M2AN. MATHEMATICAL MODELLING AND NUMERICAL ANALYSIS - MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE

P. J. LAURENT

Inf-convolution spline pour l'approximation de données discontinues

M2AN. Mathematical modelling and numerical analysis - Modélisation mathématique et analyse numérique, tome 20, n° 1 (1986), p. 89-111

<http://www.numdam.org/item?id=M2AN_1986__20_1_89_0>

MATHEMATICAL MODELLING AND NUMERICAL ANALYSIS
MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE

(Vol 20, n° 1, 1986, p 89 à 111)

INF-CONVOLUTION SPLINE POUR L'APPROXIMATION DE DONNÉES DISCONTINUES (*)

par P. J. LAURENT [1]

Résumé — *Un nouveau type de fonctions-spline est obtenu en minimisant une fonctionnelle quadratique basée sur l'opération d'inf-convolution Ces splines sont utiles pour le traitement de données correspondant à des fonctions ayant des dérivees discontinues*

Abstract — *A new type of spline-functions is obtained by minimizing a quadratic functional based on the inf-convolution operation These splines are useful for the treatment of data corresponding to functions with discontinuous derivatives*

1. INTRODUCTION

On désigne dans toute la suite par $D^{-m} L^2(I)$ (I intervalle quelconque de $\mathbb{R}$; $m \geqslant 1$) l'espace des fonctions ayant une dérivée d'ordre m (au sens des distributions) qui est une fonction de carré sommable sur I. Pour $x \in D^{-m} L^2(I)$, on pourra définir la quantité :

$$E_I^{(m)}(x) = \int_I (x^{(m)}(t))^2 \, dt \, . \tag{1.1}$$

Les fonctions spline (F.S.) d'interpolation classiques qui sont obtenues en minimisant un critère du type $E_{[a,b]}^{(m)}(x)$ avec les conditions d'interpolation $x(t_i) = f(t_i)$, $i = 1, \ldots, n$ (avec $a < t_1 < t_2 < \cdots < t_n < b$) conduisent à une approximation satisfaisante de f lorsque cette fonction est dérivable, plus particulièrement lorsque $f \in D^{-m} L^2([a, b])$.

Considérons maintenant le cas d'une fonction f continue mais ayant des discontinuités de dérivées premières en certains points $\alpha_1, \ldots, \alpha_m$ ($t_1 < \alpha_1 < \cdots < \alpha_m < t_n$). Plus précisément on suppose que f appartient à l'espace $\mathscr{H}$ des fonctions continues sur $[a, b]$ telles que la restriction f_k à l'inter-

(*) Reçu en avril 1985

[1] TIM3 Institut IMAG, Université de Grenoble, BP 68, 38402 Saint Martin d'Hères, Cedex, France

valle $I_k = [\alpha_k, \alpha_{k+1}]$, ($\alpha_0 = a, \alpha_{m+1} = b$) appartient à $D^{-2} L^2(I_k)$, $k = 0, ..., m$. La F.S. classique conduit alors à une approximation très mauvaise de f puisqu'elle a nécessairement une dérivée première continue aux points α_j. Pour remédier à cette difficulté, il est naturel d'introduire la fonction discontinuité de base en α_j pour la dérivée première :

$$p_j(t) = (t - \alpha_j)_+ = \begin{cases} (t - \alpha_j), & \text{si } t \geqslant \alpha_j \\ 0 & , \text{si } t \leqslant \alpha_j \end{cases} \tag{1.2}$$

et de minimiser par rapport à $x \in D^{-2} L^2([a, b])$ et à $d_1 \in \mathbb{R}, ..., d_m \in \mathbb{R}$ la quantité :

$$E^{(2)}_{[a,b]}(x) \tag{1.3}$$

avec les conditions :

$$x(t_i) + \sum_{j=1}^{m} d_j p_j(t_i) = f(t_i), \quad i = 1, ..., n. \tag{1.4}$$

Si l'on note $\bar{x}, \bar{d}_j, j = 1, ..., m$, la solution de ce problème (on montrera qu'elle existe et qu'elle est unique pour $n \geqslant 2$) et si l'on pose :

$$s(t) = \bar{x}(t) + \sum_{j=1}^{m} \bar{d}_j p_j(t) \tag{1.5}$$

la fonction s est caractérisée par les conditions suivantes (on suppose ici pour la simplicité que les α_j sont distincts des t_i) :

(i) s est un polynôme de degré 3 dans chaque intervalle entre les t_i ou les α_j
(ii) s est un polynôme de degré 1 dans $[a, t_1]$ et $[t_n, b]$
(iii) s, s' et s'' sont continues en t_i, $i = 1, ..., n$
(iv) s, s'' et s''' sont continues en α_j et $s''(\alpha_j) = 0, j = 1, ..., m$
(v) $s(t_i) = f(t_i)$, $i = 1, ..., n$.

(1.6)

On sait que la F.S. cubique classique correspond à la position d'équilibre d'une tige mince flexible passant par les points $[t_i, f(t_i)]$, $E^{(2)}_{[a,b]}(x)$ étant considéré comme une approximation de l'énergie de flexion d'une tige coïncidant avec le graphe de x. Dans le cas présent, la solution s ci-dessus correspond à la position d'équilibre d'une tige flexible *articulée* en chaque point α_j. Pour $x \in \mathscr{H}$, considérons l'énergie de flexion :

$$E(x) = \sum_{k=0}^{m} E^{(2)}_{I_k}(x)$$

de la tige articulée correspondant à x. La fonction s réalise alors le minimum de $E(x)$ avec les conditions d'interpolation $x(t_i) = f(t_i), i = 1, \dots, n$.

On peut encore interpréter d'une autre façon la solution s précédente : la valeur de la fonction en α_j (supposés distincts des t_i) n'est pas donnée. Fixons-nous des valeurs provisoires $f(\alpha_k) = \theta_k$, $k = 1, \dots, m$. On peut alors considérer séparément la minimisation de $E_{I_k}^{(2)}$ sur chaque I_k avec les conditions

$$x(\alpha_k) = \theta_k\,; \quad x(\alpha_{k+1}) = \theta_{k+1}$$
$$x(t_i) = f(t_i)\,, \quad t_i \in I_k$$

qui conduit à une F.S. cubique classique (ayant en particulier une dérivée seconde nulle en α_k et α_{k+1}). On minimise alors dans un second temps l'énergie totale $E(x)$ par rapport à $\theta_1, \dots, \theta_m$.

On remarque que la solution s obtenue a une dérivée seconde nulle en α_j. La fonction f qui est traitée n'ayant pas forcément cette propriété, cela peut entraîner quelques perturbations (oscillations de la dérivée seconde) au voisinage des α_j, comparables à l'effet des conditions de bout libre pour les F.S. classiques. La fonction f peut avoir une dérivée seconde continue en α_j mais non nulle, ou encore une discontinuité de dérivée seconde. Pour tenir compte de ce fait, on peut utiliser des F.S. d'ordre supérieur (quintique par exemple) ou changer les fonctions p_j.

Considérons la minimisation de $E_{[a,b]}^{(3)}(x)$ par rapport à $x \in D^{-3} L^2([a, b])$ et à $d_1, \dots, d_m$ vérifiant les conditions définies par (1.2) et (1.4). Si l'on note s la solution correspondante selon (1.5), celle-ci est caractérisée pour $n \geqslant 3$ par les conditions suivantes :

(i) s est un polynôme de degré 5 dans chaque intervalle entre les t_i ou les α_j

(ii) s est un polynôme de degré 2 dans $[a, t_1]$ et $[t_n, b]$

(iii) $s, s^{(1)}, s^{(2)}, s^{(3)}, s^{(4)}$ sont continues en t_i, $i = 1, \dots, n$

(iv) $s, s^{(2)}, s^{(3)}, s^{(4)}, s^{(5)}$ sont continues en α_j et $s^{(4)}(\alpha_j) = 0, j = 1, \dots, m$

(v) $s(t_i) = f(t_i)\,, \quad i = 1, \dots, n.$

(1.7)

Par contre, si l'on pose :

$$q_j(t) = (t - \alpha_j)_+^2/2 \tag{1.8}$$

et si l'on minimise $E_{[a,b]}^{(3)}(x)$ par rapport à $x \in D^{-3} L^2([a, b])$ et à $d_1, \dots, d_m$, $e_1, \dots, e_m$, vérifiant les conditions :

$$x(t_i) + \sum_{j=1}^{m} d_j\, p_j(t_i) + \sum_{j=1}^{m} e_j\, q_j(t_i) = f(t_i) \quad i = 1, \dots, n \tag{1.9}$$

on obtient une solution :

$$s(t) = \bar{x}(t) + \sum_{j=1}^{m} \bar{d}_j p_j(t) + \sum_{j=1}^{m} \bar{e}_j q_j(t) \tag{1.10}$$

qui est caractérisée par les conditions (1.7) en remplaçant (iv) par :

(iv′) s, $s^{(3)}$, $s^{(4)}$ et $s^{(5)}$ sont continues en α_j et

$$s^{(3)}(\alpha_j) = s^{(4)}(\alpha_j) = 0\,, \qquad j = 1, \ldots, m\,. \tag{1.11}$$

Une autre façon de procéder consiste à prendre pour p_j une fonction qui est une approximation locale de f au voisinage de α_j, par exemple une fonction continue formée de deux morceaux de polynôme du deuxième degré prenant la valeur 0 en α_j. On verra plus loin comment caractériser la solution correspondante.

Si les valeurs de la fonction f ne sont pas connues de façon exacte, c'est-à-dire si les données sont $z_i = f(t_i) + \varepsilon_i$ où les ε_i sont des erreurs, alors on peut introduire de façon similaire des F.S. de lissage. On minimisera alors, dans le cas du premier exemple :

$$E^{(2)}_{[a,b]}(x) + \rho \sum_{i=1}^{n} \left(x(t_i) + \sum_{j=1}^{m} d_j p_j(t_i) - z_i \right)^2 \tag{1.12}$$

par rapport à $x \in D^{-2} L^2([a, b])$ et à $d_1 \in \mathbb{R}, \ldots, d_m \in \mathbb{R}$ pour un coefficient $\rho > 0$ convenablement choisi. La solution $s = \bar{x} + \sum_{j=1}^{m} \bar{d}_j p_j$ obtenue est caractérisée par les conditions (1.6) dans lesquelles on remplace (v) par :

(v′) $$(s^{(3)}(t_i^+) - s^{(3)}(t_i^-)) + \rho(s(t_i) - z_i) = 0 \quad i = 1, \ldots, n\,. \tag{1.13}$$

Dans les exemples concrets, il arrive que les seules données du problème soient les $z_i = f(t_i)$, $i = 1, \ldots, r$ et le nombre m de discontinuités de f'. La position α_j de la discontinuité n'est pas connue mais on peut souvent disposer d'une localisation : $\alpha_j \in [\underline{\alpha}_j, \overline{\alpha}_j]$. En se basant sur l'image intuitive de la tige flexible articulée il est clair que l'énergie $E(s)$ de la tige à l'équilibre sera d'autant plus faible que les α_j sont choisis voisins des positions de discontinuité exacte de la fonction f. La méthode consiste donc à choisir α_j^* qui minimise $E(s)$. On verra que ceci permet de « retrouver » avec une excellente précision les positions réelles de discontinuité de la fonction f (on se place ici dans le cas où l'on se donne, pour tester la méthode, une fonction f, dont on n'utilise que l'information $f(t_i) = z_i$, $i = 1, \ldots, n$, et le nombre de discontinuités de dérivée avec leur localisation).

Enfin, le même principe peut être appliqué à l'interpolation ou au lissage de fonctions de plusieurs variables ayant des discontinuités de certains dérivées.

Nous montrerons d'abord au paragraphe 2 comment les problèmes de minimisation que nous venons d'évoquer peuvent être écrits en utilisant l'opération d'inf-convolution de deux fonctions. Nous rappellerons ensuite au paragraphe 3 les notions essentielles d'espace semi-hilbertiens et de semi-noyau pour la caractérisation de F.S. générales. Au paragraphe 4 nous montrerons alors que les F.S. par inf-convolution peuvent être de façon générale associées à un espace semi-hilbertiens dont le semi-noyau peut être évalué, ce qui permet leur caractérisation. Différents exemples sont donnés à une ou plusieurs variables. Enfin le paragraphe 5 est consacré aux algorithmes et aux résultats numériques.

2. INF-CONVOLUTION SPLINES

Considérons à nouveau l'exemple introductif (1.3), (1.4). Remarquons tout d'abord que l'on ne change pas la solution obtenue sur l'intervalle $[a, b]$ si l'on minimise $E^{(2)}_{[a',b']}(x)$ avec $[a', b'] \supset [a, b]$, ou même si l'on minimise $E^{(2)}_{\mathbb{R}}(x)$, avec x appartenant bien sûr à $D^{-2} L^2(\mathbb{R})$. Considérons l'espace $X = D^{-2} L^2(\mathbb{R})$ contenu dans l'espace $E = \mathscr{C}(\mathbb{R})$ des fonctions continues sur $\mathbb{R}$. Le sous-espace V des discontinuités engendré par les fonctions $p_j(t)$, $j = 1, ..., m$, est également contenu dans $E = \mathscr{C}(\mathbb{R})$, (mais non nécessairement dans X). Notons E' le dual topologique de E lorsque ce dernier est muni de la topologie de la convergence ponctuelle et notons l_i l'élément de E' (fonctionnelle ponctuelle en t_i) telle que :

$$\langle x, l_i \rangle = x(t_i)\,, \quad \text{pour tout} \quad x \in E\,. \tag{2.1}$$

Le problème revient alors à minimiser $E^{(2)}(x)$ par rapport à $x \in X$ et $v \in V$ vérifiant les conditions

$$\langle x + v, l_i \rangle = z_i\,, \quad \text{avec} \quad z_i = f(t_i)\,, \quad i = 1, ..., n\,. \tag{2.2}$$

Introduisons l'espace :

$$\tilde{X} = X + V = \{ \tilde{x} \in E \mid \tilde{x} = x + v, x \in X, v \in V \}\,. \tag{2.3}$$

Le problème de minimisation peut alors s'énoncer de la façon suivante :

$$\min_{\substack{\tilde{x} \in \tilde{X} \\ \langle \tilde{x}, l_i \rangle = z_i \\ i = 1,...,n}} \Big(\min_{\substack{x \in X, v \in V \\ x + v = \tilde{x}}} E^{(2)}_{\mathbb{R}}(x) \Big)\,. \tag{2.4}$$

On remarque que la solution en $\tilde{x}$ de ce problème est la solution $s = \bar{x} + \bar{v}$ caractérisée par (1.5) et (1.6).

Pour $x \in E$, posons encore :

$$\varphi(x) = \begin{cases} E_{\mathbb{R}}^{(2)}(x), & \text{si} \quad x \in X \\ +\infty \quad , & \text{sinon} \end{cases} \tag{2.5}$$

et

$$\psi(x) = \begin{cases} 0 \quad , & \text{si} \quad x \in V \\ +\infty, & \text{sinon}. \end{cases} \tag{2.6}$$

Cela permet formellement d'écrire le problème de la façon suivante :

$$\min_{\substack{\tilde{x} \in E \\ \langle \tilde{x}, l_i \rangle = z_i \\ i=1,\dots,n}} \left(\min_{\substack{x,v \in E \\ x+v=\tilde{x}}} (\varphi(x) + \psi(v)) \right) \tag{2.7}$$

On voit apparaître dans l'intérieur de la parenthèse une fonction de $\tilde{x}$ qui n'est autre que l'inf-convolution des fonctions φ et ψ (*cf.* [5]) :

$$\begin{aligned} \varphi \nabla \psi(\tilde{x}) &= \operatorname*{Inf}_{\substack{x,v \in E \\ x+v=\tilde{x}}} (\varphi(x) + \psi(v)) \\ &= \operatorname*{Inf}_{x \in E} (\varphi(x) + \psi(\tilde{x} - x)) \end{aligned} \tag{2.8}$$

Finalement le problème s'écrit donc :

$$\min_{\substack{\tilde{x} \in E \\ \langle x, l_i \rangle = z_i \\ i=1,\dots,n}} \varphi \nabla \psi(\tilde{x}) \tag{2.9}$$

On sait que l'opération d'inf-convolution correspond par polarité (au sens de Fenchel) à l'opération somme de deux fonctions, c'est-à-dire $(\varphi \nabla \psi)^* = \varphi^* + \psi^*$, *cf.* [5]. La notion d'inf-convolution a été utilisée en relation avec les F.S., par J. Duchon [4] notamment pour définir toutes les fonctionnelles quadratiques dont la minimisation conduit à des F.S. homogènes (invariantes par similitude).

3. SEMI-NOYAUX ET FONCTIONS-SPLINE

Nous donnons d'abord un bref résumé des notions de sous-espaces semi-hilbertiens et de semi-noyaux introduits par J. Duchon [3], [4], afin de poser le problème des F.S. par inf-convolution dans ce même cadre général.

a) *Sous-espaces semi-hilbertiens*

Supposons que E et E' soient deux espaces vectoriels topologiques localement convexes en dualité par la forme bilinéaire $\langle x, x' \rangle$, $x \in E$, $x' \in E'$. Un sous-espace vectoriel X de E muni d'un semi-produit scalaire noté (x, y), $x \in X$, $y \in X$ et de la semi-norme associée $|x| = (x, x)^{1/2}$ est dit *semi-hilbertien* si :

(i) l'espace nul N de la semi-norme est de dimension finie

(ii) l'espace quotient X/N muni de la norme canoniquement associée à la semi-norme est complet

(iii) l'espace quotient X/N est topologiquement inclus dans E/N (E étant muni de la topologie faible).

(3.1)

Cette dernière hypothèse est équivalente au fait que $\lim_{n \to \infty} |x_n| = 0$ entraîne $\lim_{n \to \infty} \langle x_n, l \rangle = 0$, pour tout $l \in E'$ qui est nulle sur N.

Exemples :

(i) L'espace $X = D^{-2} L^2(\mathbb{R})$ considéré au paragraphe précédent est un premier exemple de sous-espace semi-hilbertien de $E = \mathscr{C}(\mathbb{R})$, avec le semi-produit scalaire $(x, y) = \int_{\mathbb{R}} x''(t)\, y''(t)\, dt$ et la semi-norme associée

$$|x|^2 = \int_{\mathbb{R}} |x''(t)|^2\, dt. \tag{3.2}$$

(ii) Sur $\mathbb{R}^2$, l'espace $X = D^{-2} L^2(\mathbb{R}^2)$, espace des fonctions dont toutes les dérivées partielles d'ordre 2 (au sens des distributions) sont des fonctions de carré sommable sur $\mathbb{R}^2$ est un sous-espace semi-hilbertien de $\mathscr{C}(\mathbb{R}^2)$ pour le semi-produit scalaire obtenu en faisant la somme des produits scalaires de toutes les dérivées partielles d'ordre 2. La semi-norme associée s'écrit alors :

$$|x|^2 = \sum_{i,j \in \{1,2\}} \int_{\mathbb{R}^2} (D_i\, D_j\, x(t))^2\, dt \tag{3.3}$$

dans lequel D_i désigne l'opérateur de dérivation partielle par rapport à la i-ième variable.

(iii) Plus généralement sur l'espace $\mathbb{R}^d$, suivant J. Duchon [3], considérons l'espace de Hilbert $K^{\alpha}(\mathbb{R}^d)$ des distributions tempérées u dont la transformée de Fourier $\mathscr{F} u$ est une fonction localement sommable qui est de plus de carré sommable sur $\mathbb{R}^d$ par rapport à la fonction poids $\| \tau \|^{2\alpha}$, le réel α vérifiant la condition $\alpha < d/2$. Pour $u \in K^\alpha$, on pose :

$$\| u \|_\alpha^2 = \int_{\mathbb{R}^d} | \mathscr{F} u(\tau) |^2 \, \| \tau \|^{2\alpha} \, d\tau . \tag{3.4}$$

On considère alors l'espace $X_{p,\alpha}$ des distributions tempérées dont toutes les dérivées partielles d'ordre p appartiennent à K^α et on définit sur $X_{p,\alpha}$ la semi-norme :

$$| v |_{p,\alpha}^2 = \sum_{i_j \in \{1, \dots, d\}} \| D_{i_1} D_{i_2} \dots D_{i_p} v \|_\alpha^2 . \tag{3.5}$$

Lorsque l'entier k vérifie $k < p + \alpha - \frac{d}{2}$, l'espace K^α est inclus dans $\mathscr{C}^k(\mathbb{R}^d)$. Lorsque $p + \alpha - \frac{d}{2} > 0$, condition que l'on supposera toujours vérifiée dans la suite, $X_{p,\alpha}$ est un sous-espace semi-hilbertien de $E = \mathscr{C}(\mathbb{R}^d)$.

Pour $d = 1, p = 2$ et $\alpha = 0$, on retrouve l'exemple (i).

Pour $d = 2, p = 2$ et $\alpha = 0$, on retrouve l'exemple (ii).

b) *Semi-noyau*

Afin d'exprimer simplement les F.S. obtenues par minimisation d'une semi-norme quadratique, J. Duchon a introduit la notion naturelle et essentielle de noyau semi-reproduisant [3], [4]. De façon un peu plus générale, si L est un sous-espace vectoriel de E', on dira que l'application linéaire H de L dans E est un *semi-noyau* si, quel que soit $l \in N^0 \cap L$, on a :

$$\begin{aligned} &\text{(i)} \ H(l) \in X \\ &\text{(ii)} \ \langle x, l \rangle = (x, H(l)) , \quad \text{pour tout} \quad x \in X \end{aligned} \tag{3.6}$$

(N^0 désigne ici l'ensemble des éléments de E' qui sont nuls sur N).

L'application H n'est pas définie de façon unique. On voit que si P est une application linéaire de L dans N, alors $H + P$ convient encore.

L'application H a une certaine « semi-symétrie » dans le sens que pour tout $l \in L \cap N^0$ et pour tout $m \in L \cap N^0$, on a :

$$\langle Hm, l \rangle_{EE'} = (Hm, Hl) = (Hl, Hm) = \langle Hl, m \rangle_{EE'} . \tag{3.7}$$

De même elle a une certaine propriété de « semi-positivité » c'est-à-dire que pour tout $l \in L \cap N^0$, on a :

$$\langle Hl, l \rangle_{EE'} = (Hl, Hl) \geqslant 0 . \tag{3.8}$$

Mais il est important de noter que les propriétés précédentes ne sont pas nécessairement vérifiées lorsque l ou m parcourent L.

Remarque : Prenons pour un instant $L = E'$. Si l'on considère l'application linéaire $\tilde{H}$ de N^0 dans X/N topologiquement inclus dans E/N définie par :

$$\tilde{H}(l) = H(l) + N \tag{3.9}$$

alors $\tilde{H}$ est unique. Elle peut être considérée comme une application de $(E/N)'$ dans E/N et elle est alors faiblement continue, symétrique et positive au sens habituel. C'est en fait le noyau de Schwartz du sous-espace hilbertien X/N de E/N.

c) *Caractérisation des fonctions spline*

Considérons maintenant le problème général consistant à :

$$\text{minimiser} \quad | x |^2 \tag{3.10}$$

avec les conditions :

$$\left.\begin{array}{lll} & x \in X \quad \text{et} \quad \langle x, l_i \rangle = y_i\,, & i = 1, \ldots, n \\ \text{où} & l_i \in E' \quad \text{et} \quad y_i \in \mathbb{R} & , \quad i = 1, \ldots, n\,. \end{array}\right. \tag{3.11}$$

Si les conditions :

$$\langle p, l_i \rangle = 0\,, \quad i = 1, \ldots, n\,; \quad p \in N \tag{3.12}$$

entraînent $p = 0$

alors ce problème admet une solution unique que l'on appelle encore fonction-spline (F.S.).

Si les l_i appartiennent au sous-espace L pour lequel on a défini un semi-noyau H, alors on va voir que l'on peut exprimer la solution du problème à l'aide de ce semi-noyau.

En utilisant les théorèmes classiques de caractérisation d'une F.S. (*cf.* [5]), on montre que $\sigma \in X$ vérifiant $\langle \sigma, l_i \rangle = y_i$, $i = 1, \ldots, n$ est la solution du problème si et seulement s'il existe des coefficients λ_i, $i = 1, \ldots, n$ tels que :

$$(x, \sigma) = \left\langle x, \sum_{i=1}^{n} \lambda_i\, l_i \right\rangle, \quad \text{pour tout} \quad x \in X\,. \tag{3.13}$$

Comme cette condition entraîne que $\sum_{i=1}^{n} \lambda_i\, l_i \in N^0$, en utilisant la propriété de base du semi-noyau, on peut dire que $\sigma \in X$ vérifiant $\langle \sigma, l_i \rangle = y_i$, $i = 1, \ldots, n$, est solution du problème, si et seulement s'il existe des coefficients λ_i, $i = 1, \ldots, n$, tels que $\sum_{i=1}^{n} \lambda_i\, l_i \in N^0$ et

$$(x, \sigma) = \left(x, H\left(\sum_{i=1}^{n} \lambda_i\, l_i\right)\right), \quad \text{pour tout} \quad x \in X\,.$$

Ceci nous conduit au théorème de caractérisation suivant qui exprime la F.S. à l'aide du semi-noyau.

THÉORÈME 1 : *Un élément* $\sigma \in X$ *vérifiant* $\langle \sigma, l_i \rangle = y_i$, $i = 1, \dots, n$, *est solution du problème* (3.10) *et* (3.11) *si et seulement si l'on a* :

$$\left.\begin{array}{l} \sigma = \sum_{i=1}^{n} \lambda_i H(l_i) + p\,, \quad \textit{avec} \quad p \in N \\ \textit{et des } \lambda_i \textit{ vérifiant} : \\ \sum_{i=1}^{n} \lambda_i \langle \omega, l_i \rangle = 0\,, \quad \textit{pour tout} \quad \omega \in N\,. \end{array}\right\} \tag{3.14}$$

d) *Calcul des fonctions-spline*

Si $\omega_1, \dots, \omega_q$ désigne une base de N et si l'on écrit la solution sous la forme :

$$\sigma = \sum_{i=1}^{n} \lambda_i H(l_i) + \sum_{j=1}^{q} \alpha_j \omega_j \tag{3.15}$$

les coefficients λ_i et α_j sont obtenus en résolvant le système linéaire de dimension $n + q$:

$$\left.\begin{array}{l} \sum_{i=1}^{n} \lambda_i \langle H(l_i), l_k \rangle + \sum_{j=1}^{q} \alpha_j \langle \omega_j, l_k \rangle = y_k \quad k = 1, \dots, n \\ \sum_{i=1}^{n} \lambda_i \langle \omega_j, l_i \rangle = 0\,, \quad j = 1, \dots, q \end{array}\right\} \tag{3.16}$$

Si l'on note $\mathscr{H}$ la matrice $n \times n$ de terme $\langle H(l_i), l_k \rangle$ et $\mathscr{E}$ la matrice $n \times q$ de terme $\langle \omega_j, l_k \rangle$, $k = 1, \dots, n$; $j = 1, \dots, q$, le système (3.13) s'écrit :

$$\left.\begin{array}{l} \mathscr{H}\lambda + \mathscr{E}\alpha = y \\ \mathscr{E}^t \lambda = 0\,. \end{array}\right\} \tag{3.17}$$

Appelons B le sous-espace vectoriel de dimension $n - q$ de $\mathbb{R}^n$ formé des λ vérifiant $\mathscr{E}^t \lambda = 0$ et soit $b_1, \dots, b_{n-q}$ une base de B. Formons la matrice $\mathscr{R}$ de dimension $n \times (n - q)$ de termes b_j^i, $i = 1, \dots, n$; $j = 1, \dots, n - q$.

Si λ est solution de (3.17), il appartient à B et on peut poser $\lambda = \mathscr{R}\mu$. Si l'on multiplie alors par $\mathscr{R}^t$ la première équation de (3.17), on obtient, compte tenu du fait que $\mathscr{R}^t \mathscr{E}\alpha = 0$:

$$\mathscr{R}^t \mathscr{H} \mathscr{R}\mu = \mathscr{R}^t y \tag{3.18}$$

où $\mathscr{A} = \mathscr{R}^t \mathscr{K} \mathscr{R}$ est une matrice carrée $(n - q) \times (n - q)$ symétrique et non singulière. On en déduit :

$$\mu = \mathscr{A}^{-1} \mathscr{R}^t y \tag{3.19}$$

et

$$\lambda = \Omega y \tag{3.20}$$

où $\Omega = \mathscr{R}\mathscr{A}^{-1} \mathscr{R}^t$ est une matrice $n \times n$ symétrique et semi définie positive.

Le système (3.18) est plus facile à résoudre que le système (3.14) directement. On utilisera par exemple une méthode de Choleski. Cette méthode de résolution a été décrite dans le cas des F.S. à deux variables de type plaques minces (provenant de la minimisation de la semi-norme définie dans l'exemple (ii) ci-dessus) par L. Paihua [6]. Elle est en fait identique à la méthode générale décrite de façon différente dans P. M. Anselone et P. J. Laurent [1]; voir aussi [5].

4. CARACTÉRISATION DES FONCTIONS SPLINE PAR INF-CONVOLUTION

Reprenons le problème des F.S. par inf-convolution du paragraphe 2 en le formulant dans le cadre général du paragraphe 3. Pour $x \in E$, posons :

$$\varphi(x) = \begin{cases} |x|^2, & \text{si} \quad x \in X \\ +\infty, & \text{sinon} \end{cases} \tag{4.1}$$

$$\psi(x) = \begin{cases} 0, & \text{si} \quad x \in V \\ +\infty, & \text{sinon} \end{cases} \tag{4.2}$$

où X est un sous-espace semi-hilbertien de E, avec la semi-norme $|\,.\,|$, et V un sous-espace vectoriel de dimension m de E.

Considérons le problème :

$$\text{minimiser} \quad \tilde{\varphi}(\tilde{x}) \tag{4.3}$$

avec les conditions

$$\tilde{x} \in E, \quad \langle \tilde{x}, l_i \rangle = z_i, \quad i = 1, \dots, n \tag{4.4}$$

avec $\tilde{\varphi} = \varphi \, \nabla \psi$.

Pour $\tilde{x} \notin \tilde{X} = X + V$, on a $\tilde{\varphi}(\tilde{x}) = +\infty$. Pour $\tilde{x} \in \tilde{X}$, $\tilde{\varphi}(\tilde{x})$ est donné par :

$$\tilde{\varphi}(\tilde{x}) = \operatorname*{Inf}_{\substack{x \in X, v \in V \\ x + v = \tilde{x}}} \varphi(x) = \operatorname*{Inf}_{x \in X \cap (\tilde{x} - V)} \varphi(x)\,. \tag{4.5}$$

La valeur $\tilde{\varphi}(\tilde{x})$ apparaît donc comme le montant d'un sous-problème de minimisation dont les solutions ont les propriétés suivantes que nous énonçons sans démonstration :

THÉORÈME 2 :

(i) *Le problème* (4.5) *a au moins une solution* $\dot{x}$ (*et* $\dot{v} = \tilde{x} - \dot{x}$)

(ii) *L'ensemble des solutions est alors donné par* $\dot{x} + N \cap V$

(iii) $\dot{x} \in X \cap (\tilde{x} - V)$ *est solution si et seulement si* $(\dot{x}, y) = 0$, *pour tout* $y \in V \cap X$.

On peut alors montrer que l'on peut définir sur $\tilde{X}$ un semi-produit scalaire pour lequel $\tilde{X}$ est un sous-espace semi-hilbertien de E et $\tilde{\varphi}$ est égal sur $\tilde{X}$ au carré de la semi-norme associée.

Pour $\tilde{x} = x + v \in \tilde{X}$, $x \in X$, $v \in V$ et

$\tilde{y} = y + w \in \tilde{X}$, $y \in X$, $w \in W$,

on posera :

$$(\tilde{x}, \tilde{y})_{\sim} = (\dot{x}, \dot{y}) \tag{4.6}$$

où $\dot{x}$ et $\dot{y}$ sont solutions du problème (4.5) selon le théorème 2, associés respectivement à $\tilde{x}$ et $\tilde{y}$.

On remarque d'abord que si $\tilde{y} = \dot{y} + \dot{w} = y + w$, avec $y \in X$ et $w \in V$ (y n'étant pas forcément solution du problème 4.5 associé à $\tilde{y}$) alors on a :

$$(\tilde{x}, \tilde{y})_{\sim} = (\dot{x}, y) \tag{4.7}$$

(en effet, d'après le théorème 2 (iii), comme $\dot{y} - y = w - \dot{w} \in V$, on a $(\dot{x}, \dot{y} - y) = 0$).

On a alors le résultat suivant :

THÉORÈME 3 : $\tilde{X}$ *est un sous-espace semi-hilbertien de* E *pour le semi-produit scalaire* $(\tilde{x}, \tilde{y})_{\sim}$, *et la semi-norme correspondante vérifie* :

$$|\tilde{x}|_{\sim}^2 = (\tilde{x}, \tilde{x})_{\sim} = \tilde{\varphi}(\tilde{x}). \tag{4.8}$$

Son espace nul $\tilde{N}$ *est égal à* $N + V$.

Démonstration : Par la définition même (4.6) de $(.,.)_{\sim}$ on a les propriétés de positivité et de symétrie d'un semi-produit scalaire. En utilisant (4.7), on voit que l'on a la linéarité par rapport au second argument et la linéarité par rapport au premier argument résulte alors de la symétrie.

Les propriétés générales de l'inf-convolution (*cf.* [5]) entraînent que dom $(\tilde{\varphi})$ (le domaine effectif de $\tilde{\varphi}$, c'est-à-dire l'ensemble des éléments de E pour lesquels φ ne prend pas la valeur $+\infty$) est égal à dom (φ) + dom $(\psi) = X + V = \tilde{X}$.

Maintenant, pour $\tilde{x} \in \tilde{X}$, on a :

$$\tilde{\varphi}(\tilde{x}) = \varphi(\dot{x}) = (\dot{x}, \dot{x}) = (\tilde{x}, \tilde{x})_{\sim} = |\, \tilde{x} \,|_{\sim}^{2} .$$

Pour $\tilde{x} = x + v$, avec $x \in N$ et $v \in V$, il est clair que $\tilde{\varphi}(\tilde{x}) = \varphi(x) = 0$.

Pour $\tilde{x} \notin N + V$, si l'on avait $\tilde{\varphi}(\tilde{x}) = 0$, cela signifierait, d'après le théorème 2, qu'il existerait $\dot{x} \in X$ et $\dot{v} \in Y$ vérifiant $\tilde{x} = \dot{x} + \dot{v}$ et $\varphi(\dot{x}) = 0$, donc $\dot{x} \in N$, ce qui est contradictoire. L'espace nul $\tilde{N}$ de $\tilde{\varphi}$ est donc égal à $N + V$.

Il reste à montrer que $\tilde{X}/\tilde{N}$ est complet et qu'il est topologiquement inclus dans E/N.

Considérons l'application L qui à $\tilde{x} \in \tilde{X}$ associe la classe dans X/N de l'un des éléments $\dot{x}$ qui minimise $\varphi(x)$ sur $X \cap (\tilde{x} - V)$ (problème (4.5)). D'après le théorème 2 (ii) deux solutions de ce problème ont pour différence un élément de $N \cap \dot{V}$, donc appartiennent à la même classe. On vérifie facilement que L est une application linéaire de X dans X/N, que son espace nul est égal à $N + V$ et que son image Y dans X/N est fermée (d'après le théorème 2 (iii), c'est en effet l'orthogonal dans X/N de $V \cap X + N$) donc complète pour la topologie induite. De plus, on a :

$$(\tilde{x}) = \varphi(\dot{x}) = \| L\tilde{x} \|_{X/N} . \tag{4.9}$$

On note $\tilde{C}(\tilde{x})$ la classe de $\tilde{x}$ dans $\tilde{X}/(N + V)$. On peut donc définir une application $\hat{L}$ de $\tilde{X}/(N + V)$ sur Y par $\hat{L}(\tilde{C}(\tilde{x})) = L(\tilde{x})$ et il est clair que $\hat{L}$ est une bijection.

Comme :

$$\| \tilde{C}(\tilde{x}) \|_{\tilde{X}/(N+V)} = \| L(\tilde{x}) \|_{X/N} \tag{4.10}$$

on en déduit que L est aussi bicontinue. Il en résulte que $\tilde{X}/(N + V)$ est bien complet.

Montrons maintenant que $\tilde{X}/\tilde{N}$ est topologiquement inclus dans E/N : supposons que l'on ait $\lim_{n\to\infty} \tilde{\varphi}(\tilde{x}_n) = 0$. Il existe une décomposition $\tilde{x}_n = \dot{x}_n + \dot{v}_n$ selon le théorème 2 et $\varphi(\dot{x}_n) = \tilde{\varphi}(\tilde{x}_n)$ tend vers 0 lorsque n tend vers l'infini. Soit $l \in E'$ nulle sur $\tilde{N} = N + V$, donc en particulier nulle sur N et sur V. On a donc, puisque X/N est topologiquement inclus dans E/N : $\lim_{n\to\infty} \langle \dot{x}_n, l \rangle = 0$, et comme $\langle \dot{v}_n, l \rangle = 0$, on en déduit $\lim_{n\to\infty} \langle \tilde{x}_n, l \rangle = 0$.

En général, il n'existe pas un semi-noyau *unique* associé à un sous-espace semi-hilbertien $X \subset E$ et à $L \subset E'$. Mais on va voir que tout semi-noyau pour

X est un semi-noyau pour $\tilde{X}$ (associé au même $L \subset E'$), ce qui permettra d'expliciter les F.S. définies par inf-convolution.

THÉORÈME 4 : *Soit $L \subset E'$. Si l'application linéaire H de L dans E est un semi-noyau pour X muni de $(.,.)$, alors H est aussi un semi-noyau pour $\tilde{X}$ muni de* $(.,.)_{\sim}$.

Démonstration : On sait que l'espace nul de la semi-norme $|.|_{\sim}$ est égal à $N + V$. Soit $l \in L \cap (N + V)^0$. Nous avons déjà $H(l) \in X \subset \tilde{X}$. Soit $\tilde{x} = x + v \in \tilde{X}$. Décomposons $\tilde{x}$ en $\tilde{x} = \dot{x} + \dot{v}$, où $\dot{x}$ réalise le minimum de $\varphi(x)$ pour $x \in X \cap (\tilde{x} - V)$ selon le théorème 2. Comme $\tilde{x} - \dot{x} \in V$, on a :

$$\langle \tilde{x}, l \rangle = \langle \dot{x}, l \rangle .$$

Comme $\dot{x} \in H$ et H est un semi-noyau pour X, on a :

$$\langle \dot{x}, l \rangle = (\dot{x}, H(l)) .$$

Enfin, d'après la remarque (4.7) en tenant compte du fait que $H(l) \in X$, on a :

$$(\tilde{x}, H(l))_{\sim} = (\dot{x}, H(l))$$

ce qui montre que $\langle \tilde{x}, l \rangle = (\tilde{x}, H(l))_{\sim}$, pour tout $\tilde{x} \in \tilde{X}$.

La résolution du problème de la minimisation de $\tilde{\varphi}(\tilde{x})$ avec $\tilde{x} \in \tilde{X}$ vérifiant les conditions $\langle \tilde{x}, l_i \rangle = z_i$, $i = 1, \dots, n$ devient alors une application directe du théorème 1, en remarquant que seul l'espace nul N est remplacé par $N + V$:

THÉORÈME 5 : *Si*

$$(N + V) \cap \{ x \in E \mid \langle x, l_i \rangle = 0, \quad i = 1, \dots, n \} = \{ 0 \} \qquad (4.11)$$

alors il existe une solution unique $\tilde{\sigma}$ qui est donnée par

$$\tilde{\sigma} = \sum_{i=1}^{n} \lambda_i H(l_i) + p + v \quad \textit{avec} \quad p \in N \quad \textit{et} \quad v \in V \qquad (4.12)$$

où les λ_i, $i = 1, \dots, n$, p et v vérifient les conditions suivantes :

$$\left.\begin{array}{l} \sum_{i=1}^{n} \lambda_i \langle q, l_i \rangle = 0, \quad \textit{pour tout } q \in N \\ \sum_{i=1}^{n} \lambda_i \langle w, l_i \rangle = 0, \quad \textit{pour tout } w \in V \\ \langle \tilde{\sigma}, l_i \rangle = z_i, \qquad i = 1, \dots, n \end{array}\right\} \qquad (4.13)$$

La détermination de $\tilde{\sigma}$ revient donc à la résolution d'un système linéaire de dimension $n + q + m$ (q étant la dimension de N et m celle de V). On reviendra sur cette question au paragraphe suivant.

Exemples : Reprenons l'exemple de l'espace $X_{m,\alpha} \subset \mathscr{C}(\mathbb{R}^d) = E$ et prenons $L = E'$. Nous considérons l'interpolation simple c'est-à-dire le cas où les l_i sont définies par

$$\langle x, l_i \rangle = \langle x, \delta_{t_i} \rangle = x(t_i), \quad i = 1, \dots, n.$$

Dans le théorème précédent, la détermination de la valeur en t de la solution $\tilde{\sigma}$, c'est-à-dire $\langle \tilde{\sigma}, \delta_t \rangle$, suppose la connaissance de $\langle H(\delta_{t_i}), \delta_t \rangle$. On est conduit à considérer la fonction K de $\mathbb{R}^d \times \mathbb{R}^d$ dans $\mathbb{R}$ défini par :

$$K(t, t') = \langle H(\delta_{t'}), \delta_t \rangle \tag{4.14}$$

appelée *noyau semi-reproduisant*. La solution s'écrit alors :

$$\tilde{\sigma}(t) = \sum_{i=1}^{n} \lambda_i K(t, t_i) + \sum_{j=1}^{q} \alpha_j \omega_j(t) + \sum_{j=1}^{m} d_j p_j(t) \tag{4.15}$$

où $\omega_j, j = 1, \dots, q$, désigne une base de N et
p_j, $j = 1, \dots, m$, une base de V.

Remarque : Comme le semi-noyau H n'est pas unique, il en est de même du noyau semi-reproduisant K. De plus, contrairement à la notion classique de noyau reproduisant, K n'est pas forcément symétrique et positif. Toutefois quels que soient les abscisses $t_1, \dots, t_m$ et les coefficients $\lambda_1, \dots, \lambda_m$ vérifiant $\sum_{i=1}^{m} \lambda_i \delta_{t_i} \in N^0$, et quels que soient les abscisses $u_1, \dots, u_n$ et les coefficients $\mu_1, \dots, \mu_n$ vérifiant $\sum_{j=1}^{n} \mu_j \delta_{u_j} \in N^0$, on a, d'après (3.7) et (3.8) :

$$\sum_{\substack{i=1,\dots,m \\ j=1,\dots,n}} \lambda_i \mu_j K(t_i, u_j) = \sum_{\substack{i=1,\dots,m \\ j=1,\dots,n}} \lambda_i \mu_j K(u_j, t_i) \quad \text{(semi-symétrie)} \tag{4.16}$$

$$\sum_{\substack{i=1,\dots,m \\ k=1,\dots,m}} \lambda_i \lambda_k K(t_i, t_k) \geqslant 0 \quad \text{(semi-positivité)}. \tag{4.17}$$

L'intérêt essentiel de la notion de noyau semi-reproduisant réside dans le fait que l'un des noyaux de $X_{p,\alpha}$ a une expression simple qui a été donnée par J. Duchon :

$$K(t, t') = \begin{cases} k \, \| t - t' \|^\theta \operatorname{Log} \| t - t' \| , & \text{si } \theta \text{ est pair} \\ k' \, \| t - t' \|^\theta , & \text{si } \theta \text{ n'est pas pair} \end{cases} \tag{4.18}$$

où $\theta = 2p + 2\alpha - d$ ($\| \cdot \|$ désigne la norme Euclidienne dans $\mathbb{R}^d$).

La constante k (ou k') dépend de m, α et d, mais il n'est pas nécessaire de la calculer. Sa valeur peut être incluse dans les coefficients λ_i. C'est ce que nous supposerons dans la suite sans employer une nouvelle notation.

Donnons quelques valeurs à d, m et α :

1) $d = 1, p = 2, \alpha = 0\ (\theta = 3); p_j(t) = (t - \alpha_j)_+$: on obtient .

$$\tilde{\sigma}(t) = \sum_{i=1}^{n} \lambda_i \mid t - t_i \mid^3 + \alpha_1 + \alpha_2 t + \sum_{j=1}^{m} d_j(t - \alpha_j)_+$$

où les coefficients λ_i vérifient :

$$\sum_{i=1}^{n} \lambda_i (t_i)^k = 0, \qquad k = 0, 1$$

$$\sum_{i=1}^{n} \lambda_i (t_i - \alpha_j)_+ = 0, \quad j = 1, \dots, m$$

On vérifie facilement que ces conditions sont équivalentes aux conditions (1.6)(i) à (iv) données sur s.

2) $d = 1, p = 3, \alpha = 0\ (\theta = 5); p_j(t) = (t - \alpha_j)_+$: on obtient :

$$\tilde{\sigma}(t) = \sum_{i=1}^{n} \lambda_i \mid t - t_i \mid^5 + \alpha_1 + \alpha_2 t + \alpha_3 t^2 + \sum_{j=1}^{m} d_j(t - \alpha_j)_+$$

où les coefficients λ_i vérifient :

$$\sum_{i=1}^{n} \lambda_i (t_i)^k = 0, \qquad k = 0, 1, 2$$

$$\sum_{i=1}^{n} \lambda_i (t_i - \alpha_j)_+ = 0, \quad j = 1, \dots, m.$$

Ces conditions sont équivalentes à celles données en (1.7) (i) à (iv) pour la fonction s. On pourrait de même retrouver la condition (1.11) (iv') en introduisant un sous-espace V de dimension $2\,m$ engendré par $p_j(t) = (t - t_j)_+$ et $q_j(t) = \dfrac{(t - t_j)_+^2}{2}$.

3) $d = 2, p = 2, \alpha = 0\ (\theta = 2)$;

$t = (t_1, t_2) \in \mathbb{R}^2$

$t^i = (t_1^i, t_2^i) \in \mathbb{R}^2, \quad i = 1, \dots, n$

$$\tilde{\sigma}(t) = \sum_{i=1}^{n} \lambda_i \| t - t^i \|^2 \operatorname{Log} \| t - t^i \| + \alpha_1 + \alpha_2 t_1 + \alpha_3 t_2 + \sum_{j=1}^{m} d_j p_j(t)$$

où les λ_i vérifient les conditions :

$$\sum_{i=1}^{n} \lambda_i = 0\,, \quad \sum_{i=1}^{n} \lambda_i\, t_1^i = 0\,, \quad \sum_{i=1}^{n} \lambda_i\, t_2^i = 0$$

$$\sum_{i=1}^{n} \lambda_i\, p_j(t^i) = 0\,, \quad j = 1, \dots, m\,.$$

4) $d = 2, p = 2, \alpha = 0{,}5$ $(\theta = 3)$

$$\tilde{\sigma}(t) = \sum_{i=1}^{n} \lambda_i \,\| t - t^i \|^3 + \alpha_1 + \alpha_2\, t_1 + \alpha_3\, t_2 + \sum_{j=1}^{m} d_j\, p_j(t)$$

où les λ_i vérifient les mêmes conditions qu'au 3).

Dans les exemples 3) et 4) on n'a pas précisé la forme des fonctions p_j. Cela dépend évidemment du type de « singularité » dont on veut tenir compte. Cette question peut se révéler délicate ; nous ne l'aborderons pas ici.

5. ALGORITHMES

Nous aborderons d'abord le cas de l'exemple introductif (1.6) (le cas (1.7) ou (1.11) où des exemples similaires se traiteraient de la même façon), pour lequel des méthodes très rapides (de coût identique aux méthodes de calcul des F.S. cubiques classiques) peuvent être mises en œuvre. Pour simplifier l'exposé, nous supposerons qu'entre deux abscisses t_i successives on peut trouver au plus un seul α_j. La méthode dite « de transport de relations » peut être adaptée de la façon suivante : supposons qu'en t_i on ait une relation linéaire liant $s'(t_i)$ et $s''(t_i)$:

$$a_i\, s'(t_i) + b_i\, s''(t_i) = c_i\,. \tag{5.1}$$

Notons que pour $i = 1$, on dispose d'une telle relation $s''(t_1) = 0$ (c'est-à-dire $a_1 = 0, b_1 = 1, c_1 = 0$).

En tenant compte de toutes les informations que l'on a sur s, on peut « transporter » cette relation en t_{i+1}. Deux cas sont possibles :

Premier cas : Il n'y a pas de α_j entre t_i et t_{i+1} : on a alors les formules habituelles concernant les F.S. cubiques ordinaires :

$$a_{i+1} = a_i - 3\,\frac{b_i}{h_i}$$

$$b_{i+1} = b_i - \frac{a_i\, h_i}{4}$$

$$c_{i+1} = -\,\frac{c_i}{2} + 3\left(\frac{z_{i+1} + z_i}{h_i}\right)\left(\frac{a_i}{2} - \frac{b_i}{h_i}\right)$$

avec $h_i = t_{i+1} - t_i$. Il y aura lieu de faire une normalisation, par exemple de multiplier a_{i+1}, b_{i+1} et c_{i+1} par une constante convenable de sorte que la somme des valeurs absolue, soit égale à 1 (ou toute autre normalisation plus économique, *cf.* Paihua [6]).

Deuxième cas : il y a un α_j entre t_i et t_{i+1}. Si l'on pose $p = \dfrac{\alpha_j - t_i}{h_i}$ et $q = \dfrac{t_{i+1} - \alpha_j}{h_i}$, on a alors :

$$\begin{aligned} a_{i+1} &= q^2 a_i \\ b_{i+1} &= p^2 b_i - \frac{a_i h_i}{3}(1 - 3pq) \\ c_{i+1} &= -pqc_i + qa_i\left(\frac{z_{i+1} - z_i}{h_i}\right). \end{aligned} \tag{5.3}$$

On effectuera de même une normalisation.

On peut ainsi de proche en proche « transporter » la relation $s''(t_1) = 0$ jusqu'en t_n. En sens inverse, avec des formules analogues, on transporte $s''(t_n) = 0$ jusqu'en t_1. On obtient ainsi en chaque t_i, un système linéaire de dimension 2 qui fournit $s'(t_i)$ et $s''(t_i)$.

Cas général : la méthode précédente ne s'applique pas dans le cas de F.S. à plusieurs variables, ou même à une variable pour des choix plus généraux de fonctions p_j. Comme on dispose en général de procédures performantes pour le calcul de F.S. ordinaires (à une ou plusieurs variables), la méthode que nous allons décrire consiste à ramener le calcul des F.S. par inf-convolution à plusieurs calculs de F.S. ordinaires. On suppose dans la suite que $V \cap N = \{0\}$.

Parmi tous les éléments de la forme $\tilde{x} = x + \sum_{j=1}^{m} d_j p_j \in \tilde{X}$ qui vérifient $\langle \tilde{x}, l_i \rangle = z_i$, $i = 1, \dots, n$, on cherche l'élément $\tilde{\sigma} = \bar{x} + \sum_{j=1}^{m} \bar{d}_j p_j$ qui minimise la quantité $| x |^2$.

On peut d'abord effectuer la minimisation par rapport à $x \in X$ (pour $d_1, \dots, d_m$ fixés) puis minimiser le résultat par rapport à $d_1, \dots, d_m$.

Si l'on pose :

$$G(d) = \min | x |^2 \tag{5.4}$$

avec $x \in X$ et $\langle x, l_i \rangle = z_i - \sum_{j=1}^{m} d_j \langle p_j, l_i \rangle$, $i = 1, \dots, n$, $i = 1, \dots, n$, le problème revient ensuite à minimiser $G(d)$ par rapport à $d \in \mathbb{R}^m$. Notons :

$$\phi(y) = \min_{\substack{x \in X \\ \langle x, l_i \rangle = y_i \\ i = 1, \dots, n}} |x|^2 = |\sigma|^2 \tag{5.5}$$

le montant du problème de F.S. d'interpolation ordinaire (3.6), (3.7) correspondant aux données $y = (y_1, \dots, y_n)$. En utilisant l'expression (3.12) de la solution σ et les propriétés (3.9) du semi-noyau H, on voit que :

$$\begin{aligned} \phi(y) &= \left(\sum_{i=1}^{n} \lambda_i H(l_i), \sum_{i=1}^{n} \lambda_i H(l_i) \right) \\ &= \left\langle \sum_{i=1}^{n} \lambda_i H(l_i), \sum_{i=1}^{n} \lambda_i l_i \right\rangle \\ &= \sum_{i,j=1}^{n} \lambda_i \lambda_j \langle H(l_i), l_j \rangle = \langle \lambda, \mathscr{K} \lambda \rangle_n \end{aligned}$$

où $\lambda = (\lambda_1, \dots, \lambda_n)$ est associé à y selon (3.17). On en déduit facilement que :

$$\phi(y) = \langle y, \Omega y \rangle_n \tag{5.6}$$

donc que ϕ est une fonction convexe quadratique.

Si l'on note $\mathscr{P}$ la matrice $n \times m$ qui a pour terme $\langle p_j, l_i \rangle$, $j = 1, \dots, m$; $i = 1, \dots, n$, on voit que :

$$G(d) = \phi(z - \mathscr{P} d) = \langle z - \mathscr{P} d, \Omega(z - \mathscr{P} d) \rangle_n . \tag{5.7}$$

Ainsi, G est une fonction convexe dont le gradient est donné par :

$$G'(d) = 2 \mathscr{P}^t \Omega(\mathscr{P} d - z) . \tag{5.8}$$

La solution d est déterminée par $G'(d) = 0$, c'est-à-dire par :

$$(\mathscr{P}^t \Omega \mathscr{P}) d = \mathscr{P}^t \Omega z \tag{5.9}$$

qui est un système linéaire $m \times m$ dont la matrice $\mathscr{P}^t \Omega \mathscr{P}$ est symétrique non singulière.

Le système (5.9) peut être interprété de la façon suivante : dans $\mathbb{R}^n$ on définit le semi-produit scalaire

$$\langle y_1, y_2 \rangle_\Omega = \langle y_1, \Omega y_2 \rangle_n .$$

Il s'agit alors de trouver la meilleure approximation de z dans le sous-espace $\tilde{V}$ de dimension m constitué des éléments de la forme $\mathscr{P} d'$, avec $d' \in \mathbb{R}^m$. La solution d est alors caractérisée par le fait que $\mathscr{P}d - z$ doit être orthogonal au sens de $\langle\, , \rangle_\Omega$ à tout élément de la forme $\mathscr{P}d'$, avec $d' \in \mathbb{R}^m$, donc :

$$\langle \mathscr{P}d', \Omega(\mathscr{P}d - z) \rangle_n = 0\,, \quad \text{pour tout} \quad d' \in \mathbb{R}^m$$

soit

$$\langle d', \mathscr{P}^t\, \Omega(\mathscr{P}d - z) \rangle_n = 0\,, \quad \text{pour tout} \quad d' \in \mathbb{R}^m$$

ce qui donne (5.9).

Il n'est pas nécessaire de calculer la matrice Ω. Pour calculer $\tilde{\lambda} = \Omega z$, il suffit de calculer la F.S. d'interpolation ordinaire associée au vecteur z en résolvant le système (3.14), ce qui donne le vecteur $\tilde{\lambda}$ correspondant. De la même façon, chacune des m colonnes de la matrice $n \times m$: $Q = \Omega\mathscr{P}$ peut être obtenue en calculant le λ de la F.S. associée à cette colonne, c'est-à-dire dont le y est égal à cette colonne.

Ainsi, en résumé, le calcul de la F.S. par inf-convolution associée à un sous-espace V de dimension m coûte $m + 1$ calculs de F.S. d'interpolation ordinaires, pour la constitution du système (5.9), plus la résolution de ce système. Le vecteur λ vérifiant (4.13) est ensuite obtenu facilement par :

$$\lambda = \Omega(z - \mathscr{P}d) = \tilde{\lambda} - Qd. \tag{5.10}$$

6. RÉSULTATS NUMÉRIQUES

Plusieurs exemples ont été traités. Nous donnerons seulement les résultats relatifs à la fonction f définie sur l'intervalle [0, 3] par :

$$f(t) = \begin{cases} \sqrt{1 - t^2} & , \quad \text{si} \quad t \in \left[0, \dfrac{\sqrt{2}}{2}\right] \\ \sqrt{1 - (t - \sqrt{2})^2}, & \quad \text{si} \quad t \in \left[\dfrac{\sqrt{2}}{2}, \dfrac{3\sqrt{2}}{2}\right] \\ t - \sqrt{2} & , \quad \text{si} \quad t \in \left[\dfrac{3\sqrt{2}}{2}, 3\right] \end{cases}$$

qui a une discontinuité de dérivée première égale à 2 aux abscisses $\alpha_1 = \dfrac{\sqrt{2}}{2}$ et $\alpha_2 = \dfrac{3\sqrt{2}}{2}$. On prendra les abscisses,

$$t_i = 3\,\frac{(i - 1)}{n - 1} \quad \text{et} \quad z_i = f(t_i)\,, \quad i = 1, \dots, n\,.$$

Six méthodes ont été testées :

a) F.S. cubique

$$x(t) + \sum_{j=1}^{2} d_j(t - \alpha_j)_+ ; \quad cf.\,(1.6).$$

b) F.S. cubique

$$x(t) + \sum_{j=1}^{2} d_j(t - \alpha_j)_+ + \sum_{j=1}^{2} e_j \frac{(t - \alpha_j)_+^2}{2}.$$

c) F.S. cubique

$$x(t) + \sum_{j=1}^{2} d_j\, p_j(t),$$

où p_j sont des approximations locales autour de α_j de f formées de deux polynômes de degré 2.

d) F.S. quintique

$$x(t) + \sum_{j=1}^{2} d_j(t - \alpha_j)_+ ; \quad cf.\,(1.7).$$

e) F.S. quintique

$$x(t) + \sum_{j=1}^{2} d_j(t - \alpha_j)_+ + \sum_{j=1}^{2} e_j \frac{(t - \alpha_j)_+^2}{2} ; \quad cf.\,(1.11).$$

f) F.S. quintique

$$x(t) + \sum_{j=1}^{2} d_j\, p_j(t),$$

où p_j est choisi comme en *c*).

Pour $n = 100$, on obtient comme évaluations des discontinuités de dérivée première en α_1 et α_2 (qui sont égales à 2 pour f) les résultats suivants :

Méthode	Discontinuités en α_1	Discontinuités en α_2
a	1,890	1,975
b	1,914	1,987
c	1,666	1,884
d	1,992	1,998
e	1,991	1,998
f	1,992	1,998

On constate que les résultats obtenus avec les F.S. quintiques sont nettement meilleurs que ceux obtenus avec les cubiques (ce phénomène a été similaire pour les autres exemples). La raison pourrait être que dans le cas des cubiques, la minimisation introduit des conditions de type « bout libre » en α_j (par exemple $\sigma''(\alpha_j) = 0$ dans le cas *a*)) qui ne sont pas vérifiées par f. Cela produit des oscillations de σ'' autour de α_j et une imprécision sur l'estimation du saut de f' Dans le cas des quintiques, on a aussi de telles conditions, mais elles concernent des dérivées d'ordre plus élevé ($\sigma^{(4)}(\alpha_j) = 0$, dans le cas *d*)) et ainsi elles affectent moins le résultat. On peut aussi remarquer que dans les cas *b*) et *c*) il y a interférence entre le critère de lissage que l'on minimise et le sous-espace V traduisant les discontinuités : en effet l'espace nul de $\int (x''(t))^2\, dt$ est formé des polynômes de degré 1 alors que les éléments de V sont localement de degré 2 donc ont une « énergie » non nulle. Ce phénomène n'a pas lieu dans le cas *a*) ainsi que dans les trois méthodes utilisant les quintiques. Enfin, remarquons que les méthodes *e*) et *f*) permettent d'introduire une discontinuité de la dérivée seconde (alors que σ'' est continue en α_j pour la méthode *d*)) : ceci ne s'est toutefois pas traduit par une supériorité nette de ces deux méthodes par rapport à la méthode *d*). (Notons que la dérivée seconde de la fonction f est continue en α_1 mais qu'elle a une discontinuité égale à $2\sqrt{2}$ en α_2).

Le tableau suivant donne l'évolution des résultats en fonction de n dans le cas de la méthode *d*) (l'évolution est similaire pour les autres méthodes) :

n	Discontinuités en 1	Discontinuités en 2
50	1,9614	2,0319
100	1,9919	1,9980
200	1,9989	2,0099
400	1,9997	1,9972

On voit que l'évolution n'est pas monotone. Cela est dû au fait que le résultat est sensible à la proximité d'un point t_i par rapport à α_j.

Évaluation des α_j : Nous supposons maintenant que les positions exactes des points de discontinuité de f ne sont pas connues mais seulement leurs localisations (par exemple que la première discontinuité de la fonction f précédente se trouve dans l'intervalle [0,65 ; 0,75]. On choisit alors des $\alpha_j^{(0)}$ de départ, par exemple le milieu de l'intervalle, et on minimise l'énergie ($E^{(2)}_{[a,b]}(x)$ ou $E^{(3)}_{[a,b]}(x)$ selon les cas) par rapport aux α_j. On obtient ainsi des approximations α_j^* des points de discontinuité de f'. Sur tous les exemples traités on a constaté que l'on pouvait minimiser *successivement* l'énergie par rapport à chacun des α_j. Si l'on minimise l'énergie par rapport à l'un des α_j, la position optimale obtenue dépend en effet fort peu du choix même très approximatif des autres α_j. D'autre part on a aussi constaté sur tous les exemples traités que cette énergie était une fonction localement convexe dans un intervalle assez large autour de l'optimum, ce qui facilite grandement le calcul des α_j^*. Pour la première discontinuité de f' qui se trouve en $\frac{\sqrt{2}}{2} = 0{,}707106$, on a obtenu avec la méthode *a)* les résultats suivants :

n	α_1^*	Erreur
50	0,707506	0,000400
400	0,707050	− 0,000056

On constate que la précision obtenue est excellente, même avec un petit nombre de points.

Les résultats sont encore meilleurs avec les quintiques ; par exemple pour la méthode *d)* avec $n = 50$, on obtient déjà $\alpha_1^* = 0{,}707148$, c'est-à-dire une erreur de 0,000042.

Il semble donc plus facile d'obtenir une bonne précision sur la position de la discontinuité que sur son amplitude.

L'auteur tient à remercier Claire Di Crescenzo pour sa contribution efficace dans l'obtention de ces résultats numériques.

REFERENCES

[1] P. M. ANSELONE and P. J. LAURENT, *A general method for the construction of interpoling or smoothing spline-functions.* Num. Math. *12* (1968), 66-82.

[2] J. DUCHON, *Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces.* RAIRO, Analyse Numérique, *10*, n° 12 (1976), 5-12.

[3] J. DUCHON, *Splines minimizing rotation-invariant semi-norms in Sobolev spaces.* Lecture Notes in Math. 571, W. Schempp et K. Zeller Ed. (1977), 85-100.

[4] J. DUCHON, *Fonctions spline homogènes à plusieurs variables.* Thèse (Grenoble) (1980).

[5] P. J. LAURENT, *Approximations et Optimisation.* Hermann (1972), Paris.

[6] L. PAIHUA, *Quelques méthodes numériques pour les fonctions spline à une ou deux variables.* Thèse (Grenoble), (mai 1978).

[7] R. T. ROCKAFELLAR, *Convex Analysis.* Princeton University Press (1970).

[8] F. UTRERAS, *Utilisation de la méthode de la validation croisée pour le lissage par fonction-spline à une ou deux variables.* Thèse (Grenoble), (mai 1979).