Ann. sc. math. Québec, 1982, vol. 6, no 2, pp. 123-141

MONOSPLINE À OSCILLATION MINIMALE

Richard Bastien et Serge Dubuc

1. Introduction

Nous nous proposons d'analyser diverses formules de quadrature pour les fonctions de classe C^4 dont la 4^e dérivée est positive. Nous retenons un critère d'optimalité qui est nouveau, mais qui selon nous, a quand même son intérêt propre. Dans ce cadre, nous dégageons la formule optimale de quadrature. On analysera ensuite la disposition des noeuds et la répartition des poids. A part les tout premiers noeuds et les poids correspondants, on a l'impression que les noeuds de la quadrature optimale se suivent selon une progression arithmétique et que les poids sont constants. De fait, il s'agit d'une illusion numérique. L'explication de ce phénomène provient de ce qu'une certaine transformation Φ du plan en lui-même est hautement singulière. Cette transformation Φ envoie le plan sur un arc simple du plan, elle admet un point fixe x_o. D'autre part, si x est un point voisin de x_o, la suite des itérées $\Phi^n(x)$ convergent vers x_o, la convergence est excellente, elle est d'ordre 3. Les propriétés de cette formule de quadrature sont obtenues par la technique désormais classique du transfert de ces problèmes correspondants portant sur des fonctions splines. La quadrature optimale recherchée est associée à une fonction spline à oscillation minimale.

2. Critères d'optimalité pour une formule de quadrature

Soit F une classe de fonctions réelles continues définies sur l'intervalle unité $[0,1]$, on désignera par $I(f)$ l'intégrale de f sur l'intervalle

$[0,1]$. Une formule de quadrature est par définition une fonctionnelle sur F , $Q : F \to \mathbb{R}$ qui est généralement de la forme $Q(f) = \sum_{i=1}^{k} p_i f(x_i)$ où $0 \le x_1 < x_2 < \ldots < x_k \le 1$. Les valeurs $x_1, x_2, \ldots, x_k$ sont les noeuds de la quadrature. Les valeurs p_i sont les poids de la quadrature. En présence de deux formules de quadrature, Q_1 et Q_2 , qui utilisent le même nombre de noeuds, quand dira-t-on qu'une quadrature Q_1 est plus avantageuse qu'une quadrature Q_2 ? La réponse à cette question varie selon les auteurs, elle dépend fortement de la classe F que l'on a retenue et très souvent elle dépend d'une autre fonctionnelle sur F . Citons par exemple le critère de comparaison de Sard [5]. Pour un entier donné, n , supérieur ou égal à 1 , $F = C^n[0,1]$ et l'on considère sur F la semi-norme $p_n(f) = (\int_0^1 (f^{(n)}(x))^2 \, dx)^{\frac{1}{2}}$. Sard dirait qu'entre deux formules de quadrature Q_1 et Q_2 , Q_1 est meilleure que Q_2 si l'on a l'inégalité

$$\sup\{|I(f) - Q_1(f)| : p_n(f) \le 1\} \le \sup\{|I(f) - Q_2(f)| : p_n(f) \le 1\} .$$

Néanmoins ce n'est pas ce critère que nous voulons satisfaire. Soit f une fonction de classe C^n dont nous voulons évaluer numériquement l'intégrale. La formule de Taylor donne que $f(x) = P_n(x) + R_n(x)$

$$P_n(x) = \sum_{i=0}^{n-1} f^{(i)}(0) \, x^i/i! \quad , \quad R_n(x) = (\int_0^x (x-t)^{n-1} f^{(n)}(t) \, dt)/(n-1)!$$

Puisque $I(f) = I(P_n) + I(R_n)$ et que l'évaluation numérique de $I(P_n)$ ne pose aucune difficulté, nous nous limitons à l'intégration numérique du reste R_n . Enfin, avec une petite dose d'arbitraire, nous retenons la classe de fonctions $C_n = \{f : f \in C^n[0,1] , f^{(i)}(0) = 0 , i = 0,1,\ldots,n-1 \text{ et } \forall x \in [0,1] f^{(n)}(x) \ge 0\}$. Nous retenons aussi une fonctionnelle dite de contrôle:

$$p_n(f) = f^{(n-1)}(1)/n! = \int_0^1 f^{(n)}(x) \, dx/n!$$

Pour une quadrature Q , nous introduisons deux nombres $m(Q)$ et $M(Q)$:

$$m(Q) = \inf\{I(f) - Q(f) : f \in C_n , p_n(f) = 1\}$$
$$M(Q) = \sup\{I(f) - Q(f) : f \in C_n , p_n(f) = 1\} .$$

La connaissance de ces deux quantités permet de borner simplement l'intégrale de f lorsque $f \in C_n$:

$$Q(f) + m(Q)p_n(f) \leq I(f) \leq Q(f) + M(Q)p_n(f) \ .$$

L'*oscillation* de Q autour de I par rapport à la fonctionnelle de contrôle p_n est par définition le nombre $osc(Q) = M(Q) - m(Q)$. Enfin on dira qu'une quadrature Q est *optimale* si pour toute autre quadrature Q' qui utilise le même nombre de noeuds, l'oscillation de Q' n'est pas inférieure à l'oscillation de Q .

3. Formules de quadrature et fonctions splines

Nous rappelons ici la formule de Peano pour l'écart entre $I(f)$ et $Q(f)$ lorsque $Q(f)$ est la quadrature $\sum_{i=1}^{k} p_i f(x_i)$ avec $0 \leq x_1 < \ldots < x_k \leq 1$. Si les dérivées successives jusqu'à l'ordre $n-1$ de f sont nulles à l'origine et si $f^{(n)}$ existe et est continue,

$$f(x) = \int_0^1 (x-t)_+^{n-1} f^{(n)}(t)\,dt/(n-1)!$$

(L'expression u_+ désigne la partie positive de u , $u_+ = u$ si $u \geq 0$ sinon $u_+ = 0$.) D'où

$$I(f) - Q(f) = \int_0^1 N(t) f^{(n)}(t)\,dt/n!$$

où $N(t) = (1-t)^n - \sum_{i=1}^{k} np_i(x_i-t)^{n-1}$.

La fonction $N(t)$, le noyau de Peano de la formule de quadrature, est une monospline. Le noyau $N(t)$ permet de contrôler les écarts entre $I(f)$ et $Q(f)$. En particulier, les quantités $m(Q)$ et $M(Q)$ sont respectivement le minimum et le maximum du noyau $N(t)$ lorsque t varie de 0 à 1 .

Enfin l'oscillation de Q est précisément égale à l'oscillation de la fonction N sur l'intervalle $[0,1]$. Après des changements de variables tout simples, la recherche d'une formule de quadrature optimale pour la classe C_n re-

vient à trouver une fonction monospline

$$t^n - \sum_{i=1}^{k} p_i(t-x_i)_+^{n-1} \quad , \quad x_i \in [0,1]$$

dont l'oscillation sur $[0,1]$ est minimale. Pour n et k donnés, il s'agit de déterminer des nombres $p_1,p_2,\ldots,p_k$ et $x_1,x_2,\ldots,x_k$, $0 \le x_o < x_1 < \ldots < x_k \le 1$ tels que l'oscillation de $t^n - \sum_{i=1}^{k} p_i(t-x_i)_+^{n-1}$ sur l'intervalle $[0,1]$ est minimale.

4. Monosplines à oscillation minimale

Oublions les formules de quadrature, nous caractérisons maintenant la monospline réduite dont l'oscillation est minimale. Auparavant, fixons quelques notations. On désigne par $S_{n,k}$, les fonctions splines de classe (n,k) , les fonctions suivantes $S(t) = Q_n(t) + \sum_{i=1}^{k} q_i(t-x_i)_+^n$ où Q_n est un polynôme de degré inférieur ou égal à n . On désigne par $M_{n,k}$, les monosplines de classe (n,k) , les fonctions $M(t)$ de la forme $t^n + S(t)$ où S est une fonction spline de classe $(n-1,k)$. Ces notations sont usuelles. Nous désignerons par $M_{n,k}^{a,b}$ les fonctions de la forme $t^n + \sum_{i=1}^{k} q_i(t-x_i)_+^{n-1}$ où $a < x_1 < x_2 < \ldots < x_k < b$; nous dirons d'une telle fonction qu'il s'agit d'une monospline réduite sur l'intervalle $[a,b]$.

THÉORÈME 1. *Soient* n *et* k *deux entiers supérieurs ou égaux à* 1 . *Il existe une et une seule monospline réduite* $\tilde{M}$ *dont l'oscillation sur* $[0,1]$ *est minimale dans* $M_{n,k}^{0,1}$. $\tilde{M}$ *est caractérisée par les deux propriétés suivantes:*

1. $0 \le \tilde{M}(t) \le \tilde{M}(1)$ *et* $0 \le t \le 1$.

2. *Il existe deux suites de* k *points* $\{a_i\}_1^k$ *et* $\{b_i\}_1^k$ *telles que* $(\forall i)$ $\tilde{M}(a_i) = 0$ *et* $\tilde{M}(b_i) = \tilde{M}(1)$ *et* $b_1 < a_1 < b_2 < a_2 < \ldots < b_k < a_k$.

Marc Bourdeau [1] a déjà étudié les cas $n = 1$ et 2 . Nous traiterons de la situation $n \ge 3$. Notons que Johnson [2] a résolu le problème analogue de caractériser la monospline dont la norme uniforme sur $[0,1]$ est minimale dans

$M_{n,k}$. Il nous suffira d'adapter ses idées à notre contexte. En particulier, nous aurons besoin de son théorème 8 [2] que nous citons modulo quelques modifications.

THÉORÈME (Johnson, [2]). *Supposons que* $N \geq 3$. *Soit* $C = \{c = (c_1, c_2, \ldots, c_N)\}$ *la collection des* (N)*-uples de nombres réels où* $c_1 = 0 < c_2 < \ldots < c_{N-1} < 1 = c_N$. *Soit* t *un nombre réel qui varie dans* $[0,1]$ *et soit* $f(t,c)$ *une fonction de* t *et de* c *à valeurs réelles et qui remplit les conditions suivantes:*

a) $f(t,c)$ *est continue et bornée sur son domaine de définition.*

b) $\int_0^1 |f(s,c)|\,ds$ *est uniformément éloignée de* 0 , *c'est-à-dire qu'il existe un* $\delta > 0$ *qui minore chacune de ces intégrales.*

c) *Pour chaque vecteur* c *de* C , $f(c_j,c) = 0$; *pour les autres valeurs de* t , $f(t,c) \neq 0$ *et plus précisément le signe de* $f(t,c)$ *est* $(-1)^{N-j-1}$ *si* $c_{j-1} < t < c_j$, $2 \leq j \leq N$.

Soit $F(t,c) = \int_0^t f(s,c)\,ds$. *et soit* $\beta_1, \beta_2, \ldots, \beta_N$ *une suite de nombres réels telle que* $\beta_1 = 0$

$$\beta_N < \beta_{N-1} \ , \ \beta_{N-2} < \beta_{N-3} \ , \ \cdots \ , \ \beta_{N-1} > \beta_{N-2} \ , \ \beta_{N-3} > \beta_{N-4}$$

et telle que le signe de $\beta_j - \beta_{j-1}$ *est* $(-1)^{N-j-1}$. *Alors il existe un vecteur* c *de* C *et un nombre* $\lambda > 0$ *tels que* $F(c_j,c) = \lambda\beta_j$, $1 \leq j \leq N$.

Abordons la démonstration du Théorème 1. Montrons qu'il existe une monospline $M(t)$ de $M_{n,k}^{0,1}$ qui remplit les conditions 1 et 2. Dans le théorème de Johnson, posons $N = 2k + 1$. Comme fonction $f(t,c)$, utilisons l'unique monospline de classe $(n-1,k)$ qui admet pour zéros $c_1 = 0$, selon la multiplicité $n - 1$, et $c_2, c_3, \ldots$ et c_{2k+1} . Nous faisons ici appel au théorème fondamental de l'algèbre pour les splines (cf. Théorème 1.1 de [3]). Selon les mêmes arguments que ceux de Johnson, les conditions (a), (b) et (c) du théorème de Johnson sont satisfaites. Comme suite $\{\beta_i\}_1^{2k+1}$, on prend la suite alternée de 0 et de 1 : $\{0,1,0,1,\ldots,0\}$. Le théorème donne donc qu'il existe un vecteur c pour lequel $0 \leq F(t,c) \leq \lambda$, $F(c_{1+2i},c) = 0$, $F(c_{2i},c) = \lambda$, $1 \leq i \leq k$.

Si l'on pose $G(t) = nF(t,c)$, G est une monospline. Enfin, soit a la solution supérieure à un à l'équation $G(t) = n\lambda$. Alors $\tilde{M}(t) = G(at)/a^n$ est une monospline qui remplit les deux conditions 1 et 2.

Vérifions que $\tilde{M}(t)$ est d'oscillation minimale. Soit $\tilde{M}_1(t)$ une autre monospline de $M_{n,k}^{0,1}$. Posons $\alpha_1 = \inf\{\tilde{M}_1(t) : t \in [0,1]\}$, $\beta_1 = \sup\{\tilde{M}_1(t) : t \in [0,1]\}$, $L = \tilde{M}(1)$. Raisonnons par l'absurde en supposant que $\beta_1 - \alpha_1 \leq L$. Soit $\{t_i\}_{i=1}^{2k+2}$ les valeurs ordonnées par ordre croissant où la fonction $\tilde{M}$ prend successivement les valeurs 0 et L . Un calcul simple montre que

$$\tilde{M}_1(t_i) - \tilde{M}(t_i) \leq \alpha_1 \quad \text{si } i \text{ est impair}$$
$$\tilde{M}_1(t_i) - \tilde{M}(t_i) \geq \beta_1 \quad \text{si } i \text{ est pair .}$$

D'où l'équation $\tilde{M}_1(t) - \tilde{M}(t) = \alpha_1$ admet au moins $2k+1$ solutions. Si x_1 est le plus petit des noeuds utilisés dans $\tilde{M}_1$ ou dans $\tilde{M}$, on obtient que $S(t) = (\tilde{M}_1(t) - \tilde{M}(t))'$ admet au moins $2k$ zéros dans $(x_1,1)$, alors que $S(t)$ est identiquement nul sur $[0,x_1]$. Ceci est impossible si $\tilde{M}_1 \neq \tilde{M}$. Nous avons donc établi que l'oscillation de $\tilde{M}_1$ est supérieure à celle de $\tilde{M}$.

Vérifions enfin qu'il n'existe qu'une monospline qui réalise les conditions 1 et 2. En effet, si $\tilde{M}_2$ remplissait les conditions 1 et 2 et était différente de la monospline déjà considérée $\tilde{M}$, on pourrait reprendre le raisonnement du dernier paragraphe en utilisant $\tilde{M}_2$ pour $\tilde{M}$ et $\tilde{M}$ pour $\tilde{M}_1$ et l'on obtiendrait une contradiction, que l'oscillation de $\tilde{M}$ dépasse celle de $\tilde{M}_2$.

5. Détermination des monosplines à oscillation minimale $n = 1, 2$ et 3

Les monosplines minimales dans les cas $n = 1,2$ sont, selon [1],

$$\tilde{M}_{1,k} = x - \frac{1}{k+1} \sum_{1}^{k} (x - \tfrac{i}{k+1})_+^0 \quad , \quad \operatorname{osc} \tilde{M}_{1,k} = 1/(k+1)$$

$$\tilde{M}_{2,k} = x^2 - \frac{4}{1+2k} \sum_{1}^{k} (x - \tfrac{2i-1}{2k+1})_+ \quad , \quad \operatorname{osc} \tilde{M}_{2,k} = 1/(1+2k)^2 \ .$$

THÉORÈME 2. *La monospline à oscillation minimale dans* $M_{3,k}^{0,1}$ *est* $M_{3,k}(x) = x^3 - \sum_{i=1}^{k} \tilde{p}_i((x - \tilde{x}_i)_+)^2$ *où*

$$h = (2\sqrt{3}(k-1) + 5)^{-1} ,$$

$$\tilde{x}_1 = 4h/3 \quad , \quad \tilde{p}_1 = 9h ,$$

$$\tilde{x}_j = [3 + \sqrt{3}(2j-3)\ h \quad , \quad \tilde{p}_j = 6\sqrt{3}\ h , \quad j = 2,\ldots,k .$$

L'oscillation de la monospline est $4h^3$.

DÉMONSTRATION. Pour alléger la notation, nous désignons $\tilde{M}_{3,k}$ par $\tilde{M}$. Posons $b_i = (2 + 2\sqrt{3}(i-1))h$, $a_i = b_i + 2h$, $m_i = (a_i+b_i)/2$, $i = 1,2,\ldots,k$. Nous allons montrer que sur chacun des intervalles $[\tilde{x}_i,\tilde{x}_{i+1}]$,

$$\tilde{M}(x) = (x-m_i)^3 - 3h^2(x-m_i) + 2h^3 , \qquad i = 1,2,\ldots,k$$

(en sous-entendant que $\tilde{x}_{k+1} = 1$). Procédons par induction sur i . Sur l'intervalle $[\tilde{x}_1,\tilde{x}_2]$, $\tilde{M}(x) = x^3 - 9h(x-4h/3)^2 = x^3 - 9hx^2 + 24h^2x - 16h^3$. Or $m_1 = 3h$ et $(x-m_1)^3 - 3h^2(x-m_i) + 2h^3 = x^3 - 9hx^2 + 24h^2x - 16h^3$.

Posons $\tilde{N}(x) = (x-m_{i-1})^3 - 3h^2(x-m_{i-1}) + 2h^3$ si $\tilde{x}_{i-1} \le x \le \tilde{x}_i$. On vérifie aisément que $\tilde{x}_i = (a_{i-1}+b_i)/2$, $\tilde{x}_i - m_{i-1} = \sqrt{3}\ h$

$$\tilde{N}(\tilde{x}_i-) = 2h^3 = \tilde{N}(\tilde{x}_i+) ,$$

$$\tilde{N}'(\tilde{x}_i-) = 6h^2 = \tilde{N}'(\tilde{x}_i+) ,$$

$$\tilde{N}''(\tilde{x}_i-) = 6\sqrt{3}\ h \quad \text{et} \quad \tilde{N}''(\tilde{x}_i+) = -6\sqrt{3}\ h .$$

D'autre part, $\tilde{M}$ et $\tilde{M}'$ sont continues et $\tilde{M}''(\tilde{x}_i+) - \tilde{M}''(\tilde{x}_i-) = -2\tilde{p}_i = -12\sqrt{3}\ h$.

La fonction $\tilde{M}(x) - \tilde{N}(x)$ est constituée d'arcs de paraboles dont les discontinuités des dérivées sur $[\tilde{x}_1,1]$ ne peuvent être qu'aux noeuds $\tilde{x}_{i+1}$. Or selon le calcul antérieur, aucune discontinuité des dérivées de $\tilde{M}(x) - \tilde{N}(x)$ ne peut se manifester. Ce qui montre que $\tilde{M}(x) \equiv \tilde{N}(x)$ sur $[\tilde{x}_1,1]$.

On vérifie maintenant que $a_i = m_i + h$ et $b_i = m_i - h$, d'où $\tilde{M}(a_i) = 0$ et $\tilde{M}(b_i) = 4h^3$. D'autre part si $|x - m_i| \le \sqrt{3}\ h$, alors $0 \le \tilde{M}(x) \le 4h^3$. Il reste à étudier $\tilde{M}$ sur le dernier intervalle $[\tilde{x}_k,1]$. $m_k = (3 + 2(k-1)\sqrt{3})h$.

$$\tilde{M}(1) = (1-m_k)^3 - 3h^2(1-m_k) + 2h^3 .$$

Puisque $h^{-1} = 2\sqrt{3}(k-1) + 5$, $1-m_k = 2h$ et $\tilde{M}(1) = 4h^3$. Si $\tilde{x}_k \le x \le 1$, $0 \le \tilde{M}(x) \le 4h^3$.

Le Théorème 1 permet de dire maintenant que $\tilde{M}$ est la monospline d'oscillation minimale.

C.Q.F.D.

Il ne semble pas être possible d'obtenir une formule explicite pour décrire $\tilde{M}_{n,k}$ dans le cas où $n \ge 4$. Cependant, regardons par curiosité les valeurs numériques des poids et des accroissements des noeuds de $\tilde{M}_{4,10}$

i	$\tilde{p}_i$	$\tilde{x}_i - \tilde{x}_{i-1}$
1	0,27778365	indéfini
2	0,38274361	0,08792561
3	0,38353414	0,09585048
$4 \le i \le 10$	0,38353414	0,09588354

On a l'impression que tous les poids sont constants à partir du 3^e et que les noeuds à partir du 3^e suivent une progression arithmétique. Nous verrons qu'il s'agit d'une illusion numérique et nous expliquerons ce phénomène.

6. Monosplines à élongation maximale

Désignons par $\mathrm{osc}(n,k)$ l'oscillation minimale d'une monospline de $M^{0,1}_{n,k}$. Nous notons encore par $\tilde{M}_{n,k}(t)$ la monospline de $M^{0,1}_{n,k}$ dont l'oscillation est précisément $\mathrm{osc}(n,k)$. Pour mieux analyser $\tilde{M}_{n,k}$ nous allons dilater l'intervalle $[0,1]$. Soit $M^{0,\infty}_{n,k}$ la totalité des fonctions $f(t)$ de la forme $f(t) = t^n - \sum_{i=1}^{k} p_i(t-x_i)_+^{n-1}$ où $0 \le x_1 < x_2 < \ldots < x_k$, la variable indépendante t varie de 0 à ∞ . Si f appartient à $M^{0,\infty}_{n,k}$, l'élongation de f sera par définition le plus grand nombre c non négatif tel que $f(t)$ est compris entre 0 et 1 lorsque t varie de 0 à c . Si c est l'élongation de f ,

$g(t) = f(tc)/c^n$ appartient à $M_{n,k}^{0,1}$. Ainsi $c^{-n} \geq \text{osc}(n,k)$. Ainsi l'élongation de f est majorée par $\text{osc}(n,k)^{-1/n}$. D'autre part l'élongation de la fonction $\gamma^n \tilde{M}_{n,k}(t/\gamma)$ où $\gamma = (\text{osc}(n,k))^{-1/n}$ est précisément égale à γ .

La recherche de la monospline à oscillation minimale dans $M_{n,k}^{0,1}$ est donc équivalente à la recherche de la monospline à élongation maximale dans $M_{n,k}^{0,\infty}$. Ce dernier raisonnement permet de caractériser la monospline réduite d'élongation maximale.

THÉORÈME 3. *Soient* n *et* k *deux entiers supérieurs ou égaux à* 1 . *Il existe une et une seule monospline réduite de* $M_{n,k}^{0,\infty}$ *dont l'élongation est maximale.* M *est caractérisée par la propriété suivante: il existe deux suites de points* $\{y_i\}_1^k$ *et* $\{z_i\}_1^k$ *et un nombre* c *tels que*

$$0 < z_1 < y_1 < z_2 < y_2 < \ldots < z_k < y_k < c$$

$$M(y_i) = 0 \quad , \quad M(z_i) = 1 \quad , \quad 1 \leq i \leq k$$

et

$$0 \leq M(t) \leq 1 = M(c) \quad \text{si} \quad 0 \leq t \leq c \ .$$

l'élongation de M *est alors ce nombre* c .

7. Monosplines à élongation maximale de degré 4

Notre propos ici sera d'établir l'existence d'une suite de quadruplets $\{p_i,x_i,y_i,z_i\}_1$ telle que la monospline à élongation maximale de $M_{4,k}^{0,\infty}$ est $M_{4,k}(t) = t^4 - \sum_{i=1}^{k} p_i(t-x_i)_+^3$ et de fait, $M_{4,k}$ prendra la valeur 0 pour $\{y_i\}_1^k$ et la valeur 1 pour $\{z_i\}_1^k$. La clé de cette analyse est le développement de Taylor de la fonction $M_{4,k}(t)$ autour du point

$$t = y_k \ : \ M_{4,k}(t) = (t-y_k)^4 + A_k(t-y_k)^3 + B_k(t-y_k)^2 \ .$$

THÉORÈME 4. *Soient* A *et* B *deux nombres donnés où* $0 \leq B \leq 4$, $0 \leq A \leq 2\sqrt{B}$. *Alors le système d'équations*

$$y^4 + Ay^3 + By^2 - p(y-x)^3 = 0 \tag{1}$$

$$4y^3 + 3Ay^2 + 2By - 3p(y-x)^2 = 0 \tag{2}$$

$$z^4 + Az^3 + Bz^2 - p(z-x)^3 = 1 \tag{3}$$

$$4z^3 + 3Az^2 + 2Bz - 3p(z-x)^2 = 0 \tag{4}$$

sous les contraintes $0 \le x \le z \le y$ *admet une et une seule solution. De plus, si* $(w-y)^4 + C(w-y)^3 + D(w-y)^2$ *est le développement de Taylor autour de* $w = y$ *de la fonction* $w^4 + Aw^3 + Bw^2 - p(w-x)^3$ *(alors que* p , x , y *et* z *forment la solution du système), alors* (C,D) *appartient à la courbe du plan dont les équations paramétriques sont*

$$C(t) = 2(1+t^4)/t^3 \quad , \quad D(t) = (3+t^4)/t^2$$

où $1 \le t < 3^{\frac{1}{4}}$. *En particulier* $0 < D \le 4$, $0 \le C \le 2\sqrt{D}$. *Enfin si* $D = 4$, *alors* $B = 4$.

DÉMONSTRATION. a) Existence de la solution du système. Soit $y > 0$. Les équations (1) et (2) permettent de déterminer p et x en fonction de y . En effet, si $L(w)$ est la fonction $w^4 + Aw^3 + Bw^2$, $x = (y^4-By^2)/L'(y)$ et $p = (L'(y))^3/(27L^2(y))$. Faisons ce tel choix de p et de x . Puisque l'on veut que $x \ge 0$, il faudra supposer aussi que $y \ge \sqrt{B}$. Observons la fonction $f(w) = L(w) - p(w-x)^3$ sur l'intervalle $[x,y]$. Puisque f est un polynôme de degré 4 et que $f(x) > 0$, $f'(x) > 0$, $f''(x) > 0$, la règle des signes de Descartes donne que $f(w)$ et $f'(w)$ chacune admettent au plus deux zéros compte tenu de la multiplicité des racines. D'où $D(y) = f''(y)/2 > 0$ et il existe une seule valeur de z dans (x,y) telle que $f'(z) = 0$. Si $F(z) = f''(z)/2$, alors $F(z) < 0$. Effectivement, la valeur de z est la position où $f(w)$ est maximum dans l'intervalle $[x,y]$. Ce nombre z est donc déterminé par les équations (1), (2) et (4) dès que y est fixé et c'est ce choix pour z que nous faisons et posons $H(y) = f(z)$.

Pour que l'équation (3) soit satisfaite, il faut trouver un y tel que $H(y) = 1$. Faisons varier y de $\sqrt{B}$ à l'infini. Nous allons montrer que $H(\sqrt{B}) \leq 1$, que $H(y)$ est strictement croissante en fonction de y et que $\lim_{y\to\infty} H(y) = \infty$.

Remplaçons l'équation (3) par l'équation (3')

$$z^4 + Az^3 + Bz^2 - p(z-x)^3 = H(y) \ . \qquad (3')$$

En dérivant par rapport à y les équations (1), (2), (3') et (4), on obtient l'équation matricielle

$$\begin{bmatrix} -(y-x)^3 & 3p(y-x)^2 & 0 & 0 \\ -3(y-x)^2 & 6p(y-x) & 2D(y) & 0 \\ -(z-x)^3 & 3p(z-x)^2 & 0 & 0 \\ -3(z-x)^2 & 6p(z-x) & 0 & 2F(z) \end{bmatrix} \begin{bmatrix} \frac{dp}{dy} \\ \frac{dx}{dy} \\ 1 \\ \frac{dz}{dy} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ H'(y) \\ 0 \end{bmatrix} .$$

Le déterminant du système est

$$12p(y-x)^2(z-x)^2(y-z)D(y)F(z) < 0 \ .$$

Par le théorème des fonctions implicites, les fonctions p , x et z sont dérivables par rapport à y . Après calcul, on obtient la valeur de la dérivée de H

$$H'(y) = 2D(y)(z-x)^2(y-z)/(y-x)^2 \ .$$

H est donc strictement croissante. Montrons que $\lim_{y\to\infty} H(y) = \infty$. $H(y)$ est la valeur maximale de f sur l'intervalle $[x,y]$. Ainsi $H(y) \geq f(x) = L(x) \geq x^4$. De plus, $x = (y^4 - By^2)/L'(y)$ tend vers l'infini lorsque y tend vers l'infini. Ceci établit que $\lim_{y\to\infty} H(y) = \infty$.

Enfin, si $y = \sqrt{B}$, $x = 0$, $p = A + 2 + \sqrt{B}$, $z = \sqrt{B}/2$ et $H(\sqrt{B}) = B^2/16$. D'où si $B \leq 4$, $H(\sqrt{B}) \leq 1$. Ainsi il existe donc une et une seule solution au système d'équations (1), (2), (3) et (4) sous les contraintes $0 \leq x \leq z \leq y$.

b) Propriétés de coefficients (C,D) . La fonction $f(w) = w^4 + Aw^3 + Bw^2 - p(w-x)^3$ a un zéro double pour $w = y$, d'où

$$w^4 + Aw^3 + Bw^2 - p(w-x)^3 = (w-y)^4 + C(w-y)^3 + D(w-y)^2$$

où $D = f''(y)/2$ et $C = f'''(y)/6$. Reprenant les équations (3) et (4) et posant $t = (y-z)$, on a que

$$t^4 - Ct^3 + Dt^2 = 1 \tag{5}$$

$$4t^3 - 3Ct^2 + 2Dt = 0 . \tag{6}$$

La solution de ce système donne que

$$C = 2(1+t^4)/t^3 \quad , \quad D = (3+t^4)/t^2 .$$

Puisque l'on sait que $f''(z) < 0$, on obtient que $12t^2 - 6Ct + 2D < 0$ et après simplification que $t < 3^{\frac{1}{4}}$. Utilisons l'hypothèse que $A \leq 2\sqrt{B}$; ceci permet de savoir que la fonction $w^4 + Aw^3 + Bw^2$ ne prend jamais des valeurs négatives. Lorsque $w < x$, $-p(w-x)^3$ est toujours ≥ 0 . D'où $f(w) \geq 0$ pour tout w . On connait déjà deux racines à l'équation $f'(w) = 0$, il s'agit de y et de $z = y - t$. Si $y - u$ est la troisième racine, on a que $u \neq 0$, $u \neq t$

$$4u^3 - 3Cu^2 + 2Du = 0 . \tag{6'}$$

Des équations (6) et (6'), $u + t = 3C/4$ et $ut = D/2$. Si l'on veut évaluer $S = f(y-u)$, on obtient

$$S = u^4 - Cu^3 + Du^2$$
$$S = u^4 - 4(u+t)u^3/3 + 2tu^3$$
$$S = u^3 + u^3(2t-u)/3 .$$

Puisque $u = D/(2t) = (3+t^4)/(2t^3)$

$$2t - u = 3(t^4-1)/(2t^3) .$$

Ainsi $S = (3+t^4)^3(t^4-1)/(16t^{12})$. Or l'on sait que $S \geq 0$, d'où $t \geq 1$. Comme

$f \geq 0$, on saura que le discriminant de l'équation $v^2 + Cv + D = 0$ est inférieur ou égal à 0 ; $C \leq 2\sqrt{D}$. La fonction $t \to (3+t^4)/t^2$ est décroissante sur l'intervalle $(1,3^{\frac{1}{4}})$. D'où $0 < D \leq 4$.

Montrons enfin que si $D = 4$, alors $B = 4$. Si $D = 4$, alors $t = 1$ et $C = 4$.

$$f(w) = (w-y)^4 + 4(w-y)^3 + 4(w-y)^2$$
$$= (w-y)^2(w+2-y)^2 .$$

Ainsi $f(y-2) = 0$. Ceci ne peut arriver que si $x = 0$, $y = 2$. Or $x = 0$ que si $y = \sqrt{B}$. Ainsi $B = 4$.

C.Q.F.D.

Si l'on désigne par Φ la transformation $(A,B) \to (C,D)$ telle que définie par le théorème précédent sur le domaine $\{(A,B) : 0 \leq B \leq 4 , \ 0 \leq A \leq 2\sqrt{B}\}$, on peut considérer les diverses itérées Φ_n de Φ : $\Phi_{n+1} = (\Phi_n)$. Indiquons la liaison naturelle entre la transformation Φ et les monosplines à élongation maximale de degré 4 . Posons $y_o = 0$, $A_o = B_o = 0$, $(A_1,B_1) = \Phi(A_o,B_o)$, $(A_2,B_2) = \Phi(A_1,B_1)$, ... Ainsi $(A_n,B_n) = \Phi_n(0,0)$. A l'aide de la suite A_k , B_k et du **système** (1), (2), (3) et (4), nous pouvons construire par récurrence une suite $(p_k,x_k,y_k,z_k)_{k=1}$. En effet, si p , x , y , z est la solution de ce système avec $A = A_{k-1}$ et $B = B_{k-1}$, on pose $p_k = p$, $x_k = y_{k-1} + x$, $y_k = y_{k-1} + y$, $z_k = y_{k-1} + z$.

Le Théorème 3 permet d'affirmer que $M_{4,k}(t) = t^4 - \sum_{i=1}^{k} p_i(x-x_i)^3$. La suite de notre étude est d'étudier le comportement des deux suites $\{p_k\}$ et $\{x_k\}$. Nous verrons que p_k converge vers 8 et que $x_k - 2k$ converge. Cela sera une conséquence du fait que la suite (A_k,B_k) converge de façon rapide vers le point $(4,4)$. Dans un premier temps, déterminons les points fixes de la transformation Φ .

LEMME 5. *Le seul point fixe de la transformation* Φ *sur le domaine* $\{(A,B) : 0 \leq B \leq 4 , \ 0 \leq A \leq 2\sqrt{B}\}$ *est* $A = 4$, $B = 4$.

DÉMONSTRATION. Soit (A,B) un point fixe de la transformation Φ. Si $P(w) = w^4 + Aw^3 + Bw^2$ on a que $P(w) - P(w-y) = p(w-x)^3$. Dans P_4, l'espace des polynômes de degré inférieur ou égal à 4, on sait que le noyau de l'application $P(w) \to P(w) - P(w-y)$ est formé des fonctions constantes. On peut dès lors vérifier directement que la solution générale de l'équation

$$P(w) - P(w-y) = p(w-x)^3$$

est $\lambda + p(w-x)^2(w+y-x)^2/(4y)$ où λ est une constante arbitraire. Comme le coefficient de w^4 est un, on obtient $p = 4y$. Puisque $A \geq 0$, $B \geq 0$, $x \geq 0$, on obtient $P(x) \geq 0$, il faut donc que $\lambda = 0$. Puisque $P(0) = 0$, il faut aussi que $x^2(y-x)^2 + \lambda = 0$. D'où $\lambda = 0$ et $x = 0$. Le maximum de $P(w-y)$ pour $x \leq w \leq y$, doit être égal à 1 ; or ce maximum est $y^4/16$ et ainsi $y = 2$, $P(w) = w^2(w+2)^2$ et $A = B = 4$.

Du Théorème 4 et du Lemme 5, il suit que pour tout couple (A_o,B_o), $0 \leq B_o \leq 4$, $0 \leq A_o \leq 2\sqrt{B}_o$, la suite $\Phi_n(A_o,B_o)$ converge vers le point $(4,4)$. En effet Φ applique le domaine de départ sur un arc simple. La restriction de Φ à cet arc simple n'admet qu'un point fixe et ce point fixe est une extrémité de l'arc, ceci est suffisant pour établir la convergence des itérés vers ce point fixe. Pour étudier la vitesse de convergence vers ce point fixe, il est important d'étudier la transformation Φ autour de ce point fixe. C'est la prochaine étape qui est exécutée.

8. Etude locale de la transformation Φ

Pour (A,B) donné, $0 \leq B \leq 4$ et $0 \leq A$ le système d'équations (1), (2), (3) et (4) permet de définir 4 fonctions de (A,B), p, x, y et z dès que l'on exige, $0 \leq x \leq z \leq y$. Vu que le Jacobien du système n'est pas singulier, les fonctions p, x, y et z sont analytiques. Une identité polynomiale en w est associée à la transformation $C,D = \Phi(A,B)$:

$$(w-y)^4 + C(w-y)^3 + D(w-y)^2 = w^4 + Aw^3 + Bw^2 - p(w-x)^3 . \qquad (7)$$

Autrement dit,

$$C = 4y + A - p ,$$
$$D = 6y^2 + 3Ay + B - 3p(y-x) .$$

La transformation Φ est analytique. Pour A fixé ≥ 0 et pour B variable mais voisin de 4, nous voulons évaluer les premiers termes du développement de Taylor de la transformation Φ.

LEMME 6. *Soit* $A \geq 0$, *au point* $(A,4)$, *les évaluations de* $(p , x , y , z , \frac{\partial p}{\partial B} , \frac{\partial x}{\partial B} , \frac{\partial y}{\partial B} , \frac{\partial z}{\partial B})$ *donnent respectivement* $(A+4 , 0 , 2 , 1 , 0 , -(3(A+4))^{-1} , 0 , 0)$.

DÉMONSTRATION. Il est aisé de vérifier que $p = A + 4$, $x = 0$, $y = 2$ et $z = 1$ donne la solution du système (1), (2), (3) et (4) lorsque $B = 4$. En dérivant par rapport à B ces mêmes équations en faisant l'évaluation en $B = 4$, on obtient

$$\begin{bmatrix} -8 & 12(A+4) & 0 & 0 \\ -12 & 12(A+4) & 8 & 0 \\ -1 & 3(A+4) & 0 & 0 \\ -3 & 6(A+4) & 0 & -4 \end{bmatrix} \begin{bmatrix} \frac{\partial p}{\partial B} \\ \frac{\partial x}{\partial B} \\ \frac{\partial y}{\partial B} \\ \frac{\partial z}{\partial B} \end{bmatrix} = - \begin{bmatrix} 4 \\ 4 \\ 1 \\ 2 \end{bmatrix} .$$

On vérifie aisément que $\frac{\partial p}{\partial B} = \frac{\partial z}{\partial B} = \frac{\partial y}{\partial B} = 0$ et $\frac{\partial x}{\partial B} = -(3(A+4))^{-1}$ est la solution de cette équation vectorielle.

THÉORÈME 7. *Le développement de Taylor de chacune des composantes de la transformation* Φ *au point* $(A,4)$ *jusqu'à l'ordre* 3 *selon les puissances de* $(B-4)$ *est*

$$4 + (B-4)^3/(108(A+4)^2) .$$

DÉMONSTRATION. Posons $\beta = (B-4)$. Si $\Phi(A,B) = (C,D)$, faisons le calcul de C et de D en fonction de B en négligeant les puissances de β supérieures ou égales à 4. Evaluons l'identité (7) au point $w = y - 2$

$$(w-y)^4 + C(w-y)^3 + D(w-y)^2 = w^4 + (A-p)w^3 + (B+3px)w^2 - 3px^2w + px^3 .$$

Lorsque $w = y - 2$, chacun des termes du second membre est négligeable par rapport à β^3 à l'exclusion du terme px^3. D'où l'on obtient

$$16 - 8C + 4D = -\beta^3/(27(A+4)^2) .$$

Par le Théorème 4, la droite $C = D$ est tangente à la courbe paramétrique $C = C(t)$, $D = D(t)$ au point $t = 1$. Ainsi en négligeant les termes d'ordre β^4

$$C = D = 4 + \beta^3/(108(A+4)^2) .$$

THÉORÈME 8. *Soit* $(A_n,B_n) = \Phi_n(0,0)$. *Alors il existe un nombre* ξ *de* $(0,1)$ *tel que*

$$A_{n+1} = 4 - 48\sqrt{3}\,\xi^{3^n} + 0(\xi^{2\cdot 3^n})$$

$$B_{n+1} = 4 - 48\sqrt{3}\,\xi^{3^n} + 0(\xi^{2\cdot 3^n}) .$$

DÉMONSTRATION. Soit t un nombre de l'intervalle $[1,3^{\frac{1}{4}}]$. Posons $A = 2(1+t^4)/t^3$ et $B = (3+t^4)/t^2$. Si $(C,D) = \Phi(A,B)$, soit u le nombre tel que $C = 2(1+u^4)/u^3$ et $D = (3+u^4)/u^2$, notons par $u = f(t)$ cette fonction t ; celle-ci est analytique en $t = 1$. Le Théorème 7 permet de dégager le développement de Taylor de f au point 1 :

$$f(t) = 1 + (t-1)^3/432 + 0((t-1)^4) .$$

En effet $dB = -4dt$, $dD = (dB)^3/(108 \times 64)$ et $dU = -dD/4$.

Soit $F(t)$ la fonction de Bottcher associée à f (cf. Montel [4] par exemple), F est l'unique fonction analytique définie sur $[1,3^{\frac{1}{4}}]$ telle que $F(1) = 0$, $F'(1) = (432)^{-\frac{1}{2}}$ et $F(f(t)) = (F(t))^3$. On sait que si $t > 1$, $F(t)$ appartient à $(0,1)$. Si $f_n(t)$ est la n^e itérée de f et F_{-1} est l'inverse fonctionnel de F, on a que $F(f_n(t)) = (F(t))^{3^n}$ et $f_n(t) = F_{-1}((F(t))^{3^n})$. Ainsi lorsque n tend vers l'infini, $f_n(t) = 1 + \sqrt{432}(F(t))^{3^n} + 0((F(t))^{2\cdot 3^n})$.

Si $(A_n,B_n) = \Phi_n(0,0)$ et si t_o est le nombre tel que $A_1 = 2(1+t_o^4)/t_o^3$ et $B_1 = (3+t_o^4)/t_o^2$, en posant $t_n = f_n(t_o)$, on a que

$$A_{n+1} = 2(1+t_n^4)/t_n^3 \quad \text{et} \quad B_{n+1} = (3+t_n^4)/t_n^2 .$$

D'où

$$A_{n+1} = 4 - 4\sqrt{432}(F(t_o))^{3^n} + 0((F(t_o))^{2\cdot 3^n}) ,$$

$$B_{n+1} = 4 - 4\sqrt{432}(F(t_o))^{3^n} + 0((F(t_o))^{2\cdot 3^n}) .$$

Il suffit donc de poser $\xi = F(t_o)$.

9. Comportement asymptotique des monosplines à élongation maximale de degré 4

Nous en arrivons à la conclusion principale de notre étude.

THÉORÈME 9. *Soit* $\{p_i,x_i\}_{i=1}^{\infty}$ *la suite de poids et de noeuds* $0 < x_1 < x_2 < \ldots$ *et telle que pour chaque valeur entière de* n, $t^4 - \sum_{i=1}^{n} p_i(t-x)^3$ *est la monospline réduite à élongation maximale d'élongation* c_n. *Alors il existe deux constantes* γ *et* ξ, $0 < \xi < 1$ *telles que*

$$x_{n+2} = 2n + \gamma + 2 + 2\sqrt{3}\,\xi^{3^n} + 0(\xi^{2\cdot 3^n}) .$$

On obtient le comportement asymptotique suivant pour la suite p_n *et la suite* c_n :

$$p_{n+2} = 8 - 48\sqrt{3}\,\xi^{3^n} + 0(\xi^{2\cdot 3^n}) ,$$

$$c_{n+2} = 2n + \gamma + 3 + \sqrt{2} + 0(\xi^{3\cdot 3^n}) .$$

DÉMONSTRATION. Reprenons la suite $\{p_i,x_i,y_i,z_i\}_{i=1}^{\infty}$ donnée dans le Théorème 3 de la section 7 pour le cas $n = 4$. Si $(A_n,B_n) = \Phi_n(0,0)$, les quantités $(p_n,x_n-y_{n-1},y_n-y_{n-1},z_n-y_{n-1})$ forment la solution du système d'équations (1), (2), (3), (4) avec $A = A_{n-1}$ et $B = B_{n-1}$. Vu le Lemme 6, $p_{n+2} = A_{n+1} + 4 + 0((B_{n+1}-4)^2)$, $x_{n+2} - y_{n+1} = -(B_{n+1}-4)/24 + 0((B_{n+1}-4)^2)$ et $y_{n+2} - y_{n+1} = 2 + 0((B_{n+1}-4)^2)$.

Utilisons le Théorème 8 avec sa notation. Nous obtenons $p_{n+2} = 8 - 48\sqrt{3}\,\xi^{3^n} + 0(\xi^{2\cdot 3^n})$. Soit $\gamma = \sum_{n=1}^{\infty} (y_n - y_{n-1} - 2) = \lim_{n\to\infty} (y_n - 2n)$. Remarquons

$$y_{n+1} - (2n + 2) = \gamma - \sum_{k=n+2}^{\infty} (y_k - y_{k-1} - 2) ,$$

$$y_{n+1} = 2n + 2 + \gamma + 0((B_{n+1}-4)^2) .$$

Ainsi

$$x_{n+2} = 2n + 2 + \gamma - 2\sqrt{3}\,\xi^{3^n} + 0(\xi^{2\cdot 3^n}) .$$

Pour étudier c_n , on remarque d'abord que $c_{n+2} - y_{n+2}$ est la plus grande solution de l'équation $t^4 + A_{n+2}t^3 + B_{n+2}t^2 = 1$. A_{n+2} et B_{n+2} sont voisins de 4 , $-1 + \sqrt{2}$ est la plus grande solution positive de l'équation $t^4 + 4t^3 + 4t^2 = 1$. D'où

$$c_{n+2} - y_{n+1} = \sqrt{2} - 1 + 0((B_{n+2}-4))$$

$$c_{n+2} = 2n + \gamma + \sqrt{2} + 3 + 0(\xi^{3\cdot 3^n}) .$$

Le Théorème 9 est donc démontré. La détermination numérique de la constante γ est facile, $\gamma = 0{,}44424$. Le calcul de ξ est plus délicat à effectuer, $\xi = 0{,}000199$. La petitesse de ξ explique encore davantage pourquoi la suite p_n tend si vite vers 8 .

Références

[1] BOURDEAU, M., *Quadratures optimales pour certains cônes de fonctions*, Thèse de doctorat, Université de Sherbrooke, 1974.

[2] JOHNSON, R.S., *On monosplines of least deviation*, Trans. Amer. Math. Soc. 96 (1960), 458-477.

[3] KARLIN, S. and SCHUMAKER, L., *The fundamental theorem of algebra for Tchebycheffian monosplines*, J. d'anal. math. 20 (1967), 233-270.

[4] MONTEL, P., *Leçons sur les récurrences et leurs applications*, Gauthier-Villars, Paris (1957).

[5] SARD, A., *Best approximation integration formulas; best approximation formulas*, Amer. J. Math. 71 (1949), 80-91.

Département de mathématiques et de statistique
Université de Montréal
C.P. 6128, Succ. "A"
Montréal, Québec
H3C 3J7

Manuscrit reçu le 10 janvier 1980.
Revisé le 31 mai 1982.

Dans la fleur de l'âge, M. Richard Bastien est décédé en février 1980. L'article *Monosplines à oscillation minimale* a été le fruit de ma collaboration aux activités mathématiques de M. Bastien. Je tiens à rendre hommage à ses qualités: simplicité, ténacité et très grande honnêteté; il aimait bien la nature. Dans sa courte carrière d'enseignant, il a manifesté qu'il avait à coeur le bien de ses étudiants. Son départ a été une perte.

Serge Dubuc