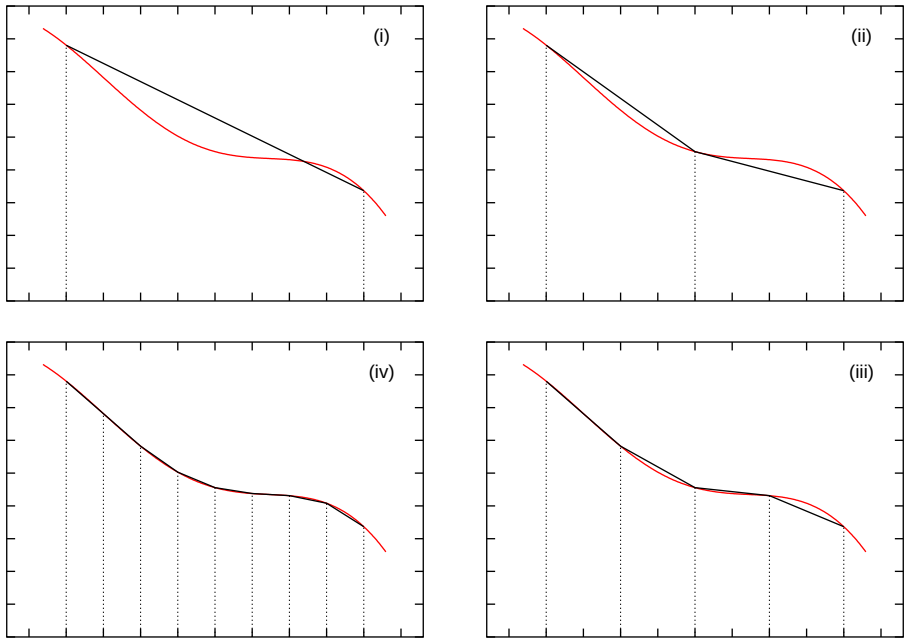
gdzie h jest długością najmniejszego podprzedziału (średnicą podziału), f_N jest wartością obliczoną w prawym krańcu przedziału całkowania, przy czym $N = 2^k$, gdzie k jest rzędem podziału. Widać zatem, że przy iteracyjnym zagęszczaniu podziałów nawet nie trzeba pamiętać wartości funkcji

w węzłach wyższego rzędu — wystarczy zapamiętać ich sumę. Procedurę iteracyjnego zagęszczania podziałów kończymy, gdy kolejne znalezione przybliżenia całki róznią się od siebie zaniedbywalnie mało, to jest gdy $|I_{k+1} - I_k| < \varepsilon$, gdzie ε jest z góry ustaloną tolerancją. Jeżeli wartość całki jest bardzo mała lub bardzo duża, należy stosować dokładność względną:

$$\frac{|I_{k+1} - I_k|}{|I_k| + \varepsilon'} < \varepsilon \,, \tag{14}$$

gdzie $0 < \varepsilon' \ll 1$ jest dodatkowym parametrem mającym chronić przed dzieleniem przez zero.

Zasadę kwadratur złożonych ilustruje poniższy rysunek. Dzięki temu, iż zagęszczeń dokonuje się przez połowienie podprzedziałów, węzły kwadratury mniej złożonej są też węzłami kwadratury bardziej złożonej. Kwadratury złożone wprowadza się też dla kwadratur wyższych rzędów (Simpsona, 3/8, Milne'a). Powiedzmy, jeśli mam kwadraturę Simpsona opartą na punktach x_0, x_1, x_2, po zagęszczeniu dostaję dwie kwadratury Simpsona oparte, odpowiednio, na punktach x_0, $(x_0+x_1)/2$, x_1 oraz x_1, $(x_1+x_2)/2$, x_2.



Ekstrapolacja Richardsona

Stosując kwadratury złożone, dostajemy cały ciąg kolejnych przybliżeń całki, odpowiadających kolejnym zagęszczeniom podziału. Czy cały ten ciąg — fakt, jakie wartości przyjmują kolejne wyrazy, nie zaś tylko wartość ostatnio obliczonego wyrazu — może być użyteczny przy numerycznym obliczaniu całki? Niekiedy tak. Najprostszym sposobem jest zastosowanie ekstrapolacji Richardsona, którą omówimy na przykładzie wzoru trapezów.

Przypuśćmy, że przybliżoną wartość całki (1) obliczamy dla dwu podziałów przedziału $[a, b]$: na n i na $2n$ podprzedziałów, odpowiednio otrzymując

$$I = I_n - \frac{(b-a)^3}{12n^2} f''(\zeta_n)\,, \tag{15a}$$

$$I = I_{2n} - \frac{(b-a)^3}{12(2n)^2} f''(\zeta_{2n})\,, \tag{15b}$$

gdzie I_n, $_{2n}$ są liczbowymi wynikami zastosowania złożonego wzoru trapezów odpowiedniego rzędu (porównaj (11)). Jeżeli założymy, że obie wartości pochodnej występujące w (15) są równe i wyeliminujemy je z tych wyrażeń, otrzymamy

$$I \simeq \frac{4I_{2n} - I_n}{3}\,. \tag{16}$$

Dobroć tego przybliżenia zależy oczywiście od tego, czy obie wartości drugiej pochodnej różnią się wystarczająco mało. Przypuśmy, że drogą zwiększania liczby podprzedziałów otrzymujemy monotoniczny ciąg przybliżeń całki: rosnący, odpowiadający kolejnym przybliżeniom niedomiarowym, dążącym do wartości całki, lub malejący, odpowiadający przybliżeniom nadmiarowym. Monotoniczność ciągu przybliżeń oznacza, że funkcja podcałkowa nie zmienia swojej wypukłości w przedziale całkowania, a zatem że jej krzywizna *być może* nie ulega znacznym zmianom, wobec czego założenie o stałości drugiej pochodnej w przedziale całkowania *być może* nie jest drastycznie złamane. Jeśli ciąg otrzymanych przybliżeń nie jest monotoniczny, stosowanie ekstrapolacji Richardsona jest wątpliwe.

Przykład ekstrapolacji Richardsona

Chcemy numerycznie obliczyć wartość całki

$$I = \int_1^2 \frac{dx}{x} \,. \tag{17}$$

Jej ścisła wartość wynosi $I = \ln 2 \simeq 0.69314718$. Funkcją podcałkową jest $f(x) = \frac{1}{x}$.

A. Zastosowanie metody trapezów daje

$$I_1 = \frac{1}{2}(f(1) + f(2)) = 0.75 \,. \tag{18a}$$

B. *Zagęszczam podział* — wprowadzam punkt pośredni $x = 1.5$. Zastosowanie złożonego wzoru trapezów (13) daje

$$I_2 = \frac{1}{2}\left(\frac{1}{2}f(1) + f(1.5) + \frac{1}{2}f(2)\right) = \frac{1}{2}\cdot\frac{17}{12} \simeq 0.70833333\,. \quad \text{(18b)}$$

C. *Zagęszczam podział* — wprowadzam punkty pośrednie $x = 1.25, x = 1.75$. Zastosowanie złożonego wzoru trapezów (13) daje

$$\begin{aligned} I_4 &= \frac{1}{4}\left(\frac{1}{2}f(1) + f(1.5) + \frac{1}{2}f(2)\right) + \frac{1}{4}\left(f(1.25) + f(1.75)\right) \\ &= \frac{1}{4}\cdot\frac{17}{12} + \frac{1}{4}\left(\frac{4}{5} + \frac{4}{7}\right) \simeq 0.69702381\,. \end{aligned} \quad \text{(18c)}$$

I_1, I_2, I_4 tworzą monotoniczny ciąg przybliżeń numerycznych poszukiwanej całki (17). Zastosowanie ekstrapolacji Richardsona (16) do I_4, I_2 daje

$$I \simeq \frac{4\cdot 0.69702381 - 0.70833333}{3} \simeq 0.69325397\,. \quad \text{(19)}$$

Metoda Romberga

Jeżeli obliczamy całkę za pomocą ciągu złożonych wzorów trapezów, możemy wyjść poza ekstrapolację Richardsona, uzyskując szczególnie efektywne przybliżenie całki. Jego podstawą jest (nieoczywisty) fakt, iż błąd metody trapezów zawiera wyłącznie parzyste potęgi średnicy podziału:

$$I = \int_a^b f(x)\,dx = h\left(\frac{1}{2}f_0 + f_1 + f_2 + \cdots + f_{N-1} + \frac{1}{2}f_N\right) + \sum_{j=1}^{\infty} \alpha_j h^{2j}\,. \tag{20}$$

Dowód tego faktu można znaleźć w podręczniku Ralstona. Nie wdając się w szczegóły wyprowadzenia, oznaczmy przez $A_{0,k}$ przybliżenie całki uzyskane złożonym wzorem trapezów z 2^k podprzedziałami.

Spodziewamy się[†], że

$$\lim_{k\to\infty} A_{0,k} = I\,. \tag{21}$$

Definiujemy teraz[‡]

$$A_{n,k} = \frac{1}{4^n - 1}\left(4^n A_{n-1,k+1} - A_{n-1,k}\right). \tag{22}$$

[†] *Jeżeli* całka (20) istnieje, to zachodzi (21), natomiast z samego faktu, iż ciąg $A_{0,k}$ jest zbieżny, nie można wnioskować, iż całka (20) istnieje.

[‡] Dziękuję panu Mateuszowi Rusowi za zwrócenie uwagi na błąd w tym wzorze!

Schemat obliczania elementów $A_{n,k}$ najłatwiej zilustrować za pomocą tabeli:

$$
\begin{array}{ccccccccccc}
A_{0,0} & & & & & & & & & & \\
 & \searrow & & & & & & & & & \\
A_{0,1} & \rightarrow & A_{1,0} & & & & & & & & \\
 & \searrow & & \searrow & & & & & & & \\
A_{0,2} & \rightarrow & A_{1,1} & \rightarrow & A_{2,0} & & & & & & \\
 & \searrow & & \searrow & & \searrow & & & & & \\
A_{0,3} & \rightarrow & A_{1,2} & \rightarrow & A_{2,1} & \rightarrow & A_{3,0} & & & & \\
 & \searrow & & \searrow & & \searrow & & \searrow & & & \\
A_{0,4} & \rightarrow & A_{1,3} & \rightarrow & A_{2,2} & \rightarrow & A_{3,1} & \rightarrow & A_{4,0} & & \\
 & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & \\
\ldots & \rightarrow & \ldots & \rightarrow & \ldots & \rightarrow & \ldots & \rightarrow & \ldots & \rightarrow & \ldots
\end{array}
\tag{23}
$$

Strzałki wskazują, które już policzone elementy są potrzebne do obliczenia kolejnego. W praktycznych zastosowaniach nie trzeba zapamietywać całej tablicy (23), a tylko jej ostatnio obliczony wiersz (a nawet mniej).

Łatwo pokazać, że

$$\begin{bmatrix} A_{00} \\ A_{10} \\ \vdots \\ A_{k0} \end{bmatrix} = \begin{bmatrix} c_{00} & 0 & 0 & \dots & 0 \\ c_{11} & c_{10} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ c_{kk} & c_{k,k-1} & c_{k,k-2} & \dots & c_{k0} \end{bmatrix} \begin{bmatrix} A_{00} \\ A_{01} \\ \vdots \\ A_{0k} \end{bmatrix}, \tag{24}$$

przy czym — co już nieco trudniej pokazać (patrz Ralston) — własności macierzy współczynników w (24) są takie, że, po pierwsze,jeśli (21) zachodzi, to zachodzi także

$$\lim_{k\to\infty} A_{k,0} = I \tag{25}$$

oraz, po drugie, zbieżność ciągu $A_{k,0}$ jest *szybsza* niż ciągu $A_{0,k}$. Porównajmy wyrażenie (22) z (16). Mówiąc niezbyt precyzyjnie, kolejne liczby $A_{n,k}$ są efektem „ekstrapolacji z ekstrapolacji” i dzięki temu zbieżność może być szybsza.

Praktyczny schemat stosowania ekstrapolacji Richardsona wygląda tak: Przypuśćmy, iż wypełniliśmy już wiersz tabeli (23) zaczynający się od elementu A_{0k}, odpowiadającemu złożonej metodzie trapezów z 2^k podprzedziałami. Naszym aktualnym przybliżeniem całki jest element A_{k0}. Teraz

- Obliczamy $A_{0,k+1}$ poprzez zastosowanie złożonej metody trapezów z 2^{k+1} podprzedziałami;
- Zapełniamy cały wiersz korzystając ze wzoru (22).

Procedurę kończymy, gdy elementy A_{k0} i $A_{k+1,0}$ różnią się o mniej niż zadana toleracja (lub gdy przekroczymy pewną graniczną ilość iteracji; to ostatnie sygnalizuje, że metoda nie jest zbieżna). W ten sposób można uzyskać taką dokładność całki, jaką wprost ze złożonego wzoru trapezów uzyskalibyśmy dopiero przy znacznie gęstszym podziale.

Przykład: Należy znaleźć wartość całki

$$I = \int_{1}^{3/2} \frac{dx}{1 + 2x^2 - \frac{1}{4}\sin(9x)} \tag{26}$$

z dokładnością do 10^{-8}. Stosujemy metodę Romberga; obliczenia przerywamy, gdy dwa kolejne wyrazy diagonalne staną się sobie równe (z zadaną dokładnością). Wyniki obliczeń można przedstawić w postaci następującej tabeli, odpowiadającej ogólnej tabeli (23):

$k=0$	0.13347528					
$k=1$	0.12398581	0.12082265				
$k=2$	0.12173305	0.12098214	0.12099277			
$k=3$	0.12118491	0.12100220	0.12100353	0.12100370		
$k=4$	0.12104904	0.12100375	0.12100385	0.12100386	0.12100386	
$k=5$	0.12101515	0.12100385	0.12100386	0.12100386	0.12100386	0.12100386
$k=6$	0.12100668	...				
$k=7$	0.12100456	...				
$k=8$	0.12100403	...				
$k=9$	0.12100390	...				
$k=10$	0.12100387	...				
$k=11$	0.12100386	...				
$k=12$	0.12100386	...				

Pierwsza kolumna macierzy (23) odpowiada kolejnym zastosowaniom metody trapezów dla zagęszczających się podziałów. Jak widzimy, do osiągnięcia zadanej dokładności potrzeba tylko $2^5 + 1 = 33$ obliczeń funkcji przy zastosowaniu metody Romberga i aż $2^{12} + 1 = 4097$ obliczeń funkcji przy zastosowaniu złożonego wzoru trapezów. Dla całek innych niż (26) relacje te mogą być inne: metoda Romberga może dawać nieco wolniejszą, lub przeciwnie, jeszcze szybszą zbieżność.

Przykład

Stosując metodę Romberga do przykładu ze strony 26 otrzymujemy

$$\begin{array}{lllll} k=0 & I_1 = 0.75 & & & \\ k=1 & I_2 = 0.70833333 & 0.69444444 & & \\ k=2 & I_4 = 0.69702381 & 0.69325397 & 0.69317461 & \end{array} \tag{27}$$

Mimo, że wykonaliśmy zaledwie pięć obliczeń funkcji podcałkowej, końcowe przybliżenie uzyskane w metodzie Romberga różni się od wartości dokładnej całki (17) o mniej niż $3 \cdot 10^{-5}$.

Całkowanie po przedziałach nieskończonych

Przy obliczaniu całek typu

$$\int_0^\infty f(x)\,dx \tag{28}$$

należy szczególnie uważać, aby numerycznie nie "obliczyć" całki, która jest rozbieżna. Podkreślamy, że kwadratury służą do znajdywania przybliżonych wartości całek, *o których wiemy, że istnieją*.

Zauważmy, że funkcja podcałkowa w (28) musi w nieskończoności zmierzać *dostatecznie szybko* do zera aby całka istniała. Możemy skorzystać z tego faktu w połączeniu z addytywnością całki:

$$\int_0^\infty f(x)\,dx = \int_0^A f(x)\,dx + \int_A^\infty f(x)\,dx \tag{29}$$

gdzie A jest dostatecznie dużą stałą dodatnią. Stałą tę dobieramy tak, aby dla $x > A$ funkcja podcałkowa spełniała $|f(x)| \leqslant B \cdot g(x)$, gdzie $g(x)$ dostatecznie szybko zmierza do zera, a jej całkę łatwo jest obliczyć **analitycznie**. Do numerycznego obliczenia pozostaje całka $\int_0^A f(x)\,dx$, a więc całka po przedziale skończonym.

Podkreślamy, że całkę "po ogonie", czyli drugą całkę po prawej stronie (29), **należy szacować analitycznie**.

Podobnie postępujemy z całkami typu $\int_{-\infty}^0$, $\int_{-\infty}^{\infty}$.

Innym sposobem jest szacowanie całek "po ogonach" przy pomocy kwadratur Gaussa, nieomawianych na tym wykładzie.

Przykład

Należy obliczyć całkę

$$\int_0^\infty \sin\left(\frac{1+\sqrt{x}}{1+x^2}\right) e^{-x}\,dx \tag{30}$$

z dokładnością do 10^{-7}. Zauważmy, że funkcja podcałkowa jest niewiększa od e^{-x} oraz że

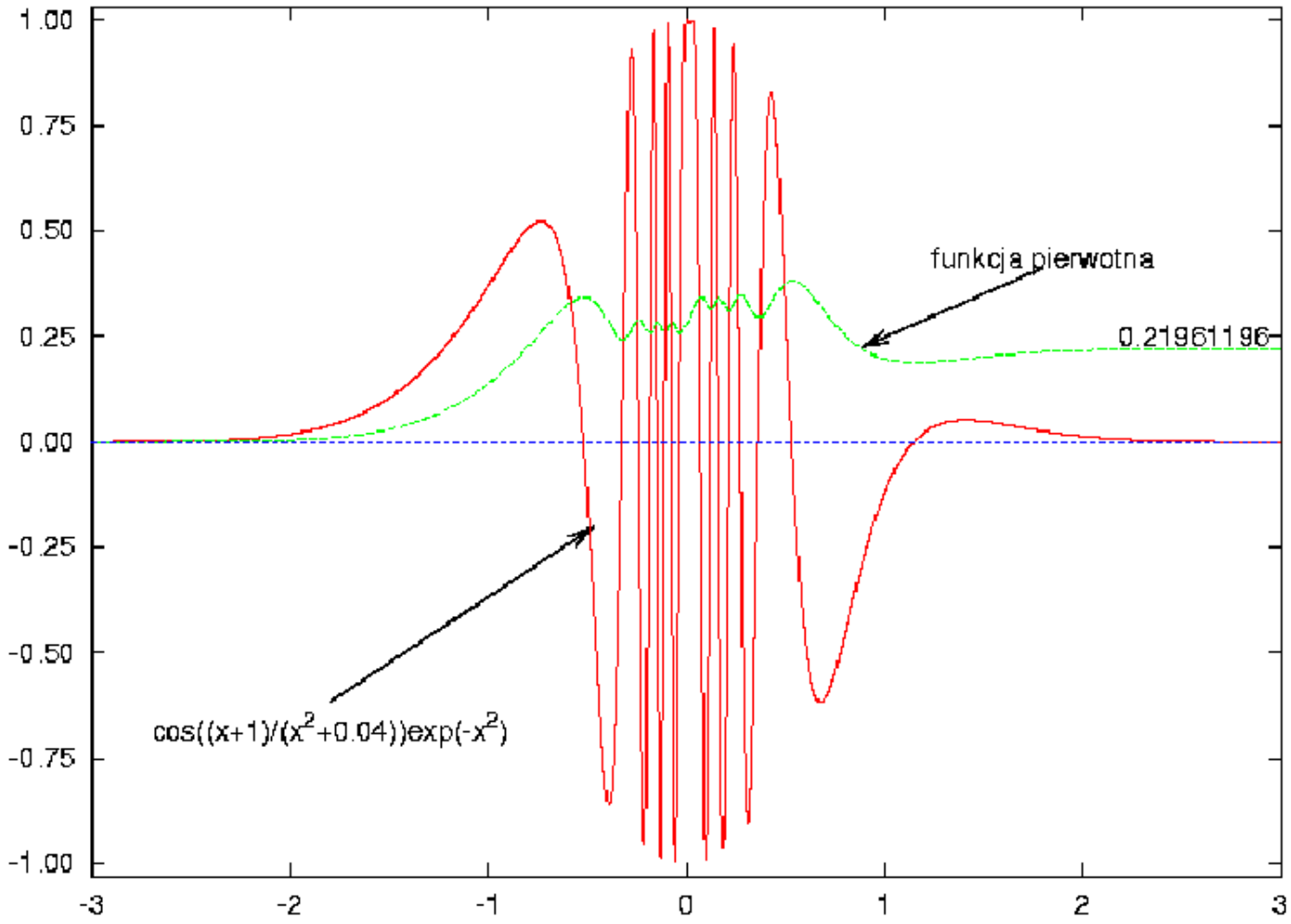
$$\int_{17}^{\infty} e^{-x}\,dx = e^{-17} \simeq 0.414 \cdot 10^{-7}$$

wobec czego wystarczy obliczyć całkę z funkcji jak we wzorze (30) po przedziale skończonym $[0, 17]$ z dokładnością nie mniejszą niż $0.586 \cdot 10^{-7}$.

Kwadratury adaptacyjne

Szczególną trudność sprawiają całki z funkcji wykazujących zlokalizowane, ale szybkie oscylacje. Stoimy przed dylematem: Albo całkujemy z dużym krokiem, ryzykując popełnienie znacznego błędu w obszarze oscylacji, albo też całkujemy z małym krokiem, dobranym do oscylacji, także w obszarze, w którym funkcja podcałkowa zmienia się powoli. Wówczas nie popełniamy dużego błędu numerycznego, ale ponosimy niepotrzebnie duży koszt numeryczny, gdyż w obszarach wolnej zmienności moglibyśmy całkować z większym krokiem.

Rozwiązaniem są *kwadratury adaptacyjne*, czyli algorytm, który lokalnie sam dobiera krok całkowania, dostosowując go do charakteru zmienności funkcji. W tym celu algorytm musi mieć dwa niezależne oszacowania całki po danym podprzedziale — ich różnica jest miarą popełnianego błędu.



Chcemy zatem znaleźć przybliżenie

$$\int_a^b f(x)\,dx \simeq I = \sum_{j=1}^{N} I_{[x_{j-1},x_j]} \tag{31}$$

gdzie pod znakiem sumy stoją numeryczne przybliżenia całki w przedziałach $[x_{j-1}, x_j]$. Dodatkowo chcemy kontrolować *globalny* błąd obliczenia całki. Jeżeli maksymalny dopuszczalny błąd wynosi τ, musimy zadbać o to, aby w każdym podprzedziale

$$\left|\mathcal{E}_{[x_{j-1},x_j]}\right| < \frac{x_j - x_{j-1}}{b-a}\tau\,, \tag{32}$$

gdzie $\mathcal{E}_{[x_{j-1},x_j]}$ oznacza błąd całkowania numerycznego po danym przedziale. Dzięki temu całkowity błąd $|\mathcal{E}| = \left|\sum_{j=1}^{N} \mathcal{E}_{[x_{j-1},x_j]}\right| \leqslant \sum_{j=1}^{N} \left|\mathcal{E}_{[x_{j-1},x_j]}\right| < \left(\sum_{j=1}^{N}(x_j - x_{j-1})\right)\tau/(b-a) = \tau$.

Strategia postępowania

Przypuśćmy, że zajmujemy się oszacowaniem całki $\int_\alpha^\beta f(x)\,dx$, gdzie $a \leqslant \alpha < \beta \leqslant b$. Mamy aktualne przybliżenie całki po tym przedziale $I_{[\alpha,\beta]}$, mamy też I, aktualną zakumulowaną sumę przyczynków do całki po przedziałach, które spełniły już kryterium (32). Teraz

1. Obliczamy całki $I_{[\alpha,(\alpha+\beta)/2]} \simeq \int_\alpha^{(\alpha+\beta)/2} f(x)\,dx$, $I_{[(\alpha+\beta)/2,\beta]} \simeq \int_{(\alpha+\beta)/2}^\beta f(x)\,dx$.

2. Mając dwa oszacowania $\int_\alpha^\beta f(x)\,dx$, a mianowicie $I_{[\alpha,\beta]}$ oraz $I_{[\alpha,(\alpha+\beta)/2]}+I_{[(\alpha+\beta)/2,\beta]}$ (całka jest addytywna!), szacujemy błąd $\mathcal{E}_{[\alpha,\beta]}$.

3. Jeżeli błąd nie spełnia oszacowania (32), dajemy na stos granice prawego podprzedziału $[(\alpha+\beta)/2, \beta]$ i aktualne oszacowanie całki po tym podprzedziale $I_{[(\alpha+\beta)/2,\beta]}$ (żebyśmy go nie musieli powtórnie obliczać), a następnie powtarzamy całą procedurę dla lewego podprzedziału $[\alpha, (\alpha+\beta)/2]$.

4. Jeżeli błąd spełnia oszacowanie (32), dodajemy obliczoną wartość całki po podprzedziale $[\alpha, \beta]$ do I, po czym ściągamy ze stosu najwyżej leżący przedział i powtarzamy dla niego całą procedurę.

Przepełnienie się stosu (lub osiągnięcie jego z góry założonej maksymalnej wysokości) jest znakiem załamania się algorytmu. Oznacza to, że albo całka jest rozbieżna, albo zażądaliśmy nierealistycznie dużej dokładności. Jeśli zachodzi ten drugi przypadek, należy *zwiększyć* maksymalny dopuszczalny błąd, τ.

Szczegóły oszacowania błędu zależą od kwadratury, za pomocą której obliczamy całki w poszczególnych podprzedziałach. Zazwyczaj stosuje się kwadratury niskich rzędów, metodę trapezów lub Simpsona.

Błąd w kwadraturze adaptacyjnej — metoda trapezów

Dla metody trapezów wiemy, że

$$\int_\alpha^\beta f(x)\,dx = I_{[\alpha,\beta]} - \frac{1}{12}(\beta-\alpha)^3 f''(\zeta)\,, \tag{33}$$

gdzie $\zeta \in [\alpha,\beta]$. Z drugiej strony

$$\begin{aligned}
\int_\alpha^\beta f(x)\,dx &= \int_\alpha^{(\alpha+\beta)/2} f(x)\,dx + \int_{(\alpha+\beta)/2}^\beta f(x)\,dx \\
&= I_{[\alpha,(\alpha+\beta)/2]} - \frac{1}{12}((\alpha+\beta)/2-\alpha)^3 f''(\zeta_1) \qquad (34)\\
&+ I_{[(\alpha+\beta)/2,\beta]} - \frac{1}{12}(\beta-(\alpha+\beta)/2)^3 f''(\zeta_2)
\end{aligned}$$

Korzystając z twierdzenia o wartości średniej możemy zapisać $\left(f''(\zeta_1)+f''(\zeta_2)\right)/2 = f''(\bar{\zeta})$, a zatem zamiast (34) mamy

$$\int_\alpha^\beta f(x) = I_{[\alpha,(\alpha+\beta)/2]} + I_{[(\alpha+\beta)/2,\beta]} - \frac{1}{4}\cdot\frac{1}{12}(\beta-\alpha)^3 f''(\bar{\zeta})\,. \quad (35)$$

(33) i (35) stanowią dwa oszacowania tej samej wielkości, $\int_\alpha^\beta f(x)\,dx$. Załóżmy teraz, że druga pochodna całkowanej funkcji nie zmienia się bardzo w przedziale $[\alpha,\beta]$, czyli $f''(\zeta) \simeq f''(\bar{\zeta})$. Eliminując drugie pochodne z wyrażeń (33), (35) otrzymujemy

$$\int_\alpha^\beta f(x)\,dx - \left(I_{[\alpha,(\alpha+\beta)/2]} + I_{[(\alpha+\beta)/2,\beta]}\right) \simeq \frac{1}{3}\left(I_{[\alpha,(\alpha+\beta)/2]} + I_{[(\alpha+\beta)/2,\beta]} - I_{[\alpha,\beta]}\right) \quad (36)$$

Po prawej stronie wzoru (36) mamy poszukiwane oszacowanie błędu, $\mathcal{E}_{[\alpha,\beta]}$, którego należy użyć w (32) (jeżeli całki obliczamy metodą trapezów!). Wobec tego za przybliżenie całki, którego używamy w punkcie 4 algorytmu na kwadratury adaptacyjne (str. 42), przyjmujemy

$$\begin{aligned}\int_\alpha^\beta f(x)\,dx &\simeq I_{[\alpha,(\alpha+\beta)/2]} + I_{[(\alpha+\beta)/2,\beta]} \\ &+ \frac{1}{3}\left(I_{[\alpha,(\alpha+\beta)/2]} + I_{[(\alpha+\beta)/2,\beta]} - I_{[\alpha,\beta]}\right) \\ &= \frac{4}{3}\left(I_{[\alpha,(\alpha+\beta)/2]} + I_{[(\alpha+\beta)/2,\beta]}\right) - \frac{1}{3}I_{[\alpha,\beta]}\,. \qquad (37)\end{aligned}$$

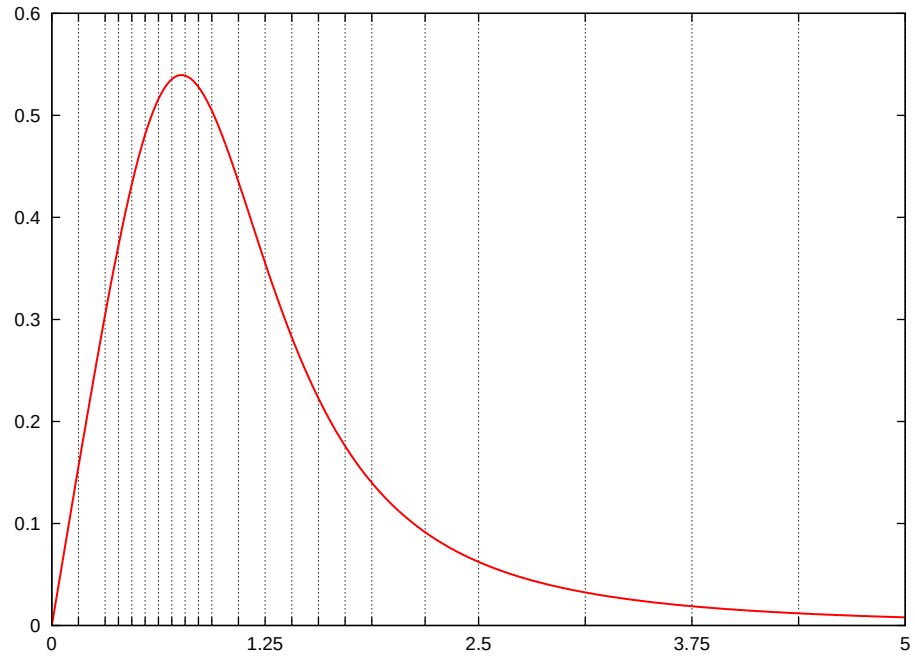
Zauważmy, że wzór (37) wykorzystuje ekstrapolację Richardsona.

Przykład

Obliczmy całkę

$$I = \int_0^5 \sin\left(\frac{x}{1+x^4}\right) dx \tag{38}$$

za pomocą kwadratury adaptacyjnej opartej na metodzie trapezów, z dokładnością do 10^{-2}. Otrzymujemy $I \simeq 0.74$, przy czym dokonano 23 obliczeń wartości funkcji podcałkowej Poniższy rysunek przedstawia podprzedziały, na jakie algorytm adaptacyjny podzielił cały przedział całkowania. Widzimy, że podział zagęszcza się w obszarze, w którym krzywizna wykresu jest największa.



Obliczenie całki (38) tą samą metodą, ale z dokładnością do 10^{-8}, wymaga 21547 obliczeń funkcji podcałkowej, ale wysokość stosu nigdy nie przekracza 15. Otrzymujemy $I \simeq 0.74482956$.

Uwaga: Kwadratury adaptacyjne **nie** są najlepszą metodą obliczanie całki (38) — metoda Romberga zastosowana do tej całki daje taką samą dokładność (i oczywiście ten sam wynik!) po 8 krokach, a więc po wyliczeniu $2^9+1=513$ wartości funkcji podcałkowej. Kwadratury adaptacyjnej użyliśmy tylko dla zilustowania tej metody — w zastosowaniach praktycznych kwadratur adaptacyjnych używa się tylko do całkowania **bardzo szybko** oscylujących funkcji.

Całki wielowymiarowe

Często zachodzi konieczność obliczania całek oznaczonych z funkcji wielu zmiennych, typu

$$\int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 \ldots \int_{a_n}^{b_n} dx_n \, f(x_1, x_2, \ldots, x_n) \tag{39}$$

Jeżeli wymiar całki $n \geqslant 3$, całki tego typu należy obliczać metodami Monte Carlo, które są wówczas najbardziej efektywne. Dla $n = 2$ metody Monte Carlo nie są konkurencyjne wobec podejścia tradycyjnego. Zastanawiamy się zatem nad sposobami obliczania całek typu

$$\iint_D f(x, y) \, dx \, dy \tag{40}$$

gdzie D oznacza pewien dwuwymiarowy obszar całkowania (niekoniecznie prostokąt!). Zakładamy przy tym, że *wiemy*, że całka (40) istnieje.

Całkowanie po siatce prostokątnej

Jeżeli mamy obliczyć całkę po obszarze prostokątnym

$$I = \int_a^b dx \int_c^d dy\, f(x,y) \tag{41}$$

możemy skorzystać z twierdzenia o iterowaniu całek i spróbować przenieść metody znane z przypadku jednowymiarowego:

$$I = \int_a^b g(x)\, dx \tag{42a}$$

gdzie

$$g(x) = \int_c^d f(x,y)\, dy \tag{42b}$$

Do obliczania całek (42a), (42b) stosujemy znane kwadratury jednowymiarowe, przy czym przy obliczaniu całki (42b) wartość x jest znana i ustalona, narzucona przez wartość aktualnego węzła kwadratury w (42a). Widać, że jeżeli znalezienie całki jednowymiarowej wymaga $\sim N$ obliczeń funkcji podcałkowej, znalezienie dwuwymiarowej całki (41) wymaga $\sim N^2$ takich obliczeń.

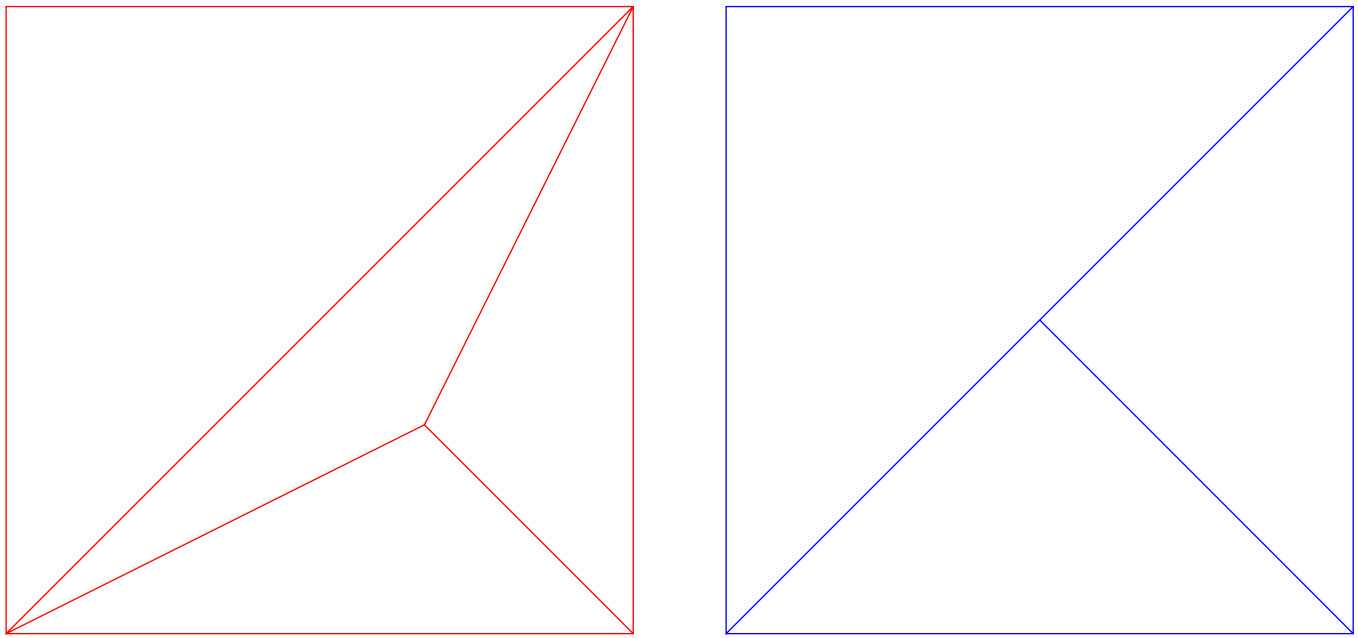
Jeżeli jest to wygodniejsze, możemy całkę po x traktować jako całkę *wewnętrzną*, całkę po y jako *zewnętrzną*, czyli obliczać całki w odwrotnej kolejności, niż w (42).

Kwadratura adaptacyjna w dwu wymiarach

Jeżeli obszar całkowania D w (40) jest dowolnym wielokątem (niekoniecznie prostokątem), narzucającym się sposobem całkowania numerycznego jest kwadratura adaptacyjna oparta o *triangulację* obszaru całkowania.

Definicja: Triangulacją wielokąta D nazywam jego skończone pokrycie trójkątami Ω_i, dla którego

1. Wielokąt jest sumą mnogościową trójkątów składowych, $D = \bigcup_i \Omega_i$.
2. Przecięcie mnogościowe dowolnych dwu elementów pokrycia, $\Omega_i \cap \Omega_j$, $i \neq j$,
   (a) jest zbiorem pustym, albo
   (b) składa się ze *wspólnego* wierzchołka, albo
   (c) składa się ze *wspólnej* krawędzi dwu trójkątów.



Przykład prawidłowej triangulacji (lewy panel) i nieprawidłowej pseudo-triangulacji (prawy panel). Stykające się trójkąty muszą mieć albo wspólne wierzchołki, albo *wspólne* krawędzie (krawędź jednego jest zarazem krawędzią drugiego).

Wstępnej triangulacji możemy dokonać “ręcznie” (nie jest to trudne nawet dla skomplikowanych wielokątów), natomiast zagęszczanie triangulacji można przeprowadzić następująco: wyznaczamy środek ciężkości trójkąta, gdyż zawsze leży on wewnątrz trójkąta (jeżeli trójkąt ma wierzchołki o współrzędnych (x_1, y_1), (x_2, y_2), (x_3, y_3), środek ciężkości ma współrzędne $((x_1 + x_2 + x_3)/3, (y_1 + y_2 + y_3)/3)$), a następnie przeprowadzamy krawędzie od środka ciężkości do wierzchołków trójkąta. W ten sposób trójkąt zostaje podzielony na trzy trójkąty potomne. Jeżeli wyjściowe pokrycie było triangulacją, także pokrycie potomne jest triangulacją.

Odpowiednikiem triangulacji w większej liczbie wymiarów byłoby pokrycie sympleksami. Na przykład sympleksami w $\mathbb{R}^3$ są czworościany.

Zauważmy, że jeśli mamy dane trzy wierzchołki trójkąta na płaszczyźnie XY, (x_1, y_1), (x_2, y_2), (x_3, y_3), to punkty w przestrzeni o odpowiednich odciętych i współrzędnej z-towej równej, odpowiednio, $f(x_1, y_1)$, $f(x_2, y_2)$, $f(x_3, y_3)$, wyznaczają płaszczyznę. Jako przybliżenie całki po trójkącie Ω o podanych wierzchołkach możemy wziąć objętość graniastosłupa ściętego o podstawie trójkątnej, wyznaczonego przez wierzchołki trójkąta-podstawy i wartości funkcji w tych punktach. Jest to dwuwymiarowy odpowiednik metody trapezów.

Co więcej, zagęszczanie triangulacji w sposób opisany powyżej, wymaga tylko jednego dodatkowego obliczenia wartości funkcji.

Algorytm kwadratur adaptacyjnych na płaszczyźnie

- Mając trójkąt Ω wyliczam przybliżenie całki zgodnie ze "wzorem graniastosłupów", I_Ω.
- Trójkąt Ω dzielę w opisany sposób na trzy trójkąty potomne Ω_i i wyliczam przybliżone calki po tych trójkątach, I_{Ω_i}, $i = 1, 2, 3$.
- Jako miarę błędu przyjmuję

$$\mathcal{E}_\Omega = I_\Omega - \left(I_{\Omega_1} + I_{\Omega_2} + I_{\Omega_3} \right) . \tag{43}$$

- Jeżeli zachodzi

$$|\mathcal{E}_\Omega| < \frac{S_\Omega}{S_D} \tau \, , \tag{44}$$

gdzie S_Ω oznacza powierzchnię aktualnego trójkąta, S_D powierzchnię całego obszaru całkowania, zaś τ jest *globalną* miarą dopuszczalnego błędu, za przybliżenie całki uznaję $I_{\Omega_1} + I_{\Omega_2} + I_{\Omega_3}$. Ściągam kolejny trójkąt ze stosu i przystępuję do jego analizy w ten sam sposób.

- Jeżeli warunek (44) nie jest spełniony, odkładam dwa trójkąty potomne na stos i przystępuję do podziałów pozostałego trójkąta potomnego.
- Procedurę kończę gdy na stosie nie zostały żadne trójkąty, ani też nie zostały żadne niezbadane trójkąty z pierwotnej triangulacji (sukces) lub gdy przekroczę dopuszczalną wysokość stosu (porażka).

Opisaną procedurę można też zastosować do całkowania po obszarze niewielokątowym, wpisując weń wielokąty coraz lepiej przybliżające poszukiwany obszar. Wpisywanie wielokątów przerywam, gdy przyczynek do całki od kolejnych poprawek staje się zaniedbywalnie mały. Jeszcza raz podkreślam, że trzeba mieć *dowód analityczny* na to, że całka istnieje. Zbieżności procedury numerycznego całkowania nie wolno uznać za "dowód numeryczny" istnienia całki.