

Polycopié

Genevois d'**A**nalyses **N**umérique



E. Hairer & G. Wanner

Université de Genève
Section de mathématiques
Case postale 64
CH-1211 Genève 4

2006/2007

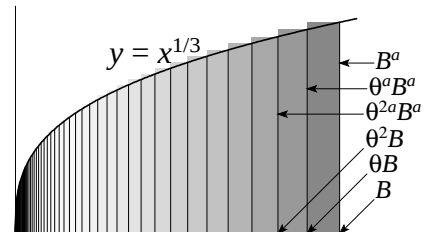
Chapitre I

Intégration Numérique

La préhistoire :

Première période. Le calcul d'aires, de surfaces, de volumes et longueurs d'arc était une des préoccupations majeures des scientifiques depuis Euclide et Archimède jusqu'à Kepler, Galilei et Fermat.

Par exemple, Fermat en 1638 a trouvé par des additions de séries géométriques que l'aire sous $y = x^a$ entre 0 et B est égal à $A = \frac{B^{a+1}}{a+1}$.



Deuxième période. Puis, avec la découverte du *Théorème fondamental du calcul différentiel* par Newton et Leibniz, ces “intégrales” n’étaient qu’une inversion de formules pour la dérivée. Par exemple, la formule de Fermat ci-dessus s’explique par le fait que la dérivée de la fonction

$$z = \frac{x^{a+1}}{a+1} \quad \text{est égal à} \quad y = x^a.$$

Ainsi, il suffit de trouver les *inverses* des formules des dérivées. Les Bernoullis et surtout Euler, dans son gigantesque traité (*Inst. Calculi Integralis* 1768, 1769, 1770, Opera XI-XIII), se sont donc mis avec enthousiasme à la recherche de formules analytiques pour les primitives.

Troisième période. Cependant, de nombreuses intégrales résistent encore et toujours à l’envahisseur (exemples $\int \frac{e^x}{x} dx$, $\int \frac{dx}{\log x}$) et l’enthousiasme laisse lentement place à un renouveau de l’intérêt pour les formules de quadrature, qui avait été commencé par Newton et ses successeurs anglais Cotes et Simpson. De nombreux calculs en astronomie (perturbations des orbites planétaires) contraignent Gauss (1814) à intensifier cette théorie.

Les premiers ordinateurs, dans les années 40 et 50, ont surtout été développés pour le calcul d’intégrales et d’équations différentielles. Ces problèmes sont les plus faciles à programmer. Pour cette même raison, nous commençons ici par le calcul d’intégrales.

Bibliographie sur ce chapitre

- P.J. Davis & P. Rabinowitz (1975): *Methods of Numerical Integration*. Academic Press, New York.
- V.I. Krylov (1959): *Priblizhennoe Vychislenie Integralov*. Goz. Izd. Fiz.-Mat. Lit., Moscow. Traduction anglaise: *Approximate calculation of integrals*. Macmillan, 1962. [MA 65/185]
- R. Piessens, E. de Doncker-Kapenga, C.W. Überhuber & D.K. Kahaner (1983): QUADPACK. *A Subroutine Package for Automatic Integration*. Springer Series in Comput. Math., vol. 1. [MA 65/210]

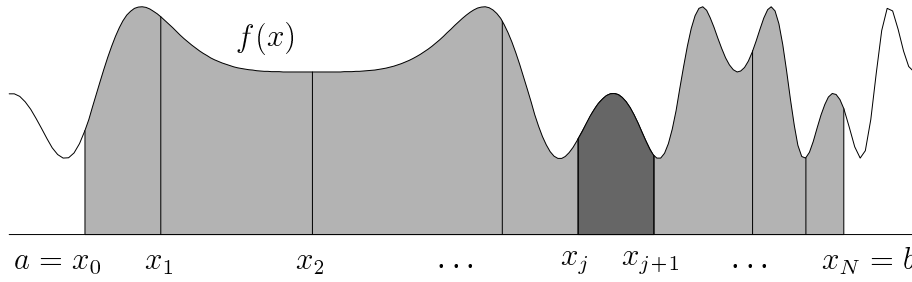


FIG. I.1: Une division d'un intervalle en sous-intervalles

A.H. Stroud (1974): *Numerical quadrature and Solution of Ordinary Differential Equations*. Springer. [MA 65/89]

I.1 Formules de quadrature et leur ordre

“If there are four ordinates at equal intervals, let A be the sum of the first and the fourth, B the sum of the second and third, and R the interval between the first and the fourth ; then ... the area between the first and the fourth ordinates will be $(A + 3B)R/8$.”

(I. Newton, *Methodus*, publ. 1711, cité d'après H.H. Goldstine, p. 76)

Problème. Etant donné une fonction continue sur un intervalle borné $f : [a, b] \rightarrow \mathbb{R}$, on cherche la valeur de l'intégrale

$$\int_a^b f(x) dx. \quad (1.1)$$

Subdivision de l'intervalle. Depuis Simpson (1743), la plupart des algorithmes numériques procèdent comme suit: on subdivise $[a, b]$ en plusieurs sous-intervalles ($a = x_0 < x_1 < x_2 < \dots < x_N = b$) et on utilise le fait que

$$\int_a^b f(x) dx = \sum_{j=0}^{N-1} \int_{x_j}^{x_{j+1}} f(x) dx. \quad (1.2)$$

De cette manière, on est ramené au calcul de plusieurs intégrales pour lesquelles la longueur de l'intervalle d'intégration est relativement petite. Prenons une de ces intégrales et notons son intervalle d'intégration par $[x_0, x_0 + h]$.

Problème 2. Donné un intervalle d'intégration $[x_0, x_0 + h]$ de longueur h petite, calculer

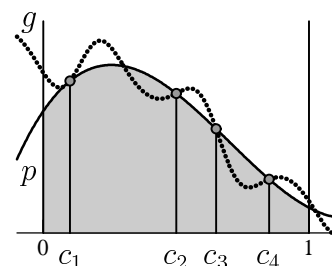
$$\int_{x_0}^{x_0+h} f(x) dx = h \int_0^1 f(x_0 + th) dt = h \int_0^1 g(t) dt \quad (1.3)$$

où $g(t) := f(x_0 + th)$.

Algorithme. On choisit un nombre s et des points $c_1, \dots, c_s$ dans (en général) $[0, 1]$. Puis on pose

$$\int_0^1 g(t) dt \approx \int_0^1 p(t) dt$$

où $p(t)$ est le polynôme de degré $s - 1$ avec $p(c_i) = g(c_i)$, $i = 1, \dots, s$.



Idée de Lagrange (1797) ¹ Si on pose les polynômes (de degré $s - 1$, voir Fig. I.2)

$$\ell_1(t) = \frac{(t - c_2)(t - c_3)(t - c_4)\dots}{(c_1 - c_2)(c_1 - c_3)(c_1 - c_4)\dots}, \quad \ell_2(t) = \frac{(t - c_1)(t - c_3)(t - c_4)\dots}{(c_2 - c_1)(c_2 - c_3)(c_2 - c_4)\dots}, \quad \text{etc.} \quad (1.4)$$

alors $\ell_1(c_1) = 1, \ell_2(c_2) = 1, \dots$ et toutes les autres valeurs aux c_i sont nulles. Par conséquent

$$p(t) = g(c_1)\ell_1(t) + g(c_2)\ell_2(t) + \dots \quad (1.5)$$

et nous avons le théorème

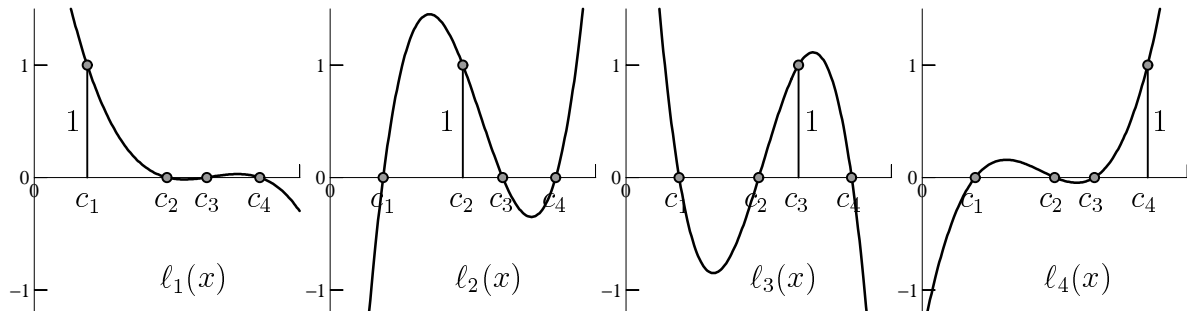


FIG. I.2: Les polynômes de Lagrange

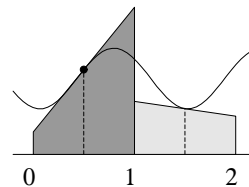
Théorème 1.1 La formule de quadrature est donnée par

$$\int_0^1 g(t) dt \approx \sum_{i=1}^s b_i g(c_i) \quad \text{où} \quad b_i = \int_0^1 \ell_i(t) dt. \quad (1.6)$$

Les c_i sont les nœuds de la formule et les b_i en sont les poids.

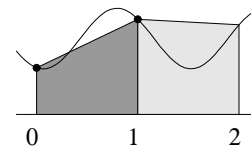
Exemples. 1. La formule du point milieu

$$\int_0^1 g(t) dt \approx g(1/2).$$



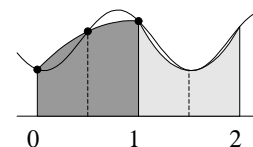
2. La formule du trapèze

$$\int_0^1 g(t) dt \approx \frac{1}{2} (g(0) + g(1)).$$



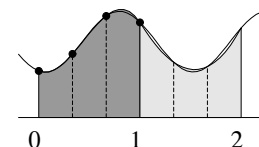
3. La formule de Simpson ; l'aire d'une parabole (polynôme de degré 2) par les trois points $(0, g(0)), (1/2, g(1/2)), (1, g(1))$:

$$\int_0^1 g(t) dt \approx \frac{1}{6} (g(0) + 4g(1/2) + g(1)).$$



4. La “pulcherrima et utilissima regula” de Newton (degré 3) :

$$\int_0^1 g(t) dt \approx \frac{1}{8} (g(0) + 3g(1/3) + 3g(2/3) + g(1)).$$



¹Archimedes, Newton, Leibniz, Euler, Gauss; il ne manquait que Lagrange pour faire le plein des plus grands mathématiciens de la période classique ...

TAB. I.1: Formules de Newton-Cotes

s	ordre	poids b_i						nom
2	2	$\frac{1}{2}$	$\frac{1}{2}$					trapèze
3	4	$\frac{1}{6}$	$\frac{4}{6}$	$\frac{1}{6}$				Simpson
4	4	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$			Newton
5	6	$\frac{7}{90}$	$\frac{32}{90}$	$\frac{12}{90}$	$\frac{32}{90}$	$\frac{7}{90}$		Boole
6	6	$\frac{19}{288}$	$\frac{75}{288}$	$\frac{50}{288}$	$\frac{50}{288}$	$\frac{75}{288}$	$\frac{19}{288}$	—
7	8	$\frac{41}{840}$	$\frac{216}{840}$	$\frac{27}{840}$	$\frac{272}{840}$	$\frac{27}{840}$	$\frac{216}{840}$	Weddle

5. En continuant ainsi (les c_i sont les points équidistants $i/(s-1)$, $i = 0, \dots, s-1$), on obtient les *formules de Newton-Cotes* (Newton 1676, Cotes 1711). Pour $s \leq 7$ ces formules sont données dans le tableau I.1. Leur dessin en figure I.3 montre que les formules “explosent” au-delà de $s = 10$. Si on veut augmenter la précision, il vaut mieux raffiner les subdivisions en (1.2) qu’augmenter le degré s .

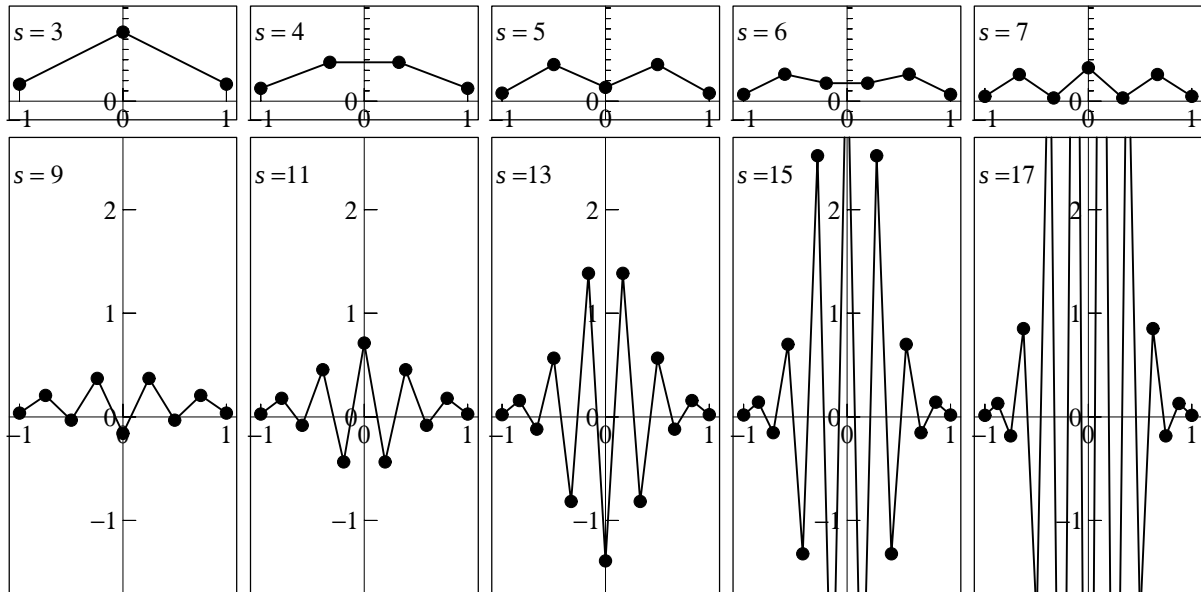


FIG. I.3: Dessin des poids des formules de Newton-Cotes

Calcul des poids par l’algèbre linéaire. L’idée est d’utiliser la propriété suivante:

Définition 1.2 On dit que l’ordre de la formule de quadrature (1.6) est p , si la formule est exacte pour tous les polynômes de degré $p-1$, c.-à-d.,

$$\int_0^1 g(t) dt = \sum_{i=1}^s b_i g(c_i) \quad \text{pour} \quad \deg g \leq p-1. \quad (1.7)$$

Théorème 1.3 La formule de quadrature a un ordre p si et seulement si

$$\sum_{i=1}^s b_i c_i^{q-1} = \frac{1}{q} \quad \text{pour} \quad q = 1, 2, \dots, p. \quad (1.8)$$

Démonstration. C'est parce que les fonctions

$$1, \quad t, \quad t^2, \quad t^3, \dots, t^{p-1}$$

forment une base des polynômes de degré $p - 1$. $\square$

En fixant les nœuds $c_1, \dots, c_s$ (distincts), la condition (1.8) avec $p = s$ représente un système linéaire pour $b_1, \dots, b_s$

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ c_1 & c_2 & \dots & c_s \\ \vdots & \vdots & & \vdots \\ c_1^{s-1} & c_2^{s-1} & \dots & c_s^{s-1} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_s \end{pmatrix} = \begin{pmatrix} 1 \\ 1/2 \\ \vdots \\ 1/s \end{pmatrix}. \quad (1.9)$$

Comme la matrice dans (1.9) est inversible (matrice de Vandermonde), la résolution de ce système nous donne une formule de quadrature d'ordre $p = s$.

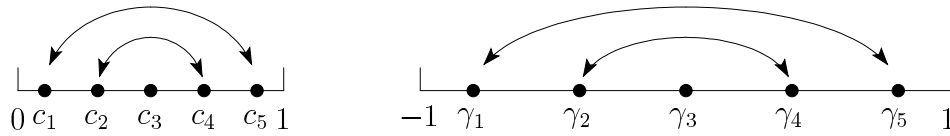


FIG. I.4: Nœuds d'une formule de quadrature symétrique

Formules symétriques. La formule est appelée *symétrique* si

$$c_i = 1 - c_{s+1-i} \quad \text{et} \quad b_i = b_{s+1-i}$$

pour tout i . Pour l'étudier, transformons l'intervalle $t \in [0, 1]$ en $\tau \in [-1, 1]$ (voir figure I.4) par les formules

$$\tau = 2t - 1, \quad d\tau = 2dt, \quad g(t) = u(\tau), \quad \gamma_i = 2c_i - 1, \quad \beta_i = 2b_i.$$

Ainsi, la symétrie devient

$$\gamma_i = -\gamma_{s+1-i} \quad (1.10)$$

et les conditions d'ordre (1.7) et (1.8) se changent en

$$\sum_{i=1}^s \beta_i \gamma_i^{q-1} = \int_{-1}^1 \tau^{q-1} d\tau = \begin{cases} \frac{2}{q} & \text{si } q \text{ impair;} \\ 0 & \text{si } q \text{ pair.} \end{cases} \quad (1.11)$$

Exemple. Calculons les β_i et les b_i pour la formule de Boole ($s = 5$) : nous avons $c_i =$

$(0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1) \Rightarrow \gamma_i = (-1, -\frac{1}{2}, 0, \frac{1}{2}, 1)$ et le système à résoudre devient

$$\begin{aligned} \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 &= 2 \\ -\beta_1 - \frac{\beta_2}{2} + \frac{\beta_4}{2} + \beta_5 &= 0 \\ \beta_1 + \frac{\beta_2}{4} + \frac{\beta_4}{4} + \beta_5 &= \frac{2}{3} \\ -\beta_1 - \frac{\beta_2}{8} + \frac{\beta_4}{8} + \beta_5 &= 0 \\ \beta_1 + \frac{\beta_2}{16} + \frac{\beta_4}{16} + \beta_5 &= \frac{2}{5} \end{aligned} \quad .$$

Les équations numéro 2 et 4 donnent

$$\begin{aligned} (\beta_5 - \beta_1) + \frac{1}{2}(\beta_4 - \beta_2) &= 0 \\ (\beta_5 - \beta_1) + \frac{1}{8}(\beta_4 - \beta_2) &= 0 \end{aligned} \Rightarrow \beta_5 = \beta_1 \text{ et } \beta_4 = \beta_2 .$$

Par conséquent, les équations 3 et 5 sont simplifiées à

$$\begin{aligned} \beta_1 + \frac{\beta_2}{4} &= \frac{1}{3} \\ \beta_1 + \frac{\beta_2}{16} &= \frac{1}{5} \end{aligned} \Rightarrow \beta_2 = \frac{32}{45}, \quad \beta_1 = \frac{7}{45}$$

et, par la première équation, nous avons finalement $\beta_3 = \frac{12}{45}$. Par construction, notre formule est d'ordre 5. Mais, à cause de la symétrie, l'équation suivante

$$-\beta_1 - \frac{\beta_2}{32} + \frac{\beta_4}{32} + \beta_5 = 0$$

est satisfaite elle aussi ! Notre formule est donc d'ordre 6. Similairement, la règle de Simpson est d'ordre 4 et la formule de Weddle d'ordre 8. Nous résumons :

Théorème 1.4 Si on satisfait pour des noeuds symétriques $c_i = 1 - c_{s+1-i}$ (pour tout i) les conditions pour ordre s , alors les poids b_i vont être symétriques ($b_i = b_{s+1-i}$). La formule aura toujours un **ordre pair**, i.e., $p \geq s + 1$ si s est impair.

“Man sieht hieraus, dass es im Allgemeinen vorthellhaft sein wird, bei Anwendung der Cotesischen Methode eine ungerade Anzahl von Ordinaten zu benutzen.” (Gauss 1814, *Werke* 3, p. 204)

I.2 Etude de l'erreur

Commençons par une *expérience numérique* :

Prenons une fonction $f(x)$, définie sur $[a, b]$, divisons l'intervalle en plusieurs sous-intervalles équidistants ($h = (b - a)/N$) et appliquons une formule de quadrature du paragraphe précédent. Ensuite, étudions l'erreur (en échelle logarithmique)

$$err = \int_a^b f(x) dx - \sum_{j=0}^{N-1} h \sum_{i=1}^s b_i f(x_j + c_i h) \quad (2.1)$$

en fonction de fe (nombre d'évaluations de la fonction $f(x)$; on a $fe = N \cdot (s - 1) + 1$ pour Newton-Cotes). Le nombre fe représente une mesure pour le travail (proportionnel au temps de calcul sur un ordinateur). La fig. I.5 montre les résultats (pour $N = 1, 2, 4, 8, 16, 32, \dots$) obtenus par les formules de Newton-Cotes pour les deux intégrales :

$$\int_0^3 \cos(x) \exp(\sin(x)) dx$$

et

$$\int_0^2 \cos(x) dx.$$

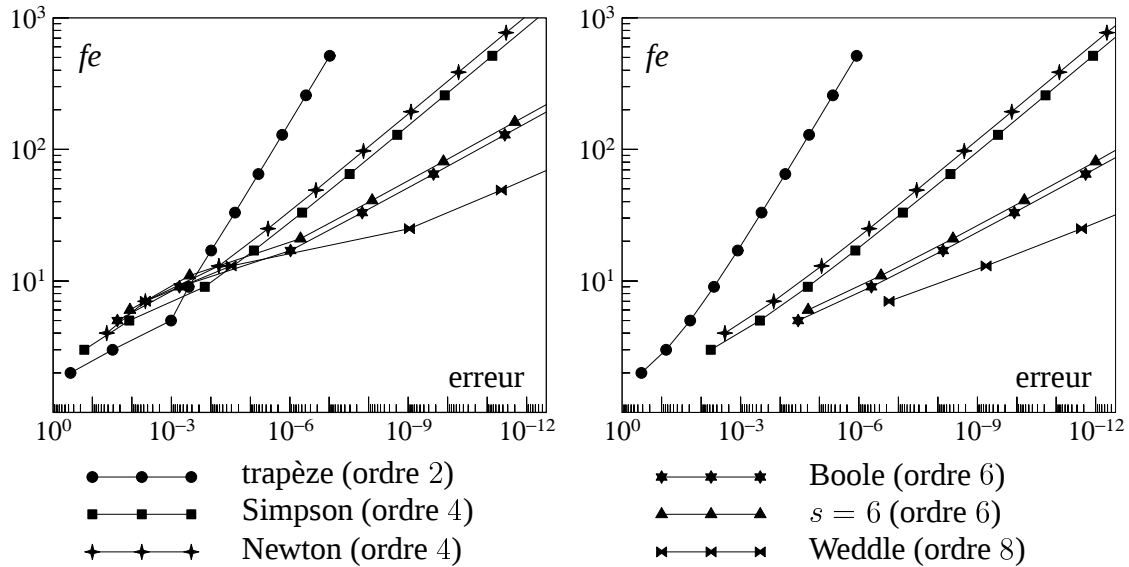


FIG. I.5: Le travail fe en fonction de l'erreur pour les formules de Newton-Cotes

En étudiant les résultats de la fig. I.5, nous constatons que :

- $\log_{10}(fe)$ dépend linéairement du nombre de chiffres exacts, donné par $-\log_{10}(err)$;
- la pente de chaque droite est $1/p$ (où p est l'ordre de la formule);
- si la précision augmente, les formules d'un ordre élevé finissent par être les meilleures.

Explication des résultats.

Étudions d'abord l'erreur faite sur *un* sous-intervalle, donc dans (1.3),

$$err_{loc} = h \left(\int_0^1 f(x_0 + th) dt - \sum_{i=1}^s b_i f(x_0 + c_i h) \right). \quad (2.2)$$

En supposant f suffisamment différentiable, on remplace $f(x_0 + th)$ par sa série de Taylor (développées autour de x_0 , voir cours d'Analyse I)

$$f(x_0 + th) = \sum_{q=0}^{p-1} \frac{h^q t^q}{q!} f^{(q)}(x_0) + \frac{h^p t^p}{p!} f^{(p)}(x_0) + \mathcal{O}(h^{p+1})$$

La première somme est un polynôme de degré $\leq p - 1$, son erreur est nulle par Def. 1.2. Donc, dans (2.2) il ne reste que

$$err_{loc} = \frac{h^{p+1}}{p!} \left(\int_0^1 t^p dt - \sum_{i=1}^s b_i c_i^p \right) f^{(p)}(x_0) + \mathcal{O}(h^{p+2}) = C h^{p+1} f^{(p)}(x_0) + \mathcal{O}(h^{p+2}) \quad (2.3)$$

où

$$C = \frac{1}{p!} \left(\frac{1}{p+1} - \sum_{i=1}^s b_i c_i^p \right) \quad (2.4)$$

s'appelle *constante de l'erreur*. Supposons que h soit petit de manière à ce que le terme $\mathcal{O}(h^{p+2})$ dans (2.3) soit négligeable par rapport au terme Ch^{p+1} .

Erreur globale. On doit calculer *la somme* de tous les erreurs locales. *Astuce:* on sépare du facteur h^{p+1} **un** h pour fabriquer une somme de Riemann. Ainsi on obtient

$$err = \sum_{j=0}^{N-1} err_{loc_j} \approx Ch^p \sum_{j=0}^{N-1} f^{(p)}(x_j) h \approx Ch^p \int_a^b f^{(p)}(x) dx = Ch^p (f^{(p-1)}(b) - f^{(p-1)}(a)). \quad (2.5)$$

Cette formule nous permet de mieux comprendre les résultats de la fig. I.5. Comme $err \approx C_1 \cdot h^p$ et $fe \approx C_2/h$, nous avons

$$-\log_{10}(err) \approx -\log_{10}(C_1) - p \cdot \log_{10}(h) \approx Const + p \cdot \log_{10}(fe).$$

Ceci montre la dépendance linéaire entre $\log_{10}(fe)$ et $-\log_{10}(err)$, et aussi le fait que la pente soit de $1/p$.

Conclusion. *Le plus on désire de la précision, le plus haut on doit choisir l'ordre.*

Estimation rigoureuse de l'erreur (noyaux de Peano).

On connaît déjà du cours d'Analyse I l'esprit très vif de G. Peano qui, de nombreuses fois, a découvert des erreurs dans les célèbres traités de l'Analyse (voir [HW97], p. 291,309,316,328,329). Rien d'étonnant que ce champion de la rigueur mathématique ait aussi trouvé une formule *exacte* de l'erreur de quadrature.

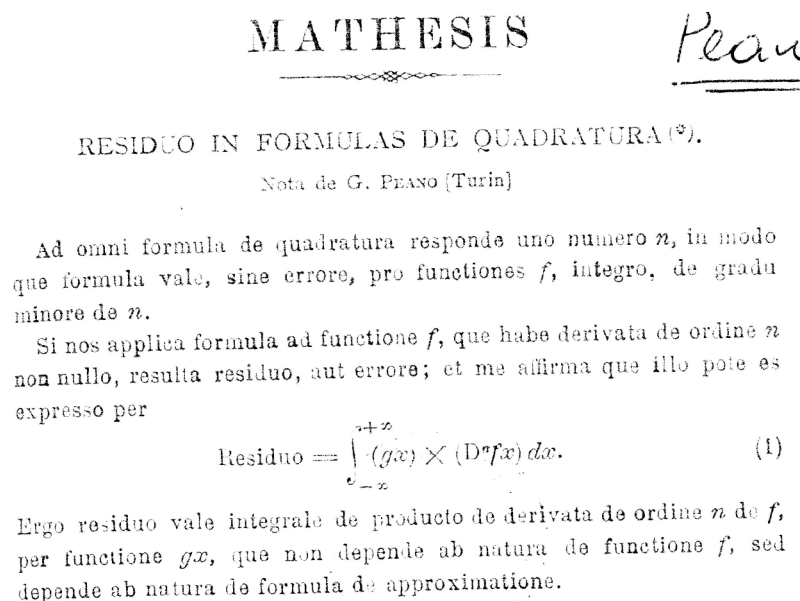


FIG. I.6: Fac-simile du début de l'article de Peano (1914) ; durant toute sa vie, Peano a lutté pour une écriture universelle en science. Cet article est écrit dans un latin simplifié “que la plupart des lecteurs ... comprennent sans peine”. Il ne pouvait pas s'imaginer que son rêve se réalisera un jour **grâce à l'anglais** !

Idée. Nous expliquons ici l'approche original de Peano à l'exemple de la formule du point milieu

$$errloc = h \left(\int_0^1 f(x_0 + \tau h) d\tau - f(x_0 + \frac{1}{2}h) \right).$$

L'idée est d'écrire le deuxième terme également sous forme d'une intégrale, ce qui donne

$$errloc = h \int_0^1 \left(1 - \delta(\tau - \frac{1}{2}) \right) \cdot f(x_0 + \tau h) d\tau \quad (2.6)$$

en introduisant une “fonction impellente” $\delta(\tau)$ très bizarre par la propriété

$$\delta(\tau) = 0 \quad \text{pour tout } \tau \neq 0 \quad \text{mais} \quad \int_{-\infty}^{\infty} \delta(\tau) d\tau = 1. \quad (2.7)$$

Cette “fonction” s'appelle de nos jours *fonction delta de Dirac* et est bien connue des “cultores de physica-mathematica”. Peano savait qu'une telle “fonction” allait contre tout bon sens des “cultores de analysi infinitesimal” et s'excuse en disant que cette définition “es commodo, sed non necessario ...”.

Ensuite, nous appelons la fonction entre parenthèses dans (2.6) par

$$N_0(\tau) = 1 - \delta(\tau - \frac{1}{2})$$

et effectuons une série d'intégrations par parties

$$errloc = h \int_0^1 N_0(\tau) f(x_0 + \tau h) d\tau = -h^2 \int_0^1 N_1(\tau) f'(x_0 + \tau h) d\tau = h^3 \int_0^1 N_2(\tau) f''(x_0 + \tau h) d\tau \quad (2.8)$$

où apparaissent des primitives $N_1(\tau) = \int_0^\tau N_0(u) du$ et $N_2(\tau) = \int_0^\tau N_1(u) du$, i.e.,

$$N_1(\tau) = \begin{cases} \tau & \text{si } \tau < 1/2 \\ \tau - 1 & \text{si } \tau \geq 1/2 \end{cases} \quad N_2(\tau) = \begin{cases} \tau^2/2 & \text{si } \tau \leq 1/2 \\ \tau^2/2 - (\tau - \frac{1}{2}) = (1 - \tau)^2/2 & \text{si } \tau \geq 1/2 \end{cases}$$

(voir la fig. I.7). Nous constatons que les deux intégrations par parties n'ont pas donné de termes intégrés, car

$$N_1(0) = 0, \quad N_1(1) = 1 - 1 = 0 \quad \text{et} \quad N_2(0) = 0, \quad N_2(1) = \frac{1}{2} - \frac{1}{2} = 0.$$

Par contre, une troisième intégration par parties donne un terme supplémentaire car $\int_0^1 N_2(\tau) d\tau \neq 0$.

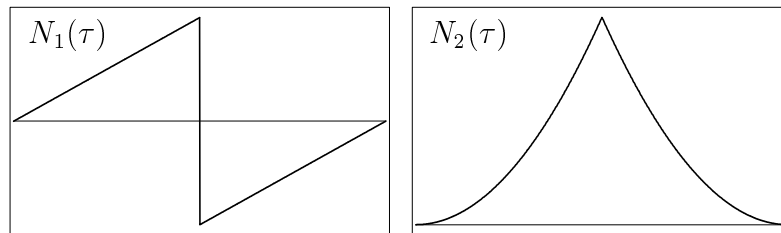
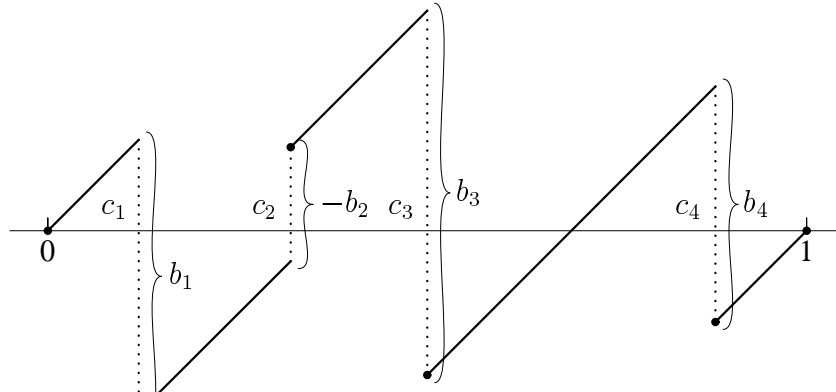


FIG. I.7: Noyaux de Peano pour la formule du point milieu

FIG. I.8. Le noyau de Peano $N_1(\tau)$ d'une formule de quadrature

Dans le **cas général** on a *plusieurs* fonctions delta centrés aux noeuds c_i et multipliés par b_i , le noyau $N_1(\tau)$ est linéaire par morceaux, de pente 1 et avec des sauts de hauteur $-b_i$ aux points c_i (voir figure I.8). On a $N_1(1) = 0$ car $b_1 + b_2 + b_3 + \dots = 1$.

Après la deuxième intégration on a

$$N_2(\tau) = \begin{cases} \tau^2/2 & \text{si } \tau \leq c_1 \\ \tau^2/2 - b_1(\tau - c_1) & \text{si } c_1 \leq \tau \leq c_2 \\ \tau^2/2 - b_1(\tau - c_1) - b_2(\tau - c_2) & \text{si } c_2 \leq \tau \leq c_3 \\ \dots & \text{etc.} \end{cases}$$

La propriété $N_2(1) = 0$ devient $\frac{1}{2} - b_1(1 - c_1) - b_2(1 - c_2) - \dots = 0$ et est la condition d'ordre (1.7) pour $g(t) = 1 - t$.

Et ainsi de suite ; aussi longtemps que $k \leq p$ on obtient $N_k(1) = 0$.

Nous pouvons résumer :

Théorème 2.1 (Peano 1914). *Considérons une formule de quadrature d'ordre p et un entier k satisfaisant $k \leq p$. Si $f : [x_0, x_0 + h] \rightarrow \mathbb{R}$ est k fois continûment différentiable, l'erreur (2.2) vérifie*

$$err_{loc} = (-1)^k h^{k+1} \int_0^1 N_k(\tau) f^{(k)}(x_0 + \tau h) d\tau \quad (2.9)$$

où $N_k(\tau)$, les **noyaux de Peano**, satisfont

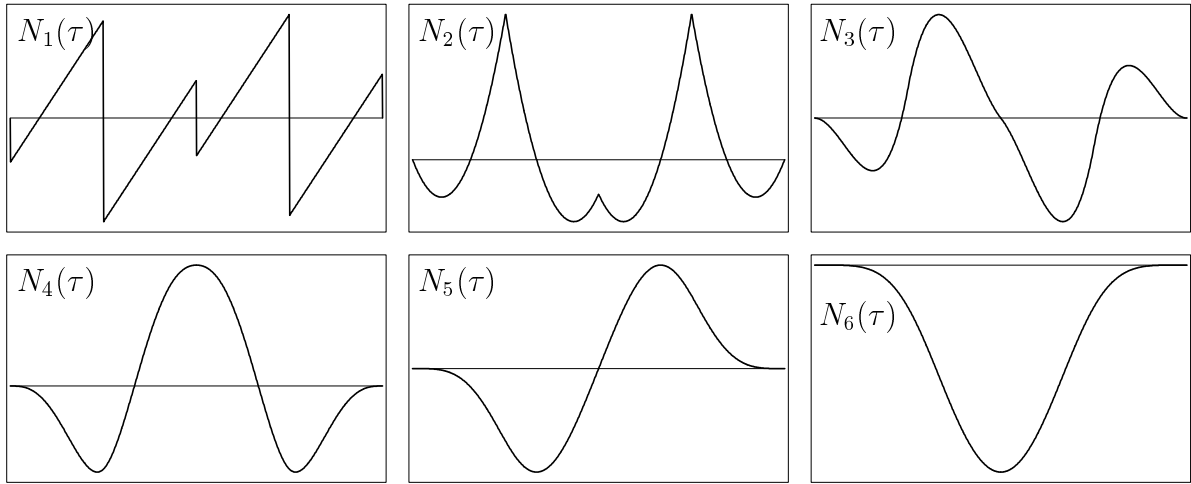
- a) $N'_k(\tau) = N_{k-1}(\tau)$ pour $k \geq 2$ (pour $\tau \neq c_i$ si $k = 2$);
- b) $N_k(1) = 0$ pour $k \geq 1$ si $c_i \leq 1$ ($i = 1, \dots, s$);
- c) $N_k(0) = 0$ pour $k \geq 2$ si $c_i \geq 0$ ($i = 1, \dots, s$);
- d) $\int_0^1 N_p(\tau) d\tau = \frac{1}{p!} \left(\frac{1}{p+1} - \sum_{i=1}^s b_i c_i^p \right) = C$ (constante de l'erreur (2.4));

□

Les noyaux de Peano pour la formule de Newton-Cotes ($s = 5$) sont dessinés dans la fig. I.9.

Grâce au résultat du théorème précédent, on peut facilement estimer l'erreur pour l'intervalle entier $[a, b]$. Pour une division arbitraire (équidistante ou non; $h_j = x_{j+1} - x_j$), notons l'erreur par

$$err = \int_a^b f(x) dx - \sum_{j=0}^{N-1} h_j \sum_{i=1}^s b_i f(x_j + c_i h_j). \quad (2.10)$$

FIG. 1.9: Noyaux de Peano pour la formule de Newton-Cotes avec $s = 5$

Théorème 2.2 Soit $f : [a, b] \rightarrow \mathbb{R}$ k fois continûment différentiable et soit l'ordre de la formule de quadrature égal à p ($p \geq k$). Alors, l'erreur (2.10) admet l'estimation

$$|err| \leq h^k \cdot (b - a) \cdot \int_0^1 |N_k(\tau)| d\tau \cdot \max_{x \in [a, b]} |f^{(k)}(x)| \quad (2.11)$$

où $h = \max_j h_j$.

Démonstration. La formule (2.9) donne

$$\begin{aligned} |err_{loc}| &\leq h^{k+1} \int_0^1 |N_k(\tau)| \cdot |f^{(k)}(x_0 + \tau h)| d\tau \\ &\leq h^{k+1} \int_0^1 |N_k(\tau)| d\tau \cdot \max_{x \in [x_0, x_0+h]} |f^{(k)}(x)|. \end{aligned}$$

Comme l'erreur (2.10) est la somme des erreurs sur les sous-intervalles de la division, on obtient

$$\begin{aligned} |err| &\leq \sum_{j=0}^{N-1} |err_{loc_j}| \leq \sum_{j=0}^{N-1} \underbrace{h_j^{k+1}}_{\leq h^k \cdot h_j} \cdot \int_0^1 |N_k(\tau)| d\tau \cdot \underbrace{\max_{x \in [x_j, x_{j+1}]} |f^{(k)}(x)|}_{\leq \max_{x \in [a, b]} |f^{(k)}(x)|} \\ &\leq h^k \cdot (b - a) \cdot \int_0^1 |N_k(\tau)| d\tau \cdot \max_{x \in [a, b]} |f^{(k)}(x)| \end{aligned}$$

ce qui montre l'assertion (2.11), car $\sum_{j=0}^{N-1} h_j = b - a$. □

Exemples. Pour la formule du point milieu, on a

$$|err| \leq h^2 \cdot (b - a) \cdot \frac{1}{24} \cdot \max_{x \in [a, b]} |f''(x)|;$$

pour la formule du trapèze

$$|err| \leq h^2 \cdot (b - a) \cdot \frac{1}{12} \cdot \max_{x \in [a, b]} |f''(x)|;$$

pour la formule de Simpson

$$|err| \leq h^4 \cdot (b - a) \cdot \frac{1}{2880} \cdot \max_{x \in [a, b]} |f^{(4)}(x)|.$$

I.3 Formules de quadrature de Gauss

Aber Gauss hat in den Göttinger Commentarien gezeigt, dass man durch schickliche Wahl der Abscissen, für welche die Ordinaten berechnet werden, den Grad der Näherung auf das Doppelte treiben kann ; ... Die grosse Einfachheit und Eleganz der Gaussischen Resultate, lässt einen einfachen Weg vermuten. (Jacobi 1826, *Crelle J.* 1, p. 302)

Si l'on fixe les nœuds $c_1, \dots, c_s$ (distincts), il existe une formule de quadrature (b_i, c_i) unique, ayant un ordre $p \geq s$. On obtient les poids b_i par les formules de la section I.1.

Question (Buzengeiger 1812). Y a-t-il un choix des c_i permettant d'avoir un ordre supérieur à s ?

La solution (difficile) de Gauss (1814, *Werke* 3, p. 165-196) qui, selon ses habitudes, ne mentionne jamais le nom de Buzengeiger, est basée sur la théorie de la série hypergéométrique (Gauss 1812)

$$F(\alpha, \beta, \gamma, x) = 1 + \frac{\alpha\beta}{1 \cdot \gamma}x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{1 \cdot 2 \cdot \gamma(\gamma+1)}x^2 + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1 \cdot 2 \cdot 3 \cdot \gamma(\gamma+1)(\gamma+2)}x^3 + etc.$$

Douze années plus tard, Jacobi (1826, *Crelle J.* vol. 1) découvre l'accès élégant suivant :

Idée. Posons

$$M(t) = (t - c_1) \cdot \dots \cdot (t - c_s) \quad (3.1)$$

("Man bilde das Product $(x - a')(x - a'')(x - a''') \dots (x - a^{(n)})$ und nenne es $\varphi(x)$ "). Soit $f(t)$ un polynôme de degré $\leq s + m - 1$. L'idée est de diviser $f(t)$ par $M(t)$, i.e., d'écrire $f(t)$ sous la forme

$$\underbrace{f(t)}_{\text{deg.} \leq s+m-1} = \underbrace{M(t)}_s \underbrace{g(t)}_{\leq m-1} + \underbrace{r(t)}_{\leq s-1}$$

Alors, l'intégrale exacte et l'approximation numérique satisfont

$$\begin{aligned} \int_0^1 f(t) dt &= \int_0^1 M(t)g(t) dt + \int_0^1 r(t) dt \\ \parallel? &\quad \parallel? \quad \parallel \\ \sum_{i=1}^s b_i f(c_i) &= \sum_{i=1}^s b_i \underbrace{M(c_i)}_{=0} g(c_i) + \sum_{i=1}^s b_i r(c_i). \end{aligned}$$

La formule de quadrature est exacte pour $r(t)$ par hypothèse (dernière colonne). Par conséquent, les deux égalités marquées "?" sont satisfaites en même temps. Nous constatons que *la formule est exacte pour $f(t)$ si et seulement si $\int_0^1 M(t)g(t) dt = 0$* . En résumé :

Théorème 3.1 (Superconvergence) Soit $(b_i, c_i)_{i=1}^s$ une formule de quadrature d'ordre $p \geq s$. Alors, l'ordre de la formule est $\geq s + m$ si et seulement si

$$\int_0^1 M(t)g(t) dt = 0 \quad \text{pour tout polynôme } g(t) \text{ de degré } \leq m - 1. \quad \square$$

Exemple 3.2 Pour qu'une formule de quadrature à $s = 2$ étages ait un ordre ≥ 3 , il faut que

$$\int_0^1 (t - c_1)(t - c_2) dt = \frac{1}{3} - (c_1 + c_2)\frac{1}{2} + c_1c_2 = 0.$$

Similairement, pour avoir l'ordre $p = 4$, il faut que

$$\begin{aligned} c_1c_2 - \frac{1}{2}(c_1 + c_2) &= -\frac{1}{3} \\ \frac{1}{2}c_1c_2 - \frac{1}{3}(c_1 + c_2) &= -\frac{1}{4} \end{aligned} \quad \Rightarrow \quad \begin{aligned} c_1 + c_2 &= 1 \\ c_1c_2 &= \frac{1}{6} \end{aligned} \quad \Rightarrow \quad \begin{aligned} c_1 &= \frac{1}{2} - \frac{\sqrt{3}}{6} \\ c_1 &= \frac{1}{2} + \frac{\sqrt{3}}{6} \end{aligned}. \quad (3.2)$$

Exemple 3.3 (Version symétrique ; GAUSS6, une formule d'ordre 6) Travaillons sur l'intervalle $[-1, 1]$ et remplaçons $M(t)$ par $P(\tau) = (\tau - \gamma_1) \dots (\tau - \gamma_s)$. Il faut satisfaire $\int_{-1}^1 \tau^{q-1} P(\tau) d\tau = 0$ pour $q = 1, 2, 3$. Si nous posons en respectant la symétrie

$$P(\tau) = \tau^3 - a \cdot \tau \quad (a \text{ constante à chercher}),$$

nous avons la première et la troisième condition automatiquement satisfaite. Il suffit de demander

$$\begin{aligned} \int_{-1}^1 \tau P(\tau) d\tau &= \int_{-1}^1 (\tau^4 - a \cdot \tau^2) d\tau = 2\left(\frac{1}{5} - a \cdot \frac{1}{3}\right) = 0 \quad \Rightarrow \quad P(\tau) = \tau^3 - \frac{3}{5} \tau \\ \Rightarrow \quad \gamma_1 &= -\frac{\sqrt{15}}{5}, \quad \gamma_2 = 0, \quad \gamma_3 = \frac{\sqrt{15}}{5}, \quad \beta_1 = \beta_3 = \frac{1}{3\gamma_1^2} = \frac{5}{9}, \quad \beta_2 = 2 - 2\beta_1 = \frac{8}{9} \end{aligned}$$

ce qui amène à une *formule d'ordre 6 basée sur $s = 3$ ordonnées*.

Exemple 3.4 (Résumé des premières formules de Gauss)

$$\begin{aligned} s = 1 : \quad \int_0^1 g(t) dt &\approx g\left(\frac{1}{2}\right) && \text{(formule du point milieu)} \\ s = 2 : \quad \int_0^1 g(t) dt &\approx \frac{1}{2}g\left(\frac{1}{2} - \frac{\sqrt{3}}{6}\right) + \frac{1}{2}g\left(\frac{1}{2} + \frac{\sqrt{3}}{6}\right) \\ s = 3 : \quad \int_0^1 g(t) dt &\approx \frac{5}{18}g\left(\frac{1}{2} - \frac{\sqrt{15}}{10}\right) + \frac{8}{18}g\left(\frac{1}{2}\right) + \frac{5}{18}g\left(\frac{1}{2} + \frac{\sqrt{15}}{10}\right) \\ s = 4 : \quad \int_0^1 g(t) dt &\approx \mu g\left(\frac{1}{2} - \delta\right) + \mu' g\left(\frac{1}{2} - \delta'\right) + \mu' g\left(\frac{1}{2} + \delta'\right) + \mu g\left(\frac{1}{2} + \delta\right) \\ s = 5 : \quad \int_0^1 g(t) dt &\approx \nu g\left(\frac{1}{2} - \epsilon\right) + \nu' g\left(\frac{1}{2} - \epsilon'\right) + \frac{64}{225}g\left(\frac{1}{2}\right) + \nu' g\left(\frac{1}{2} + \epsilon'\right) + \nu g\left(\frac{1}{2} + \epsilon\right) \\ \delta &= \frac{1}{2}\sqrt{\frac{15 + 2\sqrt{30}}{35}}, \quad \delta' = \frac{1}{2}\sqrt{\frac{15 - 2\sqrt{30}}{35}}, \quad \mu = \frac{1}{4} - \frac{\sqrt{30}}{72}, \quad \mu' = \frac{1}{4} + \frac{\sqrt{30}}{72}, \\ \epsilon &= \frac{1}{2}\sqrt{\frac{35 + 2\sqrt{70}}{63}}, \quad \epsilon' = \frac{1}{2}\sqrt{\frac{35 - 2\sqrt{70}}{63}}, \quad \nu = \frac{322 - 13\sqrt{70}}{1800}, \quad \nu' = \frac{322 + 13\sqrt{70}}{1800}. \end{aligned}$$

Expérience numérique. Après avoir trouvé de nouvelles méthodes si belles, nous sommes évidemment impatients et joyeux d'aller au combat et de refaire les intégrales $\int_0^3 \cos(x) \exp(\sin(x)) dx$ et $\int_0^2 \cos(x) dx$ de la figure I.5. Les résultats correspondants peuvent être admirés en figure I.10 ; ils montrent une claire amélioration.

Pour progresser davantage, nous avons besoin de nouvelles idées ; cherchons “un très-grand secours” auprès des “polynômes de Legendre” :

I.4 Polynômes orthogonaux de Legendre

On sait que les fonctions X_n et Y_n , introduites dans l'analyse par Legendre, sont d'un très-grand secours dans plusieurs théories importantes, en particulier dans la théorie de l'attraction des sphéroïdes et dans celle de la figure des planètes ; ...

(O. Bonnet, *J.d.math.* vol. 17, 1852, p. 265)

Problème. Trouver, pour chaque entier positif k , un polynôme $P_k(t)$ de degré k tel que

$$\int_{-1}^1 P_k(t)g(t) dt = 0 \quad \text{si } \deg g \leq k-1. \quad (4.1)$$

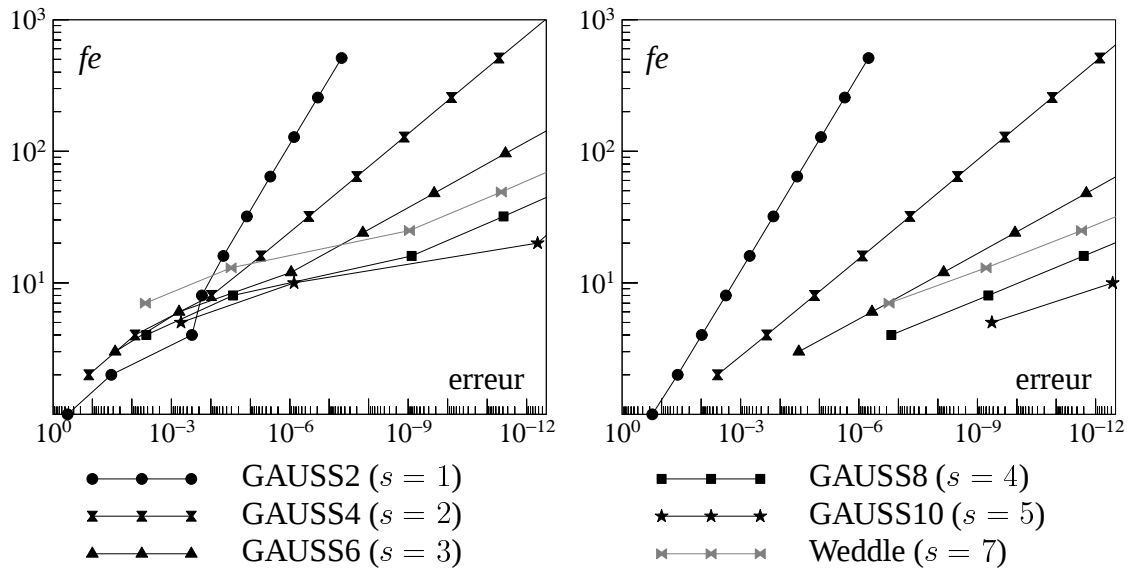


FIG. I.10: Le travail fe en fonction de l'erreur pour les formules de Gauss (en gris sont répétés les résultats pour Weddle de la figure I.5)

Ces polynômes ont été introduits en 1785 par Legendre dans un travail sur l'attraction des sphéroïdes ; ils sont d'une importance dépassant largement les formules de quadrature (voir citation). On les retrouve en chimie et en physique dans tous les calculs en coordonnées sphériques ; ils y jouent le même rôle que celui des fonctions trigonométriques dans les séries de Fourier (voir Chap. II). La construction élégante ci-dessous est due à O. Rodrigues (dans sa Thèse nonpubliée 1816 à l'École Normale), redécouverte par Jacobi (1826) :

Solution. L'idée est d'intégrer par parties k fois l'intégrale dans (4.1), en *dérivant* le facteur $g(t)$ et en *intégrant* le facteur $P_k(t)$ (rappel: $\int_a^b u'v dt = u(b)v(b) - u(a)v(a) - \int_a^b uv' dt$). Après la k -ème différentiation, le polynôme $g(t)$ disparaîtra.

Mais pour que cela marche, il faut que les termes intégrés $u(b)v(b) - u(a)v(a)$ sont tous nuls, i.e., il faut que $P_k(t)$ possède k primitives successives qui s'annulent en $t = -1$ et en $t = 1$!

Idée. Rien de plus simple : on pose

$$F(t) = (t^2 - 1)^k = (t - 1)^k \cdot (t + 1)^k$$

qui possède deux zéros de multiplicité k à ces points. Donc les $k - 1$ dérivées de $F(t)$ y seront nulles également ; la k -ème dérivée sera notre polynôme recherché

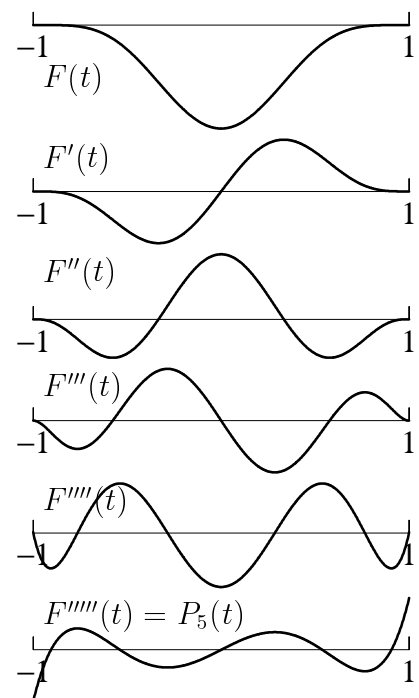
$$P_k(t) = C \cdot \left((t^2 - 1)^k \right)^{(k)}. \quad (4.2)$$

Exemple. Prenons $k = 5$ tel que

$$F(t) = (t^2 - 1)^5 = t^{10} - 5t^8 + 10t^6 - 10t^4 + 5t^2 - 1.$$

En dérivant 5 fois cela donne

$$10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot t^5 - 5 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot t^3 + 10 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot t$$



et, en divisant des facteurs communs, nous voyons que $P_5(t)$ est un multiple de $63t^5 - 70t^3 + 15t$, ce que nous avons déjà utilisé.

Théorème 4.1 *Le polynôme $P_k(t)$ défini dans (4.2) est orthogonal sur l'intervalle $[-1, 1]$ à tous les polynômes de degré $k - 1$.*

Démonstration. Comme promis, nous faisons une suite d'intégrations par parties, et ainsi l'intégrale $\int_{-1}^1 P_k(t)g(t) dt$ devient

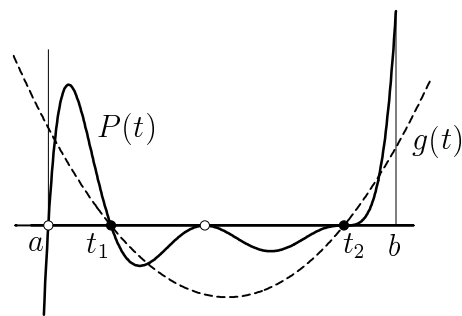
$$= C \cdot \int_{-1}^1 F^{(k)}(t)g(t) dt = -C \cdot \int_{-1}^1 F^{(k-1)}(t)g'(t) dt = \dots = \pm C \cdot \int_{-1}^1 F^{(1)}(t) \underbrace{g^{(k)}(t)}_{=0} dt = 0.$$

□

Théorème 4.2 *Toutes les racines de $P_k(t)$ sont réelles, simples et dans l'intervalle ouvert $(-1, 1)$.*

Démonstration. Cela se voit en appliquant le “théorème de Rolle” aux diverses dérivées de $F(t)$ (voir la figure ci-dessus) : entre chaque paire de racines successives (y compris les racines -1 et 1) de $F^{(j)}(t)$ doit se trouver une racine de $F^{(j+1)}(t)$. Ainsi, $P_k(t)$ doit posséder au moins k racines différentes à l'intérieur de $(-1, 1)$. Et, s'agissant d'un polynôme de degré k , il n'y en a pas d'autres. □

Donnons une deuxième preuve ! Combien de fois n'est-on pas soulagé à la fin d'une longue et pénible démonstration ? Ici, c'est le contraire : on peut choisir entre plusieurs preuves, les unes aussi belles et élégantes que les autres ! Montrons le résultat précédent à l'aide du théorème suivant :



Théorème 4.3 *Soit un polynôme $P(t)$ orthogonal sur $[a, b]$ à tous les polynômes de degré $r - 1$. Alors $P(t)$ doit posséder au moins r racines de multiplicité impaire dans l'intervalle ouvert (a, b) .*

Démonstration. Supposons, comme dans le dessin ci-contre, que $P(t)$ possède deux racines t_1 et t_2 de multiplicité impaire dans (a, b) , i.e., des racines où P change de signe. Il suffit alors de poser $g(t) = (t - t_1)(t - t_2)$, car le produit $P(t) \cdot g(t)$ ne change pas de signe dans (a, b) . L'intégrale ne peut donc pas être nulle et les deux polynômes ne sont pas orthogonaux. □

Corollaire 4.4 (Gauss 1814) *Pour chaque entier positif s , il existe une formule de quadrature unique à s noeuds d'ordre $2s$. Elle est donnée par :*

$c_1, \dots, c_s$ sont les racines de $P_s(2t - 1)$;

$b_1, \dots, b_s$ sont donnés par (1.9).

□

Normalisation des $P_k(t)$. Si l'on multiplie un polynôme orthogonal par une constante, il reste un polynôme et reste orthogonal. Donc, en principe, on peut "normaliser" les polynômes $P_k(t)$ comme on veut ; une normalisation est adoptée dans la littérature, elle demande que

$$P_k(1) = 1. \quad (4.3)$$

Comme, par la règle de Leibniz,

$$\frac{d^k}{dt^k} \left((t+1)^k (t-1)^k \right) \Big|_{t=1} = k! \cdot 2^k,$$

nous pouvons préciser la constante dans (4.2), et les polynômes de Legendre deviennent

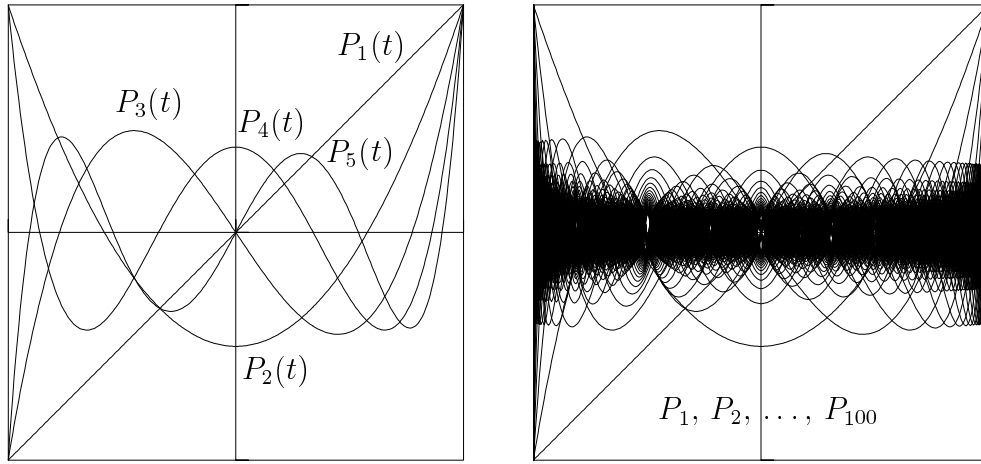


FIG. I.11: Polynômes de Legendre

$$P_k(t) = \frac{1}{k! \cdot 2^k} \cdot \frac{d^k}{dt^k} \left((t^2 - 1)^k \right) = \frac{1}{2^k} \binom{2k}{k} t^k - \frac{1}{2^k} \binom{k}{1} \binom{2k-2}{k} t^{k-2} + \dots \quad (4.4)$$

Les premiers de ces polynômes sont $P_0(t) = 1$, $P_1(t) = t$,

$$P_2(t) = \frac{3}{2}t^2 - \frac{1}{2}, \quad P_3(t) = \frac{5}{2}t^3 - \frac{3}{2}t, \quad P_4(t) = \frac{35}{8}t^4 - \frac{30}{8}t^2 + \frac{3}{8}, \quad P_5(t) = \frac{63}{8}t^5 - \frac{70}{8}t^3 + \frac{15}{8}t. \quad (4.5)$$

Ils sont dessinés dans la fig. I.11 et sont alternativement des fonctions paires et impaires :

$$\begin{aligned} P_k(t) &= P_k(-t) && \text{si } k \text{ est pair} \\ P_k(t) &= -P_k(-t) && \text{si } k \text{ est impair.} \end{aligned} \quad (4.6)$$

Formule de récurrence. La relation suivante, découverte par O. Bonnet (1852, *J. de Math.* vol. 17, p. 267), permet facilement de calculer la valeur de $P_s(x)$ pour un x donné ; et sera utile pour le calcul numérique des c_i .

Théorème 4.5 Les polynômes de Legendre satisfont la formule de récurrence

$$(k+1) P_{k+1}(t) = (2k+1) t P_k(t) - k P_{k-1}(t). \quad (4.7)$$

Démonstration. Si on multiplie $P_k(t)$ par t , il possède les mêmes puissances de t que $P_{k+1}(t)$. On peut donc soustraire, avec un facteur a bien choisi, pour faire disparaître le premier terme t^{k+1} . Le prochain terme, t^{k-1} , est éliminé par un multiple de $P_{k-1}(t)$. Le terme suivant t^{k-3} est éliminé par un multiple de $P_{k-3}(t)$ etc. Donc on peut écrire

$$P_{k+1}(t) = a \cdot t P_k(t) + b \cdot P_{k-1}(t) + c \cdot P_{k-3}(t) + d \cdot P_{k-5}(t) + \dots \quad (4.8)$$

Grande surprise : tous les coefficients $c, d, \dots$ sont nuls ! Pour voir cela, multiplions l'équation (4.8) par $P_{k-3}(t)$ et intégrons le tout de -1 à 1 . Par orthogonalité, tous les termes vont s'annuler, p.ex.,

$$\int_{-1}^1 t P_k(t) \cdot P_{k-3}(t) dt = \int_{-1}^1 P_k(t) \cdot \underbrace{t \cdot P_{k-3}(t)}_{g(t)} dt = 0,$$

sauf $\int_{-1}^1 (P_{k-3}(t))^2 dt > 0$, et c doit être zéro. La même chose s'applique à d, e etc.

En comparant le premier terme avec celui de (4.4), on trouve

$$a = \frac{2k+1}{k+1}, \quad \text{et en posant } t = 1, \text{ on trouve} \quad b = 1 - a = -\frac{k}{k+1}.$$

□

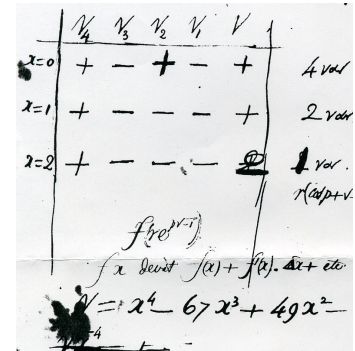
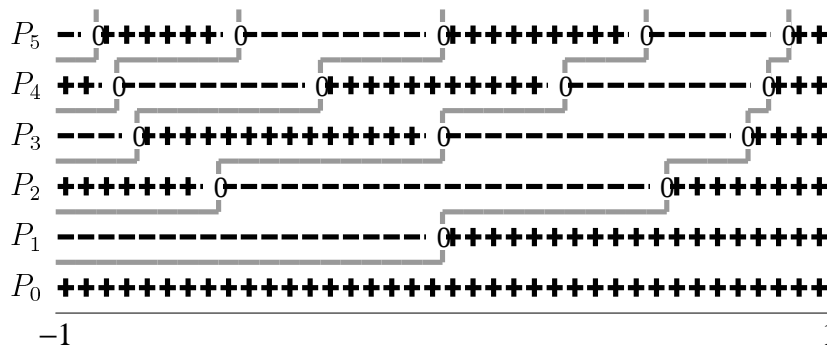


FIG. I.12: Création des racines des $P_k(t)$ par la suite de Sturm; à droite : manuscrit de Sturm (trouvé par E. Neuenschwander dans un grenier en France)

Avec ce théorème, on constate :

Si $P_k(t)$ est nul, alors $P_{k+1}(t)$ et $P_{k-1}(t)$ sont de signes opposés !

C'est la célèbre propriété des *suites de Sturm* (Charles Sturm 1829). Elle permet de prouver le théorème 4.2 une *troisième* fois, en considérant uniquement (voir figure I.12)

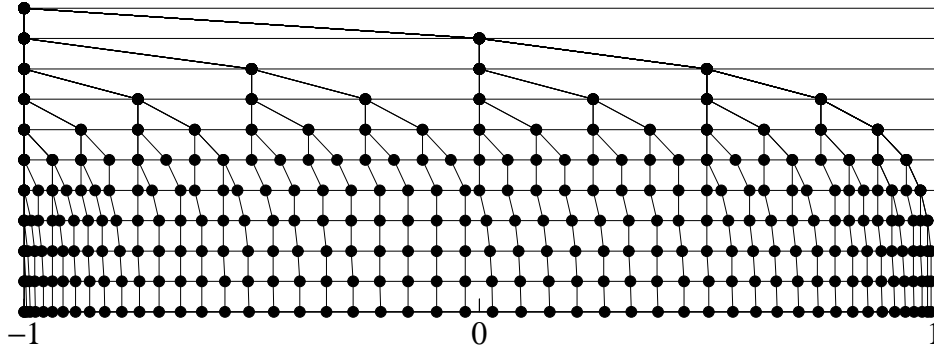
1. les valeurs des P_k en -1 et en $+1$;
2. la propriété de Sturm ;
3. et le "théorème des valeurs intermédiaires" ([HW97], p. 206).

Il y a cependant des informations complémentaires :

les racines de $P_{k-1}(t)$ sont intercalés dans les racines de $P_k(t)$!

et

chang. signe de $\{P_0(t_0), P_1(t_0), \dots, P_k(t_0)\} = \#$ de racines de $P_k(t)$ étant $> t_0$.

FIG. I.13: Calcul des racines des $P_{50}(t)$ par la méthode de bisection)

$\Rightarrow$ *algorithme de bisection* (voir Fig. I.13).

Calcul des poids. Au lieu de résoudre le système (1.9), on peut aussi utiliser la formule explicite (Christoffel, *Crelle J.* vol. 55 (1858), p. 61-82, “formule (37.)”)

$$b_i = \frac{1}{(1 - \gamma_i^2)(P'_s(\gamma_i))^2} = \frac{1 - \gamma_i^2}{s^2(P_{s-1}(\gamma_i))^2}, \quad (4.9)$$

qui est donnée ici sans démonstration (exercices).

I.5 Un programme adaptatif – TEGRAL

Posons-nous le problème d’écrire un programme

FUNCTION TEGRAL (FCN, A, B, TOL)

qui, pour une fonction $f : [a, b] \rightarrow \mathbb{R}$ donnée, calcule la valeur de $\int_a^b f(x) dx$ à une précision relative de TOL. Si l’on fixe la formule de quadrature (par exemple la formule de Gauss d’ordre 30 avec $s = 15$), il faut trouver une division $\Delta = \{a = x_0 < x_1 < \dots < x_N = b\}$ de l’intervalle (a, b) afin que l’approximation numérique I_Δ satisfasse

$$\left| I_\Delta - \int_a^b f(x) dx \right| \leq \text{TOL} \cdot \int_a^b |f(x)| dx. \quad (5.1)$$

Pour une fonction f ne changeant pas de signe sur (a, b) , la condition (5.1) signifie que l’*erreur relative* est bornée par TOL. On a mis la valeur absolue sous l’intégrale de droite pour éviter des ennuis dans le cas où $\int_a^b f(x) dx$ est très petit ou nul.

Pour écrire un tel programme, on est confronté aux deux problèmes suivants:

- choix de la division pour que (5.1) soit satisfait;
- estimation de l’erreur $I_\Delta - \int_a^b f(x) dx$.

Détermination de la division. Pour un sous-intervalle $(x_0, x_0 + h)$ de (a, b) , on sait calculer les valeurs

$$\text{res}(x_0, x_0 + h) = h \sum_{i=1}^s b_i f(x_0 + c_i h), \quad (5.2)$$

$$\text{resabs}(x_0, x_0 + h) = h \sum_{i=1}^s b_i |f(x_0 + c_i h)|. \quad (5.3)$$

Supposons, pour le moment, qu'on connaisse aussi une estimation de l'erreur

$$\text{err}(x_0, x_0 + h) \approx \text{res}(x_0, x_0 + h) - \int_{x_0}^{x_0+h} f(x) dx. \quad (5.4)$$

L'algorithme pour trouver une division convenable est le suivant:

i) on calcule $\text{res}(a, b)$, $\text{resabs}(a, b)$ et $\text{err}(a, b)$. Si

$$|\text{err}(a, b)| \leq \text{TOL} \cdot \text{resabs}(a, b),$$

on accepte $\text{res}(a, b)$ comme approximation de $\int_a^b f(x) dx$ et on arrête le calcul; sinon

ii) on subdivise (a, b) en deux sous-intervalles $I_1 = (a, (a+b)/2)$ et $I_2 = ((a+b)/2, b)$ et on calcule $\text{res}(I_1)$, $\text{resabs}(I_1)$, $\text{err}(I_1)$ et $\text{res}(I_2)$, $\text{resabs}(I_2)$, $\text{err}(I_2)$. On pose $N = 2$ et on regarde si

$$\sum_{j=1}^N |\text{err}(I_j)| \leq \text{TOL} \cdot \left(\sum_{j=1}^N \text{resabs}(I_j) \right). \quad (5.5)$$

Si (5.5) est vérifié, on accepte $\text{res}(I_1) + \text{res}(I_2)$ comme approximation de $\int_a^b f(x) dx$; sinon

iii) on pose $N := N + 1$ et on subdivise l'intervalle où l'erreur est maximale (disons I_k) en deux sous-intervalles équidistants qu'on denote par I_k et I_{N+1} . Ensuite, on calcule res , resabs et err pour ces deux intervalles. Si (5.5) est vérifié, on arrête le calcul et on accepte

$$\sum_{j=1}^N \text{res}(I_j) \approx \int_a^b f(x) dx \quad (5.6)$$

comme approximation de l'intégrale; sinon on répète la partie (iii) de cet algorithme.

Estimation de l'erreur (5.4). Malheureusement, les formules pour l'erreur, obtenues dans le paragraphe I.2, ne sont pas très utiles pour un programme général, car on ne connaît que très rarement la $p^{\text{ème}}$ dérivée de la fonction $f(x)$ (dans notre situation $p = 30$).

L'idée est d'appliquer une deuxième formule de quadrature $(\hat{b}_i, \hat{c}_i)_{i=1}^{\hat{s}}$ et d'utiliser la différence de deux approximations numériques comme estimation de l'erreur du moins bon résultat. Pour que le travail supplémentaire soit négligeable, on suppose $\hat{s} \leq s$ et on reprend les mêmes évaluations de f , c.-à-d., on suppose $\hat{c}_i = c_i$ pour tous i . Une telle formule de quadrature s'appelle *formule emboîtée*, si pour au moins un indice i on a $\hat{b}_i \neq b_i$.

Remarque. Si $(b_i, c_i)_{i=1}^s$ est une formule de quadrature d'ordre $p \geq s$, l'ordre d'une formule emboîtée est $\hat{p} \leq s - 1$. Ce résultat découle du fait que, pour une formule de quadrature d'ordre $\geq s$, les poids b_i sont uniquement déterminés par ses nœuds c_i .

Pour la formule de Gauss ($s = 15$, $p = 30$), on obtient une formule emboîtée $(\hat{b}_i, c_i)_{i=1}^{\hat{s}}$ d'ordre 14 en enlevant le point milieu $c_8 = 1/2$ (voir fig. I.14), c.-à-d., on pose $\hat{b}_8 = 0$. L'expression calculable

$$\text{ERR1} := h \sum_{i=1}^s b_i f(x_0 + c_i h) - h \sum_{i=1}^{\hat{s}} \hat{b}_i f(x_0 + c_i h) \approx C_1 h^{15} \quad (5.7)$$

est une approximation de l'erreur de la formule emboîtée, car

$$\begin{aligned} \text{ERR1} &= \underbrace{\left(h \sum_{i=1}^s b_i f(x_0 + c_i h) - \int_{x_0}^{x_0+h} f(x) dx \right)}_{= Ch^{31} + \mathcal{O}(h^{32})} + \underbrace{\left(\int_{x_0}^{x_0+h} f(x) dx - h \sum_{i=1}^{\hat{s}} \hat{b}_i f(x_0 + c_i h) \right)}_{= Ch^{15} + \mathcal{O}(h^{16})}. \end{aligned}$$

Pour le programme TEGRAL, on considère encore une deuxième formule emboîtée qui a pour nœuds $\{c_2, c_4, c_6, c_{10}, c_{12}, c_{14}\}$ et un ordre 6 (voir la fig. I.14). On dénote les poids de cette formule de quadrature par $\hat{\hat{b}}_i$ et on définit

$$\text{ERR2} := h \sum_{i=1}^s b_i f(x_0 + c_i h) - h \sum_{i=1}^s \hat{\hat{b}}_i f(x_0 + c_i h) \approx C_2 h^7. \quad (5.8)$$

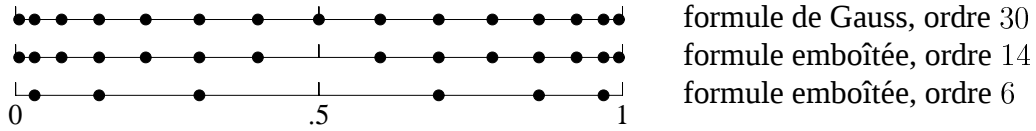


FIG. I.14: Formule de Gauss et ses formules emboîtées

Il y a plusieurs possibilités pour définir $\text{err}(x_0, x_0 + h)$.

- i) $\text{err}(x_0, x_0 + h) := \text{ERR1}$; cette estimation est trop pessimiste. En général, la formule de Gauss donne un résultat largement meilleur que la formule emboîtée d'ordre 14.
- ii) dans le programme TEGRAL, on a choisi l'approximation

$$\text{err}(x_0, x_0 + h) := \text{ERR1} \cdot \left(\frac{\text{ERR1}}{\text{ERR2}} \right)^2, \quad \left(\approx h^{15} \cdot \left(\frac{h^{15}}{h^7} \right)^2 \approx h^{31} \right)$$

ce qui donne de bons résultats.

Exemples. 1) Si l'on applique le programme TEGRAL avec $\text{TOL} = 10^{-10}$ à la fonction de la fig. I.1, on obtient le résultat avec une erreur de $2.0 \cdot 10^{-14}$. La division choisie est donnée dans la fig. I.15.

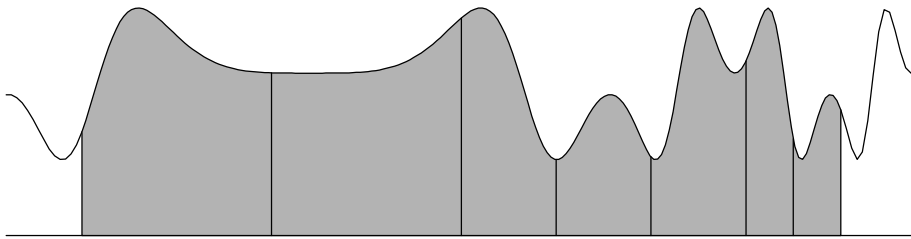


FIG. I.15: Division choisie par TEGRAL pour $f(x) = 2 + \sin(3 \cos(0.002(x-40)^2))$ sur $(10, 110)$

2) Appliquons le même programme à la fonction

$$f(x) = \sqrt{x} \cdot \log x \quad \text{sur} \quad (0, 1), \quad (5.9)$$

qui a une singularité au point 0 dans la dérivée. Les divisions successives de l'intervalle par l'algorithme sont présentées dans la fig. I.16. On voit que l'intervalle où l'erreur est maximale est toujours celui qui est tout à gauche. Les erreurs sont données dans le tableau I.2. La convergence est très lente et on se demande s'il n'y a pas de possibilité d'accélérer la convergence de la suite $\{S_N\}$.

TAB. I.2: Résultat de TEGRAL pour (5.9)

N	S_N	$err_N = S_N - \int_0^1 f(x) dx$	err_N/err_{N-1}
1	-0.4446200164956040	$-0.176 \cdot 10^{-03}$	—
2	-0.4445133092592463	$-0.689 \cdot 10^{-04}$	0.392
3	-0.4444711927155809	$-0.267 \cdot 10^{-04}$	0.388
4	-0.4444547502264998	$-0.103 \cdot 10^{-04}$	0.385
5	-0.4444483881989293	$-0.394 \cdot 10^{-05}$	0.383
6	-0.4444459448772271	$-0.150 \cdot 10^{-05}$	0.380
.	...	...	...
21	-0.4444444444449658	$-0.521 \cdot 10^{-12}$	0.366
22	-0.4444444444446352	$-0.191 \cdot 10^{-12}$	0.366

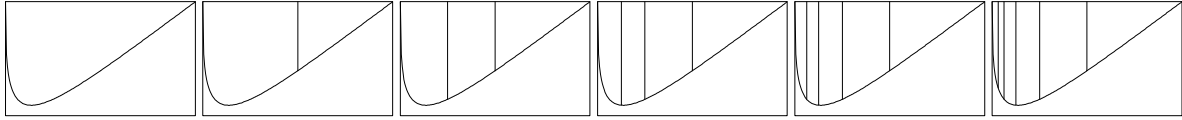


FIG. I.16: Division choisie par TEGRAL pour (5.9)

I.6 L'epsilon-algorithme

Soit donnée une suite $\{S_0, S_1, S_2, \dots\}$ qui converge lentement vers une valeur S . Le but est de trouver une autre suite avec la même limite, mais qui converge plus rapidement.

Idée. (Aitken 1926). Supposons que les S_n soient les valeurs d'une série géométrique

$$\begin{aligned}
 S_n &= S + C \cdot \rho^n \\
 S_{n+1} &= S + C \cdot \rho^{n+1} \\
 S_{n+2} &= S + C \cdot \rho^{n+2}
 \end{aligned}
 \quad
 \begin{aligned}
 \Delta S_n &= C\rho^n(\rho - 1) \\
 \Delta S_{n+1} &= C\rho^{n+1}(\rho - 1)
 \end{aligned}
 \quad
 \Delta^2 S_n = C\rho^n(\rho - 1)^2 \quad (6.1)$$

où $\Delta S_n := S_{n+1} - S_n$ et $\Delta^2 S_n = \Delta(\Delta S_n) = \Delta S_{n+1} - \Delta S_n = S_{n+2} - 2S_{n+1} + S_n$ sont des *différences finies*. On cherche à remplacer notre suite par la *limite* S , pour laquelle on obtient

$$S = S_{n+1} - C\rho^{n+1} = S_{n+1} - \frac{\Delta S_n \cdot \Delta S_{n+1}}{\Delta^2 S_n}. \quad (6.2)$$

Au cas où l'hypothèse n'est pas satisfait exactement, la valeur de S dans (6.2) va dépendre de n . On obtient ainsi une autre suite $\{S'_n\}$ définie par (procédé Δ^2 d'Aitken)

$$S'_n = S_{n+1} - \frac{\Delta S_n \cdot \Delta S_{n+1}}{\Delta^2 S_n}, \quad (6.3)$$

qui, en général, converge plus rapidement vers S que la suite originale $\{S_n\}$.

Exemple. Pour la suite $\{S_n\}$ du tableau I.2, le résultat est donné dans le tableau I.3.

Généralisation. Supposons pour une suite $\{S_n\}$

$$S_n = S + C_1 \cdot \rho_1^n + C_2 \cdot \rho_2^n. \quad (6.4)$$

TAB. I.3: Accélération de la convergence pour $\{S_n\}$ du tableau I.2

n	S_n	S'_n	S''_n
1	-0.44462001649560	-0.44444373050429	-0.44444444444445
2	-0.44451330925925	-0.44444421992844	-0.44444444444444
3	-0.44447119271558	-0.44444437296661	-0.44444444444444
4	-0.44445475022650	-0.44444442146079	-0.44444444444444
5	-0.44444838819893	-0.44444443699301	-0.44444444444444
6	-0.44444594487723	-0.44444444201188	-0.44444444444444

Cette fois, on a 5 paramètres à déterminer. La valeur de S ainsi obtenue est dénotée par S''_n . Shanks (1955) a fait ce calcul et il a trouvé la formule (nous ajoutons une formule semblable pour S'_n)

$$S'_n = \frac{\det \begin{pmatrix} S_n & S_{n+1} \\ S_{n+1} & S_{n+2} \end{pmatrix}}{\Delta^2 S_n}, \quad S''_n = \frac{\det \begin{pmatrix} S_n & S_{n+1} & S_{n+2} \\ S_{n+1} & S_{n+2} & S_{n+3} \\ S_{n+2} & S_{n+3} & S_{n+4} \end{pmatrix}}{\det \begin{pmatrix} \Delta^2 S_n & \Delta^2 S_{n+1} \\ \Delta^2 S_{n+1} & \Delta^2 S_{n+2} \end{pmatrix}}$$

(sans démonstration). Mais, pour un calcul numérique, ces formules ne sont pas très pratiques. Wynn (1956) a trouvé une formule beaucoup plus simple:

Théorème 6.1 (ϵ -algorithme) *Etant donnée la suite $\{S_0, S_1, S_2, \dots\}$. Si l'on définit $\epsilon_k^{(n)}$ par*

$$\begin{aligned} \epsilon_{-1}^{(n)} &= 0, & \epsilon_0^{(n)} &= S_n, \\ \epsilon_{k+1}^{(n)} &= \epsilon_{k-1}^{(n+1)} + \frac{1}{\epsilon_k^{(n+1)} - \epsilon_k^{(n)}}, & k \geq 0, n \geq 0, \end{aligned} \quad (6.5)$$

alors, $\epsilon_2^{(n)} = S'_n$, $\epsilon_4^{(n)} = S''_n$, $\epsilon_6^{(n)} = S'''_n, \dots$

La démonstration de $\epsilon_2^{(n)} = S'_n$ se fait par un simple calcul de $\epsilon_1^{(n)}$ et de $\epsilon_2^{(n)}$:

$$\begin{aligned} \epsilon_1^{(n)} &= 0 + \frac{1}{S_{n+1} - S_n} = \frac{1}{\Delta S_n}, \\ \epsilon_2^{(n)} &= S_{n+1} + \frac{1}{\frac{1}{\Delta S_{n+1}} - \frac{1}{\Delta S_n}} = S_{n+1} + \frac{\Delta S_n \cdot \Delta S_{n+1}}{\Delta S_n - \Delta S_{n+1}} = S'_n. \end{aligned}$$

Le cas général est moins évident et est donné sans démonstration. Tous les détails (démonstration et autres propriétés) sont donnés dans un livre de Brezinski².

Exemple. Pour la suite

$$S_n = \sum_{i=1}^n \frac{(-1)^{i+1}}{i}, \quad (6.6)$$

qui converge vers $\log(2)$, les erreurs de $\epsilon_k^{(n)}$, $k = 0, 2, 4, 6, \dots$ sont données dans la fig. I.17. L'amélioration de la convergence par l' ϵ -algorithme se voit clairement.

²C. Brezinski (1977): *Accélération de la Convergence en Analyse Numérique*. Lecture Notes in Mathematics, Nr. 584, Springer-Verlag. [MA 00.04/3 584]

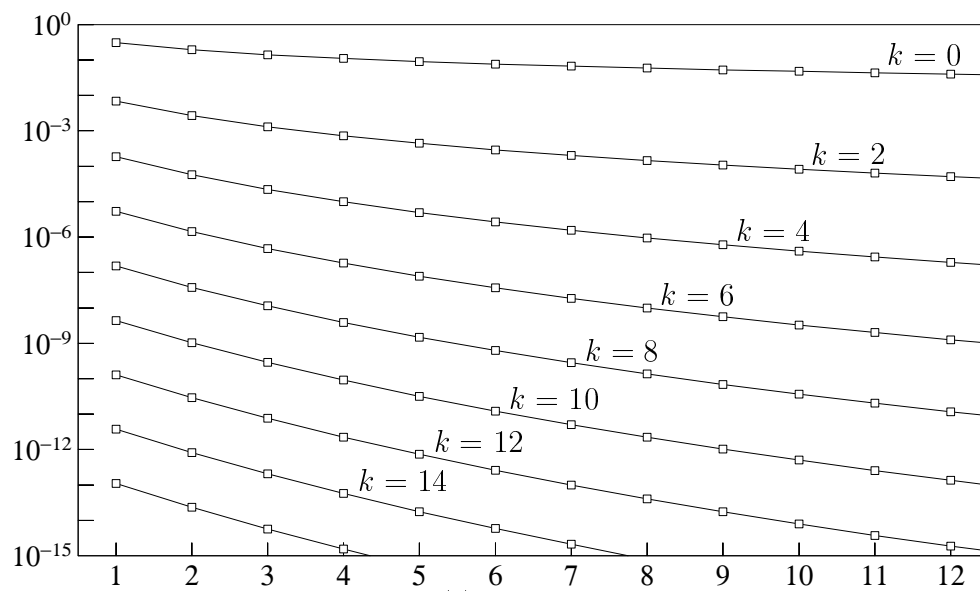


FIG. I.17: Erreur de $\epsilon_k^{(n)}$ en fonction de n pour la suite (6.6)