

## Sur le choix du paramètre d'ajustement dans le lissage par fonctions spline

F. Utreras Diaz \*

Université Scientifique et Médicale Laboratoire de Mathématiques Appliquées,  
B.P. 53, 38041 Grenoble, France

### On the Choice of the Smoothing Parameter in the Smoothing of Noisy Data by Spline Functions

**Summary.** We consider the problem of approximating an unknown function  $f$ , known with error at  $n$  equally spaced points of the real interval  $[a, b]$ .

To solve this problem, we use the natural polynomial smoothing splines. We show that the eigenvalues associated to these splines converge to the eigenvalues of a differential operator and we use this fact to obtain an algorithm, based on the Generalized Cross Validation method, to calculate the smoothing parameter.

With this algorithm, we divide by  $n$  the time used by classical methods.

*Subject Classifications:* AMS(MOS): 65D10, 65D07.

### Introduction

Considérons un intervalle réel  $[a, b]$  et soit  $f$  une fonction de l'espace de Sobolev  $H^q[a, b]$ , ( $q$  entier positif).

Notre but dans cet article est de trouver une approximation de  $f$ , à partir de la connaissance de ses valeurs entachées d'erreurs en  $n$  points, c'est-à-dire, en connaissant:

$$z_i = f(t_i) + \varepsilon_i \quad i = 1, 2, \dots, n$$

où

$$\begin{aligned} t_{i+1} - t_i &= h & i = 1, 2, \dots, n-1 \\ t_i &\in [a, b]. \end{aligned} \tag{1.1}$$

\* Adresse actuelle: Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Matemáticas, Casilla 5272 Correo 3, Santiago, Chile

Dans le cas où les  $\varepsilon_i$  peuvent être considérées comme des variables aléatoires indépendantes telles que:

$$\begin{aligned} E[\varepsilon_i] &= 0 \\ E[\varepsilon_i^2] &= v^2 \end{aligned} \quad i=1, 2, \dots, n \quad (1.2)$$

(où  $E[.]$  dénote l'espérance mathématique).

Pour approcher  $f$ , on se propose d'utiliser la fonction spline d'ajustement d'ordre  $q$  et de paramètre  $\rho$  associée à  $z$ , c'est-à-dire la fonction  $S_\rho$  ( $n$  doit être supérieur à  $q$ ) qui est la solution unique du problème:

$$\text{Inf}_{g \in H^q[a, b]} \left\{ \int_a^b [g^{(q)}(t)]^2 dt + \rho \sum_{i=1}^n (g(t_i) - z_i)^2 \right\} \quad (1.3)$$

Il reste à résoudre le problème du choix de  $\rho$ , le paramètre réalisant le compromis entre le lissage et l'approximation des données. Ce choix est extrêmement important: une valeur trop grande de  $\rho$  produit une spline qui suit trop les données et du fait des erreurs, conduit à une mauvaise approximation de  $f$ , alors qu'une valeur trop faible de  $\rho$  produit un lissage trop marqué qui risque d'éliminer les variations de la fonction.

En vue d'obtenir un procédé complètement automatique du choix de  $\rho$ , Wahba ([17, 13, 15]) a proposé la méthode de Validation Croisée Généralisée (Generalized Cross Validation) qui consiste à exploiter une idée apparue pour la première fois chez Stone [11]. Cette méthode réduit le problème du choix de  $\rho$  à celui de la minimisation de la fonction:

$$V(\rho) = \frac{1}{n} \frac{\sum_{i=1}^n [S_\rho(t_i) - z_i]^2}{\left(1 - \frac{1}{n} \text{tr}(A(\rho))\right)^2} \quad (1.4)$$

$A(\rho)$  étant la matrice  $n \times n$  qui à  $z \in \mathbb{R}^n$  associe le vecteur des valeurs de  $S_\rho$  aux points  $t_i$ ,  $i=1, \dots, n$ .

Récemment, Craven et Wahba [4] ont étudié les propriétés de ce choix de  $\rho$  et ont obtenu de très bons résultats numériques (du point de vue de l'approximation de  $f$ ); de plus ils ont donné plusieurs arguments heuristiques permettant d'envisager quelques résultats d'optimalité. Malgré ces bons résultats, la méthode restait inapplicable, son coût étant trop élevé. Le but de cet article est précisément d'éliminer cet inconvénient. On propose un algorithme permettant de diviser par  $n$  le temps de calcul de  $\rho$ .

## 2. Quelques résultats classiques

Rappelons d'abord quelques notations et propriétés classiques dans la théorie des fonctions spline.

$$H^q[a, b] = \left\{ f/f, f', \dots, f^{(q-1)} \text{ absolument continues, } \int_a^b [f^{(q)}(t)]^2 dt < +\infty \right\}$$

$$H^0[a, b] \equiv L^2[a, b].$$

$S$  sera l'espace des fonctions spline d'ordre  $q$ , c'est-à-dire l'ensemble des fonctions  $\sigma$ , telles que:

- (i)  $\sigma$  est un polynôme de degré  $q-1$  dans chacun des intervalles  $[a, t_1]$ ,  $[t_n, b]$ ,
- (ii)  $\sigma$  est un polynôme de degré  $2q-1$  dans chacun des intervalles  $[t_i, t_{i+1}]$ ,  $i = 1, 2, \dots, n-1$ ,
- (iii)  $\sigma \in C^{2q-2}[a, b]$ .

Il est bien connu que  $S$  est un espace vectoriel de dimension  $n$ , de plus, si  $x \in \mathbb{R}^n$ , il existe une et une seule fonction  $s \in S$  telle que:

$$s(t_i) = x_i \quad i = 1, 2, \dots, n \quad (2.1)$$

$s$  peut aussi s'écrire:

$$s = \sum_{i=1}^n x_i \sigma_i \quad (2.2)$$

où  $\{\sigma_1, \sigma_2, \dots, \sigma_n\}$  est la base canonique de  $S$ , avec:

$$\sigma_i(t_j) = \begin{cases} 1 & j=i \\ 0 & j \neq i \end{cases} \quad i, j = 1, 2, \dots, n. \quad (2.3)$$

On sait que cette fonction  $s$  est la solution unique du problème de la minimisation de  $\int_a^b [g^{(q)}(t)]^2 dt$  avec les contraintes

$$g(t_i) = x_i \quad i = 1, 2, \dots, n.$$

*Définition.* Pour  $\varphi, \Psi \in H^q[a, b]$

$$P(\varphi, \Psi) = \int_a^b \varphi^{(q)}(t) \Psi^{(q)}(t) dt$$

$$p(\varphi) = P(\varphi, \varphi). \quad (2.4)$$

$P$  induit sur  $\mathbb{R}^n$  une forme bilinéaire  $B(x, y)$  par la relation:

$$B(x, y) = P \left( \sum_{i=1}^n x_i \sigma_i, \sum_{i=1}^n y_i \sigma_i \right). \quad (2.5)$$

Notons  $\Omega$  la matrice  $n \times n$  associée à cette forme bilinéaire, on peut démontrer (cf. Laurent [7]) que les composantes  $\omega_{ij}$  de  $\Omega$  sont données par:

$$\omega_{ij} = (-1)^q (\sigma_i^{(2q-1)}(t_j^+) - \sigma_i^{(2q-1)}(t_j^-)). \quad (2.6)$$

Pour

$$s = \sum_{i=1}^n x_i \sigma_i.$$

On a donc:

$$p(s) = \int_a^b [s^{(q)}(t)]^2 dt = \langle x, \Omega x \rangle.$$

La détermination de la fonction spline d'ajustement  $s_\rho$  peut alors être ramené à un problème de minimisation dans  $\mathbb{R}^n$

$$\begin{aligned} \text{Min}_{x \in \mathbb{R}^n} \left[ \left( \text{Min}_{\substack{g(t_i)=x_i \\ i=1,2,\dots,n}} \int_a^b [g^{(q)}(t)]^2 dt \right) + \rho \sum_{i=1}^n (x_i - z_i)^2 \right] \\ = \text{Min}_{x \in \mathbb{R}^n} \left[ \langle x, \Omega x \rangle + \rho \sum_{i=1}^n (x_i - z_i)^2 \right] \end{aligned} \quad (2.7)$$

et la solution  $y_i = s_\rho(t_i)$ ,  $i = 1, \dots, n$ , vérifie:

$$\Omega y + \rho(y - z) = 0.$$

Soit:

$$y + \frac{1}{\rho} \Omega y = z. \quad (2.8)$$

En posant  $A(\rho) = \left( I + \frac{1}{\rho} \Omega \right)^{-1}$  on a donc:

$$y = A(\rho) z. \quad (2.9)$$

Les algorithmes disponibles pour obtenir le minimum de  $V$  nécessitent des évaluations de  $V(\rho)$  et éventuellement de  $V'(\rho)$ . Il est donc important de réduire le coût de calcul de cette fonction.

En regardant l'expression de  $V$ , on voit que l'évaluation du numérateur coûte le calcul de  $S_\rho$ , plus  $n$  multiplications et une division (les sommes ne sont pas comptées).

Soit:

$$\text{Coût numérateur} = 30n + n + 1. \quad (2.10)$$

Ce coût est linéaire en  $n$ ; par contre, l'évaluation du dénominateur est beaucoup plus compliquée: en effet, il faut calculer  $\text{tr}(A(\rho))$ . Pour cela il faut expliciter  $A(\rho)$ , où une matrice équivalente à elle.

La connaissance explicite de  $A(\rho)$  coûte le calcul de  $A(\rho) e_i$ ,  $i = 1, 2, \dots, n$ , où  $\{e_1, \dots, e_n\}$  est la base canonique de  $\mathbb{R}^n$ , d'où:

$$\text{Coût dénominateur} = n(30n) + 1. \quad (2.11)$$

On voit clairement que c'est ce calcul qui est le plus coûteux. Or, si l'on connaissait les valeurs propres de  $A(\rho)$ , on pourrait épargner ce calcul. On va donc s'efforcer d'obtenir une approximation calculable de ces valeurs propres.

On a:  $A(\rho) = \left(I + \frac{1}{\rho} \Omega\right)^{-1}$  donc:

$$\beta_i = \frac{1}{1 + \frac{1}{\rho} \alpha_i} \quad i = 1, 2, \dots, n \quad (2.12)$$

où  $\beta_i, i = 1, \dots, n$ , dénote les valeurs propres de  $A(\rho)$  et  $\alpha_i, i = 1, \dots, n$ , les valeurs propres de  $\Omega$ .

La matrice  $\Omega$  ne dépend pas de  $\rho$ ; il est clair qu'il faudrait connaître ses valeurs propres pour connaître la trace de  $A(\rho)$ .

Pour des raisons de clarté dans ce qui suivra, nous introduisons quelques notations:

$$r_i^n = \frac{1}{2n} + \frac{t_i - t_1}{h} \frac{1}{n} = \frac{1}{2n} + (i-1) \frac{1}{n} \quad i = 1, 2, \dots, n. \quad (2.13)$$

Soit  $\bar{\Omega}_n$  la matrice associée à cet ensemble de noeuds, il est facile de voir que:

$$\Omega = \frac{1}{(hn)^{2q}} \bar{\Omega}_n \quad (2.14)$$

et

$$\begin{aligned} \text{tr} \left( \left( I + \frac{1}{\rho} \Omega \right)^{-1} \right) &= \text{tr} \left( \left( I + \frac{1}{\rho(hn)^{2q}} \bar{\Omega}_n \right)^{-1} \right) \\ &= \sum_{i=1}^n \frac{1}{1 + \frac{1}{\rho(hn)^{2q}} \gamma_i^n} \end{aligned} \quad (2.15)$$

où  $\gamma_i^n, i = 1, 2, \dots, n$ , sont les valeurs propres de  $\bar{\Omega}_n$ . On étudiera donc  $\bar{\Omega}_n$ ; notons:

$$\bar{A}_n(\rho) = \left( I + \frac{1}{\rho} \bar{\Omega}_n \right)^{-1} \quad (2.16)$$

et

$$B_n(\tau) = \bar{A}_n \left( \frac{\tau}{n} \right) = \left( I + \frac{n}{\tau} \bar{\Omega}_n \right)^{-1}. \quad (2.17)$$

On a donc:

$$\text{tr} \left( \left( I + \frac{1}{\rho} \Omega \right)^{-1} \right) = \bar{A}_n(\rho(hn)^{2q}) = B_n(\tau(nh)^{2q}) \quad (2.18)$$

avec:  $\rho = \frac{\tau}{n}$ .

### 3. Expression asymptotique des valeurs propres

Le but de ce paragraphe est de démontrer le théorème suivant:

**Théorème 3.1.** Soient  $\lambda_i, i = 1, 2, \dots$ , les valeurs propres du problème:

$$\begin{aligned} D^{(2q)} \varphi &= (-1)^q \lambda \varphi \\ \varphi^{(j)}(0) &= \varphi^{(j)}(1) = 0 \quad j = q, \dots, 2q - 1 \end{aligned} \quad (3.1)$$

alors, si  $q \geq 2$  on a:

$$\text{Tr}(B_n(\tau)) = \sum_{i=1}^n \frac{1}{1 + \frac{\lambda_i}{\tau}} + O(n^{-1}), \quad \text{avec } hn = \text{cte.} \quad (3.2)$$

Pour démontrer ce théorème, on aura besoin du résultat classique suivant:

**Lemme 3.2.** Si  $g \in H^q[0, 1]$ ,  $q \geq 2$ , alors:

$$\left| \frac{1}{n} \sum_{i=1}^n g(r_i^n) - \int_0^1 g(t) dt \right| \leq \frac{c}{n^2} \|g\|_{H^q} \quad (3.3)$$

avec

$$r_i^n = \frac{2i-1}{2} \frac{1}{n} \quad i = 1, 2, \dots, n$$

*Notations.* 1) Si  $y \in \mathbb{R}^n$ , on notera  $I_n(y)$  la fonction  $\sigma \in S_n$  ( $S_n$  étant l'espace des fonctions spline associées aux abscisses  $r_i^n$ ) telle que:

$$\sigma(r_i^n) = y_i \quad i = 1, 2, \dots, n.$$

2) Si  $g \in H^q[0, 1]$  on notera  $P_n(g)$  l'élément  $y \in \mathbb{R}^n$  tel que:

$$y_i = g(r_i^n) \quad i = 1, 2, \dots, n.$$

Pour démontrer le théorème, on aura besoin de la formulation variationnelle du problème 3.1, donnée par:

**Lemme 3.3.** Les deux problèmes suivants sont équivalents:

$$\begin{aligned} D^{(2q)} \varphi &= (-1)^q \lambda \varphi \\ \varphi^{(j)}(0) &= \varphi^{(j)}(1) = 0 \quad j = q, \dots, 2q - 1, \end{aligned} \quad (3.4)$$

$$P(\varphi, \Psi) = \lambda(\varphi, \Psi) \quad \forall \Psi \in H^q[0, 1] \quad (3.5)$$

où

$$(\varphi, \Psi) = \int_0^1 \varphi(t) \Psi(t) dt.$$

Il est bien connu que les valeurs propres du problème (3.5) peuvent être calculées avec les propriétés de minimax (cf. [18]) à partir du quotient de Rayleigh

$$Q(\varphi) = \frac{P(\varphi, \varphi)}{(\varphi, \varphi)}. \quad (3.6)$$

Pour obtenir les valeurs propres associées à ce quotient, on peut appliquer la méthode de Rayleigh-Ritz (cf. [2, 18]) associée à l'espace  $S_n$ . On aura donc à trouver les valeurs propres du quotient

$$\bar{Q}(x) = \frac{\langle x, \bar{Q}_n x \rangle}{(\sigma, \sigma)} \quad (3.7)$$

où

$$x_i = \sigma(r_i^n) \quad i = 1, 2, \dots, n$$

le dénominateur de  $\bar{Q}$  étant difficile à expliciter, on remplace  $(\sigma, \sigma)$  par  $\frac{1}{n} \sum_{i=1}^n \sigma^2(r_i^n)$  pour obtenir

$$Q_n(x) = \frac{\langle x, \bar{Q}_n x \rangle}{\frac{1}{n} \langle x, x \rangle} \quad (3.8)$$

les valeurs propres associées à  $Q_n$  sont les solutions de

$$\bar{Q}_n x = \theta \frac{1}{n} x \quad (3.9)$$

valeurs propres qu'on notera  $\lambda_i^n$ ,  $i = 1, \dots, n$ , et on a  $\lambda_i^n = n \gamma_i^n$ .

Grâce aux résultats de Fix [5], on obtient que les valeurs propres  $\lambda_i^n$  convergent vers les valeurs propres  $\lambda_i$  du problème (3.5), de plus:

$$|\lambda_i^n - \lambda_i| = O(n^{-2}). \quad (3.10)$$

Dénotons  $R(\tau)$  la résolvante de l'opérateur associé, c'est-à-dire l'opérateur

$$R_n(\tau): S_n \rightarrow S_n$$

$$u \rightarrow \sigma$$

avec

$$P(\sigma, w) + \tau(\sigma, w) = (u, w) \quad \forall w \in S_n.$$

On notera  $R_n(\tau)$  aussi sa prolongation sur  $L^2[0, 1]$ . Finalement, dénotons par  $\bar{R}_n(\tau)$  la prolongation sur  $L^2[0, 1]$  de l'opérateur

$$\begin{aligned} \bar{R}_n(\tau): S_n &\rightarrow S_n \\ u &\rightarrow \sigma \end{aligned}$$

avec

$$\sigma = I_n((I + n\bar{\Omega}_n)^{-1} P_n(u)).$$

Les résultats obtenus par Fix et le fait que  $R$ ,  $R_n$ ,  $\bar{R}_n$  soient autoadjoints nous garantissent que:

$$\begin{aligned} \|R(\tau) - R_n(\tau)\| &= O(n^{-2}) \\ \|R_n(\tau) - \bar{R}_n(\tau)\| &= O(n^{-2}). \end{aligned} \quad (3.11)$$

Avec ceci, on va démontrer le résultat du théorème 3.1, en effet:

$$\text{tr}(R(\tau)) = \sum_{i \in \mathbb{N}} \frac{1}{\tau + \lambda_i} = \frac{1}{\tau} \sum_{i \in \mathbb{N}} \frac{1}{1 + \lambda_i/\tau}. \quad (3.12)$$

On a aussi:

$$\text{tr}(R_n(\tau)) = \sum_{i \in \mathbb{N}} \frac{1}{\tau + \eta_i^n} \quad (3.13)$$

où  $\eta_i^n$  sont les valeurs propres obtenues par la méthode de Rayleigh-Ritz, donc, on sait que (cf. [2, 18])

$$\eta_i^n \geq \lambda_i \quad \forall i, n. \quad (3.14)$$

Soit  $\delta > 0$ , on peut choisir  $J$  tel que:

$$\sum_{i=J}^{\infty} \frac{1}{\tau + \lambda_i} < \delta. \quad (3.15)$$

On peut aussi choisir  $N$  tel que pour  $n \geq N$  on ait:

$$\left| \sum_{i=1}^J \frac{1}{\tau + \lambda_i} - \sum_{i=1}^J \frac{1}{\tau + \eta_i^n} \right| < \delta. \quad (3.16)$$

Finalement, à cause de l'inégalité (3.14)

$$\sum_{i=J}^n \frac{1}{\eta_i^n + \tau} \leq \sum_{i=J}^{\infty} \frac{1}{\lambda_i + \tau} < \delta. \quad (3.17)$$

On a donc:

$$|\text{tr}(R(\tau)) - \text{tr}(R_n(\tau))| < 3\delta$$

on peut démontrer aussi que:

$$|\text{tr}(R(\tau)) - \text{tr}(R_n(\tau))| = O(n^{-1}). \quad (3.18)$$

Finalement:

$$\begin{aligned} & |\text{tr}(R_n(\tau)) - \text{tr}(\bar{R}_n(\tau))| \\ &= \left| \sum_{i=1}^n (S_i, (R_n(\tau) - \bar{R}_n(\tau)) S_i) \right| \\ &\leq n \|R_n(\tau) - \bar{R}_n(\tau)\| \\ &\leq O(n^{-1}) \end{aligned}$$

où  $\{S_1, \dots, S_n\}$  est une base orthonormale (pour le produit de  $L^2$ ) de  $S_n$ .  
 Pour obtenir l'énoncé du théorème, il suffit de voir que:

$$\sum_{i=n}^{\infty} \frac{1}{\lambda_i + \tau} = O\left(\frac{1}{\lambda_n}\right) = O(n^{-2q}).$$

#### 4. Splines cubiques ( $q = 2$ ) et quintiques ( $q = 3$ )

Pour les splines cubiques, le problème auquel on est amené est:

$$\begin{aligned} \varphi^{(iv)} &= \lambda \varphi \\ \varphi''(0) &= \varphi''(1) = 0 \\ \varphi'''(0) &= \varphi'''(1) = 0. \end{aligned} \tag{4.1}$$

Ce problème correspond aux vibrations libres d'une baguette (cf. Courant [3]). Ses valeurs propres sont données par

$$\lambda_i = p_i^4 \tag{4.2}$$

où  $p_i \in \mathbb{R}^+$  est solution de l'équation

$$\cos p \cosh p = 1. \tag{4.3}$$

Dans le tableau No.1 on donne les douze premières racines de l'équation (4.3) divisées par  $\pi$ . On constate que, à partir de la douzième,  $\frac{p_k}{\pi} = k - 2 + \frac{1}{2}$ ,  $k \geq 10$ ,  $k \in \mathbb{N}$ , (à la précision machine près).

**Tableau No.1**

$p_1/\pi$	0	$p_7/\pi$	5.50000001994390
$p_2/\pi$	0	$p_8/\pi$	6.49999999913815
$p_3/\pi$	1.50561873114194	$p_9/\pi$	7.50000000003725
$p_4/\pi$	2.49975267007397	$p_{10}/\pi$	8.49999999999838
$p_5/\pi$	3.50001067943591	$p_{11}/\pi$	9.50000000000007
$p_6/\pi$	4.49999953848358	$p_{12}/\pi$	10.5000000000000

Pour les splines d'ordre 3 (quintiques), on a résoudre

$$\begin{aligned}
 \varphi^{(vi)} &= \lambda \varphi \\
 \varphi'''(0) &= \varphi'''(1) = 0 \\
 \varphi^{(iv)}(0) &= \varphi^{(iv)}(1) = 0 \\
 \varphi^{(v)}(0) &= \varphi^{(v)}(1) = 0.
 \end{aligned}
 \tag{4.4}$$

Ses valeurs propres se calculent avec l'expression

$$\lambda_i = p_i^6 \tag{4.5}$$

où  $p_i \in \mathbb{R}^+ \cup \{0\}$  est racine de l'équation

$$\sin \frac{p}{2} \left[ \sin^2 \frac{p}{2} \cos \frac{p}{2} + \cos \frac{p}{2} sh^2 \frac{\sqrt{3}}{2} p + 2 \cos \frac{p}{2} - 2 ch \frac{\sqrt{3}}{2} p \right] = 0. \tag{4.6}$$

Les racines d'indice impair de cette équation sont données par le tableau No.2 pour les 11 premières, les autres étant (à précision machine près)

$$p_{2k+1} = (2k - 1)\pi \quad k > 11$$

et les racines paires correspondent à celles du sinus, multipliées par 2.

**Tableau No. 2**

		6	28.27433388212233
1	0	7	34.55751918948855
2	0	8	40.84070449666733
3	2.527482153303429	9	47.12388980384693
4	9.427052987008248	10	53.40707511102652
5	21.99114861798320	11	59.69026041820610

### 5. Résultats numériques

Dans ce paragraphe, on présente un ensemble d'expériences numériques qui permettent d'illustrer les propriétés du paramètre de lissage choisi par la méthode de Validation Croisée. On discute aussi la méthode de calcul du minimum de  $V$ .

Dans le tableau No.3 ci-dessous, on peut trouver le nombre d'opérations nécessaires pour évaluer  $V$  en utilisant un algorithme de calcul des fonctions spline d'ajustement dû à Paihua [9] (on ne compte que les multiplications et divisions). On y voit clairement que le coût d'évaluation de  $V$  est divisé par  $n$ .

**Tableau No. 3**

Méthode directe	$30n(n+1) + n + 3$
Méthode proposée	$30n + 3n + 3$

Pour minimiser  $V$ , on a testé quatre méthodes:

- 1) Regula Falsi sur l'équation  $V'(\rho)=0$
- 2) Newton sur l'équation  $V'(\rho)=0$
- 3) Dichotomie sur l'équation  $V'(\rho)=0$
- 4) Section Dorée (cf. [10], [12]) pour la minimisation directe de  $V$ .

Les méthodes de Newton et Regula Falsi coûtent très cher pour ce problème et ne sont pas compétitives. Parmi les deux dernières, la méthode de la section dorée est celle qui donne les meilleurs temps de calcul (à peu près 25% de moins que la méthode 3).

Le tableau No. 4 donne les temps de calcul (sur l'ordinateur IBM/360/67 du C.I.C.G.) pour une précision (erreur relative) de  $10^{-5}$ .

Tableau No. 4. Temps de calcul de  $\hat{\rho}$

n	Temps (s)
50	0.95
100	1.89
200	3.70
300	5.56

En fait, on a observé que pour la plupart des cas, la précision relative de  $10^{-5}$  n'est pas nécessaire, et on peut se contenter d'une précision de  $10^{-3}$  sans perdre significativement dans l'approximation de  $f$ .

La deuxième partie des expériences a pour but d'illustrer le comportement de la fonction spline d'ajustement ainsi trouvée, en tant qu'approximation de la fonction cherchée  $f$ .

Supposons donnés  $f \in H^q[a, b]$  (dans les expériences  $a = -1$ ,  $b = 1$ ), un nombre de points  $n$  et un échantillon d'erreurs  $\varepsilon_i$  de variance  $v^2$  et moyenne nulle. Dénotons:

$$T(\rho) = \left( \frac{1}{n} \sum_{i=1}^n (\sigma_{\rho, n}(t_i) - f(t_i))^2 \right)^{1/2}$$

$\hat{\rho}$  Valeur de  $\rho$  qui minimise  $V$

$$\hat{T} = T(\hat{\rho})$$

$\bar{\rho}$  Valeur de  $\rho$  qui minimise  $T$

$$\bar{T} = T(\bar{\rho})$$

$$I = \frac{\hat{T}}{\bar{T}} \quad (\text{Inefficacité relative de la méthode.})$$

La quantité  $I$ , mesure d'une certaine façon l'inefficacité de la méthode par rapport au « meilleur » résultat qu'on pourrait obtenir ( $\bar{T}$ ). Evidemment, la fonction  $T(\rho)$  ne peut pas être calculée dans la pratique puisque l'on ne connaît pas  $f$ . Dans le tableau No. 5 on donne la quantité  $I$  trouvée comme moyenne pour différentes valeurs de  $n$  et  $\sigma$ , sur un ensemble de 400 essais: on remarque que cette quantité est toujours voisine de 1.

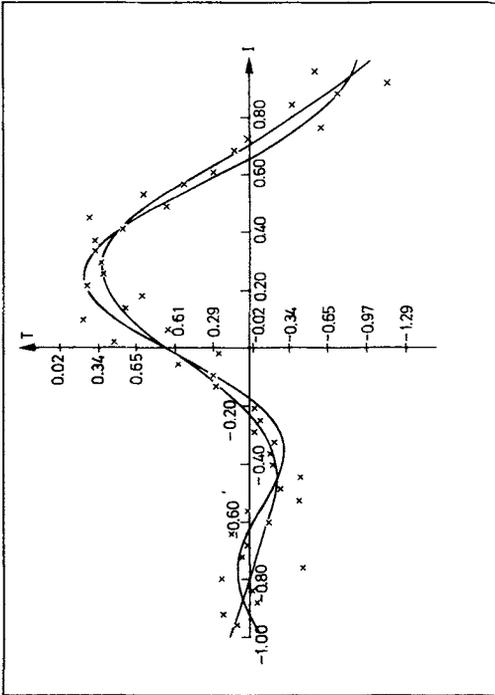


Figure 1 - 50 points

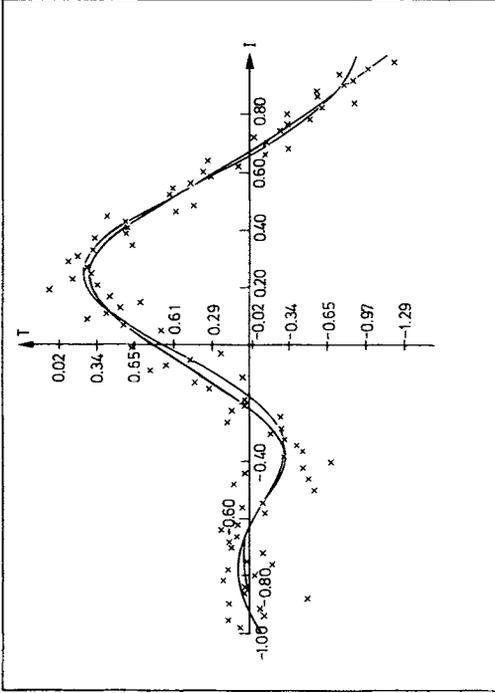


Figure 2 - 100 points

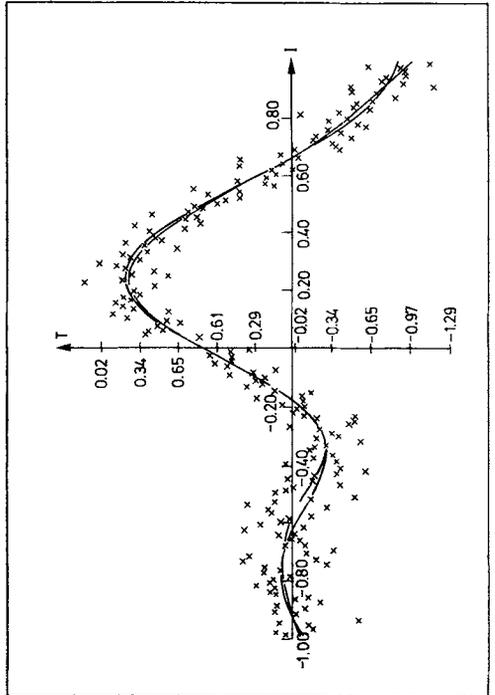


Figure 3 - 200 points

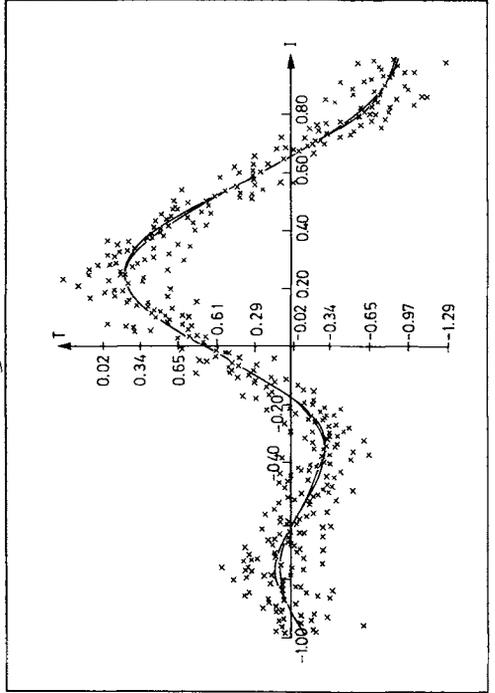


Figure 4 - 400 points

**Tableau No. 5.** Inefficacité relative

$\sigma$	$n$			
	50	100	200	300
0.5	1.276	1.186	1.093	1.091
0.1	1.197	1.143	1.242	1.062
0.05	1.176	1.112	1.075	1.149
0.01	1.038	1.031	1.039	1.019

On constate la convergence de  $T$  vers  $\bar{T}$  lorsque  $n$  tend vers l'infini. Cette constatation de convergence, on peut la faire aussi dans les dessins n° 1, 2, 3 et 4 qui montrent les splines pour 50, 100, 200 et 400 points, les nuages de points et la fonction  $f$  initiale (cf. Fig. 1-4).

## Conclusion

La méthode de Validation Croisée Généralisée s'est avérée une très bonne méthode pour choisir  $\rho$  dans la pratique, grâce à l'algorithme présenté ici.

Les personnes désirant obtenir les programmes de calcul de  $\rho$  pour les splines cubiques doivent écrire à l'auteur.

## Références

1. Ahlberg, J.H., Nilson, E.N., Walsh: The theory of splines and their applications. New-York: Academic (1967)
2. Chatelin, F., Lemordant, J.: La méthode de Rayleigh-Ritz appliquée à des opérateurs différentiels elliptiques, ordres de convergence des éléments propres. Numer. Math. **23**, 215-222 (1975)
3. Courant, R., Hilbert, D.: Methods of Mathematical Physics, New-York: Interscience Publishers (1953)
4. Craven, P., Wahba, G.: Smoothing noisy data with spline functions estimating the correct degree of smoothing by the method of generalized cross-validation. TR # 445 (october 1977) University of Winsconsin-Madison
5. Fix, G.: Effects of quadrature errors in finite element approximation of steady state, eigenvalue and parabolic problems. In: Foundations of the finite element method with applications to partial differential equations. A.K. Aziz, ed. New York: Academic (1972)
6. Joly, J.L.: Théorèmes de convergence pour les fonctions spline générales d'interpolation et d'ajustement. C.R. Acad. Sci. Paris 264, ser. A, 126-128 (1970)
7. Laurent, P.J.: Approximation et optimisation. Paris: Hermann (1972)
8. Necas, J.: Les méthodes directes en théorie des equations elliptiques. Paris: Masson (1967)
9. Paihua, L.: Quelques méthodes numériques pour le calcul de fonctions spline à une et plusieurs variables. Thèse, Université Scientifique et Médicale de Grenoble. Grenoble (Mai 1978)
10. Robert, F.: Analyse numérique itérative. Cours I.N.P.G., Grenoble (1974)
11. Stone, M.: Cross-validators choice and assesement of statistical prediction. J. Roy. Statist. Soc. Ser. B, **36**, 111-147 (1974)
12. Sutti, C.: One dimensional minimisation methods. In: Minimisation algorithms. G.P. Szegő, ed. London: Academic (1972)

13. Wahba, G., Wold, S.: A completely automatic french curve: fitting spline functions by cross-validation. *Commun. Statist.* **4**, 1-17 (1975)
14. Wahba, G.: A survey of some smoothing problems and the method of generalized cross-validation for solving them. TR # 457 (July 1976) University of Wisconsin-Madison
15. Wahba, G.: Smoothing Noisy Data with Spline Functions. *Numer. Math.* **24**, 383-393 (1975)
16. Wahba, G.: Optimal smoothing of density estimates. TR #469 (october 1976), University of Wisconsin-Madison
17. Wahba, G.: Practical approximate solutions to linear operator equations when the data are noisy. *SIAM, J. Num. Analysis*, **14**, n° 4 (1977)
18. Weinberger, H.F.: Variational methods for eigenvalue approximations. SIAM, Philadelphia (1974)

Reçu le 9 septembre 1978