# La formule de Simpson avec reste intégral

Jean-François Burnol, septembre 2016

On cherche à approcher l'intégrale $\int_a^b f(t)\,dt$ par une combinaison linéaire

$$\lambda f(a) + \mu f(\frac{a+b}{2}) + \nu f(b)$$

On va tout d'abord prendre $a = -1$ et $b = 1$. On choisit $\lambda, \mu, \nu$ pour que la formule $\lambda f(-1) + \mu f(0) + \nu f(1)$ soit exacte pour les polynômes de degrés au plus deux, ce qui donne trois conditions sur nos trois inconnues, en prenant successivement $f(t) = 1, t, t^2$ :

$$\begin{aligned} 2 &= \lambda + \mu + \nu \\ 0 &= -\lambda + \nu \\ \frac{2}{3} &= \lambda + \nu \end{aligned}$$

Le système se résout aisément et donne $\lambda = \nu = \frac{1}{3}$ et $\mu = \frac{4}{3}$. D'où la forme linéaire

$$\mathrm{S}(f) = \frac{f(-1) + 4f(0) + f(1)}{3}$$

Pour l'intervalle $[\,0,1\,]$ on écrit : $\int_0^1 f(t)\,dt = \int_{-1}^1 g(u)\,du$ avec $g(u)$ obtenu par changement de variable et valant après calcul $\frac{1}{2} f(\frac{1}{2}(1+u))$. Ainsi $\mathrm{S}(g) = \frac{g(-1)+4g(0)+g(1)}{3}$ s'exprime avec $f$ via

$$\mathrm{S}_{[\,0,1\,]}(f) = \frac{f(0) + 4f(\frac{1}{2}) + f(1)}{6}$$

Une fonction $h$ sur un intervalle $[\,a,b\,]$ se ramène à une fonction $f$ sur $[\,0,1\,]$ par $f(u) = (b-a)h(a+(b-a)u)$, $\int_a^b h(t)\,dt = \int_0^1 f(u)\,du$, d'où notre choix d'approximation :

$$\mathrm{S}_{[\,a,b\,]}(h) = (b-a)\frac{h(a) + 4h(\frac{a+b}{2}) + h(b)}{6}$$

On veut une formule pour l'erreur $\Delta = \int_a^b f(t)\,dt - \mathrm{S}_{[\,a,b\,]}(f)$. À nouveau on travaille d'abord avec $a = -1$, $b = 1$. On pose donc :

$$\Delta(f) = \int_{-1}^1 f(t)\,dt - \frac{f(-1) + 4f(0) + f(1)}{3}$$

On sait que ceci est nul si $f$ est une fonction polynomiale de degré au plus deux. De plus, si $f$ est impaire alors à la fois l'intégrale et l'approximation sont nulles. Donc en fait la formule est exacte pour les polynômes de degrés au plus trois et cela suggère qu'on devrait pouvoir exprimer l'erreur en fonction de la dérivée quatrième de $f$. À partir de maintenant on suppose donc que $f$ est de classe $\mathrm{C}^4$.

L'idée est la suivante : par la formule de Taylor avec reste intégral on a

$$f(x) = \underbrace{f(0) + f'(0)x + f''(0)\frac{x^2}{2} + f^{(3)}(0)\frac{x^3}{6}}_{\mathrm{T}(x)} + \underbrace{\int_0^x \frac{(x-t)^3}{6} f^{(4)}(t)\,dt}_{g(x)}$$

Le premier membre $\mathrm{T}(x)$ est un polynôme de degré au plus trois. On a $f = \mathrm{T} + g$ donc $\Delta(f) = \Delta(\mathrm{T}) + \Delta(g)$ par linéarité. Au passage je signale que $g$ est aussi $\mathrm{C}^4$ puisque T est $\mathrm{C}^\infty$. Bref :

$$\Delta(f) = \underbrace{\Delta(\mathrm{T})}_{=0} + \Delta(g) = \int_{-1}^{1} g(x)\,dx - \frac{g(-1) + 4g(0) + g(1)}{3}$$

Avant de poursuivre, on rappelle que pour $f$ impaire on sait déjà que $\Delta(f) = 0$, donc on va supposer $f$ paire. Dans ce cas le polynôme de Taylor T était aussi pair, et par conséquent $g$ est paire. De plus $g(0) = 0$ donc il nous faut évaluer :

$$\Delta(f) = 2\int_{0}^{1} g(x)\,dx - \frac{2}{3}g(1)$$

On pourrait calculer $\int_0^1 g(x)\,dx$ par une intégrale double sur le domaine $0 \leqslant t \leqslant x \leqslant 1$ mais si on n'est pas à l'aise avec cela il y a une astuce.

Il suffit d'écrire la formule de Taylor avec reste intégral pour la fonction $\mathrm{C}^5$, $\mathrm{F}(y) = \int_0^y f(x)\,dx$.

$$\mathrm{F}(y) = \mathrm{F}(0) + \mathrm{F}'(0)y + \mathrm{F}''(0)\frac{y^2}{2} + \mathrm{F}^{(3)}(0)\frac{y^3}{6} + \mathrm{F}^{(4)}(0)\frac{y^4}{24} + \int_0^y \frac{(y-t)^4}{24}\mathrm{F}^{(5)}(t)\,dt$$

$$\mathrm{F}(y) = \underbrace{f(0)y + f'(0)\frac{y^2}{2} + f^{(2)}(0)\frac{y^3}{6} + f^{(3)}(0)\frac{y^4}{24}}_{\mathrm{Q}(y)} + \underbrace{\int_0^y \frac{(y-t)^4}{24} f^{(4)}(t)\,dt}_{\mathrm{G}(y)}$$

En dérivant par rapport à $y$ (F est $\mathrm{C}^5$, donc G aussi) :

$$f(y) = \mathrm{F}'(y) = f(0) + f'(0)y + f''(0)\frac{y^2}{2} + f^{(3)}(0)\frac{y^3}{6} + \mathrm{G}'(y)$$

Par comparaison avec la formule $f(x) = \mathrm{T}(x) + g(x)$ cela donne $\boxed{\mathrm{G}'(x) = g(x)}$. On a ainsi une formule pour $\int_0^1 g(x)\,dx$ :

$$\boxed{\int_0^1 g(x)\,dx = \mathrm{G}(1) - \mathrm{G}(0) = \int_0^1 \frac{(1-t)^4}{24} f^{(4)}(t)\,dt}$$

On a fait tout ce raisonnement [1], sans plus supposer que $f$ était paire, pour montrer la généralité. Mais maintenant je suppose à nouveau $f$ paire et je termine le calcul.

$$\begin{aligned}
\Delta(f) = 2\int_0^1 g(x)\,dx - \frac{2}{3}g(1) &= \int_0^1 \frac{(1-t)^4}{12} f^{(4)}(t)\,dt - \int_0^1 \frac{(1-t)^3}{9} f^{(4)}(t)\,dt \\
&= \frac{1}{36}\int_0^1 (3(1-t)-4)(1-t)^3 f^{(4)}(t)\,dt \\
&= -\frac{1}{36}\int_0^1 (1+3t)(1-t)^3 f^{(4)}(t)\,dt
\end{aligned}$$

1. Une autre méthode consiste à appliquer les règles de dérivation à $\mathrm{H}(y,z) = \int_0^y (z-t)^4 f^{(4)}(t)\,dt$, $\frac{d}{dx}\mathrm{H}(x,x) = (\frac{\partial}{\partial y} + \frac{\partial}{\partial z})\,|_{(y,z)=(x,x)}\,\mathrm{H}(y,z)$. La contribution de la première dérivée partielle est nulle.

Il ne reste plus qu'à étendre l'intégrale sur tout $[-1,1]$ pour avoir une formule qui sera valable pour les fonctions $C^4$ paires ou impaires, donc pour toutes : [2]

$$\boxed{\Delta(f) = -\frac{1}{72}\int_{-1}^{1}(1+3|t|)(1-|t|)^3 f^{(4)}(t)\,dt}$$

On a obtenu comme désiré une formule exacte utilisant les dérivées quatrièmes de $f$. C'est la seule formule possible de ce type, on ne peut pas se débarrasser des valeurs absolues par des bidouillages, on ne peut pas trouver un polynôme qui ferait le job à la place : si un autre intégrande fonctionnait la différence devrait être « perpendiculaire » à tous les $f^{(4)}$ donc à toutes les fonctions continues, et par conséquent serait nulle (sauf en un nombre fini de points, si on postule quelque chose de continu par morceaux au départ).

Venons-en maintenant à la majoration :

$$|\Delta(f)| \leqslant \underbrace{\frac{1}{72}\int_{-1}^{1}(1+3|t|)(1-|t|)^3\,dt}_{\mathrm{C}} \cdot \mathrm{M}_4(f) = \frac{1}{90}\mathrm{M}_4(f)$$

L'intégrale C se calcule en effet facilement (par parité puis $t \to 1-t$ par exemple) mais on peut aussi appliquer la formule encadrée à $f(x) = x^4$, d'où :

$$\frac{2}{5} - \frac{2}{3} = -24\mathrm{C} \implies \mathrm{C} = \frac{1}{90}$$

Finalement si on revient à une fonction $h$ sur un intervalle $[a,b]$, il faut appliquer ce qui précède à $f(t) = \frac{b-a}{2}h(\frac{a+b}{2} + \frac{b-a}{2}t)$ ce qui donne (bien sûr on suppose $a < b$) la formule bien connue pour la majoration de l'erreur dans la méthode de Simpson :

$$|\Delta_{[a,b]}(h)| = |\Delta_{[-1,1]}(f)| \leqslant \frac{1}{90}\mathrm{M}_4(f) = \frac{1}{90}\frac{b-a}{2}\left(\frac{b-a}{2}\right)^4 \mathrm{M}_4(h) = \frac{(b-a)^5}{2880}\mathrm{M}_4(h)$$

On peut aussi écrire explicitement la formule exacte :

$$\begin{aligned}
\Delta_{[a,b]}(h) &= \int_a^b h(t)\,dt - (b-a)\frac{h(a) + 4h(\frac{a+b}{2}) + h(b)}{6} \\
&= \Delta_{[-1,1]}(f) \\
&= -\frac{1}{72}\int_{-1}^{1}(1+3|t|)(1-|t|)^3\left(\frac{b-a}{2}\right)^4 h^{(4)}(\frac{a+b}{2} + \frac{b-a}{2}t)\,\frac{b-a}{2}dt \\
&= -\frac{1}{72}\int_a^b \left(\frac{b-a}{2} + 3\left|x - \frac{a+b}{2}\right|\right)\left(\frac{b-a}{2} - \left|x - \frac{a+b}{2}\right|\right)^3 h^{(4)}(x)\,dx
\end{aligned}$$

Note : pour éviter les confusions je précise que dans mes commentaires de leçons présentées le 23 septembre, j'avais proposé de simplifier certains calculs en utilisant des formules de Taylor avec reste intégral, mais ce qui est fait ici est encore autre chose.

2. Bien sûr si $f$ est impaire alors $f^{(4)}$ l'est aussi et donc notre intégrale est bien nulle. Et toute fonction sur $[-1,1]$ est de manière unique somme d'une fonction paire et d'une fonction impaire. Et nous travaillons avec des formes linéaires.

Un dernier mot [3] : rien ne nous oblige à utiliser $f^{(4)}$ on peut aussi donner une formule avec $f^{(3)}$. Il suffit de reprendre la technique mais en allant un cran moins loin dans la formule de Taylor initiale :

$$f(x) = \underbrace{f(0) + f'(0)x + f''(0)\frac{x^2}{2}}_{\mathrm{T}_1(x)} + \underbrace{\int_0^x \frac{(x-t)^2}{2} f^{(3)}(t)\,dt}_{g_1(x)}$$

À nouveau $\Delta(f) = \Delta(\mathrm{T}_1) + \Delta(g_1) = \int_{-1}^{1} g(x)\,dx - \frac{g(-1)+4g(0)+g(1)}{3}$.

On suppose à nouveau que $f$, donc $g_1$ est paire. Cette fois-ci on obtient :

$$\Delta(f) = 2\int_0^1 g_1(x)\,dx - \frac{2}{3}g_1(1) = 2\int_0^1 \frac{(1-t)^3}{6} f^{(3)}(t)\,dt - \frac{2}{3}\int_0^1 \frac{(1-t)^2}{2} f^{(3)}(t)\,dt = -\frac{1}{3}\int_0^1 t(1-t)^2 f^{(3)}(t)\,dt$$

Si on écrit cette formule sous la forme (attention que si $f$ est paire alors $f^{(3)}$ est impaire, donc $tf^{(3)}(t)$ est paire et c'est pour cela qu'il n'y a pas de valeurs absolues autour du premier $t$ ci-dessous) :

$$\Delta(f) = -\frac{1}{6}\int_{-1}^{1} t(1-|t|)^2 f^{(3)}(t)\,dt$$

On calcule maintenant $\frac{1}{3}\int_0^1 t(1-t)^2\,dt = \frac{1}{3}(\frac{1}{3} - \frac{1}{4}) = \frac{1}{36}$ et donc $|\Delta_{[-1,1]}(f)| \leqslant \frac{1}{36}\mathrm{M}_3(f)$ (ou aussi la variante $\frac{1}{9}\sup|tf^{(3)}(t)|$). Sur un intervalle $[a,b]$ la majoration sera

$$|\Delta_{[a,b]}(f)| \leqslant \frac{1}{36}\frac{(b-a)^4}{16}\mathrm{M}_3(f) = \frac{(b-a)^4}{576}\mathrm{M}_3(f)$$

On ne voit pas immédiatement un désavantage flagrant par rapport à l'autre majoration $\frac{(b-a)^5}{2880}\mathrm{M}_4(f)$, à part le fait de rater que l'approximation est exacte pour les polynômes de degrés trois. Prenons $a = 0$, $b = 1$, et $f(t) = t^n$ la majoration classique nous dit

$$\frac{n(n-1)(n-2)(n-3)}{2880}$$

et celle avec la derivée tierce nous dit

$$\frac{n(n-1)(n-2)}{576}$$

qui est *meilleure* dès que $n-3 > 5$, $n > 8$ !

Mais bon, on est en train de comparer $\int_0^1 t^n\,dt = \frac{1}{n+1}$ avec $(4\cdot 2^{-n}+1)/6$, inutile de préciser que de toute façon c'est catastrophique pour $n$ grand comme approximation ! Et de plus l'erreur véritable est proche de 1/6 tandis que nos majorations, que ce soit la classique ou celle avec $\mathrm{M}_3$ tendent vers l'infini avec $n$ !

Si j'essaie avec $a = 0$, $b$ général, et toujours $f(t) = t^n$, $n$ grand, la majoration avec $\mathrm{M}_4$ donne $\frac{n(n-1)(n-2)(n-3)}{2880}b^{n-4}b^5$ et celle avec $\mathrm{M}_3$ donne $\frac{n(n-1)(n-2)}{576}b^{n-3}b^4$ ce qui est meilleur sous la même condition $n-3 > 5$, $n > 8$ ! Mais la valeur exacte $b^{n+1}/(n+1)$ diffère de toute façon de l'approximation $\frac{1}{6}b^{n+1}(1+4\cdot 2^{-n})$ par bien moins que ce que ces estimations prétendent…

L'idée est donc peut-être que si la fonction varie trop brutalement la majoration de l'erreur avec $\mathrm{M}_3$ est meilleure ; mais que l'estimation de l'intégrale est de toute façon dans ces circonstances probablement médiocre.

3. en fait c'est plutôt l'avant avant dernier mot.

Mais il y a bien un **désavantage majeur** de la majoration avec $M_3$ c'est qu'en cas de découpage préalable en N sous-intervalles on obtient une majoration de l'erreur globale en $N^{-3}$, et pas en $N^{-4}$ qui est l'estimation correcte. Comme nous allons le justifier maintenant *grâce à notre formule exacte*.

Je rappelle sur $[-1,1]$ :

$$\boxed{\Delta(f) = -\frac{1}{72}\int_{-1}^{1}(1+3|t|)(1-|t|)^3 f^{(4)}(t)\,dt}$$

Comme le poids est positif, par la première formule de la moyenne (à poids), on a :

$$\exists \xi \in ]-1,1[ : \Delta(f) = -\frac{1}{72}\int_{-1}^{1}(1+3|t|)(1-|t|)^3\,dt \cdot f^{(4)}(\xi) = -\frac{1}{90}f^{(4)}(\xi)$$

Et sur un intervalle $[a,b]$ la formule sera :

$$\exists \xi \in ]a,b[ : \Delta_{[a,b]}(f) = -\frac{(b-a)^5}{2880}f^{(4)}(\xi)$$

Considérons maintenant un intervalle $[A,B]$, que l'on subdivise en N sous-intervalles de consécutifs de mêmes longueurs $(B-A)/N$. La formule pour l'erreur totale $\Delta_N$ commise en appliquant Simpson à chaque sous-intervalle sera :

$$\Delta_N = -\sum_{k=0}^{N-1}\frac{(B-A)^5}{2880\cdot N^5}f^{(4)}(\xi_k)$$

avec $A + k(B-A)/N < \xi_k < A + (k+1)(B-A)/N$. On reconnaît une somme de Riemann pour l'intégrale de $f^{(4)}$ :

$$\frac{2880\cdot N^4}{(B-A)^4}\Delta_N = -\frac{B-A}{N}\sum_{k=0}^{N-1}f^{(4)}(\xi_k) \to -\int_A^B f^{(4)}(t)\,dt = f^{(3)}(A) - f^{(3)}(B)$$

On a donc, sauf si $f^{(3)}(B) = f^{(3)}(A)$, un *équivalent* pour l'erreur commise dans cette méthode d'approximation :

$$\boxed{\Delta_N \underset{N\to\infty}{\sim} \frac{(B-A)^4\left(f^{(3)}(A) - f^{(3)}(B)\right)}{2880\cdot N^4}}$$

Ceci montre que, sauf cas exceptionnel, la décroissance est exactement en $N^{-4}$. On ne peut pas obtenir ce résultat si on a seulement une majoration de l'erreur dans la méthode de Simpson.

Comme je l'ai indiqué pendant la séance, si par contre la fonction sur $[A,B]$ se prolonge par périodicité de période $B-A$, en étant $C^\infty$ sur la droite réelle, alors $N^{-4}$ n'est plus du tout un bon estimé de la décroissance de l'erreur $\Delta_N$ car celle-ci sera en fait un $\mathcal{O}(N^{-k})$ pour tout $k$.