Fra Bernoullis forunderlige polynomier og tal via Euler–Maclaurins sumformel til Rombergs integralapproximation

Anden del.

Mogens Esrom Larsen
Københavns universitets matematiske institut

I første del stiftede vi bekendtskab med *Bernoullipolynomierne,* de første seks ser fredelige ud,

$$\begin{aligned} B(x,0) &= 1 \\ B(x,1) &= x - \tfrac{1}{2} \\ B(x,2) &= x^2 - x + \tfrac{1}{6} \\ B(x,3) &= x(x-1)\left(x - \tfrac{1}{2}\right) \\ B(x,4) &= x^4 - 2x^3 + x^2 - \tfrac{1}{30} \\ B(x,5) &= x(x-1)\left(x - \tfrac{1}{2}\right)\left(x^2 - x - \tfrac{1}{3}\right) \\ B(x,6) &= x^6 - 3x^5 + \tfrac{5}{2}x^4 - \tfrac{1}{2}x^2 + \tfrac{1}{42} \end{aligned} \tag{13}$$

og den almindelige formel er

$$B(x,m) = \sum_{k=0}^{m} \binom{m}{k} B_{m-k} x^k \tag{10}$$

hvori indgår *Bernoullitallene,* der defineres rekursivt ved

$$\begin{aligned} B_0 &= 1 \\ \textstyle\sum_{j=0}^{i-1} \binom{i}{j} B_j &= 0 \text{ for } i > 1 \end{aligned} \tag{11}$$

De første Bernoullital så meget uskyldige ud:

$$B_0 = 1,\ B_1 = -\tfrac{1}{2},\ B_2 = \tfrac{1}{6},\ B_4 = -\tfrac{1}{30},\ B_6 = \tfrac{1}{42},\ B_8 = -\tfrac{1}{30},\ B_{10} = \tfrac{5}{66},\ B_{12} = -\tfrac{691}{2730},\ B_{14} = \tfrac{7}{6} \tag{12}$$

men de vokser meget hurtigt, faktisk ligner de $(2(2n)!)/((2\pi)^{2n})$. F. eks. er nogle af dem

$$\begin{aligned} B_{20} &= -\tfrac{174611}{330},\ B_{40} = -\tfrac{261082718496449122051}{13530}, \\ B_{60} &= -\tfrac{1215233140483755572040304994079820246041491}{56786730} \end{aligned}$$

Bernoullipolynomierne havde den vigtige egenskab, at

$$B'(x,m) = mB(x,m-1) \tag{6}$$

så de kunne indgå bekvemt i en række partielle integrationer. Dette gav os *Euler–MacLaurin's* sumformel,

$$\begin{aligned} \textstyle\int_{x_1}^{x_2} f(t)dt = &\ \tfrac{x_2-x_1}{2}(f(x_1) + f(x_2)) + \textstyle\sum_{j=1}^{n-1} \tfrac{B_{2j}}{(2j)!}(x_2 - x_1)^{2j}\left(f^{(2j-1)}(x_1) - f^{(2j-1)}(x_2)\right) \\ &- \tfrac{(x_2-x_1)^{2n+1}}{(2n)!} B_{2n} f^{(2n)}(\xi) \end{aligned}$$

som her skal vise sig at indgå på afgørende måde i moderne integrationsteknik.........

Romberg integration

Hvis vi nu skal beregne et integral over et interval, $[a, b]$, så deler vi det i N lige store delintervaller, der altså hver får længden $h = \frac{b-a}{N}$. Delepunkterne betegner vi $a_0 = a, a_1 = a+h, \cdots, a_i = a+ih, a_N = a+Nh = b$. Vi vil foretrække at vælge N stor nok til, at $h \leq \frac{1}{2}$. Når vi så anvender Euler–Maclaurins formel på hvert delinterval, og lægger dem alle sammen sammen, så går alle de afledede i delepunkterne ud mod hinanden. Derved fås *Euler–Maclaurins sumformel:*

$$
\begin{aligned}
\int_a^b f(t)dt = \; & h\left(\tfrac{1}{2}(f(a) + \textstyle\sum_{\nu=1}^{N-1} f(a_\nu)) + \tfrac{1}{2}f(b)\right) + \\
& + \textstyle\sum_{j=1}^{n-1} \frac{B_{2j}}{(2j)!} h^{2j} \left(f^{(2j-1)}(a) - f^{(2j-1)}(b)\right) - \\
& - \textstyle\sum_{\nu=1}^{N} \frac{B_{2n}}{(2n)!} h^{2n+1} f^{(2n)}(\xi_\nu)
\end{aligned}
\tag{25}
$$

hvor $\xi_\nu \in]a_{\nu-1}, a_\nu[$. Da $h = \frac{b-a}{N}$ og $f^{(2n)}$ er kontinuert, kan vi skrive

$$
\sum_{\nu=1}^{N} h f^{(2n)}(\xi_\nu) = (b-a) \sum_{\nu=1}^{N} \frac{f^{(2n)}(\xi_\nu)}{N} = (b-a) f^{(2n)}(\xi) \tag{26}
$$

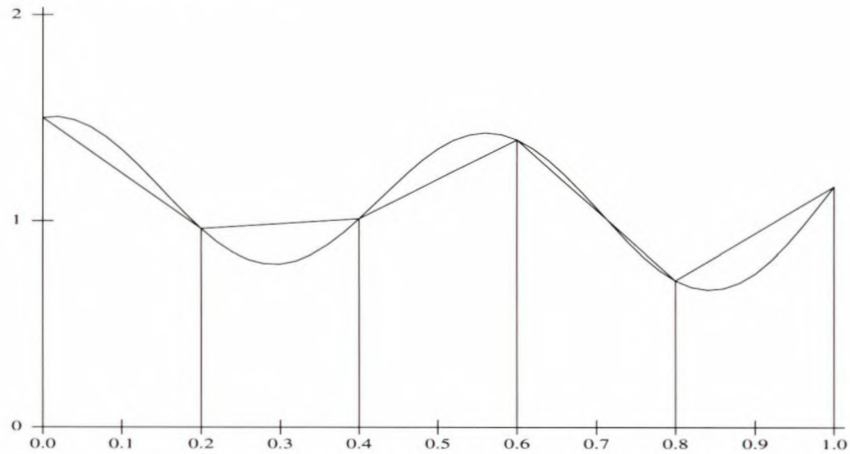
idet gennemsnitsværdien må antages i et punkt, $\xi \in]a, b[$. Indsættes (26) i (25), fås

$$
\begin{aligned}
\int_a^b f(t)dt = \; & h\left(\tfrac{1}{2}(f(a) + \textstyle\sum_{\nu=1}^{N-1} f(a_\nu)) + \tfrac{1}{2}f(b)\right) + \\
& + \textstyle\sum_{j=1}^{n-1} \frac{B_{2j}}{(2j)!} h^{2j} \left(f^{(2j-1)}(a) - f^{(2j-1)}(b)\right) - \\
& -(b-a) \frac{B_{2n}}{(2n)!} h^{2n} f^{(2n)}(\xi)
\end{aligned}
\tag{27}
$$

Formlens første led,

$$K_1 = h\left(\frac{1}{2}(f(a) + \sum_{\nu=1}^{N-1} f(a_\nu) + \frac{1}{2}f(b)\right) \tag{28}$$

kaldes *kordetrapezformlen,* fordi det er den integralapproximation, man får ud af at approximere arealet under funktionens graf med trapezer med højde h og siderne lig med funktonsværdierne i endepunkterne. Resten af formlen kan så siges at være udtryk for fejlen ved denne kordetrapezformel.



Approximation med kordetrapezer

Det bemærkelsesværdige ved denne integrationsformel er, at koefficienterne til h^{2j} kun afhænger af intervallets endepunkter, ikke af h eller N.
Det betyder, at vi har en integrationsformel af formen

$$\int_a^b f(t)dt = K_1 + \sum_{j=1}^{n-1} c_j h^{2j} - (b-a)\frac{B_{2n}}{(2n)!}h^{2n} f^{(2n)}(\xi_1)$$

Hvis vi nu fordobler antallet af delintervaller og dermed halverer skridtlængden, h, så får vi en formel af udseendet:

$$\int_a^b f(t)dt = K_2 + \sum_{j=1}^{n-1} c_j \left(\frac{h}{2}\right)^{2j} - (b-a)\frac{B_{2n}}{(2n)!}\left(\frac{h}{2}\right)^{2n} f^{(2n)}(\xi_2)$$

Beregning af K_2 kræver blot beregning af funktionsværdierne i alle intervallernes midtpunkter, de andre har vi i forvejen. Af disse to formler kan vi eliminere leddene, der indeholder koefficienten c_1. Vi tager 4 af den anden og trækker den første fra. Efter deling med 3 fås:

$$\begin{aligned} \int_a^b f(t)dt &= \tfrac{4K_2-K_1}{3} + \textstyle\sum_{j=2}^{n-1} c_j h^{2j}\,\tfrac{2^{2(1-j)}-1}{3} + \\ &+ (b-a)\tfrac{B_{2n}}{(2n)!}\tfrac{h^{2n}}{3}(f^{(2n)}(\xi_1) - 2^{2(1-n)}f^{(2n)}(\xi_2)) = \\ &= K_1^1 + \textstyle\sum_{j=2}^{n-1} c_j' h^{2j} + (b-a)\tfrac{B_{2n}}{(2n)!}\tfrac{h^{2n}}{3}f^{(2n)}(\xi) \end{aligned} \tag{29}$$

idet vi definerer:

$$\begin{aligned} K_1^1 &= \tfrac{4K_2-K_1}{3} \\ c_j' &= c_j\tfrac{2^{2(1-j)}-1}{3} \end{aligned} \tag{30}$$

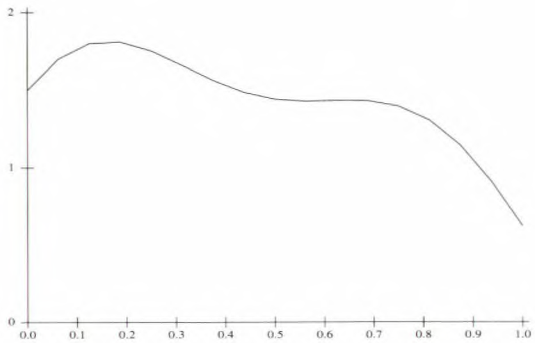
og regner med, at der findes en værdi, ξ, så

$$f^{(2n)}(\xi) = (f^{(2n)}(\xi_1) - 2^{2(1-n)}f^{(2n)}(\xi_2))$$

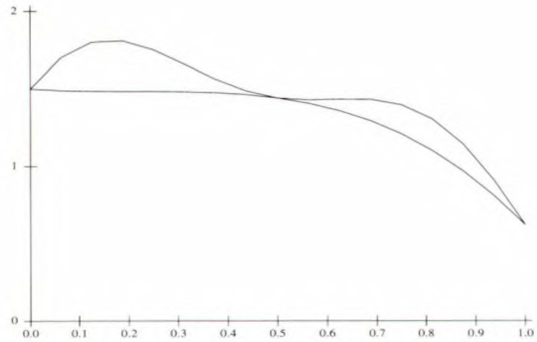
Dermed er fejlen på K_2, der var af størrelsesorden $\frac{h^2}{4}$, ændret til en fejl på K_1^1, der er af størrelsesorden h^4. Når blot $h < \frac{1}{2}$, er det i sig selv en forbedring.

Den ny integralapproximation, (30), svarer til, at vi over et interval, $]a_{\nu-1}, a_\nu[$, med midtpunkt c_ν, bruger approximationen

$$\frac{h}{6}\left(f(a_{\nu-1}) + 4f(c_\nu) + f(a_\nu)\right)$$



Funktion



Trediegradsapproximation

Denne approximation er genkendelig som *Keplers tønderegel* efter Johannes Kepler (1571–1630), som i 1615 i sin berømte bog, *Stereometria doliorum vinorum,* brugte den til måling af indholdet af et halvfyldt vinfad, eller som *Simpsons formel* efter Thomas Simpson (1710–1761), der nævner den i en lærebog. Tanken er simpelthen den, at for givne tre punkter, e.g., $-1, 0, 1$, har vi de givne funktionsværdier, y_{-1}, y_0, y_1, samt i midtpunktet værdien af funktionens afledede, y_0'. Nu tænker vi os, at vi har lagt et polynomium af tredie grad gennem punkterne, så det har afledet i 0 lig med funktionens:

$$\begin{aligned} p(x) &= \alpha x^3 + \beta x^2 + \gamma x + \delta \\ p'(x) &= 3\alpha x^2 + 2\beta x + \gamma \end{aligned}$$

Når vi integrerer polynomiet fra -1 til 1, får vi jo

$$\int_{-1}^{1} p(x)dx = \left[\alpha\frac{x^4}{4} + \beta\frac{x^3}{3} + \gamma\frac{x^2}{2} + \delta x\right]_{-1}^{1} = \frac{2}{3}\beta + 2\delta \tag{31}$$

Vi behøver altså kun at finde β og δ. Da $p(0) = y_0$, har vi åbenbart $\delta = y_0$. Sætter vi $x = \pm 1$, får vi ligningerne

$$\begin{aligned} \alpha + \beta + \gamma + \delta &= y_1 \\ -\alpha + \beta - \gamma + \delta &= y_{-1} \end{aligned}$$

Lægger vi sammen og indfører $\delta = y_0$, får vi $2\beta = y_1 + y_{-1} - 2y_0$. Indsættes disse værdier i (31), får vi integralapproximationen for et interval af længde 2:

$$\int_{-1}^{1} p(x)dx = \frac{y_1 + y_{-1} - 2y_0}{3} + 2y_0 = \frac{y_1 + y_{-1} + 4y_0}{3} \tag{32}$$

Korrigeres til skridtlængden h, fås (30).
Vi kan gøre det samme igen, halvere skridtlængden og beregne K_3. Ud fra K_2 og K_3 kan vi så forbedre approksimationen til Keplers tønderegel, $K_2^1 = \frac{1}{3}(4K_3 - K_2)$, hvis fejl er af størrelsesorden $\frac{h^4}{16}$. Også her gælder det, at koefficienterne til hhv. h^4 og $\frac{h^4}{16}$ er den samme. Så vi kan eliminere disse led og få en fejl af størrelsesorden h^6 i stedet for, ved at approksimere med

$$K_1^2 = \frac{16K_2^1 - K_1^1}{15}$$

Har vi ved sukcessiv halvering af skridtlængden dannet approximationerne
$K_1, K_2, \cdots, K_n$, kan vi danne et mønster af approximationer efter opskriften:

$$K_j^i = \frac{2^{2i}K_{j+1}^{i-1} - K_j^{i-1}}{2^{2i} - 1} \qquad i = 1, \cdots, n-1, \quad , j = 1, \cdots, n-i \tag{33}$$

Vi kan overskueligt skrive dem i et trekantet mønster, hvor meningen er, at man først beregner approximationerne i søjlen til venstre, derefter ved hjælp af (33) udregner de øvrige, altid ved hjælp af værdien til venstre og den ovenover:

$$\begin{array}{ccccccccc}
K_1 & & & & & & & & \\
 & \searrow & & & & & & & \\
K_2 & \rightarrow & K_1^1 & & & & & & \\
 & \searrow & & \searrow & & & & & \\
K_3 & \rightarrow & K_2^1 & \rightarrow & K_1^2 & & & & \\
\cdot & & \cdot & & \cdot & & & & \\
\cdot & & \cdot & & \cdot & & & & \\
\cdot & \searrow & \cdot & \searrow & & & \searrow & & \\
K_n & \rightarrow & K_{n-1}^1 & \rightarrow & K_{n-2}^2 & \cdot\ \cdot & \rightarrow & K_1^{n-1} &
\end{array} \tag{34}$$

Når vi har elimineret alle leddene i summen, når vi til approximationen K_1^{n-1}, hvis fejl bliver

$$\frac{|B_{2n}|}{(2n)!} \cdot \frac{h^{2n}}{3 \cdot 15 \cdots (2^{2(n-1)} - 1)} \cdot (b-a)|f^{(2n)}(\xi)| \tag{35}$$

For små værdier af n kan vi jo regne koefficienten ud, og for store værdier kan vi for det første vurdere nævneren i den anden brøk ned til

$$3 \cdot 15 \cdots (2^{2(n-1)} - 1) > 2 \cdot 2^3 \cdots 2^{2n-3} = 2^{(n-1)^2} \tag{36}$$

for det andet benytte den ovenfor nævnte approximation,

$$\frac{B_{2n}}{(2n)!} \simeq \frac{2}{(2\pi)^{2n}} \simeq \frac{2}{40^n} \tag{37}$$

Indsættes (36) og (37) i (35), fås fejlvurderingen:

$$\frac{2h^{2n}}{40^n 2^{(n-1)^2}} \cdot (b-a)|f^{(2n)}(\xi)| = \frac{h^{2n}}{10^n 4^n 2^{n(n-2)}} \cdot (b-a)|f^{(2n)}(\xi)| = \frac{h^{2n}}{10^n 2^{n^2}} \cdot (b-a)|f^{(2n)}(\xi)| \tag{38}$$

Hvis vi nu for eksempel vil beregne π med 10 decimaler, altså med en fejl, der højst er 10^{-12}, kan vi jo approximere integralet

$$\int_0^1 \frac{4}{1+x^2}dx = 4 \cdot \int_0^{\frac{\pi}{4}} \frac{1+tg^2v}{1+tg^2v}dv = \pi \tag{39}$$

Vælger vi nu skridtlængden $h = \frac{1}{4}$, ser vi, at for $n = 4$ bliver fejlen (35) af størrelse

$$\begin{aligned} \frac{|B_8|}{8!}\frac{\left(\frac{1}{4}\right)^8}{3 \cdot 15 \cdot 63} \cdot (1-0)|f^{(8)}(\xi)| &= \frac{1}{30 \cdot 40320 \cdot 65536 \cdot 2835} \cdot |f^{(8)}(\xi)| = \\ &= \frac{|f^{(8)}(\xi)|}{224737099776000} \simeq 10^{-14} \cdot |f^{(8)}(\xi)| \end{aligned}$$

Med brug af (38) er det derimod hovedregning. Fejlen vurderes ved

$$\frac{\left(\frac{1}{4}\right)^8}{10^4 2^{16}} \cdot |f^{(2n)}(\xi)| = \frac{|f^{(2n)}(\xi)|}{10^4 2^{32}} \simeq 10^{-13} \cdot |f^{(2n)}(\xi)|$$

idet vi blot har benyttet, at $2^{10} = 1024 \simeq 10^3$.
Udføres beregningerne i skemaet ovenfor, fås

$$\begin{array}{ccccccc}
3.131176470588 & & & & & & \\
& \searrow & & & & & \\
3.138988494491 & \to & 3.141592502459 & & & & \\
& \searrow & & \searrow & & & \\
3.140941612041 & \to & 3.141592651225 & \to & 3.141592661143 & & \\
& \searrow & & \searrow & & \searrow & \\
3.141429893175 & \to & 3.141592653553 & \to & 3.141592653708 & \to & 3.141592653590
\end{array}$$

Til sammenligning er tabelværdien $\pi = 3.14159265358979$.

Denne metode til at forbedre integralapproximationen er fra 1955 og kaldes *Romberg integration* efter Werner Romberg (1909–).

Teknikken til elimination af fejl af lav orden er, som det også fremgår af eksemplet, langt mere økonomisk end den banale, at forbedre værdien ved sukcessive halveringer. Romberg integration har ikke alene stor praktisk betydning ved integralberegninger, men ideen er videreudviklet af Josef Stoer (1934–) og Roland Bulirsch (1932–) i 1971 til den hidtil bedste metode til numerisk løsning af differentialligninger.

Mogens Esrom Larsen er redaktør af matematikken i KVANT. Han er ansat som lektor ved Københavns Universitet.