



Des algorithmes rapides pour la validation croisée sur des problèmes d'approximation non linéaires

Lalita-Narhar Deshpande

► To cite this version:

Lalita-Narhar Deshpande. Des algorithmes rapides pour la validation croisée sur des problèmes d'approximation non linéaires. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1990. Français. NNT: . tel-00337860

HAL Id: tel-00337860

<https://theses.hal.science/tel-00337860>

Submitted on 10 Nov 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TU10111

THESE

présentée par

DESHPANDE Lalita Narhar

pour obtenir le titre de

DOCTEUR

de l'Université Joseph Fourier - Grenoble I

(arrêté ministériel du 5 juillet 1984)

Spécialité : Mathématiques Appliquées

DES ALGORITHMES RAPIDES POUR LA VALIDATION CROISEE SUR DES PROBLEMES D'APPROXIMATION NON LINEAIRES

Date de soutenance : 4 Octobre 1990

Composition du jury :

J. BLUM

Président

A. ANTONIADIS

R. ARCANGELI

D. GIRARD

P.J. LAURENT

Examineurs

Thèse préparée au sein du laboratoire LMC

Je voudrais exprimer mes remerciements et toute ma gratitude à Monsieur Pierre-Jean LAURENT, responsable de ma thèse, pour m'avoir fait confiance en m'accueillant au sein de l'équipe Géométrie Algorithmique, pour ses conseils pertinents et ses encouragements, qui m'ont été précieux pour ce travail et mon séjour en France.

J'exprime toute ma reconnaissance et mes remerciements les plus sincères à Monsieur Didier GIRARD qui a dirigé cette thèse; la sympathie qu'il m'a témoignée, ses conseils et ses encouragements constants ont contribué énormément à l'accomplissement de ce travail.

Monsieur R. ARCANGELI, professeur à Université de Pau, a eu l'amabilité de se déplacer pour juger cette thèse, je l'en remercie.

Je tiens à remercier aussi, Monsieur A. ANTONIADIS pour avoir accepté de juger mon travail.

Je remercie Monsieur J. BLUM d'avoir accepté de présider le jury.

Je remercie également tous les membres de l'équipe de Géométrie Algorithmique pour leur aide et sympathie, et en particulier mes amies Ritha RUIZ LOJA et Rekha P. KULKARNI.

Je suis extrêmement reconnaissante à Madame Guizou LAURENT; son accueil et son amitié chaleureuse a fait de ma vie à Grenoble une expérience très agréable.

Table des matières

Introduction générale	1
Chapitre I : Rappels sur les problèmes classiques de moindres carrés pénalisés; notations et préliminaires	5
1.1 Introduction	5
1.2 Rappels sur les fonctions splines classiques polynômiales	7
1.3 Choix du paramètre de lissage	9
1.3.1 La méthode de validation croisée	10
1.4 Calcul de la trace	12
1.4.1 Estimateur de type Monte-Carlo	13
1.5 Estimation d'autres paramètres non linéaires	13
1.6 Notations et préliminaires	14
1.6.1 Itérations de type Newton	17
1.6.2 Itérations de Gauss-Newton	19
Chapitre II : GCV pour les problème de moindres carrés non linéaires	23
2.1 Introduction	23
2.2 Choix de p par GCV	24
2.2.1 Expression analytique pour $(B_p')_z$	26
2.3 Calcul de la trace par Monte-Carlo	27
2.4 Expériences numériques	34
Chapitre III : GCV quand le critère de moindres carrés est remplacé par une "distance convexe"	39
3.1 Introduction	39
3.1.1 Observation de type Poisson	40
3.1.2 Splines robustes	41
3.2 Choix de p par GCV	43

3.2.1	Approximation quadratique de $\phi(\mu_i, z_i)$ par rapport à μ_i	43
3.2.2	Expression analytique pour $(B_p)_z$	46
3.2.3	Approximation quadratique de $F(y_i, z_i)$ par rapport à y_i	47
3.3	Calcul de la trace par Monte-Carlo	48
3.3.1	Observation de type Poisson	49
3.3.2	Splines Robustes	52
3.4	Une comparaison des méthodes	53
3.4.1	Observations de type Poisson	54
3.4.2	Splines Robustes	57
 Chapitre IV : Détection de discontinuités par Inf-Convolution spline et GCV dans un cadre non linéaire		 65
4.1	Introduction	65
4.2	L'approche ICS: un lissage non quadratique préservant les discontinuités	66
4.2.1	Modélisation d'une fonction lisse par morceaux	66
4.2.2	Calcul des estimateurs à P et p fixés	67
4.3	Choix de p par GCV	71
4.4	Choix de p et de la position d'une discontinuité par GCV	72
4.5	Expériences numériques	73
 Chapitre V : Choix du paramètre de lissage pour un problème d'enveloppe		 79
5.1	Introduction	79
5.2	Formulation du problème	81
5.3	Estimateurs pour différentes valeurs de p	83
5.4	Choix de p par GCV	88
 Annexe		 95
 Bibliographie		 97

Introduction générale

Le problème classique de lissage est d'approcher une fonction "lisse", disons g , ses valeurs $z=(z_1, \dots, z_n)$ en n points étant supposées connues avec des erreurs aléatoires indépendantes de moyenne nulle et de variance v^2 . Pour résoudre ce problème, on considère la fonction spline de lissage de paramètre p . Le choix de ce paramètre p est extrêmement important.

Dans le présent travail nous donnons une approche unifiée de validation croisée pour le choix de p dans les problèmes de lissage non linéaire.

Le premier chapitre est un rappel consacré au cas de spline de lissage classique i.e. lissage avec des moindres carrés linéaires. Dans ce cas si la variance v^2 des erreurs est connue, le paramètre p est déterminé par v^2 et peut être calculé en fonction des données z_i : c'est la méthode de Reinsch. Deux autres méthodes de choix de p sont aussi présentées: une méthode statistique décrite par Mallows qui nécessite de connaître (à peu près) la variance des erreurs, et qui consiste à choisir p qui minimise une estimation calculée en fonction de p , les données, la variance v^2 et l'opérateur de lissage linéaire A_p ; et une troisième méthode, la validation croisée qui a l'avantage de ne pas nécessiter la connaissance de v^2 .

Dans le cadre linéaire classique le calcul de la fonction de validation croisée généralisée (GCV) nécessite le calcul de la trace de l'opérateur A_p . Un algorithme rapide de Monte-Carlo a été proposé par Girard pour calculer un estimateur efficace de la trace d'une matrice. La méthode de GCV a été aussi utilisée pour l'estimation d'autres paramètres non linéaires comme la position d'une discontinuité ou d'un pic Gaussien et la période de la fonction, etc.

La dernière section de chapitre I est consacrée aux notations et à des résultats préliminaires.

En pratique on peut disposer d'observations indirectes sur certaines fonctions non linéaires appliquées à la fonction g recherchée, ou on suppose une information a priori sur des contraintes de forme, par exemple de positivité de la fonction g . Nous sommes alors conduit à résoudre des problèmes de lissage de moindres carrés non linéaires. Pour cela nous utilisons la méthode itérative de Gauss-Newton. Dans le chapitre II nous proposons une approche générale pour le calcul de GCV pour de tels problèmes. L'expression de GCV utilise la dérivée de l'opérateur non linéaire de lissage B_p par rapport à z . Dans l'annexe nous donnons une expression explicite de cette dérivée pour le cas plus général.

Nous proposons deux méthodes de Monte-Carlo pour le calcul de la trace des opérateurs intervenant dans la GCV. La première est une extension directe de la méthode de Girard proposée dans le cas linéaire, et la deuxième est une formule originale dont l'évaluation ne nécessite que le calcul d'une spline supplémentaire associée à des données perturbées. Aussi nous faisons une comparaison des différentes méthodes pour le calcul de la GCV par des expériences numériques effectuées sur des données fictives.

Dans le chapitre III nous considérons le cas où la classique distance aux données (somme des carrés) est remplacée par une certaine "distance" convexe. Les deux exemples suivant sont traités ici: données suivant une loi de Poisson ou données avec quelques points aberrants. La "distance" utilisée dans chacun de ces exemples est obtenue par le classique principe du maximum de vraisemblance. Alors, nous avons un problème de lissage non linéaire à résoudre. Pour cela nous utilisons la méthode itérative de Newton relaxée. Nous proposons de choisir le paramètre p par le principe de validation croisée. Plusieurs façons de généraliser la fonction de validation croisée ordinaire sont suggérées.

Une étude comparative de ces différentes méthodes de GCV et de leur approximation par Monte-Carlo est ensuite présentée. Des exemples numériques sont traités pour deux types de données, précisément: données suivant une loi de Poisson et des données avec quelques points aberrants.

L'expression de la GCV obtenue dans le chapitre III reste valable pour un terme de régularisation quelconque. Dans le chapitre IV nous supposons que la fonction g est lisse par morceaux, c'est-à-dire qu'elle présente un certain nombre de discontinuités et les données bruitées sont avec des erreurs de type Poisson. Le but

principal est de montrer que des résultats analogues à ceux déjà obtenus ailleurs dans le cadre gaussien, sont aussi valables ici. C'est-à-dire que le principe de GCV permet de choisir le meilleur paramètre de lissage, et aussi connaissant la présence d'une discontinuité de déterminer sa position.

Dans le dernier chapitre nous considérons un exemple de problème de maximum de vraisemblance plus général, c'est-à-dire avec une distance non convexe. Les données sont du type: $z_i = z_i^0 + \varepsilon_i$, $i=1, \dots, n$ où z_i^0 suit une loi uniforme sur $[0, g(t_i)]$ pour $i=1, \dots, n$ et $\varepsilon_1, \dots, \varepsilon_n$ sont des variables aléatoires normales indépendantes de moyenne nulle et de variance v^2 . Dans le problème d'estimation de g la somme des carrés est remplacée par la fonction négative de log-vraisemblance correspondant au modèle ci-dessus. Le choix de p est proposé par GCV. Les résultats numériques obtenus sont favorables et étonnants.

Tout au long de cette thèse nous utilisons un terme de lissage quadratique (par exemple $\int (f'')^2$) mais toutes les expressions de GCV proposées ici restent valables pour d'autres termes de pénalité et des problèmes à plusieurs variables.

Chapitre I

Rappels sur les problèmes classiques de moindres carrés pénalisés; notations et préliminaires

1.1 Introduction

Soit g une fonction à valeurs réelles définie sur un intervalle $[a,b]$ de la droite réelle. Soient aussi $a \leq t_1 < \dots < t_n \leq b$ n points de l'intervalle $[a,b]$.

Le problème classique de lissage est d'approcher g , ses valeurs aux points $t_1, \dots, t_n$ étant supposées connues avec des erreurs $\varepsilon_1, \dots, \varepsilon_n$; c'est-à-dire, en connaissant:

$$(1.1) \quad z_i = g(t_i) + \varepsilon_i, \quad i=1, \dots, n.$$

On suppose que les erreurs ε_i , $i=1, \dots, n$ sont des réalisations des variables aléatoires indépendantes de moyenne nulle et de variance v_i^2 , i.e. $E(\varepsilon_i)=0$, $E(\varepsilon_i \varepsilon_j)=v_i^2 \delta_{ij}$ et $E(\varepsilon_i \varepsilon_j)=0$ pour $i \neq j$, $i, j=1, \dots, n$.

On suppose que g est "lisse". Ceci veut dire que g appartient à $H^q[a,b]$ (q entier positif), où:

$$H^q[a,b] = \left\{ f: [a,b] \rightarrow \mathbb{R} / \begin{array}{l} f, f', \dots, f^{[q-1]} \text{ absolument continues} \\ \text{et } \int_a^b [f^{[q]}(t)]^2 dt < \infty \end{array} \right\}.$$

Pour résoudre ce problème, on considère la solution unique ($n \geq q$) du problème:

$$(1.2) \quad \min_{f \in H^q[a,b]} \left\{ \int_a^b [f^{(q)}(t)]^2 dt + \rho \sum_{i=1}^n \alpha_i^2 [f(t_i) - z_i]^2 \right\},$$

avec $\alpha_i = \frac{\beta}{v_i}$, $i=1, \dots, n$ où β est une constante de proportionnalité

choisie pour avoir $\frac{1}{n} \sum_{i=1}^n \alpha_i^2 = 1$. Il faut remarquer que ceci n'implique pas la connaissance des v_i , mais de quantités proportionnelles, où la constante de proportionnalité est la même.

Il est bien connu que la solution de (1.2) est une fonction spline polynômiale de degré $2q-1$ (par ex. Laurent [72]). Reste le problème de choisir le paramètre de lissage ρ . Le paramètre de lissage ρ contrôle le compromis entre la fidélité aux données

mesurée par $\sum_{i=1}^n \alpha_i^2 [f(t_i) - z_i]^2$ et le caractère lisse de la solution

mesuré par $\int_a^b [f^{(q)}(t)]^2 dt$.

Le choix de ce paramètre est extrêmement important, en effet, une valeur de ρ trop importante fait passer la spline très près des données, et du fait des erreurs, produit une mauvaise approximation de g ; par contre, un ρ trop faible produit un lissage trop marqué, et on risque d'éliminer les variations de la fonction. Il y a eu de nombreuses tentatives de détermination du paramètre ρ . Dans la Section 1.3, on va rappeler celles qui conduisent au calcul effectif de ce paramètre. Ces méthodes connues peuvent se diviser en deux groupes: celles qui supposent qu'une bonne estimation de v^2 est connue, et celles qui facilitent un choix automatique de ρ sans avoir aucune idée sur le bruit, c'est la méthode appelée validation croisée.

Une bonne estimation du paramètre ρ par la méthode de validation croisée nécessite de calculer la trace de la matrice de lissage. Le calcul exact de cette trace est très coûteux. Utreras [80] a proposé une approximation calculable des valeurs propres de la matrice de lissage, dans le cas de t_i équidistants. Récemment, Girard [87c] et [87d] a proposé un algorithme rapide de Monte-Carlo pour calculer un estimateur efficace de la trace d'une matrice, qui est présenté dans la section 1.4.

Dans la Section 1.5, nous citons les exemples d'utilisation de la

méthode de validation croisée pour déterminer des paramètres non linéaires comme la position d'une discontinuité ou d'un pic, la période de la fonction, etc.

La Section 1.6 donne les notations et des résultats préliminaires utilisés dans cette thèse.

1.2 Rappels sur les fonctions splines classiques polynômiales

Soient $a \leq t_1 < \dots < t_n \leq b$ n points de l'intervalle $[a, b]$ et $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ alors si $n \geq q$ (q entier positif) le problème

$$(1.3) \quad \begin{array}{l} \text{Min} \\ f \in H^q[a, b] \\ f(t_i) = x_i, \quad i=1, \dots, n \end{array} \left\{ \int_a^b [f^{(q)}(t)]^2 dt \right\}$$

a une et une seule solution $\sigma \in H^q[a, b]$. De plus, il est bien connu (cf. par exemple Laurent [72]) que σ satisfait:

(1.4) σ est un polynôme de degré $2q-1$ dans chacun des intervalles $[t_i, t_{i+1}]$, $i=1, \dots, n-1$,

(1.5) σ est un polynôme de degré $q-1$ dans chacun des intervalles $[a, t_1]$ et $[t_n, b]$,

(1.6) $\sigma \in C^{2q-2}$,

(1.7) $\sigma(t_i) = x_i$, $i=1, \dots, n$.

Notons S_n^q l'espace de toutes les fonctions satisfaisant aux conditions (1.4), (1.5), (1.6). On appellera S_n^q l'espace des fonctions splines d'ordre q basées sur les noeuds $t_1, \dots, t_n$. Il est connu que S_n^q est un espace vectoriel de dimension n . La solution de (1.3) sera alors le seul élément de S_n^q qui satisfait aux conditions d'interpolation (1.7).

On appellera base canonique de S_n^q , la base composée d'éléments σ_i $i=1, \dots, n$ tels que:

$$\sigma_i(t_j) = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Il est facile de voir que dans cette base, σ s'écrit:

$$\sigma(t) = \sum_{i=1}^n x_i \sigma_i(t).$$

Soient $\sigma, s \in S_n^q$ et considérons la forme bilinéaire

$$B(\sigma, s) = \int_a^b \sigma^{[q]}(t) s^{[q]}(t) dt.$$

Alors si $\sigma(t_i) = x_i$, $s(t_i) = y_i$, $i=1, \dots, n$ on aura:

$$\begin{aligned} B(\sigma, s) &= \int_a^b \left(\sum_{i=1}^n x_i \sigma_i(t) \right)^{[q]} \left(\sum_{j=1}^n y_j \sigma_j(t) \right)^{[q]} dt \\ &= \sum_{i=1}^n \sum_{j=1}^n x_i y_j \int_a^b \sigma_i^{[q]}(t) \sigma_j^{[q]}(t) dt. \end{aligned}$$

Nous définissons:

$$\omega_{ij} = \int_a^b \sigma_i^{[q]}(t) \sigma_j^{[q]}(t) dt$$

$$\Omega = (\omega_{ij}) \begin{matrix} i=1, \dots, n \\ j=1, \dots, n \end{matrix}$$

On a alors:

$$B(\sigma, s) = \langle x, \Omega y \rangle$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire dans $\mathbb{R}^n$.

La matrice Ω est symétrique semi-définie positive.

Soient maintenant $z_1, \dots, z_n$ n nombres réels. On définit la fonction spline d'ajustement d'ordre q et de paramètre p pour les données $z_1, \dots, z_n$ associées à des poids $\alpha_1^2, \dots, \alpha_n^2$ comme la solution unique du problème:

$$(1.8) \quad \min_{f \in H^q[a, b]} \left\{ \int_a^b [f^{[q]}(t)]^2 dt + p \sum_{i=1}^n \alpha_i^2 [f(t_i) - z_i]^2 \right\}.$$

Soit $g_{p,z}$ la solution unique du problème (1.8). Elle satisfait aussi

(1.4), (1.5) et (1.6). Si l'on appelle $y_{p,z}$ la solution du problème:

$$\min_{y \in \mathbb{R}^n} \left\{ \langle y, \Omega y \rangle + \rho \sum_{i=1}^n \alpha_i^2 [y_i - z_i]^2 \right\},$$

on voit que le vecteur $y_{p,z} = (y_1, \dots, y_n)^T$ avec $y_i = g_{p,z}(t_i)$, $i=1, \dots, n$ satisfait donc: $y_{p,z} = A_p z$ avec $A_p = (\Omega + \rho W)^{-1} \rho W$ où W est la matrice diagonale qui a comme éléments diagonaux $w_i = \alpha_i^2$, $i=1, \dots, n$.

Rappelons qu'il existe des algorithmes efficaces pour le calcul de $g_{p,z}$ qui ne nécessitent pas le calcul explicite de la matrice Ω (e.g. Reinsch [71], Paihua [78]). Leur coût est linéaire en n .

1.3 Choix du paramètre de lissage

Nous allons maintenant décrire quelques méthodes pratiques pour le choix du paramètre de lissage.

La première méthode a été suggérée par Reinsch [67], qui obtient des résultats numériques satisfaisants dans le cas où $v_i = v$ est supposé connu. Reinsch propose de choisir ρ tel que

$$\frac{1}{n} \sum_{i=1}^n [g_{p,z}(t_i) - z_i]^2 = v^2$$

où $g_{p,z}$ est la solution du problème (1.8).

Dans le cas où g et la spline sont périodiques, Wahba [75] démontre que le choix 'optimal' est donné par:

$$\frac{1}{n} \sum_{i=1}^n [g_{p,z}(t_i) - z_i]^2 = kv^2$$

où k est une constante inférieure à 1. Reinsch [71] suggère aussi cette possibilité en conseillant de choisir k dans l'intervalle $\left[1 - \sqrt{\frac{2}{n}}, 1 + \sqrt{\frac{2}{n}}\right]$. Wold [74] observe aussi sur un certain nombre d'exemples que k devrait être choisi dans l'intervalle $[0.7, 0.95]$. Malheureusement, personne ne connaît exactement la bonne valeur de k , ce qui nous place devant un problème équivalent au premier. Le deuxième inconvénient de cette méthode est le fait qu'on ne connaît pas toujours une bonne estimation de v^2 .

Si l'estimation de v_i^2 ne pose pas de problème, il est possible d'obtenir une bonne valeur de ρ en utilisant une méthode décrite

par Mallows [73]. Pour définir cette méthode de calcul de ρ , considérons la valeur moyenne du carré des écarts entre $g_{\rho,z}$ et g

aux points t_i (pondérés par $\alpha_i^2 = \frac{\beta^2}{v_i^2}$):

$$\begin{aligned} T^2(\rho) &\equiv \frac{1}{n} \sum_{i=1}^n \alpha_i^2 [g_{\rho,z}(t_i) - g(t_i)]^2 \\ &= \frac{1}{n} \|W^{1/2}(A_\rho(z) - \bar{g})\|^2 \end{aligned}$$

où $\|\cdot\|$ est la norme Euclidienne dans $\mathbb{R}^n$ et $\bar{g} = (g(t_1), \dots, g(t_n))^T$.

Il est très facile de calculer l'espérance de l'erreur vraie $T^2(\rho)$:

$$ET^2(\rho) = \frac{1}{n} \|W^{1/2}(A_\rho(\bar{g}) - \bar{g})\|^2 + \frac{\beta^2}{n} \text{Tr}(A_\rho^2).$$

Or, si l'on pouvait minimiser $ET^2(\rho)$, le choix serait vraisemblablement très bon. Malheureusement, cette quantité étant inconnue, on ne peut le faire, mais on connaît un estimateur sans biais de $ET^2(\rho)$, donné par $\hat{T}^2(\rho)$ comme:

$$\hat{T}^2(\rho) = \frac{1}{n} \|W^{1/2}((I - A_\rho)z)\|^2 - \frac{\beta^2}{n} \text{Tr}(I - A_\rho)^2 + \frac{\beta^2}{n} \text{Tr} A_\rho^2$$

c'est-à-dire que

$$E\hat{T}^2(\rho) = ET^2(\rho).$$

Comme Mallows pour le cas d'une régression linéaire paramétrique, Craven et Wahba [79] propose donc d'estimer ρ en minimisant $\hat{T}^2(\rho)$.

1.3.1 La méthode de validation croisée

La première version de cette méthode a été proposée par Wahba et Wold [75] suivant une idée de Allen, reprise par Stone [74] qui avaient introduit la Validation Croisée Ordinaire (ou PRESS) pour le choix du paramètre de lissage dans l'estimation de modèles linéaires. Cette première version est basée sur l'idée très simple suivante: elle consiste à supprimer une donnée et à étudier la validité de différents ρ par l'aptitude des estimateurs $g_{\rho,z}$ associés à prédire la donnée non utilisée. On choisit alors le ρ qui

assure la meilleure prédiction en moyenne sur les n suppressions possibles d'une donnée. Pour simplifier la présentation, ici on supposera que $v_i=v$, $i=1,\dots,n$.

Soit $g_p^{(k)}$ l'estimation obtenue quand on laisse de côté la donnée z_k , i.e. $g_p^{(k)}$ est la solution unique du problème suivant:

$$(1.9) \quad \min_{f \in H^q[a,b]} \left\{ \int_a^b [f^{[q]}(t)]^2 dt + \rho \sum_{i=1, i \neq k}^n [f(t_i) - z_i]^2 \right\}.$$

La quantité $[g_p^{(k)}(t_k) - z_k]^2$ mesure l'aptitude de $g_p^{(k)}$ à prédire z_k . Alors le ρ optimum au sens de la validation croisée ordinaire est celui qui minimise:

$$V_0(\rho) = \frac{1}{n} \sum_{k=1}^n [g_p^{(k)}(t_k) - z_k]^2.$$

Heureusement l'évaluation de $V_0(\rho)$ ne nécessite pas la résolution des n sous-problèmes (1.9) mais seulement la connaissance de A_ρ . Pour le montrer, il faut d'abord observer que $g_p^{(k)}$ est aussi la solution du problème initial avec $g_p^{(k)}(t_k)$ à la place de z_k . On écrit alors que la k ème valeur prédite par $g_p^{(k)}$ est la k ème ligne du produit de $A_\rho = (a_{ij})_{i=1,\dots,n, j=1,\dots,n}$ par ce nouveau vecteur de données:

$$g_p^{(k)}(t_k) = \sum_{\substack{i=1 \\ i \neq k}}^n a_{ki} z_i + a_{kk} g_p^{(k)}(t_k),$$

ce qui s'écrit aussi après une simple manipulation algébrique:

$$(1 - a_{kk}) [g_p^{(k)} - z_k] = \left[\sum_{i=1}^n a_{ki} z_i \right] - z_k.$$

Donc on a

$$V_0(\rho) = \frac{1}{n} \sum_{k=1}^n \frac{[(A_\rho z)_k - z_k]^2}{(1 - a_{kk})^2}.$$

Quand toutes les données interviennent de manière symétrique, c'est-à-dire lorsqu'une permutation circulaire des coefficients de z ne modifie le vecteur lissé $A_p(z)$ que par cette même permutation i.e. nous sommes dans le cas où A_p est circulante, donc $1 - a_{kk} = \frac{1}{n} \text{Tr}(I - A_p)$ et on obtient $V(p)$ la validation croisée généralisée (GCV) :

$$(1.10) \quad V(p) = \frac{\frac{1}{n} \| (I - A_p) z \|^2}{\left[\frac{1}{n} \text{Tr}(I - A_p) \right]^2}.$$

Dans le cas où il n'y a pas cette symétrie dans les données, on peut théoriquement circulariser le problème et considérer la validation croisée ordinaire (cf. Craven et Wahba [79]).

Pour obtenir un bon choix du paramètre de lissage, on sera amené à minimiser $V(p)$ sur l'ensemble des réels positifs.

Les propriétés asymptotiques de la GCV et de l'estimateur de Mallows sont largement étudiées (cf. Craven et Wahba [79], Golub et al. [79], Speckman [85]). Les résultats de convergence pour le paramètre actuel donné par la GCV et l'estimateur de Mallows sont démontrés par Li [85], [86].

1.4 Calcul de la trace

Tous les algorithmes disponibles pour obtenir le minimum de V nécessitent des évaluations de $V(p)$. Il est donc très important d'étudier un algorithme rapide pour calculer $V(p)$. En regardant l'expression (1.10) pour $V(p)$, on voit que l'évaluation du numérateur coûte essentiellement le calcul de $A_p z$. Il est bien connu que ce coût est linéaire en n ; par contre, l'évaluation du dénominateur est beaucoup plus compliquée: en effet, il faut calculer $\text{Tr}(A_p)$. Pour cela, il faut expliciter A_p ou une matrice qui lui est semblable.

La connaissance explicite de A_p coûte le calcul de $A_p e_i$, $i=1, \dots, n$, où $\{e_1, \dots, e_n\}$ est la base canonique de $\mathbb{R}^n$. On voit clairement que c'est ce calcul qui est le plus coûteux. Or, si on connaissait les valeurs propres de Ω , on pourrait s'épargner ce calcul: en effet, si β_{in} dénote les valeurs propres de A_p , on aura:

$$\beta_{in} = \frac{p}{p + \lambda_{in}}, \quad i=1, \dots, n$$

où λ_{in} , $i=1,\dots,n$ sont les valeurs propres de Ω . Craven et Wahba [79] utilisent un algorithme de Reinsch pour calculer exactement les valeurs propres, mais ceci est encore très coûteux puisque le coût est encore $O(n^3)$, même s'il est fait une seule fois.

Utreras [81] a proposé une bonne approximation des valeurs propres qui permet de les calculer en ne faisant que $O(n)$ opérations si les pas sont équidistants et les poids tous égaux, et $O(n^2)$ opérations dans le cas contraire. Plus tard, Hutchinson et deHoog [85] ont exploité le fait que le calcul de la diagonale de A_p peut se ramener à celui des 5 diagonales principales de l'inverse d'une matrice bande et ils ont utilisé les résultats de Erisman et Tinney [75]. Signalons aussi l'approximation de Silverman [84].

Girard a proposé une approximation générale de la trace pour n assez grand. On donne une description de cette méthode au paragraphe suivant.

1.4.1 Estimateur de type Monte-Carlo

Un algorithme rapide de Monte-Carlo a été proposé par Girard (cf. Girard [87c], [87d]) pour calculer un estimateur efficace de la trace d'une matrice. Cet estimateur de Monte-Carlo de la trace d'une matrice A d'ordre n consiste à générer m vecteurs pseudo-aléatoire $w_1, \dots, w_m \in \mathbb{R}^n$ de bruit blanc normal de variance 1, à calculer Aw_i , $i=1, \dots, m$ puis à prendre comme estimateur de

$(1/n)\text{Tr}(A)$ la moyenne $\frac{1}{m} \sum_{i=1}^m \frac{w_i^T A w_i}{w_i^T w_i}$ où m est un entier positif.

Le coût d'une simulation de Monte-Carlo, c'est-à-dire pour $m=1$, est essentiellement le même que le calcul du résidu. Pour des propriétés théoriques et une justification numérique voir Girard [87d]. Les résultats de convergence asymptotique pour le paramètre actuel donné par GCV, et l'estimateur de Mallows quand les termes de trace sont remplacées par les estimateurs de Monte-Carlo (avec $m=1$), sont démontrés par Girard dans [89].

1.5 Estimation d'autres paramètres non linéaires

Supposons que les données soient de type (1.1). De plus on suppose qu'une certaine information a priori sur l'allure de la fonction g est connue. En général l'information de ce type sera qualitative. Pour préciser nous décrivons les trois exemples typiques considérés par Girard et Laurent [88]:

1. Le premier exemple est le suivant: on sait seulement que la fonction g est lisse partout avec probablement des discontinuités de la fonction (ou bien de la première dérivée) en

- certain points. Le nombre et les positions inconnus de ces discontinuités seront notés respectivement par m et α .
2. Dans le deuxième exemple, l'information a priori est que la fonction g est la somme d'une fonction lisse et d'un pic Gaussien "étroit". Le paramètre α inconnu peut être la position du pic, sa largeur ou les deux.
 3. Le troisième exemple est associé aux fonctions périodiques pour lesquelles les données correspondent à plusieurs périodes. Ici le paramètre non linéaire inconnu est la période α .

Le problème important est le suivant:

Supposons que les données soient de la forme (1.1) et que l'aspect essentiel de g n'est pas déterminé, i.e. le paramètre α est inconnu. Nous cherchons une fonction lisse à partir des données et simultanément un estimateur de α . On peut distinguer pour ce problème trois étapes principales. Dans les deux premières étapes on suppose que l'aspect essentiel de g est connu, c'est-à-dire que la vraie valeur de α , est connue. Alors la première étape consiste à lisser les données en préservant l'aspect essentiel α pour une valeur de ρ donnée. Ceci est fait par l'approche Inf-Convolution Spline (ICS) (Laurent [81], Laurent et Utreras [86], Girard et Laurent [88]). La deuxième étape consiste à choisir le paramètre de lissage. Le principe de GCV a été utilisé pour le choix optimal de ρ dans de tels problèmes (cf. Laurent et Utreras [86], Shiau et al [86], Girard et Laurent [88]). La troisième étape est le choix du paramètre α .

Dans le cas d'une spline de lissage polynômiale d'ordre q , la méthode de GCV a été proposée pour le choix de q (cf. Wahba [77a], [86]). Pour l'exemple 1 Girard [88a] a proposé la méthode de GCV pour détecter la présence de discontinuités et déterminer leurs positions. Notons que pour le même exemple Girard [90] a aussi proposé un estimateur de la position de la discontinuité par minimisation de l'énergie associée à ce problème de lissage. Pour les trois exemples ci-dessus Girard et Laurent [88] ont employé la méthode de GCV pour déterminer le paramètre α .

1.6 Notations et préliminaires

Dans cette section nous précisons certaines notations utilisées tout au long de cette thèse ainsi que les méthodes itératives utilisées pour résoudre les systèmes non linéaires résultant des problèmes de minimisation non quadratique.

Soit g une fonction à valeurs réelles définie sur un intervalle $[0,1]$ de la droite réelle. Soient aussi $0 \leq t_1 < \dots < t_n \leq 1$ n points de

l'intervalle $[0,1]$. On étudie le problème d'approcher g , en connaissant ses mesures $z=(z_1, \dots, z_n)^T$. Les mesures z sont soit les valeurs de g aux points $t_1, \dots, t_n$ avec des erreurs $\varepsilon_1, \dots, \varepsilon_n$ indépendantes, soit des observations indirectes sur g , c'est-à-dire, des z_i du type $z_i = \eta_i(g) + \varepsilon_i$ où η_i , $i=1, \dots, n$ sont des fonctionnelles non linéaires et connues. Ces fonctionnelles η_i peuvent être connues implicitement (par exemple dans les problèmes d'identification de systèmes, O'Sullivan [86]). Le second modèle permet aussi de traiter des contraintes de forme, par exemple de positivité en prenant $\eta_i(g) = \exp(g(t_i))$, la solution étant alors $\exp(g)$. Un exemple en restauration d'image (où g est fonction de 2 variables) est celui où la fonctionnelle η_i modélise un flou uniforme par bougé i.e. chaque pixel observé est en fait la moyenne de certains pixels voisins. Cela est une fonctionnelle linéaire, mais quand on impose la contrainte de positivité, on a une fonction non linéaire η_i qui dépend non seulement de $g(t_i)$ mais aussi de ses valeurs voisines (voir par ex. Harouche et al. [85]). Pour simplifier la présentation nous supposons que $\eta_i(g)$ ne dépend que de la valeur de g en $t_1, \dots, t_n$. On utilisera aussi l'abus de notation $\eta_i(g) = \eta_i(g(t_1), \dots, g(t_n))$. Les erreurs $\varepsilon_1, \dots, \varepsilon_n$ ne suivent pas toujours une loi gaussienne de variance uniforme, mais, par exemple, les données sont avec quelques points aberrants ou bien suivent une loi de Poisson, une loi binomiale, etc. Dans le cas classique chaque z_i est la réalisation d'une variable aléatoire Z_i de loi gaussienne centrée à $g(t_i)$ et de variance v^2 . C'est-à-dire la densité de probabilité de Z_i est:

$$d_{Z_i}(z_i) = \frac{1}{v\sqrt{2\pi}} \exp\left(-\frac{(g(t_i) - z_i)^2}{2v^2}\right).$$

Rappelons que la fonction de log-vraisemblance correspondant aux données z est alors $\ln\left(\prod_{i=1}^n d_{Z_i}(z_i)\right)$ car les données sont supposées indépendantes. Maximiser la fonction de log-vraisemblance par rapport à g (i.e. le principe de maximum de vraisemblance) est équivalent à minimiser la fonction négative de log-vraisemblance. Alors dans ce cas la fonction négative de log-vraisemblance correspondant aux données z est $\sum_{i=1}^n \phi(g(t_i), z_i) = \sum_{i=1}^n (g(t_i) - z_i)^2$.

En général, on notera la fonction négative de log-vraisemblance correspondant au modèle des données par $\sum_{i=1}^n \phi(\mu_i, z_i)$. Les erreurs ne suivent pas toujours une loi gaussienne et la fonction ϕ correspondante est non quadratique. Cette "non quadrativité" peut avoir d'origine suivante:

1. Soit les données sont des observations indirectes (non linéaires) sur g et les erreurs suivent une loi gaussienne. Alors la fonction ϕ n'est pas directement une fonction quadratique de $g(t_i)$ mais de $\eta_i(g)$ i.e. $\phi(\eta_i(g), z_i) = (\eta_i(g) - z_i)^2$, où la fonctionnelle η_i est non linéaire. Ce modèle permet de traiter des contraintes de forme aussi.
2. Soit les données sont avec des erreurs gaussienne mais avec quelques points aberrants. Alors la "distance" de Huber utilisée pour de telles données est

$$\phi(g(t_i), z_i) = \begin{cases} \frac{(g(t_i) - z_i)^2}{2\tau^2\alpha^2} & , |g(t_i) - z_i| \leq \tau\alpha, \\ \frac{|g(t_i) - z_i|}{\tau\alpha} - \frac{1}{2} & , |g(t_i) - z_i| \geq \tau\alpha. \end{cases}$$

où τ et α sont des constantes qui dépendent des données. Ici la fonction ϕ est directement une fonction de $g(t_i)$ mais non quadratique. Un autre exemple de ce type est quand les erreurs suivent une loi de Poisson et on impose les contraintes de positivité par changement de variable $\eta_i(g) = \exp(g(t_i))$. Alors on a $\phi(\eta_i(g), z_i) = \exp(g(t_i)) - z_i g(t_i)$. Nous verrons que ces deux exemples peuvent être traités par des approches similaires. Nous verrons au chapitre V un autre exemple qui nécessite un traitement différent.

Alors en général on va supposer que ϕ est une fonction non quadratique de $\eta_i(g)$ et aussi z_i .

D'autre part, on suppose que g est "lisse". Ceci veut dire, par exemple, que g appartient à $H^2[0,1]$ où:

$$H^2[0,1] = \left\{ \begin{array}{l} f: [0,1] \rightarrow \mathbb{R} / f, f' \text{ absolument continues} \\ \text{et } \int_0^1 [f''(t)]^2 dt < \infty \end{array} \right\}.$$

Pour résoudre ce problème, on considère la fonction spline cubique d'ajustement de paramètre ρ qui est la solution du problème suivant:

$$(1.11) \quad \min_{f \in H^2[0,1]} \left\{ \int_0^1 [f''(t)]^2 dt + \rho \sum_{i=1}^n \phi(\eta_i(f), z_i) \right\}.$$

C'est le principe du maximum de vraisemblance pénalisé. Comme il a été décrit dans la section 1.2, pour $n \geq 2$ le problème

$$\min_{\substack{f \in H^2[0,1] \\ f(t_i) = z_i, i=1, \dots, n}} \left\{ \int_0^1 [f''(t)]^2 dt \right\}$$

a une et une seule solution, que nous notons par $g_{\infty, z}$. De plus

$$\min_{\substack{f \in H^2[0,1] \\ f(t_i) = z_i, i=1, \dots, n}} \left\{ \int_0^1 [f''(t)]^2 dt \right\} = z^T \Omega z.$$

Toute solution (si elle existe) de (1.11) est donc une spline cubique naturelle $g_{\rho, z}$ avec $y_{\rho, z} = (g_{\rho, z}(t_1), \dots, g_{\rho, z}(t_n))^T$ solution de:

$$(1.12) \quad \min_{y \in \mathbb{R}^n} \left\{ y^T \Omega y + \rho \sum_{i=1}^n \phi(\eta_i(y), z_i) \right\}.$$

Supposons que ce problème ait une solution unique.

Pour des raisons de simplicité, $y_{\rho, z}$ sera notée soit y_ρ , soit y_z et l'opérateur non linéaire de lissage $z \rightarrow (\eta_1(y_z), \dots, \eta_n(y_z))^T$ sera noté B_ρ , i.e. nous avons $B_\rho(z) = (\eta_1(y_z), \dots, \eta_n(y_z))^T$.

De tels problèmes de minimisation (pour une valeur fixée de ρ) sont résolus par des méthodes itératives. Nous rappelons ici les itérations de Gauss-Newton et autres itérations de type Newton utilisées dans ce travail.

1.6.1 Itérations de type Newton

Nous rappelons les itérations de type Newton utilisées pour minimiser la fonction objective $\Psi_{p,z}(y)$ (dans (1.12)) donnée par

$$(1.13) \quad \Psi_{p,z}(y) = y^T \Omega y + \rho \sum_{i=1}^n \phi(\eta_i(y), z_i).$$

Itération de Newton:

Soit y^k le $k^{\text{ième}}$ itéré pour minimiser $\Psi_{p,z}(y)$. Considérons l'approximation quadratique du terme sous la sommation dans (1.13), qui exprime la fidélité aux données. Notons $f_i(y, z_i) = \phi(\eta_i(y), z_i)$.

Alors, l'approximation quadratique de $f_i(y, z_i)$ par rapport à y autour y^k est:

$$f_i(y, z_i) \approx f_i(y^k, z_i) + J_i(y^k, z_i)^T (y - y^k) + \frac{1}{2} (y - y^k)^T H_i(y^k, z_i) (y - y^k)$$

où $J_i(y, z_i)$ est le vecteur Jacobien de $f_i(y, z_i)$, i.e.

$$J_i(y, z_i) \equiv \nabla_y f_i(y, z_i) = \left(\frac{\partial f_i}{\partial y_1}(y, z_i), \dots, \frac{\partial f_i}{\partial y_n}(y, z_i) \right)^T$$

et $H_i(y, z_i)$ est la matrice Hessienne de $f_i(y, z_i)$ donnée par

$$H_i(y, z_i) \equiv \nabla_y^2 f_i(y, z_i) = \begin{bmatrix} \frac{\partial^2 f_i}{\partial y_1^2}(y, z_i) & \dots & \frac{\partial^2 f_i}{\partial y_1 \partial y_n}(y, z_i) \\ \vdots & & \vdots \\ \frac{\partial^2 f_i}{\partial y_1 \partial y_n}(y, z_i) & \dots & \frac{\partial^2 f_i}{\partial y_n^2}(y, z_i) \end{bmatrix}.$$

Alors, l'itéré y^{k+1} de Newton est défini comme l'élément qui minimise:

$$y^T \Omega y + \rho \sum_{i=1}^n \left[J_i(y^k, z_i)^T (y - y^k) + \frac{1}{2} (y - y^k)^T H_i(y^k, z_i) (y - y^k) \right].$$

Posons $W^k = \frac{1}{2} \sum_{i=1}^n H_i(y^k, z_i)$, $u^k = \frac{1}{2} \sum_{i=1}^n J_i(y^k, z_i)$ et $z^k = W^k y^k - u^k$. Ainsi

l'itération y^{k+1} est donnée par:

$$y^{k+1} = (\Omega + \rho W^k)^{-1} \rho z^k.$$

Si la matrice W^k est inversible,

$$y^{k+1} = (\Omega + \rho W^k)^{-1} \rho W^k \hat{z}^k$$

où $\hat{z}^k = y^k - W^{k-1} u^k$. Dans tous les exemples que nous allons considérer, W^k sera diagonale et donc le calcul de y^{k+1} est exactement le calcul d'une certaine spline avec poids (comme (1.8)).

Itération de Newton relaxée (Damped Newton):

L'itération de Newton relaxée est alors:

$$y_{DN}^{k+1} = y_{DN}^k + \tau_k (y_N^{k+1} - y_{DN}^k), \quad k=1,2,\dots$$

Ici, y_N^{k+1} désigne l'itéré de Newton ordinaire obtenu à partir de y_{DN}^k et τ_k est la première valeur prise dans la séquence $\left\{1, \frac{1}{2}, \frac{1}{4}, \dots\right\}$ qui fait diminuer la valeur de la fonction objective $\Psi_{\rho,z}(y)$. Ces itérations sont utilisées par O'Sullivan dans [88a], [88b] dans le contexte de lissage.

Itération de Newton Modifiée :

Pour les exemples considérés dans le chapitre II, nous avons utilisé l'itération de Newton modifiée décrite ci-dessous:

Si pour une valeur de k , W^k n'est pas une matrice définie positive, on ajoute une matrice D^k telle que $W_D^k = W^k + D^k$ soit définie positive.

L'itération de Newton Modifiée est alors:

$$y^{k+1} = (\Omega + \rho W_D^k)^{-1} \rho W_D^k \hat{z}^k$$

où $\hat{z}^k = y^k - W_D^{k-1} u^k$. Cette itération est localement convergente sous certaines conditions sur D^k (voir Ortega et Rheinboldt [70]). Lorsque les itérations convergent, nous notons y_z sa limite et définissons l'opérateur A_p^{MN} par:

$$A_p^{MN} = (\Omega + \rho W)^{-1} \rho W$$

où $W = \frac{1}{2} \sum_{i=1}^n H_i(y_z, z_i)$.

1.6.2 Itérations de Gauss-Newton

Si $\phi(\eta_i(y), z_i) = (\eta_i(y) - z_i)^2$, nous pouvons appliquer l'iteration de Gauss-Newton comme le font O'Sullivan et Wahba [84] pour minimiser la fonction objective suivante:

$$\Psi_{p,z}(y) = y^T \Omega y + p \sum_{i=1}^n (\eta_i(y) - z_i)^2.$$

Soit y^k le $k^{\text{ième}}$ itéré pour minimiser $\Psi_{p,z}(y)$. Nous dénotons $\bar{f}(y, z) = (\eta_1(y) - z_1, \dots, \eta_n(y) - z_n)^T$. Alors, $\Psi_{p,z}(y) = y^T \Omega y + p \|\bar{f}(y, z)\|_2^2$. On suppose que les deux premières dérivées partielles de chaque η_i sont absolument continues. L'approche classique, appelée méthode de Gauss-Newton consiste à considérer l'approximation linéaire de chaque η_i dans l'expression de f , en utilisant: $\bar{f}(y, z) \approx \bar{f}(y^k, z) + J(y^k)(y - y^k)$, où J est la matrice Jacobienne de $\bar{f}$ donnée par

$$J(y) \equiv \nabla_y \bar{f}(y, z) = \begin{bmatrix} \frac{\partial \eta_1}{\partial y_1} & \dots & \frac{\partial \eta_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial \eta_n}{\partial y_1} & \dots & \frac{\partial \eta_n}{\partial y_n} \end{bmatrix}.$$

L'itéré y^{k+1} est obtenu en minimisant:

$$y^T \Omega y + p \|\bar{f}(y^k, z) + J(y^k)(y - y^k)\|_2^2$$

i.e.

$$y^T \Omega y + p [h(y^k)^T (y - y^k) + (J(y^k)(y - y^k))^T (J(y^k)(y - y^k))]$$

où $h(y) \equiv \nabla_y (\bar{f}^T \bar{f}) = 2J(y)^T \bar{f}(y, z)$. Alors, cet itéré est défini par

$$y^{k+1} = \left(\Omega + p \Theta^k \Theta^k \right)^{-1} p \Theta^k \tilde{z}^k$$

où $\Theta^k = J(y^k)$ et $\tilde{z}^k = \Theta^k y^k - \bar{f}(y^k, z)$. Le procédé pour le calcul d'une certaine spline avec poids (comme (1.8)) peut être facilement adapté pour le calcul de y^{k+1} . Donc le coût d'une itération reste le même que d'une certaine spline avec poids. Lorsque ces itérations convergent, nous notons y_z sa limite et l'opérateur de lissage correspondant A_p^{GN} par

$$A_p^{GN} = \Theta (\Omega + \rho \Theta^T \Theta)^{-1} \rho \Theta^T$$

où $\Theta = J(y_z)$.

Pour obtenir ensuite la valeur de ρ qui minimise $V(\rho)$, on utilise la méthode de la section dorée qui est une méthode de recherche globale si la fonction à minimiser est unimodale. Cette méthode de minimisation est une méthode qui permet d'obtenir à chaque pas un intervalle où se trouve la valeur optimale. Il reste à choisir la précision pour laquelle on considère que le calcul est terminé. On a utilisé le critère suivant pour arrêter la procédure:

$$\frac{\rho_2 - \rho_1}{(\rho_1 + \rho_2)/2} < 10^{-3}$$

où $[\rho_1, \rho_2]$ est l'intervalle fourni par l'algorithme. Les nombres pseudo-aléatoires de loi normale de moyenne nulle et de variance v^2 sont générés par la méthode directe (avec l'approximation rationnelle classique cf. Abramowitz et Stegun [72], formule 26.2.23 pour l'inverse de la fonction de répartition).

Tous les calculs présentés dans cette thèse sont faits sur Macintosh II en double précision utilisant Think Pascal.

Chapitre II

GCV pour les problèmes de moindres carrés non linéaires

2.1 Introduction

Dans certaines branches des sciences physiques et biologiques, le phénomène désiré ne peut être observé directement. Cependant on peut disposer d'observations indirectes sur certaines fonctions appliquées à la fonction g recherchée. (par exemple O'Sullivan et Wahba [84], O'Sullivan [86] et les références citées dans cette article) Nous considérons ici le modèle $z_i = \eta_i(g) + \varepsilon_i$, $i=1, \dots, n$ pour g dans $H^2[0,1]$, où η_i , $i=1, \dots, n$ sont certaines fonctionnelles non linéaires données et $\varepsilon_1, \dots, \varepsilon_n$ sont n variables aléatoires indépendantes de moyenne nulle et de variance v^2 , laquelle peut être inconnue. Le but est de trouver une estimation de g à partir des données z . Ceci est obtenu en cherchant $g_{p,z}$ qui minimise

$$(2.1) \quad \int_0^1 [f''(t)]^2 dt + \rho \sum_{i=1}^n [\eta_i(f) - z_i]^2$$

pour $f \in H^2[0,1]$ et $\rho > 0$ donné. Le paramètre de lissage ρ contrôle le compromis entre la fidélité aux données mesurées par $\sum_{i=1}^n [\eta_i(f) - z_i]^2$ et le caractère lisse de la solution mesurée par

$$\int_0^1 [f''(t)]^2 dt.$$

Remarquons que cette formulation non quadratique du problème de lissage est une approche classique pour estimer une fonction

positive ou bornée à partir de données erronées. En pratique, le choix de ρ est tout à fait crucial. Dans le chapitre précédent nous avons vu que pour le problème classique de lissage par des splines, ce choix est souvent fait par la méthode de validation croisée généralisée (GCV). Dans ce chapitre nous proposons une approche générale pour le calcul de la GCV pour des problèmes de moindres carrés non linéaires comme (2.1).

Quand $\eta_i(f)$ est une fonction uniquement de $f(t_1), \dots, f(t_n)$, toute solution (si elle existe) du problème de la minimisation de la fonction objective donnée par (2.1) est une spline cubique naturelle $g_{\rho,z}$ telle que $y_{\rho,z} = (g_{\rho,z}(t_1), \dots, g_{\rho,z}(t_n))^T$ est la solution de

$$\min_{y \in \mathbb{R}^n} \left\{ y^T \Omega y + \rho \sum_{i=1}^n [\eta_i(y) - z_i]^2 \right\}.$$

Ce problème non quadratique de minimisation pour un ρ donné est résolu en utilisant l'itération de Gauss-Newton rappelée dans le chapitre I, section 1.6.2.

La section 2.2 est consacrée au choix du paramètre de lissage par la méthode de GCV. L'expression de la fonction GCV utilise la dérivée de l'opérateur non linéaire de lissage B_ρ par rapport à z . Dans l'annexe, nous donnons une expression explicite de cette dérivée pour le cas plus général.

Nous proposons deux méthodes de Monte-Carlo pour le calcul de la trace de la dérivée de B_ρ , intervenant dans la GCV. Si on connaît l'expression explicite de la dérivée de B_ρ , la première méthode est une extension directe de la méthode de Girard proposée dans le cas linéaire. La deuxième est une formule originale qui utilise l'approximation de la dérivée par des différences finis dont l'évaluation ne nécessite que le calcul d'une spline supplémentaire associée à des données perturbées. Dans la section 2.3, on compare les traces exactes des opérateurs intervenant dans la GCV et leurs estimateurs de Monte-Carlo.

Les expériences numériques effectuées et les résultats obtenus sont présentés dans la section 2.4.

2.2 Choix du ρ par validation croisée généralisée

Dans ce paragraphe nous développons la méthode de validation croisée pour choisir le paramètre de lissage associé aux problèmes de lissage de moindres carrés non linéaires du type (2.1).

L'idée de base est comme dans le cas classique décrit dans la section 1.3. Soit $y_p^{(k)}$ la solution calculée à partir de toutes les données excepté z_k en utilisant la valeur p pour le paramètre de lissage, i.e. $y_p^{(k)}$ est le vecteur qui réalise le minimum de:

$$y^T \Omega y + \rho \sum_{\substack{i=1 \\ i \neq k}}^n [\eta_i(y) - z_i]^2.$$

Alors, la fonction de validation croisée ordinaire $V_o(p)$ est définie par:

$$(2.2) \quad V_o(p) = \frac{1}{n} \sum_{k=1}^n [\eta_k(y_p^{(k)}) - z_k]^2.$$

On peut voir facilement que $y_p^{(k)}$ réalise également le minimum de:

$$y^T \Omega y + \rho \sum_{\substack{i=1 \\ i \neq k}}^n [\eta_i(y) - z_i]^2 + \rho [\eta_k(y) - \eta_k(y_p^{(k)})]^2.$$

On a donc:

$$B_p(z_1, \dots, z_{k-1}, \eta_k(y_p^{(k)}), z_{k+1}, \dots, z_n)^T = (\eta_1(y_p^{(k)}), \dots, \eta_n(y_p^{(k)}))^T$$

i.e.

$$(2.3) \quad \eta_k(y_p^{(k)}) = [B_p(z + \delta_k e^k)]_k$$

où e^k est le $k^{\text{ième}}$ vecteur de la base canonique de $\mathbb{R}^n$, et $\delta_k = \eta_k(y_p^{(k)}) - z_k$. En substituant (2.3) dans (2.2) nous obtenons:

$$V_o(p) = \frac{1}{n} \sum_{k=1}^n \frac{[\eta_k(y_p) - z_k]^2}{[1 - a_{kk}^*(p)]^2}$$

où

$$a_{kk}^*(p) = \frac{[B_p(z + \delta_k e^k)]_k - [B_p(z)]_k}{\delta_k}.$$

Par analogie avec le cas de la spline de lissage classique, nous

remplaçons $\dot{a}_{kk}(\rho)$ par $\frac{1}{n} \sum_{j=1}^n \dot{a}_{jj}(\rho)$ pour obtenir la validation croisée généralisée (GCV) donnée par:

$$V(\rho) = \frac{\frac{1}{n} \sum_{k=1}^n [(B_\rho(z))_k - z_k]^2}{\left[\frac{1}{n} \left(n - \sum_{k=1}^n \dot{a}_{kk}(\rho) \right) \right]^2}.$$

Lorsque ρ est bien choisi, la quantité $\delta_k = \eta_k(y_\rho^{(k)}) - z_k$ devrait être petite. Ainsi, $\dot{a}_{kk}(\rho)$ peut être approché par $\frac{\partial (B_\rho(z))_k}{\partial z_k}$ et la fonction GCV par:

$$V(\rho) \equiv \frac{\frac{1}{n} \sum_{k=1}^n [(B_\rho(z))_k - z_k]^2}{\left[\frac{1}{n} \text{Tr} \left(I - (B'_\rho)_z \right) \right]^2}.$$

O'Sullivan et Wahba [84] ont approché $(B'_\rho)_z$ par la matrice de lissage associée au dernier problème quadratique à la fin des itérations. Une autre façon est d'appliquer le principe de GCV classique à la forme pseudo linéarisée i.e. au dernier problème quadratique que l'on résout lors des itérations. Ceci est fait dans O'Sullivan et al [86] et par O'Sullivan [88b]. Dans l'expression de validation croisée ordinaire donné par (2.2), O'Sullivan [88a] approche la solution $y_\rho^{(k)}$ par une itération de Newton Raphson à partir de y_ρ . Cette approximation donne une expression pour la GCV en terme de la trace de la matrice de lissage obtenue à la fin des itérations.

2.2.1 Expression analytique pour $(B'_\rho)_z$

L'expression analytique pour $(B'_\rho)_z$ est donnée dans l'annexe pour le problème général quand on minimise $\Psi_{\rho,z}(y)$ donné par:

$$y^T \Omega y + \rho \sum_{i=1}^n \phi(\eta_i(y), z_i).$$

Ici on a $\phi(\eta_i(y), z_i) = f_i(y, z_i) = (\eta_i(y) - z_i)^2$ et on suppose que $\eta_i(y) = \psi(y_i)$

pour $\psi: \mathbb{R} \rightarrow \mathbb{R}$. Donc on a: $\Theta = \text{diag}(\psi'(y_{z,1}), \dots, \psi'(y_{z,n}))$, $V = -2\Theta$,
 $W = \text{diag}(w_1, \dots, w_n)$ où $w_i = (\psi'(y_{z,i}))^2 + (\psi(y_{z,i}) - z_i) \psi''(y_{z,i})$ et

$$(B'_p)_z = 2\Theta(\Omega + \rho W)^{-1} \rho \Theta.$$

Nous allons utiliser cette expression pour le calcul de la trace dans la fonction de GCV. Si la matrice W est inversible et diagonale, le calcul de $(B'_p)_z(\cdot)$ est exactement le calcul d'une certaine spline avec poids. Au contraire, si la matrice W n'est pas inversible, l'algorithme pour la spline de lissage avec poids peut être facilement modifié pour le problème

$$\min_{y \in \mathbb{R}^n} \left\{ y^T \Omega y + \rho \sum_{i=1}^n [w_i (y_i - z_i)^2 + \beta_i y_i] \right\}.$$

Le coût de calcul de cet algorithme modifié reste linéaire en n .

2.3 Calcul de la trace par Monte-Carlo

Dans ce paragraphe on présente une comparaison entre le calcul exact de la trace et ses estimateurs de Monte-Carlo que l'on a proposés dans Deshpande et Girard [89].

La méthode de GCV nécessite le calcul de $\frac{1}{n} \text{Tr}(I - (B'_p)_z)$. Dans le paragraphe précédent nous avons donné une expression explicite de la dérivée de B_p . Nous notons ici:

$$T_{DE}(\rho) = \frac{1}{n} \text{Tr}(I - (B'_p)_z).$$

Dans la littérature, $(B'_p)_z$ est approché par la matrice de lissage obtenue à la fin des itérations. Pour faire une comparaison on considère aussi les deux traces suivantes:

$$T_{NM}(\rho) = \frac{1}{n} \text{Tr}(I - A_p^{NM}), \quad T_{GN}(\rho) = \frac{1}{n} \text{Tr}(I - A_p^{GN}).$$

On ne considère pas la méthode de Newton non modifiée car même dans les exemples simples considérées dans la suite, la fonctionnelle à minimiser n'est pas partout convexe.

Maintenant, nous proposons deux estimateurs de Monte-Carlo pour ce calcul de la trace. La première est une extension directe de la méthode de Girard proposé dans le cas linéaire. Pour cela soit w

un vecteur de n variables aléatoires indépendantes normales de moyenne nulle et de variance 1. Alors les estimateurs de Monte-Carlo de $T_{DE}(\rho)$, $T_{NM}(\rho)$ et $T_{GN}(\rho)$ sont donnés par:

$$T_{DE}^*(\rho) = \frac{w^T (I - (B'_\rho)_z) w}{w^T w},$$

$$T_{NM}^*(\rho) = \frac{w^T (I - A_\rho^{NM}) w}{w^T w}, \quad T_{GN}^*(\rho) = \frac{w^T (I - A_\rho^{GN}) w}{w^T w}.$$

On propose maintenant un autre estimateur de Monte-Carlo de T_{DE} . Il est basé sur l'approximation de la dérivée $(B'_\rho)_z w$ par des différences finies de la manière suivante:

$$(B'_\rho)_z w \approx \frac{B_\rho(z+hw) - B_\rho(z)}{h} \quad \text{pour } h>0 \text{ convenablement choisi.}$$

Alors, le deuxième estimateur de la trace est obtenue en utilisant ces différences finies:

$$T_{DF}^*(\rho) = \frac{w^T}{w^T w} \left(w - \frac{B_\rho(z+hw) - B_\rho(z)}{h} \right).$$

On remarque que cet estimateur de Monte-Carlo T_{DF}^* est très utile quand on ne connaît pas l'expression explicite de $(B'_\rho)_z$ ou quand le calcul de $(B'_\rho)_z(\cdot)$ est très coûteux.

Pour les expériences numériques nous avons choisi les points t_i , $i=1, \dots, n$ équidistants et $n=200$. Les données fictives sont obtenues comme décrit dans la section 2.4 avec la fonction $\eta(g_1)$. Nous calculons les traces exactes T_{DE} , T_{NM} , T_{GN} et 6 simulations de leur estimateur de Monte-Carlo pour 28 valeurs de ρ (équidistantes en échelle logarithmique) dans l'intervalle $[1, 10^9]$. Pour chaque simulation le vecteur w était régénéré.

Sur les figures 2.1 et 2.2, on représente les trois traces exactes, fonctions de ρ , pour les deux choix de η : respectivement $\eta_i(y) = \exp(y_i)$ et $\eta_i(y) = y_i^2$. On représente ensuite la trace exacte et 6 simulations de son estimation par Monte-Carlo. Sur ces figures les valeurs du paramètre ρ et les traces associées sont présentées en échelles logarithmiques.

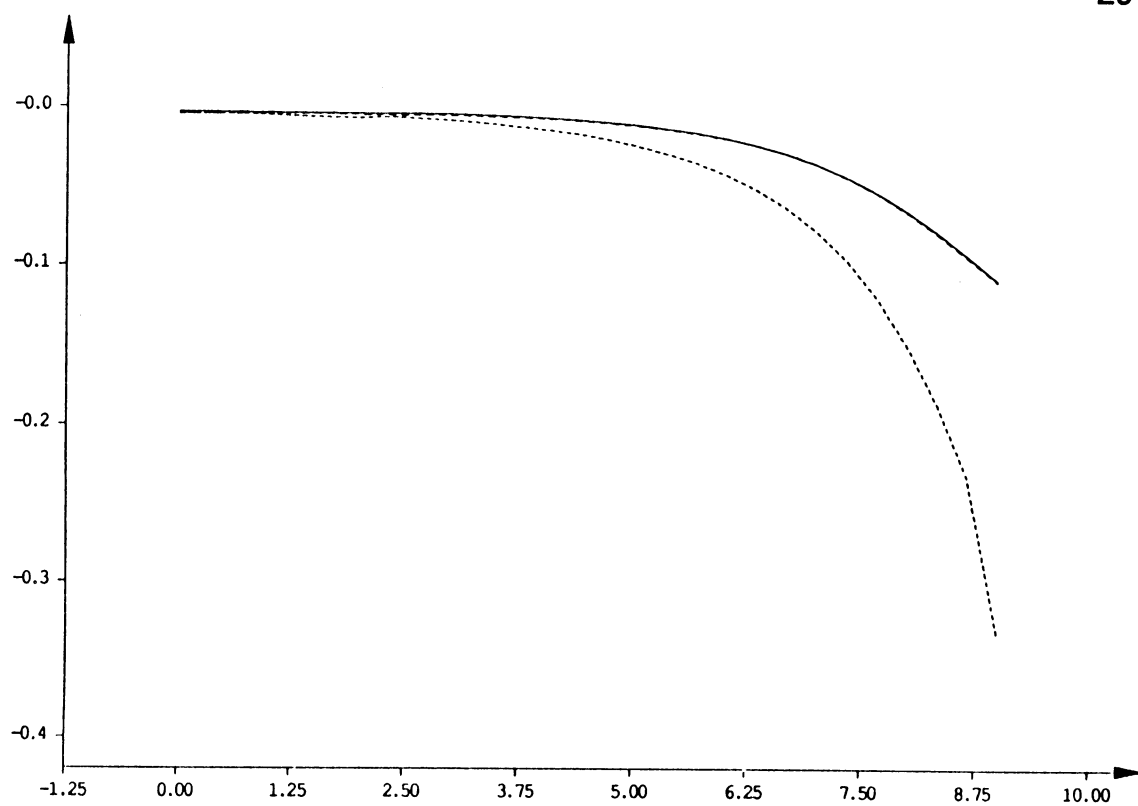


Figure 2.1 $\eta_i(y) = \exp(y_i)$

— : T_{DE} , - - - : T_{NM} , : T_{GN}

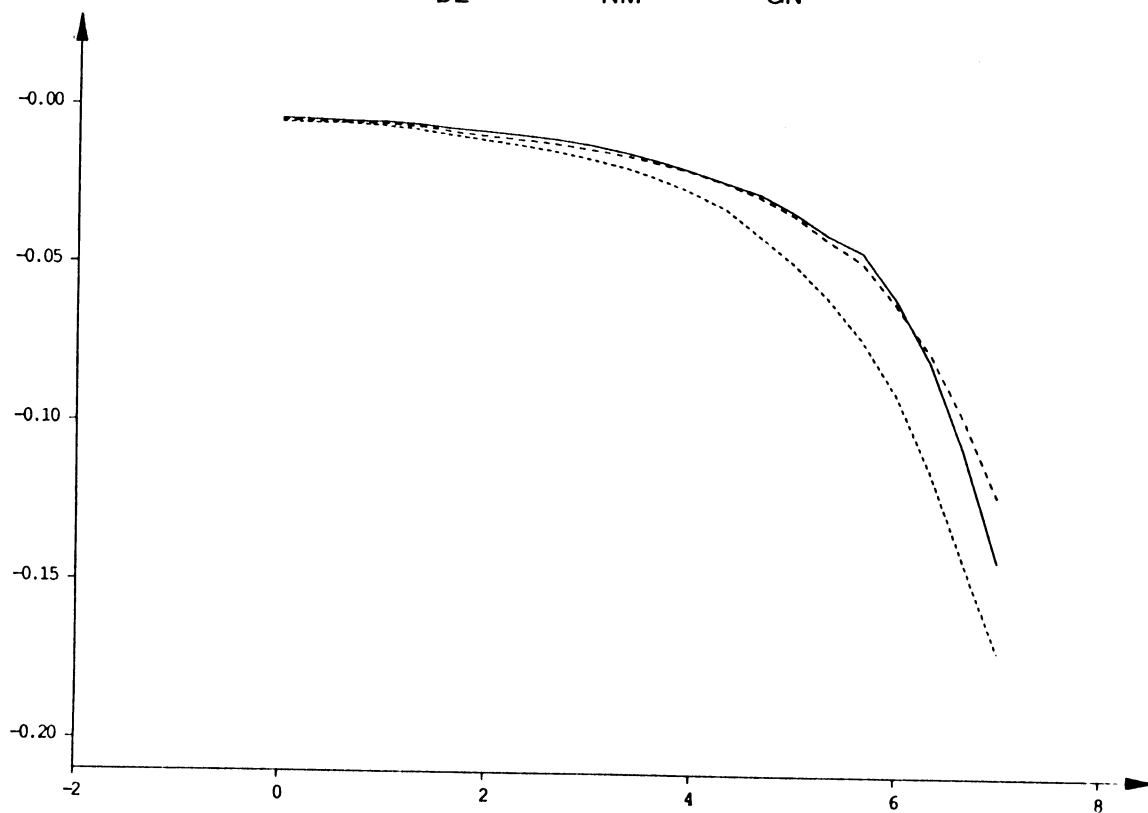


Figure 2.2 $\eta_i(y) = y_i^2$

— : T_{DE} , - - - : T_{NM} , : T_{GN}

30

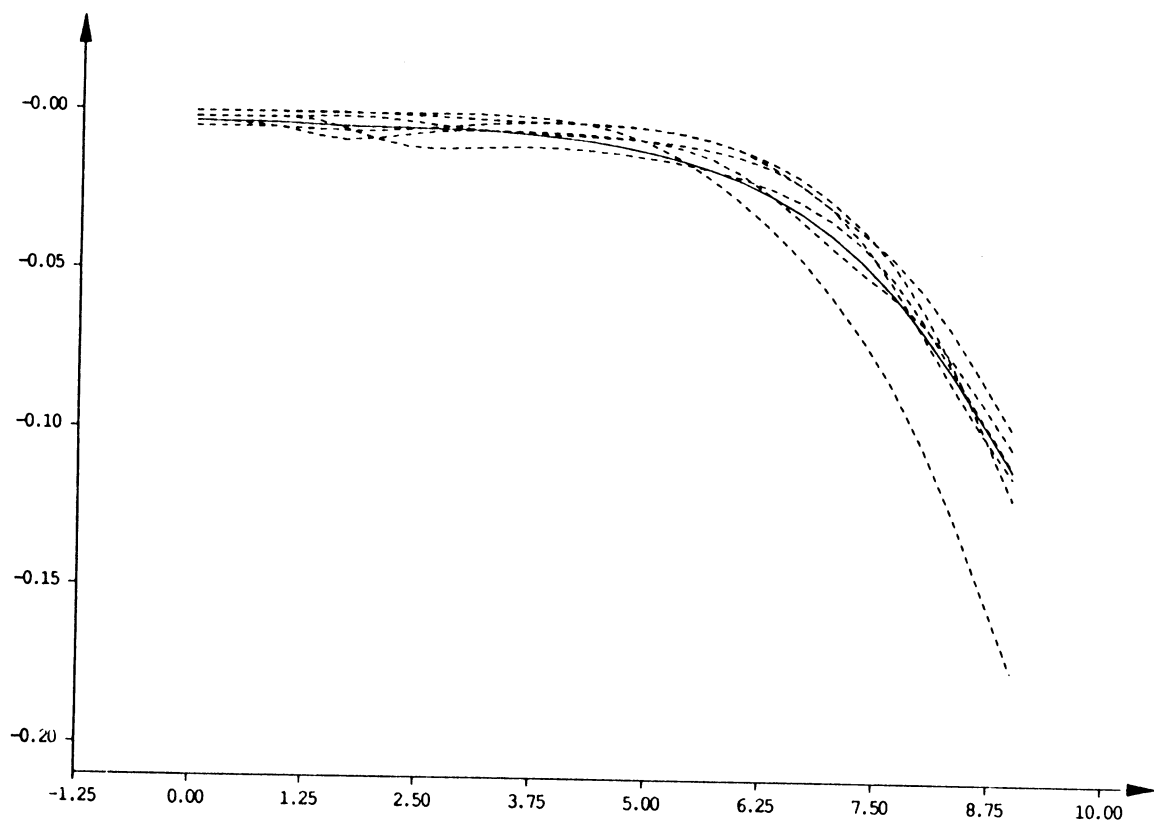


Figure 2.3a $\eta_i(y)=\exp(y_i)$

— : T_{DE} , - - - : T_{DE}^*

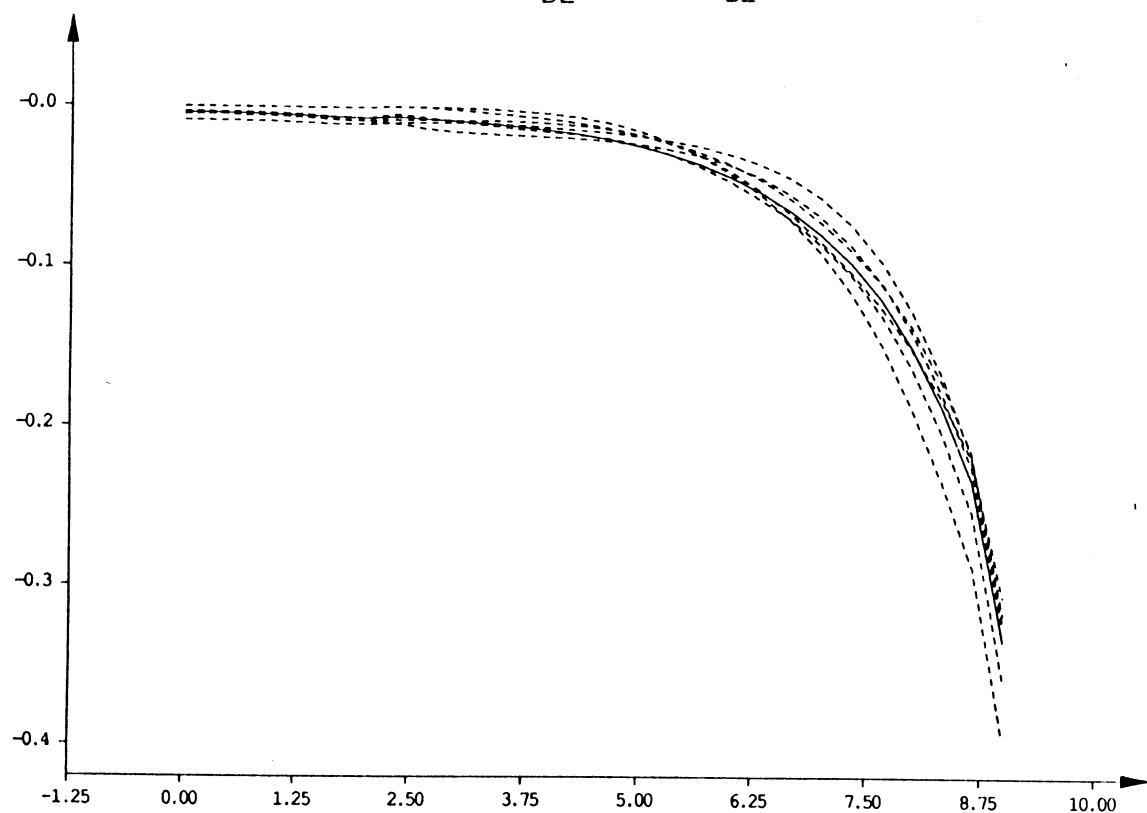


Figure 2.3b $\eta_i(y)=\exp(y_i)$

— : T_{NM} , - - - : T_{NM}^*

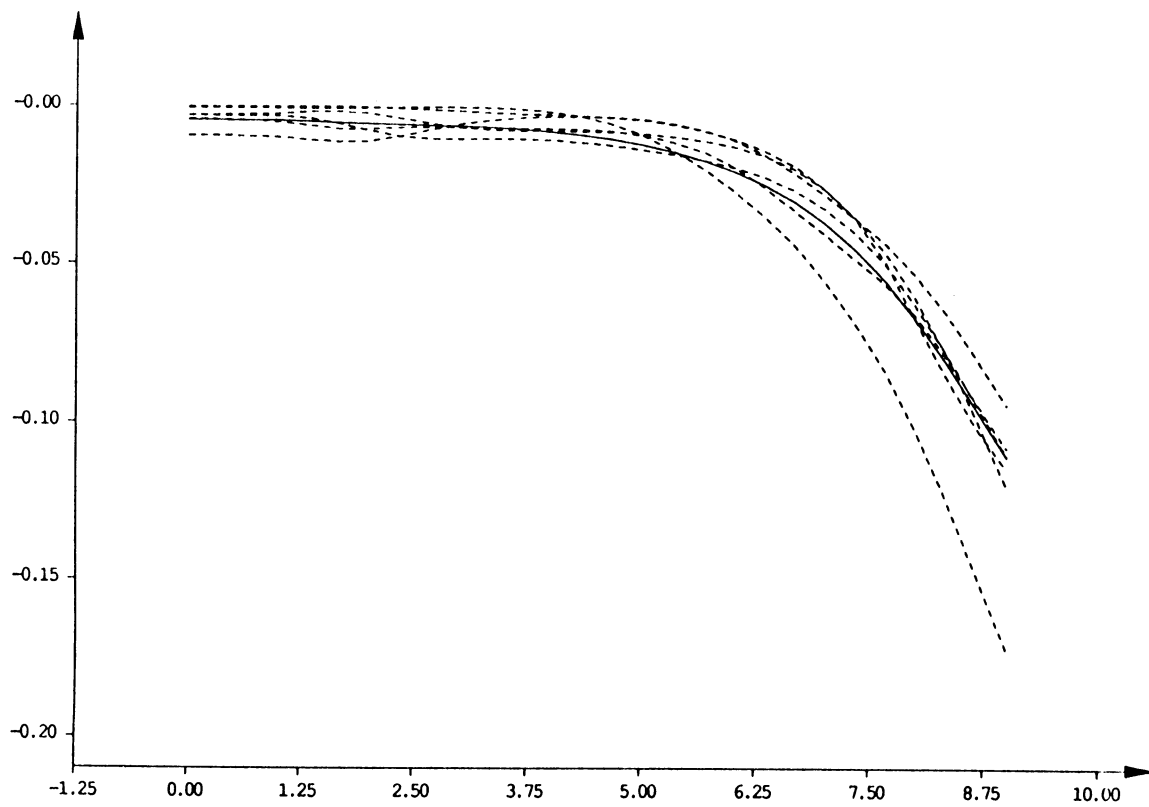


Figure 2.3c $\eta_i(y) = \exp(y_i)$

— : T_{GN} , - - - : T_{GN}^*

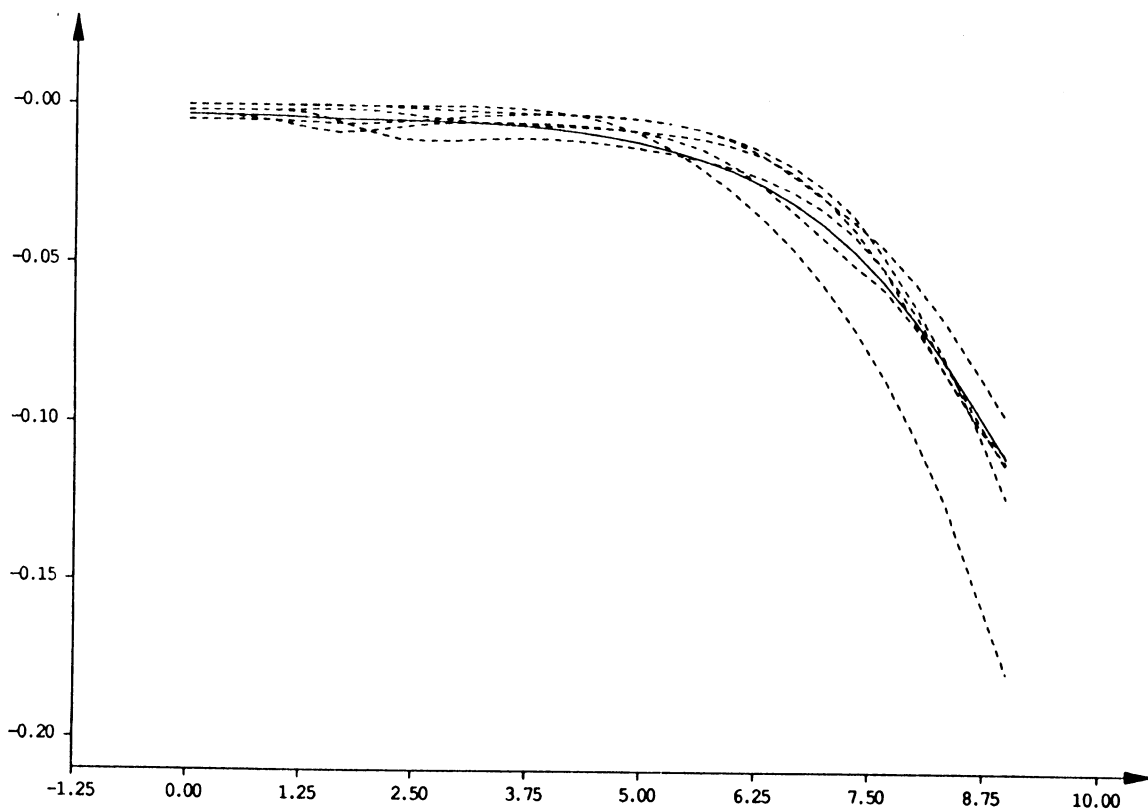


Figure 2.3d $\eta_i(y) = \exp(y_i)$

— : T_{DE} , - - - : T_{DF}

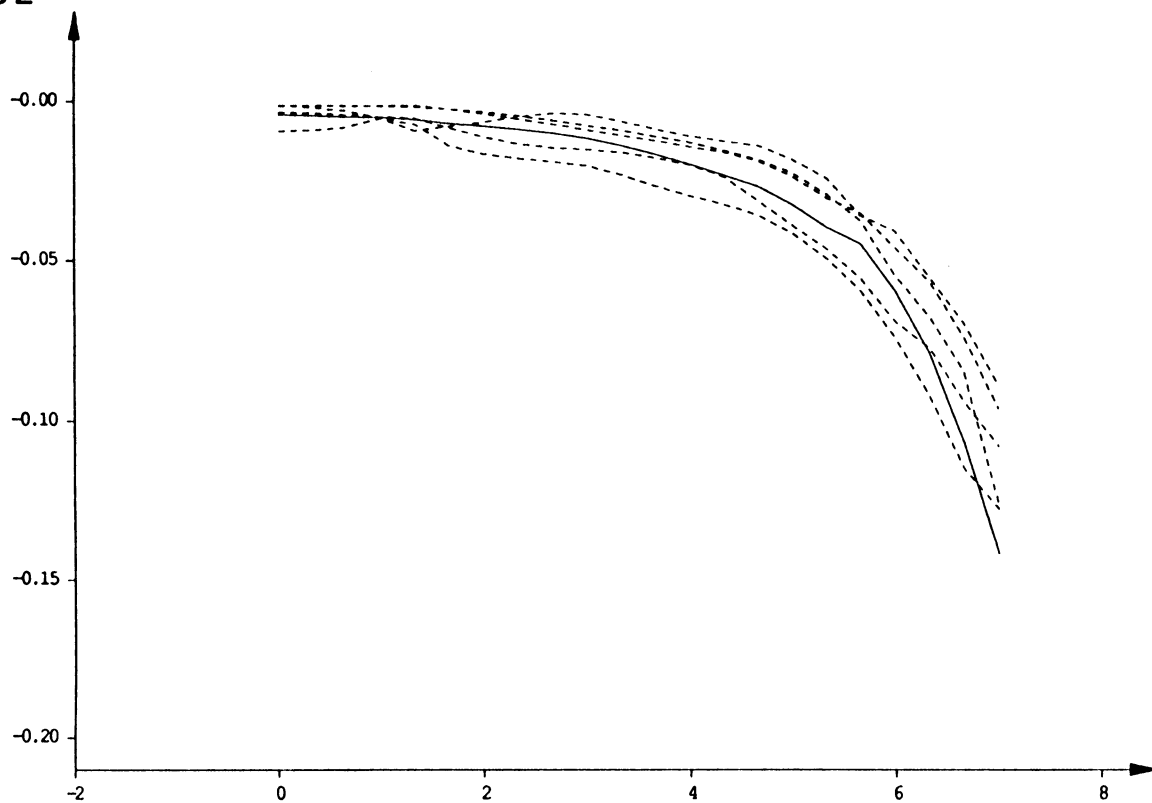


Figure 2.4a $\eta_i(y) = y_i^2$
 — : T_{DE} , - - : T_{DE}^*

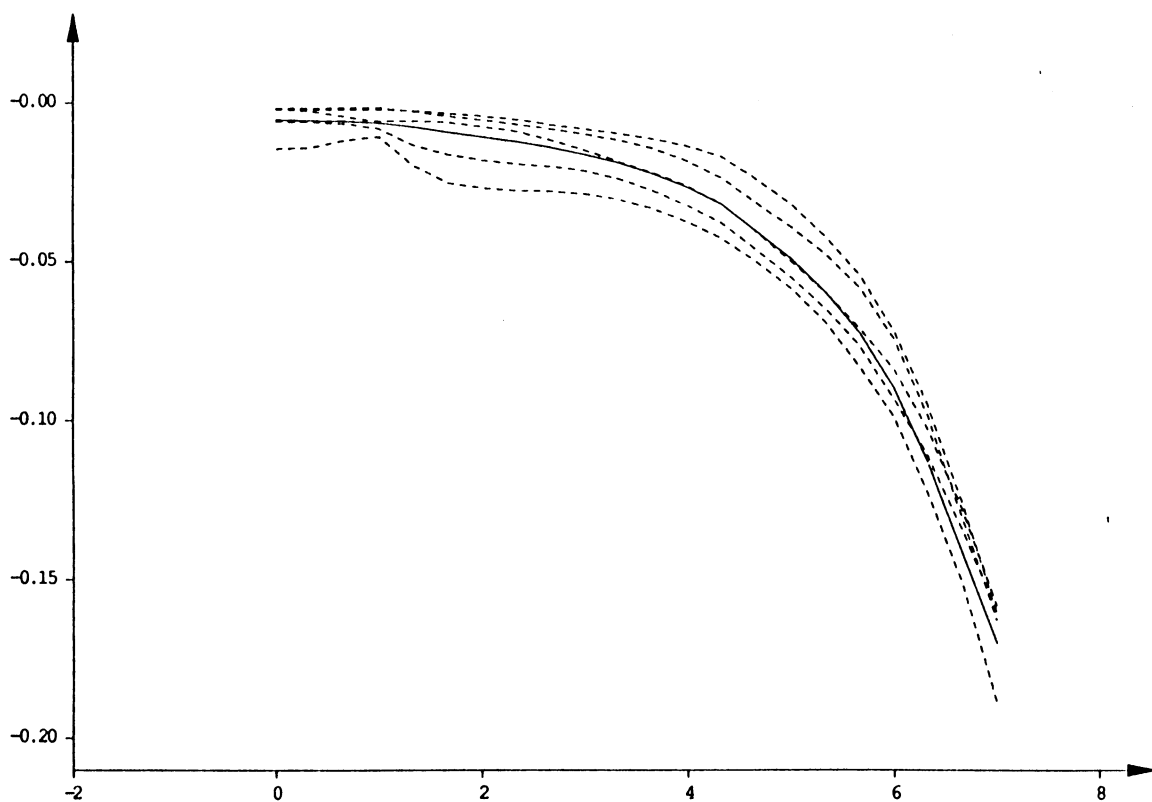


Figure 2.4b $\eta_i(y) = y_i^2$
 — : T_{NM} , - - : T_{NM}^*

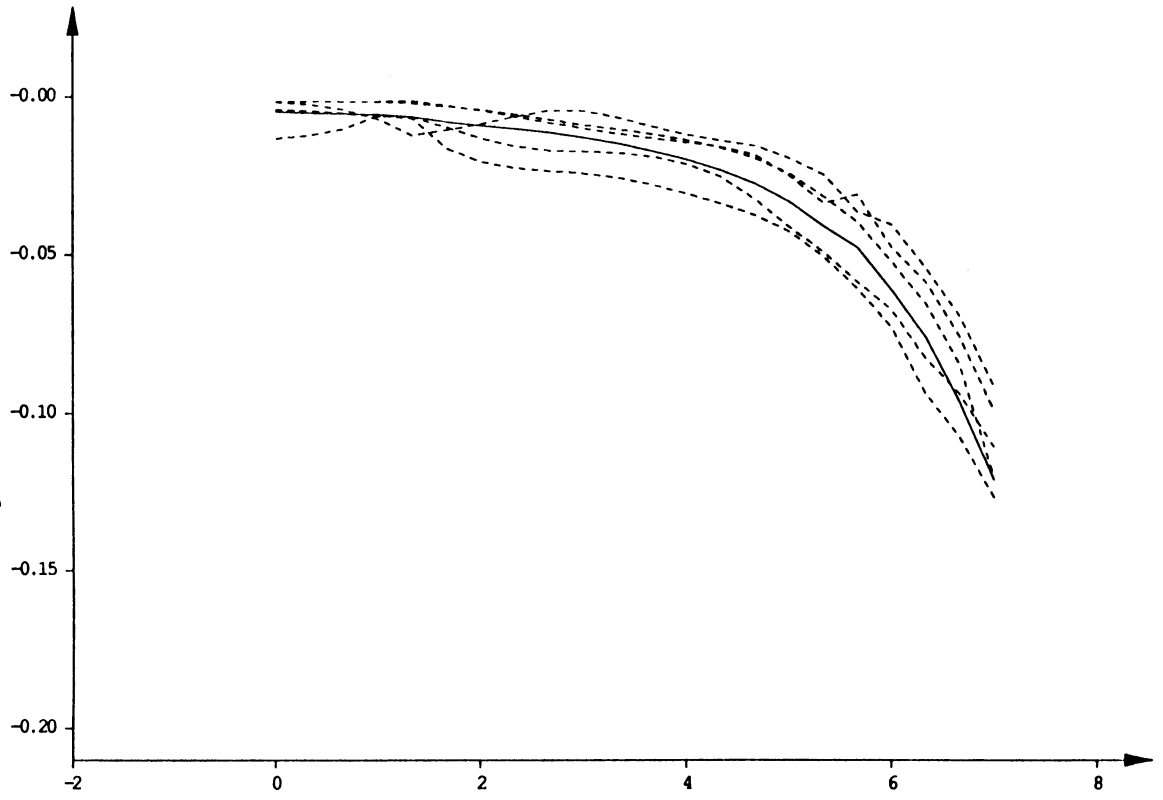


Figure 2.4c $\eta_i(y) = y_i^2$
 — : T_{GN} , - - : $\dot{T}_{GN}$

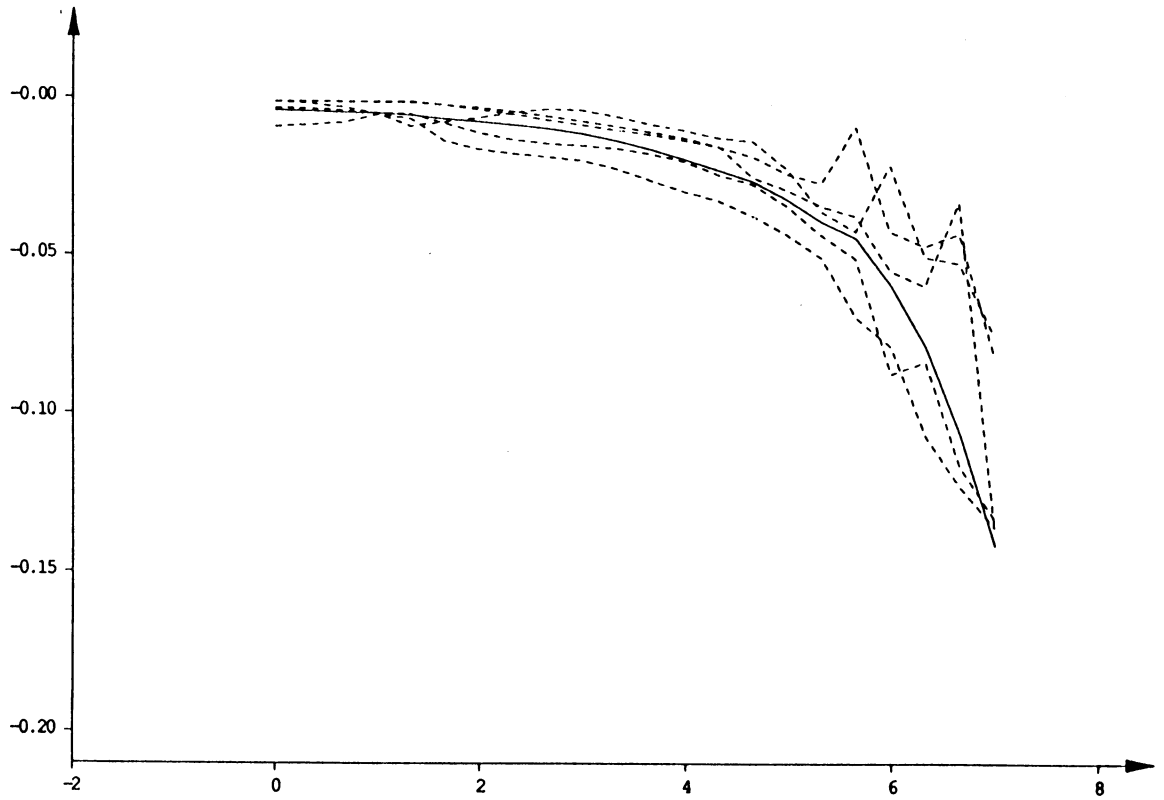


Figure 2.4d $\eta_i(y) = y_i^2$
 — : T_{DE} , - - : $\dot{T}_{DF}$

On observe que la variation dans les estimateurs de Monte-Carlo par rapport au vecteur aléatoire paraît grande, c'est-à-dire qu'une seule simulation de Monte-Carlo n'est pas une 'très bonne' approximation de la trace. Ainsi dans la section 2.4 pour le calcul de la GCV, l'estimateur de Monte-Carlo est calculé par la moyenne de 5 simulations. Contrairement au cas classique, ici les estimateurs de Monte-Carlo sont des fonctions de p moins 'régulières' que les traces exactes (qui paraissent monotones). Ici nous avons pris $h=10^{-2}$. Une remarque sur le choix de h est fait à la fin de la section 2.4.

2.4 Expériences numériques

Le but de ce paragraphe est d'illustrer à l'aide d'expériences numériques le comportement de la méthode de validation croisée. Pour réduire le temps de calcul et aussi la complexité nous nous restreignons ici au cas 'diagonal', c'est-à-dire $\eta_i(y)$ est une fonction seulement de y_i . Les deux exemples de η_i considérés dans ce paragraphe sont $\eta_i(y)=\exp(y_i)$ et $\eta_i(y)=y_i^2$, $i=1,\dots,n$. Donc notons $\eta_i(y)=\psi(y_i)$ pour $\psi:\mathbb{R}\rightarrow\mathbb{R}$. Ces deux choix de η_i rend l'estimateur de g positive.

Pour tous les exemples de ce paragraphe on considère que les points t_i , $i=1,\dots,n$ sont équidistants et $n=200$. Les données fictives sont générées par:

$$z_i = \psi(g(t_i)) + \varepsilon_i, \quad i=1,\dots,n,$$

où g est une fonction connue.

La méthode de GCV consiste à minimiser la fonction $V(\rho)$ donnée par:

$$V(\rho) = \frac{\frac{1}{n} \sum_{k=1}^n [(B_\rho(z))_k - z_k]^2}{\left[\frac{1}{n} \text{Tr} (I - (B'_\rho)_z) \right]^2}.$$

Soit $\text{Num} = \frac{1}{n} \sum_{k=1}^n [(B_\rho(z))_k - z_k]^2$. Selon la méthode de calcul de la trace on a:

$$V_{DE}(\rho) = \frac{\text{Num}}{[T_{DE}(\rho)]^2}, \quad V_{DE}^*(\rho) = \frac{\text{Num}}{[T_{DE}^*(\rho)]^2}$$

$$V_{NM}(\rho) = \frac{\text{Num}}{[T_{NM}(\rho)]^2}, \quad V_{NM}^*(\rho) = \frac{\text{Num}}{[T_{NM}^*(\rho)]^2}$$

$$V_{GN}(\rho) = \frac{\text{Num}}{[T_{GN}(\rho)]^2}, \quad V_{GN}^*(\rho) = \frac{\text{Num}}{[T_{GN}^*(\rho)]^2}$$

et

$$V_{DF}^*(\rho) = \frac{\text{Num}}{[T_{DF}^*(\rho)]^2}.$$

Les estimateurs de Monte-Carlo sont obtenus par la moyenne de 5 simulations. Soient ρ_{DE} , ρ_{DE}^* , ρ_{NM} , ρ_{NM}^* , ρ_{GN} , ρ_{GN}^* et ρ_{DF}^* les valeurs de ρ qui minimisent respectivement $V_{DE}(\rho)$, $V_{DE}^*(\rho)$, $V_{NM}(\rho)$, $V_{NM}^*(\rho)$, $V_{GN}(\rho)$, $V_{GN}^*(\rho)$ et $V_{DF}^*(\rho)$.

Pour trouver la solution $B_p(z)$ pour un ρ donné on utilise la méthode itérative de Gauss-Newton décrite dans la section 1.6.2. Pour obtenir ensuite la valeur de ρ qui minimise la fonction GCV on utilise la méthode de section dorée.

Exemple d'utilisation

Pour illustrer le comportement de la méthode de GCV nous considérons le problème d'approcher une fonction positive à partir de données erronées. On considère la fonction

$$\psi(g_1(t)) = \exp\left(\frac{-(t-0.5)^2}{2 \times 0.06^2}\right), \quad t \in [0,1].$$

avec un écart type de $v=0.1$ pour les erreurs.

Pour un choix de ψ les splines de lissage obtenues par rapport à la valeur de $\rho \in \{\rho_{DE}, \rho_{DE}^*, \rho_{NM}, \rho_{NM}^*, \rho_{GN}, \rho_{GN}^*, \rho_{DF}^*\}$ sont visuellement les mêmes. Sur la figure 2.5 on représente les données (par +), la fonction théorique (par la ligne continue) et les fonctions splines obtenues par la GCV avec les deux fonctions ψ (e.g. $\psi(y)=\exp(y)$ par -----, et $\psi(y)=y^2$ par ----). Pour faire une comparaison sur la figure 2.6 on représente les données, la fonction théorique et la spline classique avec ρ choisi par la GCV. On voit bien que la fonction spline obtenue par les moindres carrés non linéaires est

une meilleur approximation.

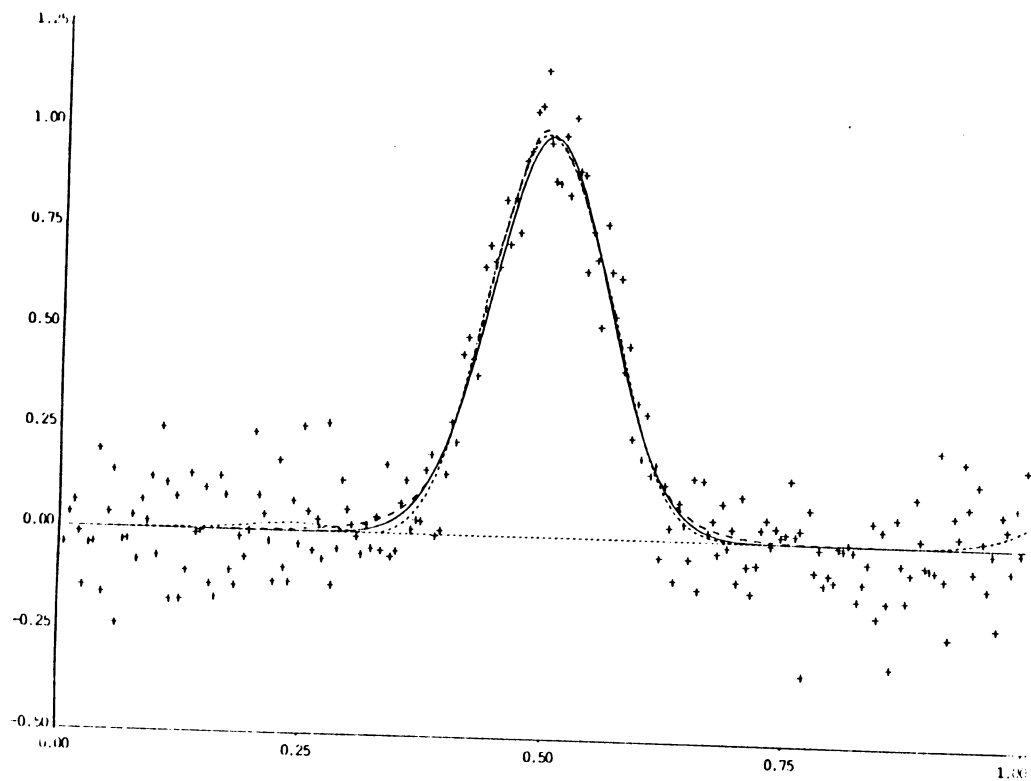


Figure 2.5

----- : $\psi(y) = \exp(y)$, ----- : $\psi(y) = y^2$

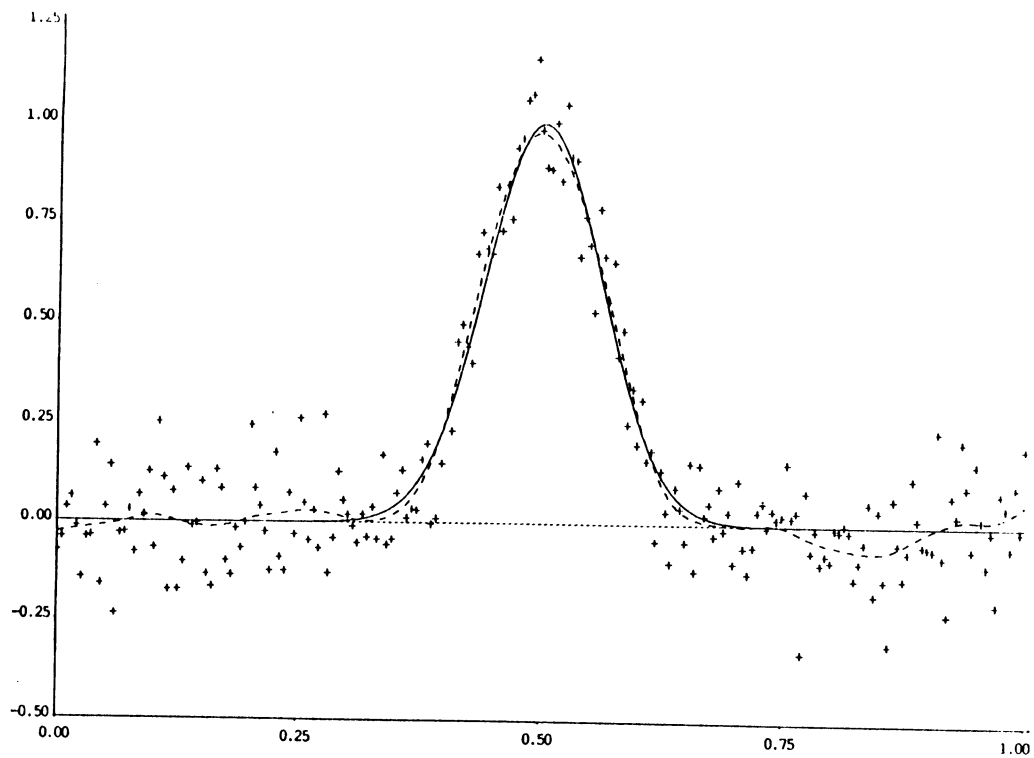


Figure 2.6

spline classique

Pour une validation expérimentale plus complète, nous avons considéré en plus le deuxième exemple suivant avec les valeurs d'écart type $v \in \{0.04, 0.06, 0.1\}$ pour chaque exemple.

$$\psi(g_2(t)) = \begin{cases} \exp\left(\frac{-(t-0.5)^2}{2 \times 0.06^2}\right) & , t \in [0, 0.5] \\ 0.6 \times 256 \times t^4 (1-t)^4 & , t \in [0.5, 1] \end{cases}$$

Pour chaque couple (g, v) avec $\psi(y) = \exp(y)$ on a simulé 20 fois un ensemble de données. Notons ici qu'à chaque exécution les vecteurs ε et w sont régénérés. Par analogie avec le cas classique on espère que la valeur de ρ qui minimise la fonction GCV est proche de celle qui minimise l'erreur quadratique donnée par:

$$(2.4) \quad R(\rho) = \frac{1}{n} \sum_{i=1}^n [(B_\rho(z))_i - \eta_i(g)]^2.$$

Pour faire une comparaison des différentes méthodes de calcul de la GCV données ci-dessus, on définit l'inefficacité de la procédure X par:

$$\text{inefficacité}(\text{procédure } X) = \frac{R(\rho_X)}{R(\hat{\rho})}$$

où $\hat{\rho}$ est la valeur qui minimise l'erreur $R(\rho)$ donné par (2.4) et

$$\rho_X \in \{\rho_{DE}, \dot{\rho}_{DE}, \rho_{NM}, \dot{\rho}_{NM}, \rho_{GN}, \dot{\rho}_{GN}, \dot{\rho}_{DF}\}.$$

On a observé qu'une valeur d'inefficacité proche de 1 n'est pas suffisante pour conclure que le choix de ρ par la GCV est une bonne méthode. Mais si en plus la valeur minimum $R(\hat{\rho})$ est 'petite' par rapport à v^2 , alors la méthode sera bonne. On a observé que $R(\hat{\rho})$ était de l'ordre de $v/10$ ce qui semble bonne. La distribution des inefficacités obtenues sur 120 cas est indiquée dans le tableau 2.1 suivant.

Tableau 2.1
 $\psi(y) = \exp(y)$

inefficacité	V_{DE}	V_{DE}^*	V_{DF}^*	V_{GN}	V_{GN}^*	V_{NM}	V_{NM}^*
(2.50, ∞]	4	3	2	4	3	1	1
(1.66, 2.50]	5	5	8	2	4	11	11
(1.25, 1.66]	23	18	19	22	17	24	18
[1.00, 1.25]	88	93	91	92	96	84	90

Le critère pour arrêter les itérations était

$$\left(\frac{\|y^k - y^{k-1}\|^2}{\|y^k\|^2} < 10^{-8} \right) \text{ ou } k > 200.$$

Le nombre d'itérations nécessaires pour la convergence, pour une valeur de p dans un intervalle convenable, était en moyenne 15. Comme précisé à la fin des sections 1.6.1 et 1.6.2, une itération de type Newton ou Gauss-Newton coûte une certaine spline avec poids. C'est-à-dire le calcul de $B_p(z)$, pour une valeur de p dans un intervalle convenable, coûte en moyenne 15 splines avec poids. Souvent la convergence était lente pour des valeurs de p qui correspondaient au lissage faible.

Remarque 2.4.1

On a aussi fait des expériences pour plusieurs valeurs de h . Nous avons observé que le minimum de $V_{DF}^*(p)$ ne varie pas beaucoup par rapport à h dans un intervalle convenable, c'est-à-dire quand h est à peu près 1 ou 10% de l'écart type v . Pour les résultats numériques présentés ici nous avons pris $h=10^{-2}$.

CHAPITRE III

GCV quand le critère de moindres carrés est remplacé par une "distance convexe"

3.1 Introduction

Il est bien connu que les techniques classiques d'approximation aux moindres carrés peuvent être interprétées comme des techniques de maximum de vraisemblance si l'erreur d'observation est supposée avoir une répartition gaussienne (aussi appelée normale). Il existe de nombreuses applications où l'on sait que les erreurs ne sont pas gaussiennes de variance uniforme. Pour ces problèmes le principe de maximum de vraisemblance conduit à remplacer la somme des carrés des résidus par une autre "distance". On est alors conduit à résoudre des problèmes de lissage non linéaires et à choisir un bon paramètre de lissage pour ceux-ci.

Soit $z=(z_1, \dots, z_n)^T$ le vecteur des observations aux points $t_1, \dots, t_n$ de la fonction $g(t)$ et soit $L(g, z)$ la fonction de log-vraisemblance correspondant au modèle du bruit. C'est-à-dire que $e^{L(g, z)}$ est la densité de probabilité de z sachant que la fonction qui a généré ces observations est g . Appliquer brutalement le principe du maximum de vraisemblance (c'est-à-dire maximiser $L(\mu, z)$) peut conduire à des estimateurs extrêmement oscillants de g . De la manière classique, nous introduisons donc un terme de régularisation (ou pénalité de lissage) $R(\mu)$ pour définir comme estimateur de g celui qui minimise $\Phi_{p, z}(\mu)$ définie par

$$\Phi_{p, z}(\mu) = R(\mu) - pL(\mu, z)$$

où $\mu=(f(t_1),\dots,f(t_n))$ et où f est l'approximation cherchée de g .

Ici ρ est un paramètre de lissage qui contrôle le compromis entre la fidélité aux données mesurée par la fonction de log-vraisemblance $L(\mu,z)$ et l'aspect lisse de la solution mesuré par $R(\mu)$.

Dans ce chapitre, nous supposons que la "distance" notée par $-L(\mu,z)$ est convexe et atteint son minimum par rapport à μ en $\mu=z$. On utilise la méthode itérative de Newton relaxée rappelée dans le chapitre I, section 1.6.1 pour résoudre les problèmes de lissage non linéaire pour une valeur de ρ donnée, et un bon choix de ρ est fait par validation croisée. Des exemples numériques ont été exécutés pour deux types de données, précisément: données suivant une loi de Poisson et des données avec des points aberrants.

3.1.1 Observation de type Poisson

Un aspect commun à tous les types de tomographie par émission est que les observations sont bien modélisées par un processus de Poisson. Dans la littérature, les algorithmes de maximum de vraisemblance (expectation-maximization: EM) pénalisé sont utilisés (voir par exemple Snyder et Miller [85], Silverman et al [90] et Green [90]). Ici nous considérons le problème important du choix du paramètre de lissage.

Nous supposons que chaque observation z_i suit une loi de Poisson de paramètre $g(t_i)$, notons $z_i \sim \text{Poisson}(g(t_i))$. Rappelons que cela signifie que la fonction de probabilité est

$$P(z_i=k) = \frac{1}{e^{g(t_i)}} \frac{(g(t_i))^k}{k!}.$$

Alors la fonction de log-vraisemblance vaut

$$(3.1) \quad L(\mu,z) = \sum_{i=1}^n (z_i \ln(\mu_i) - \mu_i),$$

où $\mu=(\mu_1,\dots,\mu_n)=(f(t_1),\dots,f(t_n))$ et où f est l'approximation cherchée de g . Pour des raisons classiques (changement de variable $f(t)=\exp(\theta(t))$ pour assurer la positivité de f), nous introduisons un terme de régularisation $R(\mu)$ suivant:

$$R(\mu) = (\ln \mu)^T \Omega (\ln \mu)$$

avec $\ln \mu = (\ln(\mu_1), \dots, \ln(\mu_n))^T$.

Notons:

$$\Phi_{\rho, z}(\mu) = R(\mu) - \rho L(\mu, z).$$

Alors, la minimisation de $\Phi_{\rho, z}(\mu)$ qui doit se faire sous la contrainte de positivité est équivalent à minimiser $\Psi_{\rho, z}(y) = \Phi_{\rho, z}(\psi(y_1), \dots, \psi(y_n))$ par rapport à y avec $\psi(y_i) = \exp(y_i)$. En conséquence, l'estimateur de g , $g_{\rho, z}$, est l'exponentielle de la spline cubique naturelle qui interpole $y_{\rho, z, i}$, $i=1, \dots, n$ en $t_1, \dots, t_n$ avec $y_{\rho, z}$ solution de:

$$\min_y \left\{ y^T \Omega y + \rho \sum_{i=1}^n \phi(\psi(y_i), z_i) \right\}$$

pour une valeur ρ de paramètre de lissage. Dans l'expression ci-dessus $\phi(\psi(y_i), z_i) = \psi(y_i) - z_i \ln(\psi(y_i))$. Notons que par rapport aux notations introduites dans la section 1.6, ici on a $\eta_i(y) = \psi(y_i)$.

3.1.2 Splines robustes

Durant les dernières années il y a eu beaucoup de travaux consacrés à l'utilisation de critères d'approximation plus "robustes" à la place des critères de moindres carrés. Ces alternatives ont pour objectif de produire des estimateurs plus résistants à la mauvaise influence de quelques points aberrants (par exemple Huber [73], [81]). Parmi les critères robustes, le plus classique est sans doute celui construit avec la fonction de Huber:

$$H_\tau(t) = \begin{cases} \frac{t^2}{2} & , |t| \leq \tau, \\ \tau |t| - \frac{\tau^2}{2} & , |t| \geq \tau. \end{cases}$$

La somme classique des carrés $\sum_{i=1}^n (y_i - z_i)^2$, mesurant la proximité de y aux données z , est alors remplacée par $\sum_{i=1}^n H_\tau\left(\frac{y_i - z_i}{v}\right)$ où v est un paramètre d'échelle. Si la constante τ dans la fonction de Huber est très grande, alors on aura la méthode de moindres carrés. Au

contraire, si τ tends vers zero, on approche l'approximation I_1 . Pour discuter le choix de τ convenable, on va rappeler l'interprétation statistique de la fonction de Huber (cf. Ekblom et Madsen [89]).

Le critère de Huber est basé sur le modèle statistique de données: soit les données suivent une distribution de type $(1-\epsilon)G+\epsilon H$, où G est une distribution normale centré à zero et la variance v^2 , et H est une distribution qui perturbe les données. On suppose que H est symétrique et centré à zero, mais on a aucune autre information sur H . Souvent la variance v^2 est inconnue et on suppose qu'elle peut être estimée, notons son estimateur par α^2 .

Dans ce modèle la partie des points aberrants est ϵ . La constante τ est déterminée comme solution d'une équation non linéaire en ϵ (cf. Huber [81]). Dans les cas extrême $\epsilon=0$ correspond à $\tau=\infty$ (i.e. la méthode de moindres carrés), et $\epsilon=1$ à $\tau=0$ (i.e. l'approximation I_1). Le tableau des valeurs de τ en fonction de ϵ est donné dans Ekblom et Madsen [89]. Pour être complet, nous allons le rappeler ici.

Tableau 3.1

ϵ	0%	1%	5%	10%	20%	100%
τ	∞	1.95	1.40	1.14	0.86	0

Ici nous considérons alors le problème non quadratique suivant dont la solution sera la spline de lissage robuste de paramètre ρ :

$$\text{Min}_y \left\{ y^T \Omega y + \rho \sum_{i=1}^n \phi(y_i, z_i) \right\},$$

où

$$\begin{aligned} \phi(y_i, z_i) &= \frac{1}{\tau^2} H_\tau \left(\frac{y_i - z_i}{\alpha} \right) \\ (3.2) \quad &= \begin{cases} \frac{(y_i - z_i)^2}{2\tau^2\alpha^2} & , |y_i - z_i| \leq \tau\alpha, \\ \frac{|y_i - z_i|}{\tau\alpha} - \frac{1}{2} & , |y_i - z_i| \geq \tau\alpha. \end{cases} \end{aligned}$$

La section 3.2 est consacrée au choix du paramètre de lissage par le principe de validation croisée. Plusieurs façons de généraliser

la fonction de validation croisée ordinaire sont suggérées. L'étude des estimateurs de Monte-Carlo est présentée dans la section 3.3. Dans la section 3.4, on présente des expériences numériques. La sous section 3.4.3 est consacré à une procédure automatique pour le choix des paramètres de la fonction de Huber pour le cas des splines robustes.

L'expression de la GCV obtenue dans ce chapitre reste valable pour les problèmes à plusieurs variables. Voir Deshpande et Girard [90] pour les résultats en deux dimensions sur les splines robustes.

3.2 Choix de ρ par validation croisée généralisée

Dans ce paragraphe nous développons la méthode de validation croisée ordinaire et nous la généralisons pour des problèmes non linéaires tels que ceux présentés dans la section 3.1.

Soit $\mu_{\rho,i}^{(k)} = \psi(y_{\rho,i}^{(k)})$, $i=1, \dots, n$ où $y_{\rho}^{(k)}$ est la solution correspondant à tous les points de données excepté z_k , pour une valeur fixée ρ du paramètre de lissage, i.e. $y_{\rho}^{(k)}$ est la solution du problème suivant:

$$(3.3) \quad \text{Min}_y \left\{ y^T \Omega y + \rho \sum_{\substack{i=1 \\ i \neq k}}^n \phi(\psi(y_i), z_i) \right\}.$$

Ici la quantité $\phi(\mu_{\rho,k}^{(k)}, z_k)$ mesure l'aptitude de $\mu_{\rho,k}^{(k)}$ à prédire z_k . Alors, on définit la fonction de validation croisée ordinaire de la manière suivante:

$$(3.4) \quad V_0(\rho) = \frac{1}{n} \sum_{k=1}^n \phi(\mu_{\rho,k}^{(k)}, z_k).$$

Pour développer une validation croisée généralisée, nous avons besoin d'une approximation quadratique de $\phi(\psi(y_i), z_i)$. Une telle approximation peut être considérée par rapport à μ_i ou y_i . Nous traiterons ici les deux alternatives.

3.2.1 Approximation quadratique de $\phi(\mu_i, z_i)$ par rapport à μ_i

On supposera que la fonction $\mu_i \rightarrow \phi(\mu_i, z_i)$ est deux fois différentiable par rapport à μ_i . Alors on peut considérer l'approximation quadratique de $\phi(\mu_i, z_i)$ par rapport à μ_i autour $\mu_i = z_i$:

$$\phi(\mu_i, z_i) \approx \phi(z_i, z_i) + (\mu_i - z_i) \frac{\partial \phi}{\partial \mu_i}(z_i, z_i) + \frac{1}{2} (\mu_i - z_i)^2 \frac{\partial^2 \phi}{\partial \mu_i^2}(z_i, z_i).$$

Nous notons que pour les deux exemples considérés dans la section 3.1, $\phi(\mu_i, z_i)$ a un minimum par rapport à μ_i à $\mu_i = z_i$, c'est-à-dire $\frac{\partial \phi}{\partial \mu_i}(z_i, z_i) = 0$. On supposera aussi que la fonction $\mu_i \rightarrow \phi(\mu_i, z_i)$ est convexe. Dans le cas des splines robuste, la fonction n'est pas deux fois différentiable partout par rapport à μ_i , mais presque partout. Les deux points de discontinuités ne pose pas de problème en pratique.

Alors

$$\phi(\mu_i, z_i) \approx \phi(z_i, z_i) + \frac{\gamma_i^2}{2} (\mu_i - z_i)^2$$

où $\gamma_i^2 = \frac{\partial^2 \phi}{\partial \mu_i^2}(z_i, z_i)$, $i=1, \dots, n$.

Donc le minimiseur de $V_0(\rho)$ peut être approché par le minimiseur de $\tilde{V}_0(\rho)$ défini par:

$$\begin{aligned} \tilde{V}_0(\rho) &= \frac{1}{n} \sum_{k=1}^n \gamma_k^2 [\mu_{\rho, k}^{(k)} - z_k]^2 \\ (3.5) \quad &= \frac{1}{n} \sum_{k=1}^n \gamma_k^2 [\psi(y_{\rho, k}^{(k)}) - z_k]^2. \end{aligned}$$

Comme en pratique il est presque impossible de calculer $y_{\rho}^{(k)}$ pour $k=1, \dots, n$, nous généralisons la fonction de validation croisée ordinaire notée par $\tilde{V}_0(\rho)$. Ceci est fait de la façon suivante:

A: On peut voir facilement que $x = y_{\rho}^{(k)}$ solution de (3.3) réalise également le minimum de:

$$\text{Min}_y \left\{ y^T \Omega y + \rho \sum_{\substack{i=1 \\ i \neq k}}^n \phi(\psi(y_i), z_i) + \rho \phi(\psi(y_k), \psi(x_k)) \right\}.$$

Bien que le raisonnement soit similaire à celui indiqué dans la section 2.2, pour être complet, nous allons le rappeler ici.

On a donc

$$B_p(z_1, \dots, z_{k-1}, \psi(x_k), z_{k+1}, \dots, z_n)^T = (\psi(x_1), \dots, \psi(x_n))^T,$$

i.e.,

$$(3.6) \quad \psi(x_k) = [B_p(z + \delta_k e^k)]_k$$

où e^k est le $k^{\text{ième}}$ vecteur de la base canonique de $\mathbb{R}^n$, et $\delta_k = \psi(x_k) - z_k$. En substituant (3.6) dans (3.5) nous obtenons:

$$\tilde{V}_0(\rho) = \frac{1}{n} \sum_{k=1}^n \frac{\gamma_k^2 [\psi(y_{\rho,k}) - z_k]^2}{[1 - a_{kk}^*(\rho)]^2}$$

où

$$a_{kk}^*(\rho) = \frac{[B_p(z + \delta_k e^k)]_k - [B_p(z)]_k}{\delta_k}.$$

Par analogie avec le cas de la spline de lissage classique, nous remplaçons $a_{kk}^*(\rho)$ par $\frac{1}{n} \sum_{j=1}^n a_{jj}^*(\rho)$ pour obtenir la fonction de validation croisée généralisée (GCV) donnée par:

$$V(\rho) = \frac{\frac{1}{n} \sum_{k=1}^n \gamma_k^2 [(B_p(z))_k - z_k]^2}{\left[\frac{1}{n} \left(n - \sum_{k=1}^n a_{kk}^*(\rho) \right) \right]^2}.$$

Lorsque ρ est bien choisi, la quantité $\delta_k = \psi(x_k) - z_k$ devrait être petite. Ainsi, $a_{kk}^*(\rho)$ peut être approché par $\frac{\partial (B_p(z))_k}{\partial z_k}$ et la fonction GCV par:

$$V(\rho) \cong \frac{\frac{1}{n} \sum_{k=1}^n \gamma_k^2 [(B_p(z))_k - z_k]^2}{\left[\frac{1}{n} \text{tr} \left(I - (B_p)_z \right) \right]^2}.$$

B: Une deuxième approche consiste à réduire l'expression pour $\tilde{V}_0(\rho)$ donnée par (3.5) à celle considérée dans le chapitre II, (2.2). Pour cela on suppose que $\gamma_k \neq 0$ pour $k=1, \dots, n$. Soient $\Gamma = \text{diag}(\gamma_1, \dots, \gamma_n)$,

$\tilde{z} = \Gamma z$, $\tilde{\mu}_z = \Gamma \mu_z$ et $\tilde{B}_p = \Gamma B_p \Gamma^{-1}$. Alors, on a $\tilde{B}_p: \tilde{z} \rightarrow \tilde{\mu}_z$ et $\tilde{V}_0(\rho)$ peut s'écrire comme suit:

$$\tilde{V}_0(\rho) = \frac{1}{n} \sum_{k=1}^n \left[\tilde{\mu}_{p,k}^{(k)} - \tilde{z}_k \right]^2.$$

Par un raisonnement similaire à celui de la section 2.2 et en remplaçant z par $\tilde{z}$ et B_p par $\tilde{B}_p$, on obtient

$$\begin{aligned} V(\rho) &\equiv \frac{\frac{1}{n} \sum_{k=1}^n \left[(\tilde{B}_p(z))_k - \tilde{z}_k \right]^2}{\left[\frac{1}{n} \mathbf{1}^T \mathbf{I} \left(\tilde{B}_p(z) \right) \right]^2} \\ &= \frac{\frac{1}{n} \sum_{k=1}^n \gamma_k^2 \left[(B_p(z))_k - z_k \right]^2}{\left[\frac{1}{n} \mathbf{1}^T \mathbf{I} \left(B_p(z) \right) \right]^2}. \end{aligned}$$

3.2.2 Expression analytique pour $(B_p)_z$

Nous rappelons que l'expression analytique pour $(B_p)_z$ est donnée dans l'annexe. Notons $F(y_i, z_i) = f_i(y, z) = \phi(\psi(y_i), z_i)$. Alors l'expression pour $(B_p)_z$ est:

$$(B_p)_z = -\Theta(\Omega + \rho W)^{-1} \rho V$$

où $\Theta = \text{diag}(\psi'(y_{z,1}), \dots, \psi'(y_{z,n}))$, $W = \text{diag}\left(\frac{1}{2} \frac{\partial^2 F}{\partial y_1^2}(y_{z,1}, z_1), \dots, \frac{1}{2} \frac{\partial^2 F}{\partial y_n^2}(y_{z,n}, z_n)\right)$

et $V = \text{diag}\left(\frac{1}{2} \frac{\partial^2 F}{\partial z_1 \partial y_1}(y_{z,1}, z_1), \dots, \frac{1}{2} \frac{\partial^2 F}{\partial z_n \partial y_n}(y_{z,n}, z_n)\right)$.

Cas I: Observation de type Poisson

Dans ce cas $\phi(\psi(y_i), z_i) = \psi(y_i) - z_i \ln(\psi(y_i))$ et $\psi(y) = \exp(y)$. Ainsi, $F(y_i, z_i) = \exp(y_i) - z_i y_i$, $\Theta = \text{diag}(\exp(y_{z,1}), \dots, \exp(y_{z,n}))$, $W = (1/2)\Theta$, et $V = -(1/2)\mathbf{I}$. Donc on a

$$(B_p)_z = W(\Omega + \rho W)^{-1} \rho \mathbf{I}.$$

Cas II: Splines robustes

Ici on a $\psi(y)=y$ et $F(y_i, z_i)=\phi(y_i, z_i)$ donnée par (3.2).

Alors, $\Theta=I$, $W=\text{diag}(w_1, \dots, w_n)$ et $V=-W$ où

$$w_i = \begin{cases} \frac{1}{2}(\tau\alpha)^2, & |y_{z,i}-z_i| \leq \tau\alpha \\ 0 & \text{sinon} \end{cases}, \text{ pour } i=1, \dots, n.$$

Donc on a

$$(B'_\rho)_z = (\Omega + \rho W)^{-1} \rho W.$$

Nous remarquons ici que dans les deux cas ci-dessus on a $\text{tr}((B'_\rho)_z) = \text{tr}(A_\rho^{NM})$.

3.2.3 Approximation quadratique de $F(y_i, z_i)$ par rapport à y_i

On supposera que la fonction $y_i \rightarrow F(y_i, z_i) = \phi(\psi(y_i), z_i)$ est deux fois différentiable par rapport à y_i . Alors on peut considérer l'approximation quadratique de $F(y_i, z_i)$ par rapport à y_i autour de $y_i = y_{z,i}$ comme:

$$F(y_i, z_i) \approx F(y_{z,i}, z_i) + (y_i - y_{z,i}) \frac{\partial F}{\partial y_i}(y_{z,i}, z_i) + \frac{1}{2} (y_i - y_{z,i})^2 \frac{\partial^2 F}{\partial y_i^2}(y_{z,i}, z_i).$$

Soient $\alpha_i = F(y_{z,i}, z_i)$, $\beta_i = \frac{\partial F}{\partial y_i}(y_{z,i}, z_i)$ et $w_i = \frac{1}{2} \frac{\partial^2 F}{\partial y_i^2}(y_{z,i}, z_i)$ pour $i=1, \dots, n$.

Donc le minimiseur de $V_0(\rho)$ (défini par (3.4)) peut être approché par le minimiseur de $\bar{V}_0(\rho)$ défini par:

$$(3.7) \quad \bar{V}_0(\rho) = \frac{1}{n} \sum_{k=1}^n \left[\alpha_k + (y_{z,k}^{(k)} - y_{z,k}) \beta_k + (y_{z,k}^{(k)} - y_{z,k})^2 w_k \right]$$

où $y_z^{(k)}$ réalise le minimum de (3.3).

Supposons que $w_i \neq 0$, $i=1, \dots, n$. On approche le vecteur $y_z^{(k)}$ par une itération de Newton à partir de y_z , i.e. par le vecteur qui réalise le minimum de:

$$y^T \Omega y + \rho \sum_{\substack{i=1 \\ i \neq k}}^n w_i \left[y_i - \left(y_{z,i} - \frac{\beta_i}{2w_i} \right) \right]^2,$$

i.e.

$$y^T \Omega y + \rho \sum_{\substack{i=1 \\ i \neq k}}^n w_i \left[y_i - \left(y_{z,i} - \frac{\beta_i}{2w_i} \right) \right]^2 + \rho w_k \left[y_k - y_{z,k}^{(k)} \right]^2.$$

Donc on a

$$y_z^{(k)} = (\Omega + \rho W)^{-1} \rho W (\bar{z} + \delta_k e^k)$$

où $W = \text{diag}(w_1, \dots, w_n)$, $\bar{z}_i = y_{z,i} - \beta_i / 2w_i$, $i=1, \dots, n$ et $\delta_k = y_{z,k}^{(k)} - y_{z,k} + \beta_k / 2w_k$.
Il est clair que

$$(3.8) \quad y_{z,k}^{(k)} - y_{z,k} = \frac{\beta_k / 2w_k}{(1 - a_{kk}) / a_{kk}}$$

où $A_p^N = (\Omega + \rho W)^{-1} \rho W = (a_{ij})_{\substack{i=1, \dots, n \\ j=1, \dots, n}}$. En substituant (3.8) dans (3.7) nous obtenons

$$\bar{V}_0(\rho) = \frac{1}{n} \sum_{k=1}^n \left[\alpha_k + \frac{\beta_k^2}{2w_k} \left(\frac{a_{kk}}{1 - a_{kk}} \right) + \frac{\beta_k^2}{4w_k} \left(\frac{a_{kk}}{1 - a_{kk}} \right)^2 \right].$$

Comme dans le cas de la spline de lissage classique, nous remplaçons a_{kk} par $\frac{1}{n} \sum_{j=1}^n a_{jj}$ pour obtenir la fonction de validation croisée généralisée donnée par:

$$V(\rho) = \frac{1}{n} \sum_{k=1}^n \alpha_k + \frac{1}{n} \left(\sum_{k=1}^n \frac{\beta_k^2}{2w_k} \right) \left(\frac{\frac{1}{n} \text{tr}(A_p^{NM})}{\frac{1}{n} \text{tr}(I - A_p^{NM})} \right) \times \left[1 + \frac{1}{2} \left(\frac{\frac{1}{n} \text{tr}(A_p^{NM})}{\frac{1}{n} \text{tr}(I - A_p^{NM})} \right) \right].$$

3.3 Calcul de la trace par Monte-Carlo

Dans ce paragraphe on présente une comparaison entre le calcul exact de la trace et ses estimateurs de Monte-Carlo.

Dans la section 3.2.2 nous avons donné une expression explicite de la dérivée de B_p . Nous notons ici:

$$T_{DE}(\rho) = \frac{1}{n} T \left(I - (B_p)'_z \right).$$

Soit w un vecteur de n variables aléatoires normales indépendantes de moyenne nulle et variance 1. Alors l'estimateur de Monte-Carlo de $T_{DE}(\rho)$ est donné par:

$$\dot{T}_{DE}(\rho) = \frac{w^T (I - (B'_\rho)_z) w}{w^T w}.$$

On approche les dérivées $(B'_\rho)_z w$ et $(\tilde{B}'_\rho)_z w$ par les différences finies ce qui conduit respectivement à:

$$\dot{T}_{DF}(\rho) = \frac{w^T}{w^T w} \left(w - \frac{B_\rho(z+hw) - B_\rho(z)}{h} \right),$$

et

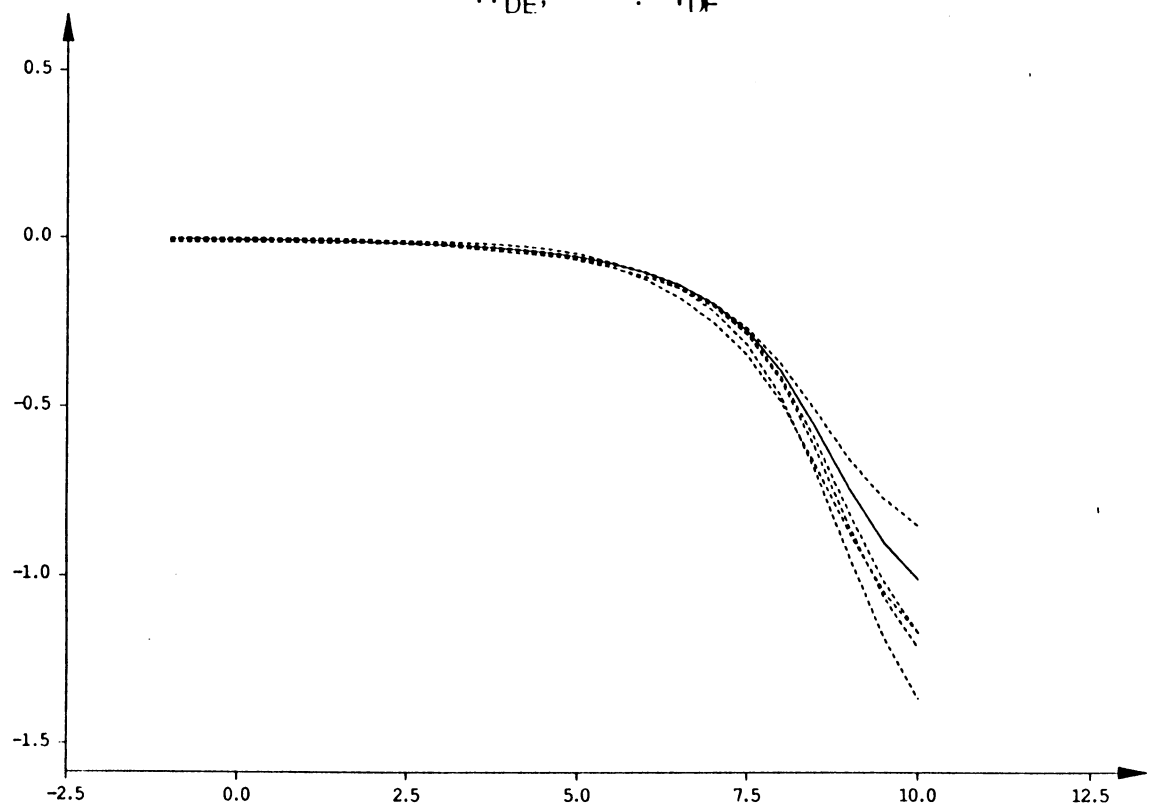
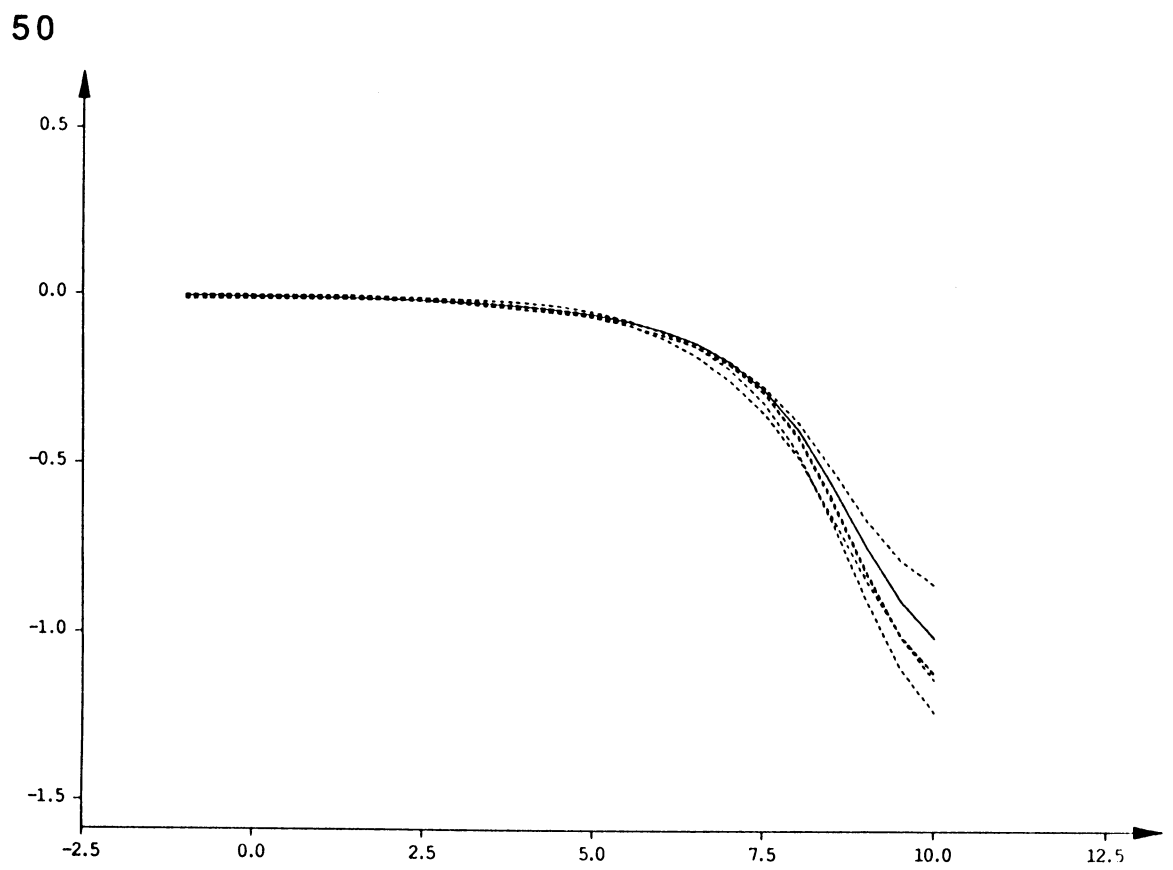
$$\dot{\tilde{T}}_{DF}(\rho) = \frac{w^T}{w^T w} \left(w - \frac{\tilde{B}_\rho(\tilde{z}+hw) - \tilde{B}_\rho(\tilde{z})}{h} \right).$$

Dans les deux sous-sections suivantes, nous donnons les observations expérimentales pour des données avec un bruit de type Poisson et des données avec des points aberrants. Pour les expériences numériques nous avons choisi les points t_i , $i=1, \dots, n$ équidistants. Sur toutes les figures de cette section, on représente les traces en fonction du paramètre de lissage en coordonnées logarithmiques.

3.3.1 Observation de type Poisson

Pour ce type de données on a réalisé un ensemble de simulations numériques pour $n=200$. Les données fictives sont générées avec la fonction g_1 précisée dans la section 3.4.1. Nous calculons les traces pour 23 valeurs de ρ (équidistantes en échelle logarithmique) dans l'intervalle $[10^{-1}, 10^{10}]$.

Sur les figures 3.1, 3.2 et 3.3 on représente la trace exacte T_{DE} et respectivement cinq simulations de $\dot{T}_{DE}$, $\dot{T}_{DF}$ et $\dot{\tilde{T}}_{DF}$. Pour chaque simulation de Monte-Carlo un nouveau vecteur de bruit blanc a été généré. Dans ces figures en coordonnées logarithmiques, on observe que la variation dans les estimateurs de Monte-Carlo par rapport au vecteur aléatoire ne paraît pas grande. Aussi, on a observé que la variation dans le minimum de la GCV par rapport au vecteur aléatoire n'est pas importante. Ainsi, dans la section 3.4.1, l'estimateur de la trace par Monte-Carlo est calculé par une seule simulation.



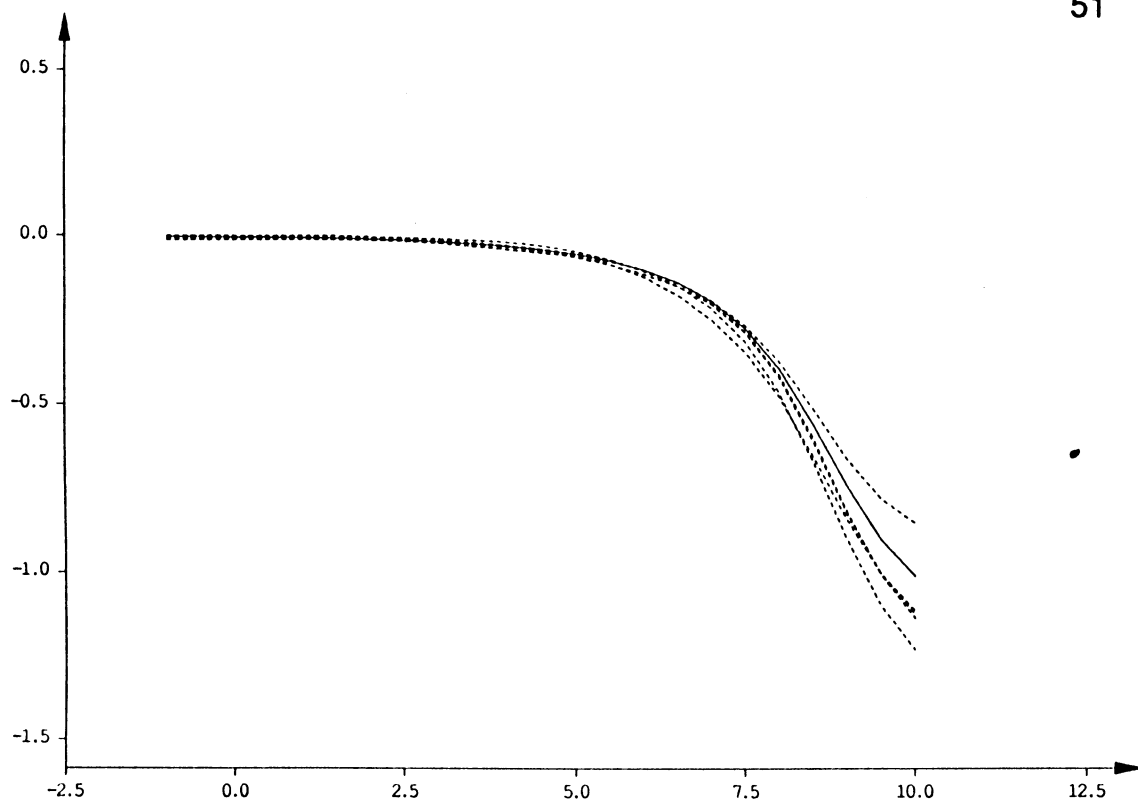


Figure 3.3 Observation de type Poisson

— : T_{DE} , - - - : T_{DF}

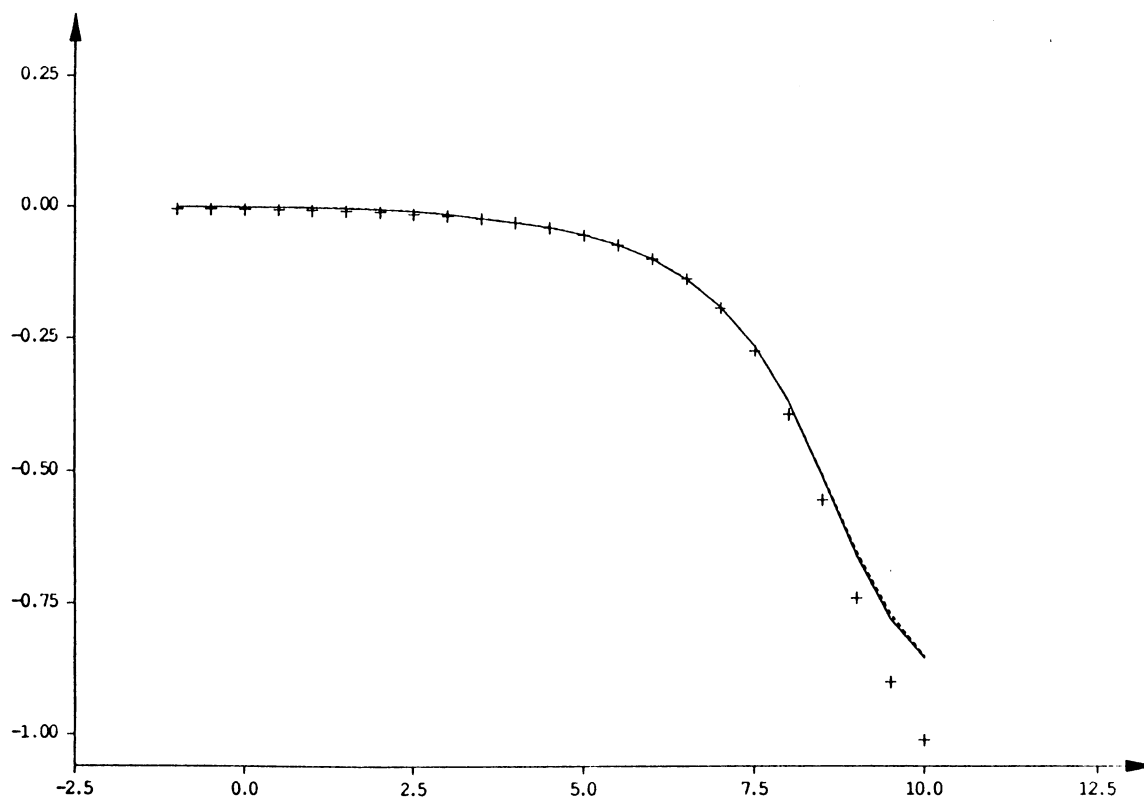


Figure 3.4 Observation de type Poisson

+ : T_{DE} , — : T_{DE}^* , - - - : T_{DF}^*

Pour comparer les différentes traces, sur la figure 3.4 on représente la trace exacte T_{DE} et une simulation de chacun des estimateurs $\dot{T}_{DE}$, $\dot{T}_{DF}$ et $\tilde{\dot{T}}_{DF}$. La remarque sur le calcul de $\dot{T}_{DF}$ (et aussi $\tilde{\dot{T}}_{DF}$) faite à la fin de la section 2.3, est valable ici aussi.

3.3.2 Splines robustes

Les données fictives sont générées comme décrit dans la section 3.4.2 avec la fonction g_3 pour $n=100$. Ici les traces sont calculées pour 31 valeurs de ρ (équidistantes en échelle logarithmique) dans l'intervalle $[10^2, 10^{12}]$.

Dans ce cas la fonction est quadratique sauf aux points aberrants. Alors l'opérateur B_ρ est presque linéaire "par morceaux", ce qui explique les résultats. Pour $h=1e-3$ et un vecteur w , les traces $\dot{T}_{DE}(\rho)$ et $\dot{T}_{DF}(\rho)$ sont visuellement les mêmes. Sur la figure 3.5 on représente la trace exacte T_{DE} et cinq simulations de $\dot{T}_{DE}$. Pour chaque simulation de Monte-Carlo un nouveau vecteur de bruit blanc a été généré. Dans cette figure en coordonnées logarithmiques, on observe que la variation dans les estimateurs de Monte-Carlo par rapport au vecteur aléatoire est très petite.

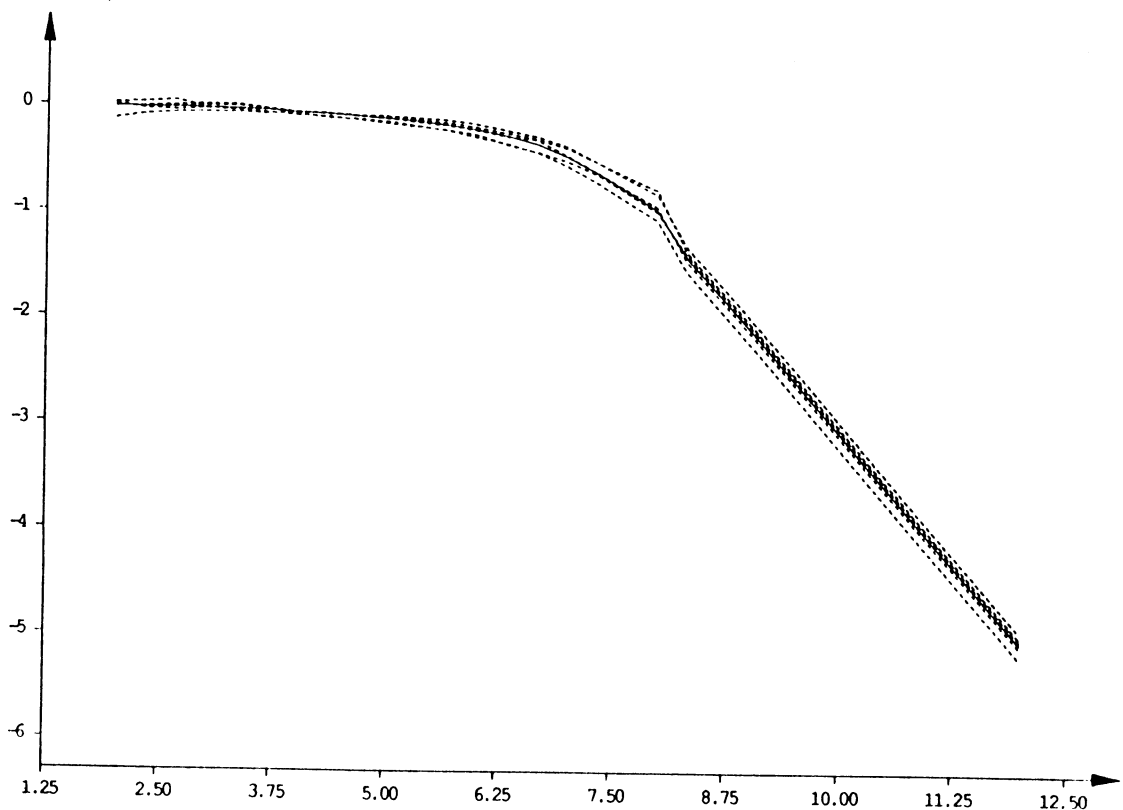


Figure 3.5 Splines robustes

— : T_{DE} , - - - : $\dot{T}_{DE}$

3.4 Une comparaison des méthodes

Dans ce paragraphe on présente une comparaison des méthodes décrites dans la section 3.2.

Notons $\text{Num} = \frac{1}{n} \sum_{k=1}^n \gamma_k^2 [(B_p(z))_k - z_k]^2$. Selon la méthode de calcul de la trace nous désignons les GCV par:

$$V_{DE}(\rho) = \frac{\text{Num}}{[T_{DE}(\rho)]^2}, \quad V_{DE}^*(\rho) = \frac{\text{Num}}{[T_{DE}^*(\rho)]^2},$$

$$V_{DF}^*(\rho) = \frac{\text{Num}}{[T_{DF}^*(\rho)]^2}, \quad V_{DF}^{\sim*}(\rho) = \frac{\text{Num}}{[T_{DF}^{\sim*}(\rho)]^2}.$$

Dans l'expression de GCV au numérateur au lieu de prendre l'approximation quadratique de $\phi(y_{z,i}, z_i)$, on garde l'expression exacte et on définit:

$$V_{EN}^*(\rho) = \frac{\sum_{i=1}^n \phi((B_p(z))_i, z_i) - \sum_{i=1}^n \phi(z_i, z_i)}{[T_{DE}^*(\rho)]^2}.$$

Désignons enfin par $V_Y^*(\rho)$ la fonction de GCV par rapport à l'expression obtenue dans la sous-section 3.2.3:

$$V_Y^*(\rho) = \frac{1}{n} \sum_{k=1}^n \alpha_k + \frac{1}{n} \left(\sum_{k=1}^n \frac{\beta_k^2}{2w_k} \right) \left(\frac{1+T_{DE}^*(\rho)}{T_{DE}^*(\rho)} \right) \times \left(1 + \frac{1}{2} \frac{1+T_{DE}^*(\rho)}{T_{DE}^*(\rho)} \right),$$

où $\alpha_i = F(y_{z,i}, z_i)$, $\beta_i = \frac{\partial F}{\partial y_i}(y_{z,i}, z_i)$ et $w_i = \frac{1}{2} \frac{\partial^2 F}{\partial y_i^2}(y_{z,i}, z_i)$ pour $i=1, \dots, n$.

Soient ρ_{DE} , ρ_{DE}^* , ρ_{DF}^* , $\rho_{DF}^{\sim*}$, ρ_{EN}^* et ρ_Y^* les valeurs de ρ qui minimisent respectivement $V_{DE}(\rho)$, $V_{DE}^*(\rho)$, $V_{DF}^*(\rho)$, $V_{DF}^{\sim*}(\rho)$, $V_{EN}^*(\rho)$ et $V_Y^*(\rho)$. En général, la variance du bruit n'est pas uniforme. Alors, on définit l'erreur quadratique pondérée par:

$$R(\rho) = \frac{1}{n} \sum_{i=1}^n \gamma_i^2 [\psi(y_{z,i}) - g(t_i)]^2,$$

où les poids γ_i^2 'approchent' l'inverse de la variance des données z_i . Par analogie avec le cas classique on espère que la valeur de p qui minimise la fonction GCV est proche de celle qui minimise l'erreur quadratique $R(p)$. Pour faire une comparaison des différentes méthodes de calcul de GCV données ci-dessus, on définit l'inefficacité de la procédure X par:

$$\text{inefficacité}(\text{procédure } X) = \frac{R(\rho_X)}{R(\hat{p})}$$

où $\hat{p}$ est la valeur qui minimise l'erreur $R(p)$ définie ci-dessus et $\rho_X \in \{\rho_{DE}, \rho_{DE}^*, \rho_{DF}^*, \rho_{DF}^{\sim}, \rho_{EN}^*, \rho_Y^*\}$.

3.4.1 Observations de type Poisson

D'abord on illustre le comportement de la méthode de GCV avec deux exemples qui sont les suivants. Les données z_i , $i=1, \dots, n$ suivent une loi de Poisson et sont générées avec les deux fonctions suivantes:

$$g_1(t) = 1 + 10 \times \exp\left(\frac{-(t-0.5)^2}{2 \times 0.06^2}\right), t \in [0, 1].$$

$$g_2(t) = \begin{cases} 1 + 20 \times \exp\left(\frac{-(t-0.5)^2}{2 \times 0.06^2}\right), t \in [0, 0.5] \\ 1 + 10 \times \exp\left(\frac{-(t-0.5)^2}{2 \times 0.06^2}\right), t \in [0.5, 1] \end{cases}$$

Sur la figure 3.6 (et 3.7) on représente les données (par +), la fonction g_1 (respectivement g_2) (par la ligne continue) et la fonction spline obtenue avec le paramètre de lissage ρ_{DE} qui minimise $V_{DE}(\rho)$.

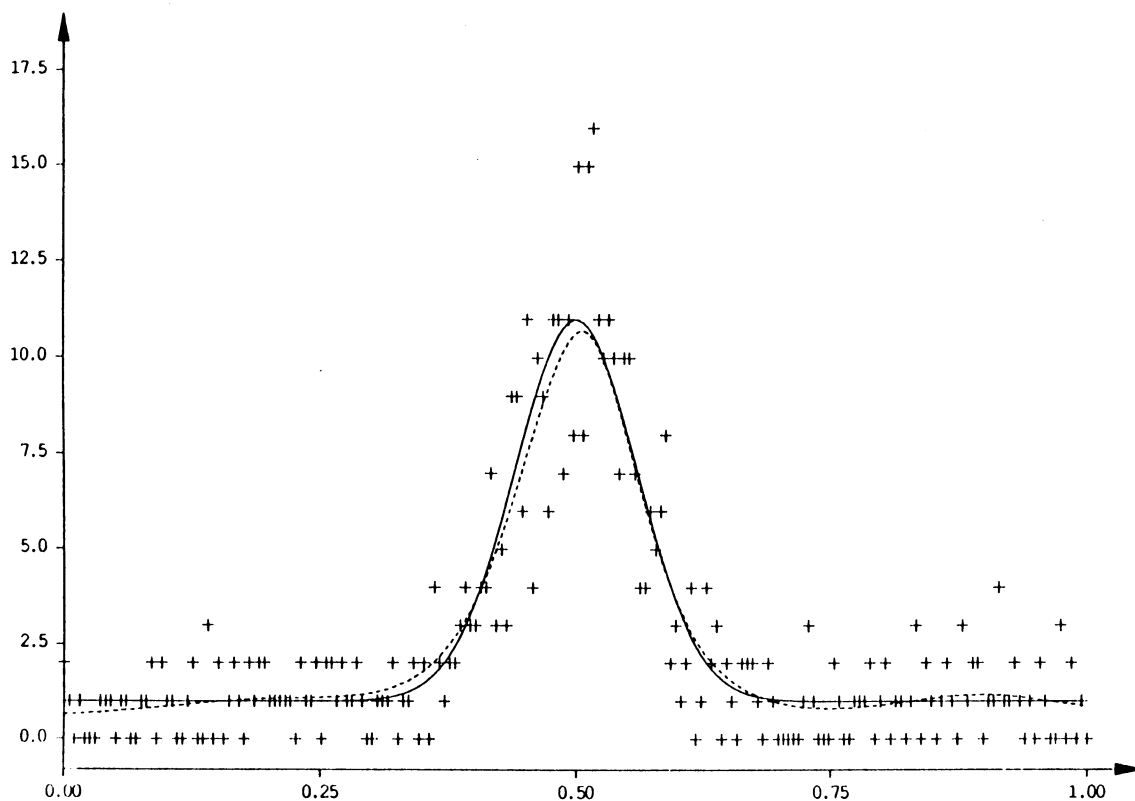


Figure 3.6 Observation de type Poisson, g_1

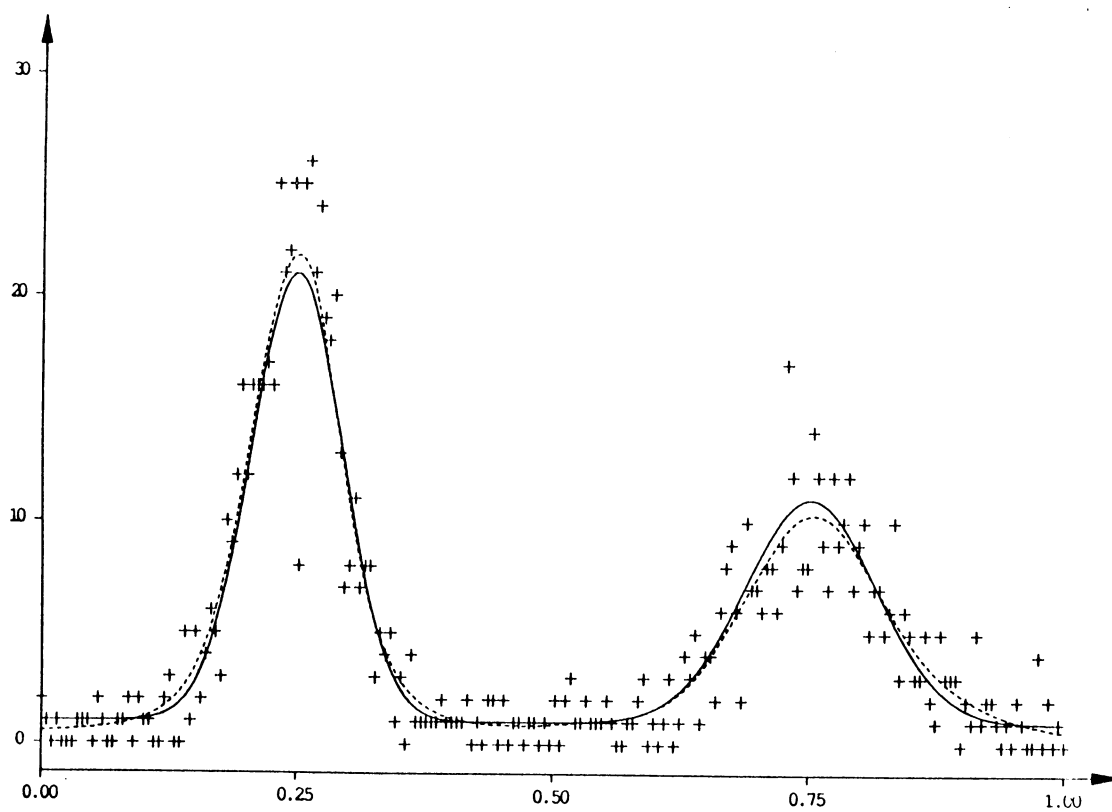


Figure 3.7 Observation de type Poisson, g_2

Pendant l'exécution on a noté des inefficacités qui sont présentées dans le tableau 3.2. On a aussi observé que les splines de lissage obtenues avec un paramètre de lissage $\rho \in \{\rho_{DE}, \rho_{DE}^*, \rho_{DF}^*, \rho_{DF}^{\sim}, \rho_{EN}^*, \rho_y^*\}$ sont visuellement presque les mêmes, ce qui est démontré par les inefficacités obtenues.

Tableau 3.2
inefficacités

fonction	V_{DE}	V_{DE}^*	V_{DF}^*	$V_{DF}^{\sim}$	V_{EN}^*	V_y^*
g_1	1.01	1.01	1.01	1.02	1.01	1.00
g_2	1.15	1.15	1.13	1.09	1.13	1.35

Pour une validation expérimentale plus complète, on a simulé 21 fois un exemple de données avec les deux fonctions g_1 et g_2 déjà explicitées. Notons ici qu'à chaque exécution les vecteurs z et w sont régénérés. La distribution des inefficacités obtenues sur 42 cas est indiquée dans le tableau 3.3 suivant.

Tableau 3.3

inefficacité	V_{DE}	V_{DE}^*	V_{DF}^*	$V_{DF}^{\sim}$	V_{EN}^*	V_y^*
(2.50,∞]	1	4	3	4	3	3
(1.66,2.50]	6	6	7	7	1	4
(1.25,1.66]	5	7	8	7	9	12
[1.00,1.25]	30	25	24	24	29	23

Les résultats du tableau 3.2 permettent de dire que les méthodes $V_{DE}^*(\rho)$, $V_{DF}^*(\rho)$, $V_{DF}^{\sim}(\rho)$, $V_{EN}^*(\rho)$ et $V_y^*(\rho)$ sont presque aussi bonnes que V_{DE} .

Pour les résultats présentés ici nous avons pris $h=1e-2$. Le critère pour arrêter les itérations est

$$\left(\frac{\|y^k - y^{k-1}\|^2}{\|y^k\|^2} < 1e-10 \right) \text{ ou } k > 50.$$

Le nombre d'itérations nécessaires pour la convergence était en

moyenne 7. C'est-à-dire le calcul de $B_p(z)$ pour un p fixé coûte en moyenne 7 splines avec poids car une itération de type Newton coûte une certaine spline avec poids.

3.4.2 Splines robustes

On présente ici deux exemples de données fictives utilisant les fonctions suivantes:

$$g_3(t) = 5\sin(6\pi t), t \in [0,1].$$

$$g_4(t) = 5\exp(2t)\sin(4\pi t), t \in [0,1].$$

Les données z_i , $i=1, \dots, n$ sont de la forme $z_i = g(t_i) + \varepsilon_i$, $i=1, \dots, n$ où ε_i , $i=1, \dots, n$ sont n nombres pseudo-aléatoires qui suivent une loi normale de moyenne nulle et d'écart type $\alpha=0.5$ pour g_3 et $\alpha=0.8$ pour g_4 , et quelques z_i sont perturbés par des ε_i aberrants. Pour les exemples numériques présentés dans cette sous section, on a supposé que α est donné, et le pourcentage des points aberrants (i.e. le τ dans (3.2)) est connu.

Dans ce cas, comme les poids γ_i^2 , $i=1, \dots, n$ sont constants par rapport à p dans un intervalle convenable et $\psi(y)=y$, on définit l'erreur quadratique par:

$$R(p) = \frac{1}{n} \sum_{i=1}^n [y_{z,i} - g(t_i)]^2.$$

Pour la même raison, on n'a pas besoin de traiter le cas V_{DF}^* . Puisque les w_i correspondant au points aberrants sont nuls, on n'a pas le choix de V_y^* .

Sur la figure 3.8a on représente les données (par +), la fonction g_3 (par la ligne continue) et la spline cubique classique avec le paramètre de lissage choisi par la GCV classique. Les figures 3.8b et 3.8c montrent les données, la fonction g_3 et la spline robuste avec le paramètre de lissage respectivement ρ_{DE}^* et ρ_{EN}^* qui minimise $V_{DE}^*(p)$ et $V_{EN}^*(p)$. De la même façon sur les figures 3.9a 3.9b et 3.9c on a tracé les résultats pour la fonction g_4 . Nous avons observé que les splines robustes avec les paramètre de lissage ρ_{DE} , ρ_{DE}^* et ρ_{DF}^* sont visuellement les mêmes.

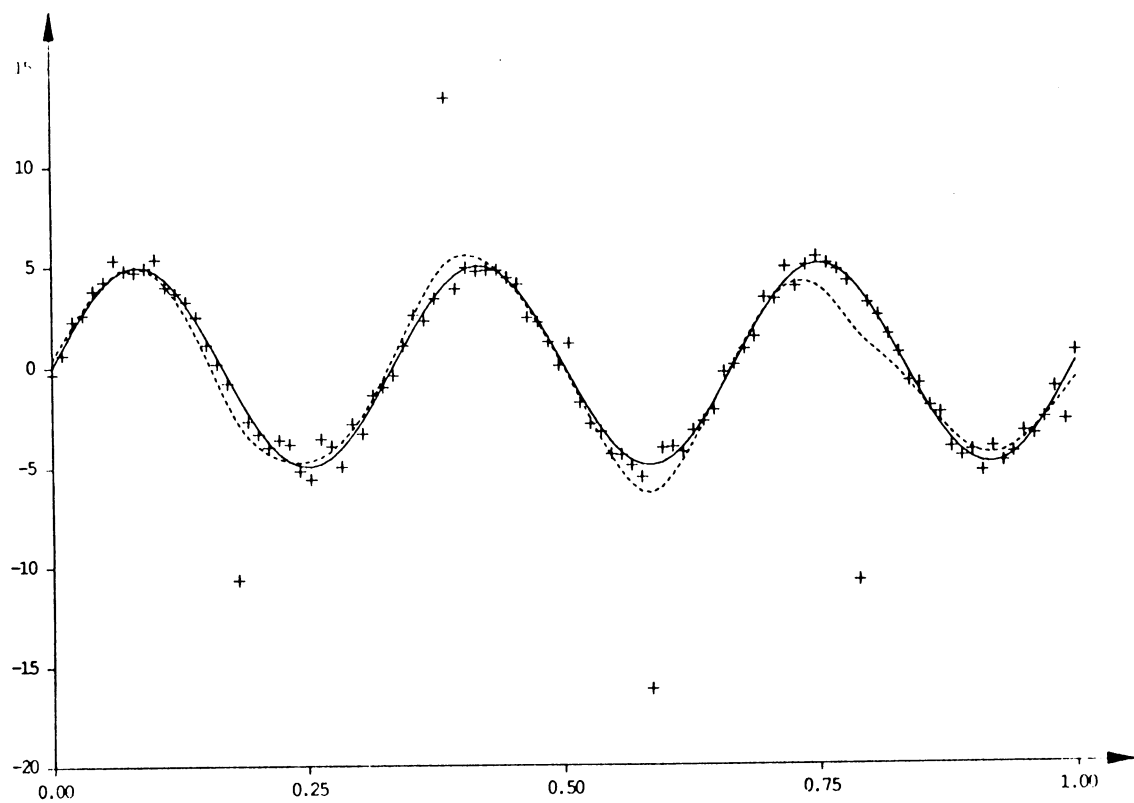


Figure 3.8a Spline classique, g_3

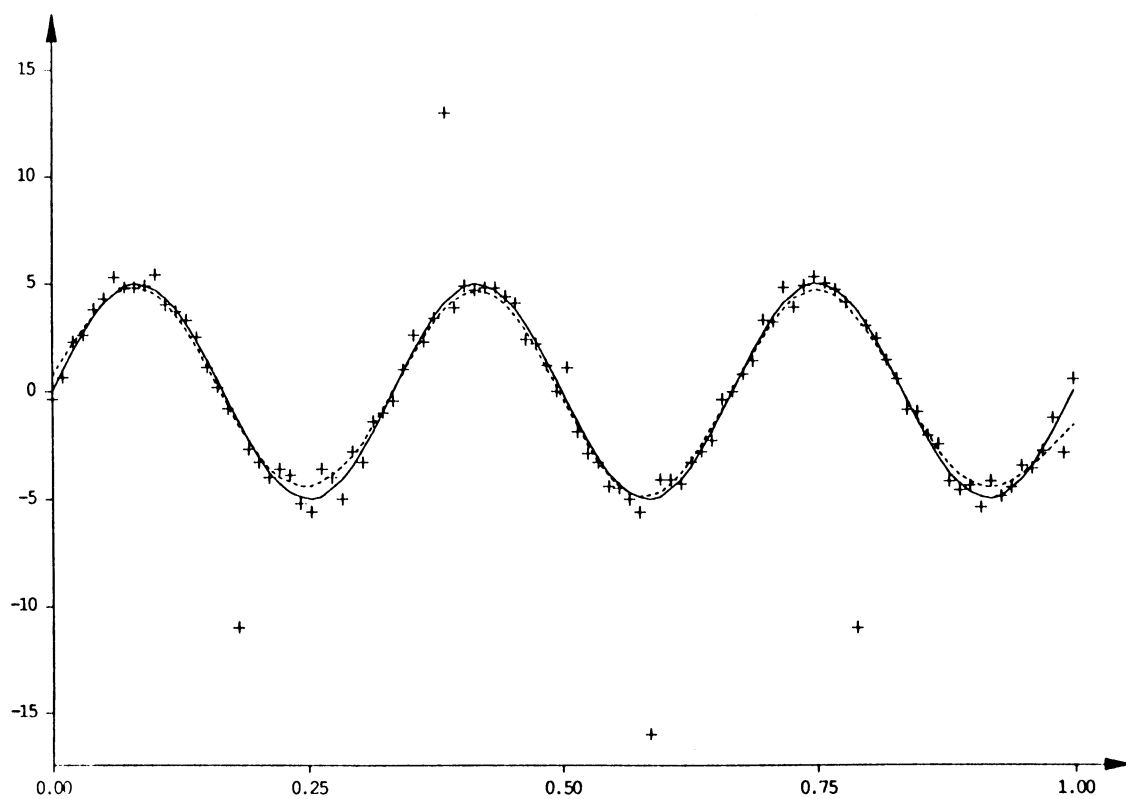


Figure 3.8b Spline robuste, g_3 avec ρ_{DE}^*

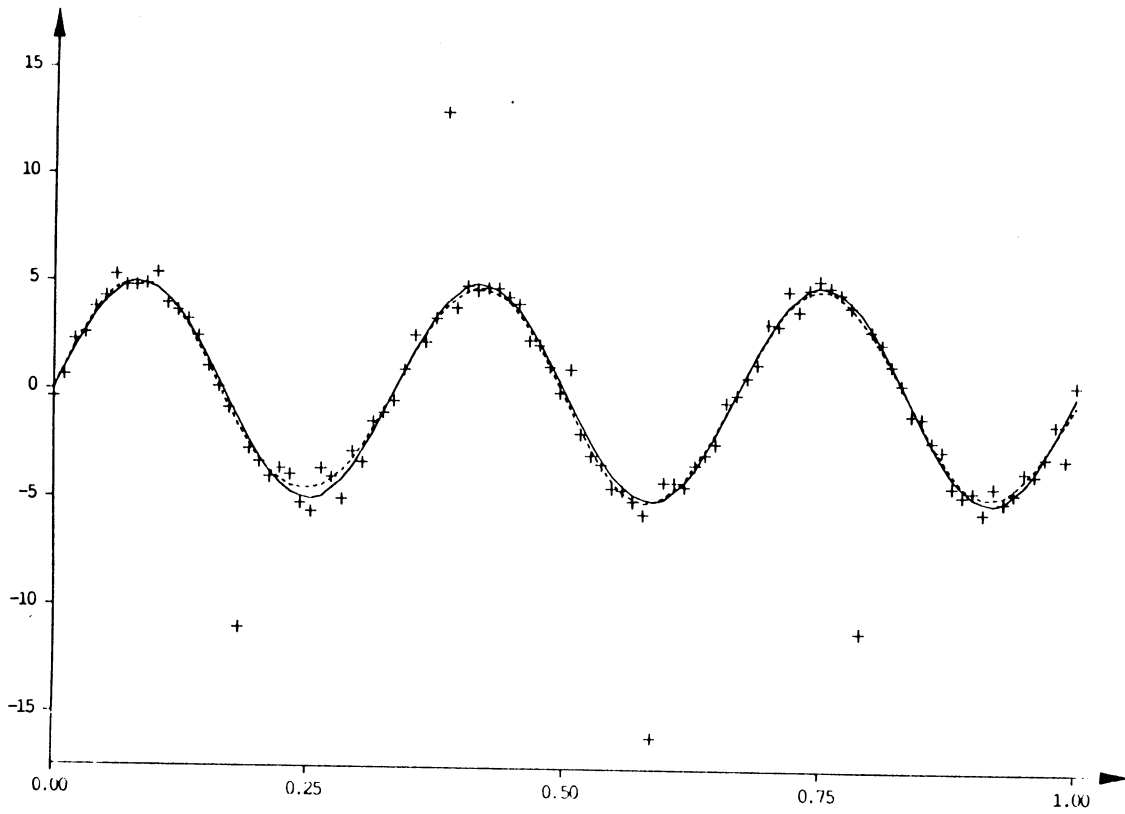


Figure 3.8c Spline robuste, g_3 avec ρ_{EN}^*

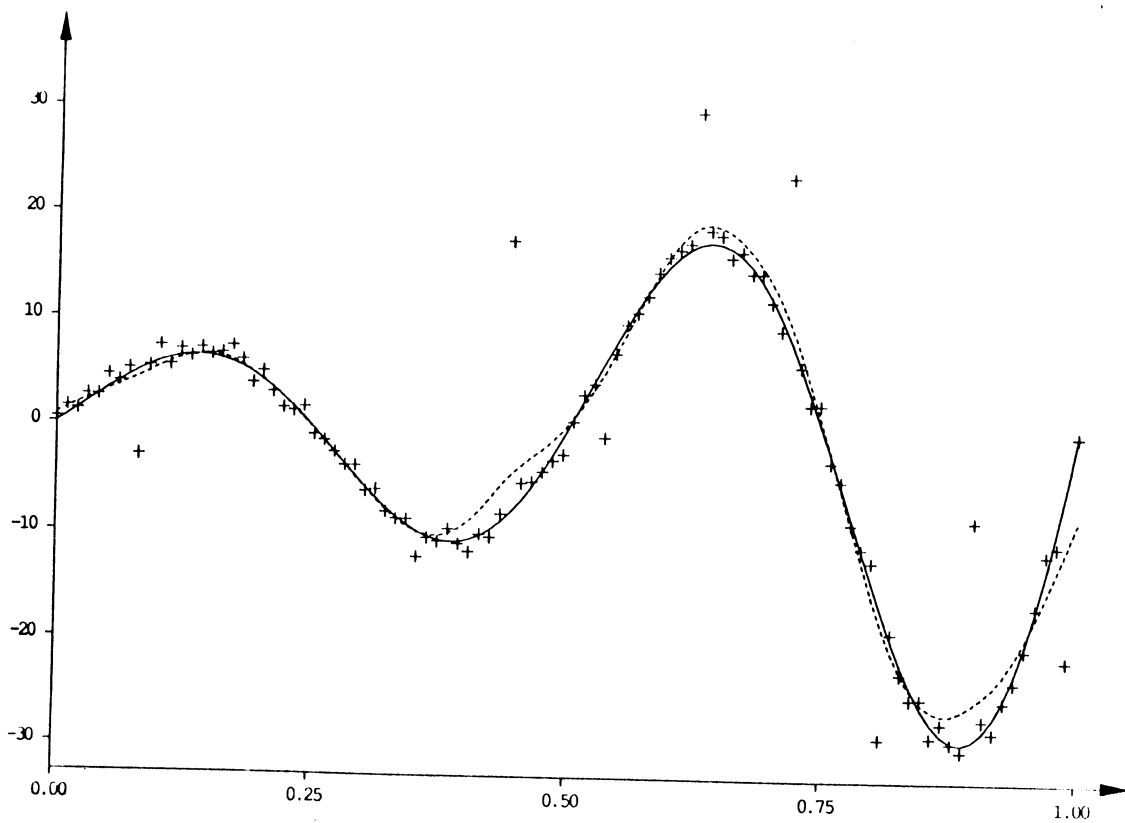


Figure 3.9a Spline classique, g_4

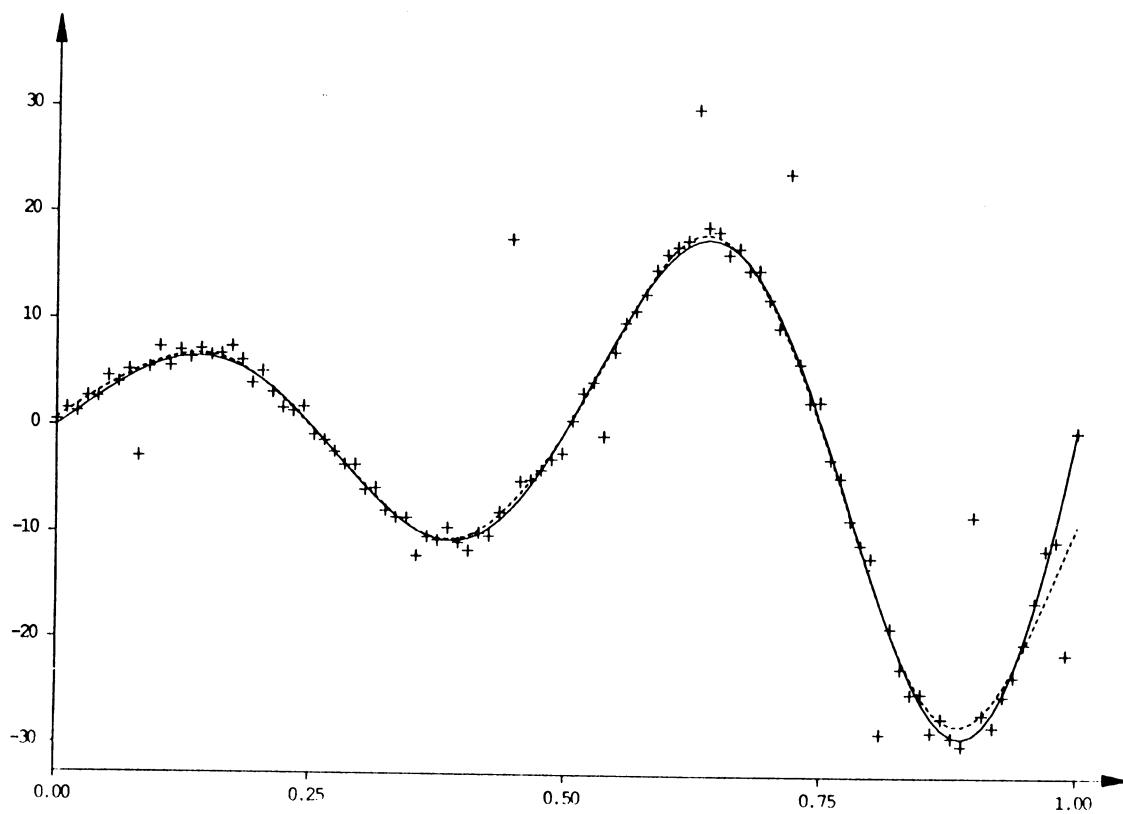


Figure 3.9b Spline robuste, g_4 avec ρ_{DE}^*

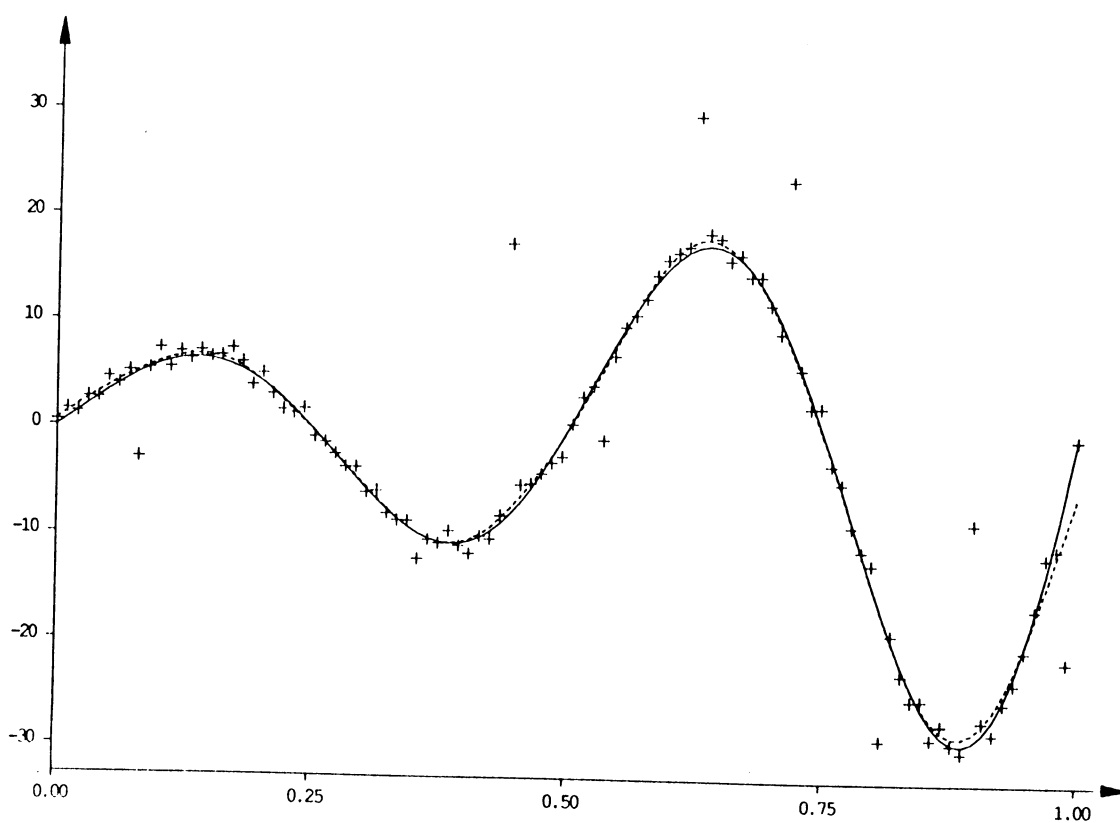


Figure 3.9c Spline robuste, g_4 avec ρ_{EN}^*

Pendant les exécutions, on a noté les inefficacités. Les inefficacités observées sont données dans les tableaux 3.4 et 3.5 suivants.

Tableau 3.4
splines classiques

fonction	inefficacité de GCV	$\text{Min}_\rho R(\rho)$
g_3	1.07	5.30e-1
g_4	1.01	2.95e+0

Tableau 3.5
splines robustes

fonction	inefficacité de			$\text{Min}_\rho R(\rho)$
	V_{DE}	V_{DE}^*	V_{EN}^*	
g_3	4.72	4.84	1.90	3.14e-2
g_4	9.95	7.54	3.22	2.13e-1

Dans le cas des splines robustes, même si les inefficacités sont assez élevés par rapport à ceux des splines classiques, on observe (figures 3.8 et 3.9) que les splines robustes sont meilleures. En plus on observe que V_{EN}^* fait mieux que d'autres. Pour les résultats présentés ici nous avons pris $h=1e-3$. Le critère pour arrêter les itérations est

$$\left(\frac{\|y^k - y^{k-1}\|^2}{\|y^k\|^2} < 1e-10 \right) \text{ ou } k > 50.$$

Le nombre d'itérations nécessaires pour la convergence était en moyenne 9. C'est-à-dire le calcul de $B_\rho(z)$ pour un ρ fixé coûte en moyenne 9 splines avec poids car une itération de type Newton coûte une certaine spline avec poids.

3.4.3 Estimation des paramètres dans la fonction de Huber

Le critère robuste déterminé par la fonction de Huber dépend de deux paramètres: le premier α qui est un estimateur d'écart type des données et le deuxième τ qui est une fonction non linéaire du pourcentage des points aberrants dans les données.

D'abord on cherche un estimateur de la variance des données. Le premier estimateur s^2 de la variance est donné par la formule (3.9) (par ex. Girard [88b]) dans le cas des splines classiques. Alors,

$$(3.9) \quad s^2 = \frac{\|(I - A_{\hat{p}})z\|^2}{\text{tr}(I - A_{\hat{p}})^2},$$

où $A_{\hat{p}}$ est l'opérateur de lissage classique (moindres carrés) associé au paramètre de lissage $\hat{p}$ choisie par GCV classique. Cet estimateur s^2 peut être un sur-estimateur de v^2 à cause de la présence des points aberrants. La deuxième étape est d'améliorer cet estimateur s^2 . On le fait comme suit: nous définissons l'ensemble des données non aberrantes par rapport à l'estimateur s^2 par

$$\mathcal{J} = \{i \in \{1, \dots, n\} / |z_i - (A_{\hat{p}}z)_i| < s\}$$

et soit $\hat{n}$ son cardinal. Alors, l'estimateur final est

$$(3.10) \quad \alpha^2 = \frac{\sum_{i \in \mathcal{J}} (z_i - (A_{\hat{p}}z)_i)^2}{\text{tr}(I - A_{\hat{p}})^2}.$$

Le deuxième paramètre τ est obtenu par le tableau 3.1, où ϵ est le pourcentage des points aberrants dans les données donc $\epsilon = \frac{(n - \hat{n}) \times 100}{n}$.

Pour les deux exemples présentés dans la section 3.4.2, on a fait le choix automatique des paramètres de Huber par la procédure proposé ci-dessus et les résultats obtenus sont présentés dans le tableau 3.6.

Tableau 3.6
splines robustes avec choix automatique des paramètres α et τ

fonction	inefficacité de			$\min_{\rho} R(\rho)$
	V_{DE}	V_{DE}^*	V_{EN}^*	
g_3	4.90	3.37	1.94	3.24e-2
g_4	8.04	8.21	3.65	2.20e-1

Chapitre IV

Détection de discontinuités par Inf-Convolution spline et GCV dans un cadre non linéaire

4.1 Introduction

Dans le chapitre I, section 1.5 on a vu que pour déterminer le paramètre non linéaire α qui correspond à la position de la discontinuité, la méthode de GCV a été proposé avec succès. Dans ce chapitre on va considérer le même problème de la détermination de la position de la discontinuité mais pour un modèle de données tel que le problème de lissage associé est non quadratique et convexe. Un exemple d'application est le suivant: on observe en n points de l'intervalle $[0,1]$, la valeur approchée d'une fonction réelle g inconnue, c'est-à-dire on observe $z_1, \dots, z_n$ tels que

$$(4.1) \quad z_i = g(t_i) + \varepsilon_i, \quad i=1, \dots, n,$$

où les abscisses t_i sont connues et ε_i , $i=1, \dots, n$ sont des erreurs inconnues qui suivent une loi de Poisson. Aussi on suppose que la fonction g est lisse par morceaux.

Alors, l'objet principal de ce chapitre est de montrer que des résultats analogues à ceux de Girard et Laurent [88] sont aussi valables ici. C'est-à-dire que le principe de GCV permet de choisir le meilleur paramètre de lissage, et aussi connaissant la présence de la discontinuité de déterminer sa position.

Dans la section 4.2 nous décrivons la méthode itérative de Newton pour résoudre ce problème de lissage non linéaire pour un p et α fixés. A chaque itération la méthode d'Inf-Convolution Spline

(ICS) est utilisée. La section 4.3 est consacrée au choix de p par la méthode de GCV. Maintenant il reste à déterminer α , ceci est fait dans la section 4.4 par le principe de GCV. Les résultats des expériences numériques sont donnés dans la section 4.5.

4.2 L'approche ICS: un lissage non quadratique préservant les discontinuités

4.2.1 Modélisation d'une fonction lisse par morceaux

Supposons qu'on ait des données de forme (4.1) avec un bruit de type Poisson. Dans ce chapitre nous supposons aussi que la fonction $g:[0,1] \rightarrow \mathbb{R}$ sous-jacente aux données a une représentation de la forme:

$$g(t) = x^0(t) + \sum_{k=1}^m \lambda_k^0 p_k(t)$$

où $x^0 \in H^2[0,1]$. Deux types de fonctions singularités p_k seront utilisées ici. Le premier est la fonction saut d'ordre 0 en α , notée par d_α^0 et définie par

$$d_\alpha^0(t) = \begin{cases} 1, & t \geq \alpha \\ 0, & t < \alpha \end{cases}$$

Le deuxième type est la fonction saut d'ordre 1 en α , définie par

$$d_\alpha^1(t) = \begin{cases} (t-\alpha), & t \geq \alpha \\ 0, & \text{sinon} \end{cases}$$

Notons qu'une rupture "complète" en α , c'est-à-dire saut d'ordre 0 et 1 en même point, dans une fonction lisse ailleurs, nécessite d'introduire en même temps les deux singularités de base d_α^0 et d_α^1 . Nous noterons $\mathbf{p}_j = (p_j(t_1), \dots, p_j(t_n))^T$ le vecteur colonne de ces valeurs, et $P = (\mathbf{p}_1, \dots, \mathbf{p}_m)^T$ la matrice formée de ces m colonnes. Donc $\bar{g} = \bar{x}^0 + P\lambda^0$, où $\lambda^0 = (\lambda_1, \dots, \lambda_m)^T$ et $\bar{g}$, $\bar{x}^0$ sont les vecteurs de ces valeurs.

4.2.2 Calcul des estimateurs à P et ρ fixés

Dans le chapitre précédent on a vu que pour un ρ fixé l'estimateur de $g \in H^2[0,1]$ est donné par une spline cubique naturelle $g_{\rho,z}$ telle que $y_{\rho,z} = (g_{\rho,z}(t_1), \dots, g_{\rho,z}(t_n))^T$ soit la solution du problème suivant:

$$(4.2) \quad \min_{y \in \mathbb{R}^n} \left\{ y^T \Omega y + \rho \sum_{i=1}^n \phi(\psi(y_i), z_i) \right\}.$$

Dans l'exemple d'application on a

$$\psi(y_i) = \exp(y_i) \text{ et } \phi(\psi(y_i), z_i) = \exp(y_i) - y_i z_i.$$

Maintenant on considère le problème d'estimation de g en préservant des discontinuités en nombre connue. On suppose pour l'instant que les m fonctions $p_1, \dots, p_m$ (définies sur $[0,1]$), qui jouent le rôle des singularités, sont fixées. Alors le problème (4.2) de lissage associé à ρ et à la matrice des singularités P est équivalent au problème:

chercher le couple, que l'on notera $(x_{\rho,P} \in \mathbb{R}^n, \lambda_{\rho,P} \in \mathbb{R}^m)$, solution de:

$$(4.3) \quad \min_{\substack{x \in \mathbb{R}^n \\ \lambda \in \mathbb{R}^m}} \left\{ x^T \Omega x + \rho \sum_{i=1}^n \phi(\psi(y_i), z_i) \right\}$$

où $y_i = x_i + (P\lambda)_i$. Nous noterons $y_{\rho,P}$ le résultat final de ce lissage ICS de z, associé à P et ρ. Donc $y_{\rho,P}$ est obtenu en ajoutant la composante lisse et la composante singulière estimée, c'est-à-dire $y_{\rho,P} = x_{\rho,P} + P\lambda_{\rho,P}$.

Pour résoudre ce problème (4.3) on utilise la méthode itérative de Newton rappelée dans la section 1.6. Ici on considère les expressions simplifiées car $\eta_i(y) = \psi(y_i)$ pour $i=1, \dots, n$. Soit $F(y_i, z_i) = \phi(\psi(y_i), z_i)$, $i=1, \dots, n$. Supposons aussi que y^k , k^{ième} estimateur de $y_{\rho,P}$ soit égal à $y^k = x^k + P\lambda^k$. Pour définir y^{k+1} , c'est-à-dire (x^{k+1}, λ^{k+1}) , on considère l'approximation quadratique de $F(y_i, z_i)$ par rapport à y_i autour y_i^k :

$$F(y_i, z_i) \approx F(y_i^k, z_i) + \frac{\partial F}{\partial y_i}(y_i^k, z_i)(y_i - y_i^k) + \frac{1}{2} \frac{\partial^2 F}{\partial y_i^2}(y_i^k, z_i)(y_i - y_i^k)^2.$$

Alors la $(k+1)$ ème itération (x^{k+1}, λ^{k+1}) est la solution du problème:

$$\min_{\substack{x \in \mathbb{R}^n \\ \lambda \in \mathbb{R}^m}} \left\{ x^T \Omega x + \rho \sum_{i=1}^n \left[\beta_i^k (x_i + (P\lambda)_i) + w_i^k (x_i + (P\lambda)_i - y_i^k)^2 \right] \right\},$$

c'est-à-dire:

$$\min_{\lambda \in \mathbb{R}^m} \left\{ \min_{x \in \mathbb{R}^n} \left(x^T \Omega x + \rho \sum_{i=1}^n w_i^k \left[x_i - \left(y_i^k - (P\lambda)_i - \frac{\beta_i^k}{2w_i^k} \right) \right]^2 \right) \right\}$$

où $\beta_i^k = \frac{\partial F}{\partial y_i}(y_i^k, z_i)$ et $w_i^k = \frac{1}{2} \frac{\partial^2 F}{\partial y_i^2}(y_i^k, z_i) > 0$ pour $i=1, \dots, n$. Soient

$W^k = \text{diag}(w_1^k, \dots, w_n^k)$ et $z_i^k = y_i^k - \frac{\beta_i^k}{2w_i^k}$, $i=1, \dots, n$. Supposons d'abord λ

fixé, alors la minimisation intérieure devient

$$\min_{x \in \mathbb{R}^n} \left(x^T \Omega x + \rho \sum_{i=1}^n w_i^k [x_i - (z_i^k - (P\lambda)_i)]^2 \right),$$

ce qui correspond à un problème de spline classique, qui a donc pour solution

$$(4.4) \quad x^{k+1} = A_{\rho, P}^k (z^k - P\lambda)$$

où $A_{\rho, P}^k = (\Omega + \rho W^k)^{-1} \rho W^k$.

Notons ici que x^{k+1} dépend de λ . Une simple manipulation algébrique utilisant la relation $\rho W^k (I - A_{\rho, P}^k) = \Omega A_{\rho, P}^k$, permet de voir que la solution en λ est donnée par:

$$\min_{\lambda \in \mathbb{R}^m} \left\{ (z^k - P\lambda)^T W^k (I - A_{\rho, P}^k) (z^k - P\lambda) \right\}.$$

Le calcul des m amplitudes λ^{k+1} se ramène ainsi à la résolution du système linéaire $m \times m$:

$$(4.5) \quad P^T W^k (I - A_{\rho, P}^k) P \lambda^{k+1} = P^T W^k (I - A_{\rho, P}^k) z^k.$$

On note que l'on est assuré de l'unicité de λ^{k+1} dès que les vecteurs colonne de P sont indépendants, et que l'intersection de l'espace qu'ils engendrent (i.e. l'espace des singularités) et du noyau de $W^k(I-A_{p,p}^k)$ est réduite au vecteur nul.

Donc le lissage ICS de z^k est $y^{k+1}=x^{k+1}+P\lambda^{k+1}$ où x^{k+1} et $P\lambda^{k+1}$ sont donnés respectivement par (4.4) et (4.5). Alors $S_{p,p}^k$, opérateur de lissage ICS qui transforme les pseudo-données z^k en

$y^{k+1}=S_{p,p}^k(z^k)$, est donné par

$$S_{p,p}^k = A_{p,p}^k (I - R_{p,p}^k) + R_{p,p}^k$$

où $R_{p,p}^k$ est l'opérateur qui transforme z^k en la partie singulière $P\lambda^{k+1}$:

$$R_{p,p}^k = P \left[P^T W^k (I - A_{p,p}^k) P \right]^{-1} P^T W^k (I - A_{p,p}^k).$$

Donc on a le simple algorithme suivant pour le calcul d'une itération par inf-convolution spline (Girard et Laurent [88]):

1. calcul d'une matrice $n \times m$, $Q^k = A_{p,p}^k P$ (calcul des m splines)
2. calcul de $A_{p,p}^k z^k$ (une spline)
3. calcul d'une matrice $m \times m$, $U^k = P^T W^k (P - A_{p,p}^k P)$ et
 $b^k = P^T W^k (z^k - A_{p,p}^k z^k)$ ($m^2 n + n m^2 + m n$ opérations)
4. calcul de λ^{k+1} comme solution de $U^k \lambda = b^k$ (résolution d'une système linéaire $m \times m$)
5. $y^{k+1} = x^{k+1} + P\lambda^{k+1} = A_{p,p}^k z^k - A_{p,p}^k P\lambda^{k+1} + P\lambda^{k+1}$ ($m n$ opérations).

Alors l'implémentation de cet algorithme est possible dès qu'on sait calculer une spline avec des poids. On aperçoit qu'en pratique ($m < n$), le coût essentiel de cet algorithme est $m+1$ splines avec poids.

Lorsque l'itération (x^k, λ^k) converge, nous notons sa limite par $(x_{p,p}, \lambda_{p,p})$ et définissons l'opérateur $S_{p,p}$ par:

$$(4.6) \quad S_{p,p} = A_{p,p} (I - R_{p,p}) + R_{p,p}$$

où

$$w_i = \frac{1}{2} \frac{\partial^2 F}{\partial y_i^2}(y_{p,p,i}, z_i), \quad W = \text{diag}(w_1, \dots, w_n), \quad A_{p,p} = (\Omega + \rho W)^{-1} \rho W \quad \text{et}$$

$$(4.7) \quad R_{p,p} = P \left[P^T W (I - A_{p,p}) P \right]^{-1} P^T W (I - A_{p,p}).$$

Notons les opérateurs non linéaires $C_{p,p}$ et $L_{p,p}$ tels que $x_{p,p} = C_{p,p}(z)$ et $P\lambda_{p,p} = L_{p,p}(z)$. Soit aussi $y_{p,p} = x_{p,p} + P\lambda_{p,p}$. Comme dans les chapitres précédents, on note

$$B_{p,p}(z) = (\psi(y_{p,p,1}), \dots, \psi(y_{p,p,n}))^T.$$

Remarque 4.2.1: De (4.6) et (4.7) nous avons

$$S_{p,p} - A_{p,p} = (I - A_{p,p}) P \left[P^T W (I - A_{p,p}) P \right]^{-1} P^T W (I - A_{p,p}).$$

Nous notons la racine carrée de W par V et $\tilde{A}_{p,p} = V A_{p,p} V^{-1}$, $\tilde{P} = V P$, $\tilde{R}_{p,p} = V R_{p,p} V^{-1}$, $\tilde{S}_{p,p} = V S_{p,p} V^{-1}$. Alors $\tilde{A}_{p,p}$ est symétrique et $\tilde{S}_{p,p} = \tilde{A}_{p,p} (I - \tilde{R}_{p,p}) + \tilde{R}_{p,p}$. Maintenant, sachant qu'on a $\text{tr}(AB) = \text{tr}(BA)$ dès que A et B sont deux matrices quelconques telles que A et B^T soient de même dimension, on a alors

$$(4.8) \quad \begin{aligned} \text{tr}(S_{p,p}) - \text{tr}(A_{p,p}) &= \text{tr}(\tilde{S}_{p,p}) - \text{tr}(\tilde{A}_{p,p}) \\ &= \text{tr} \left(\left[\tilde{P}^T (I - \tilde{A}_{p,p}) \tilde{P} \right]^{-1} \tilde{P}^T (I - \tilde{A}_{p,p}) (I - \tilde{A}_{p,p}) \tilde{P} \right). \end{aligned}$$

Dans le cas d'une singularité ($m=1$), la matrice à droite de (4.8) est d'ordre 1 et la relation entre les deux traces devient:

$$\text{tr}(S_{p,p}) - \text{tr}(A_{p,p}) = \frac{\|(I - \tilde{A}_{p,p}) \tilde{P}\|^2}{\tilde{P}^T (I - \tilde{A}_{p,p}) \tilde{P}}.$$

Comme dans Girard [88a] il est facile de voir que ce rapport peut se mettre sous la forme d'un quotient de Rayleigh $\frac{q^T (I - A_p) q}{\|q\|^2}$.

On peut montrer cela en prenant $q = Q\tilde{P}$ où Q est la racine carrée symétrique de $I - \tilde{A}_{p,p}$. Ce rapport est donc minoré par la plus petite des valeurs propres de $I - \tilde{A}_{p,p}$ et majoré par la plus grande.

Or il est bien connu que les valeurs propres de $I-\tilde{A}_{\rho,P}$ sont toujours comprises entre 0 et 1 (par exemple Girard [87a]). La différence entre les 2 traces est donc comprise entre 0 et 1. On peut montrer en fait que, dans le cas général, toutes les valeurs propres de la matrice $m \times m$ à droite de (4.8) sont comprises entre 0 et 1. On a donc dans le cas général:

$$(4.9) \quad 0 \leq \text{tr}(S_{\rho,P}) - \text{tr}(A_{\rho,P}) \leq m.$$

Remarque 4.2.2: On peut faire une remarque sur la proximité relative des 2 traces quand n est grand. Si $m \ll n$ (ce qui est souvent vrai en pratique pour les grands problèmes) alors (4.9) implique que l'écart entre $(1/n)\text{tr}(S_{\rho,P})$ et $(1/n)\text{tr}(A_{\rho,P})$ (qui sont deux quantités normalisées entre 0 et 1, et très souvent proches de 1 quand n est grand et ρ est bien choisi cf. Craven et Wahba [79]) est d'ordre $(1/n)$.

4.3 Choix de ρ par GCV

On admet dans ce paragraphe que les singularités présentes dans le signal sont connues à leur amplitude près, c'est-à-dire que la matrice P est convenablement fixée.

Nous avons vu dans la section 3.2.1 que la méthode de validation croisée généralisée consiste à prendre pour ρ la valeur qui minimise la fonction $V(\rho, P)$ définie par:

$$V(\rho, P) = \frac{\frac{1}{n} \sum_{k=1}^n \gamma_k^2 [(B_{\rho, P}(z))_k - z_k]^2}{\left[\frac{1}{n} \text{tr} \left(I - (B'_{\rho, P})_z \right) \right]^2}$$

où $\gamma_k^2 = \frac{\partial^2 \phi}{\partial \mu_k^2}(z_k, z_k)$ et $\phi(\mu_k, z_k) = \mu_k - z_k \ln(\mu_k)$ pour $k=1, \dots, n$.

D'après l'étude comparative faite dans le dernière chapitre, ici on calcule la trace seulement en deux façons différentes. La première façon consiste à approcher la dérivée de $B_{\rho, P}$ par l'opérateur de lissage ICS obtenu à la fin des itérations qui est noté par $S_{\rho, P}$ et prendre son estimateur de Monte-Carlo. Maintenant, la remarque 4.2.1 sur le calcul de la différence entre la trace de $S_{\rho, P}$ et celle de $A_{\rho, P}$ est très intéressante. Donc en pratique on peut estimer la trace de $S_{\rho, P}$ par la trace de $A_{\rho, P}$.

Notons

$$\text{Num} = \frac{1}{n} \sum_{k=1}^n \gamma_k^2 [(B_{\rho,P}(z))_k - z_k]^2$$

et

$$V_1^*(\rho, P) = \frac{\text{Num}}{\left[\frac{W^T}{W^T W} (I - A_{\rho, P}) W \right]^2}.$$

La deuxième façon consiste à calculer la trace par l'estimateur de Monte-Carlo qui utilise les différences finies (suggéré et utilisé dans les chapitres II et III). Notons donc

$$V_2^*(\rho, P) = \frac{\text{Num}}{\left[\frac{W^T}{W^T W} \left(W - \frac{B_{\rho, P}(z+hw) - B_{\rho, P}(z)}{h} \right) \right]^2}.$$

La minimisation à P fixé de $V_1^*(\rho, P)$ et $V_2^*(\rho, P)$ par rapport à ρ est donc praticable.

4.4 Choix de ρ et de la position d'une discontinuité par GCV

Ici on suppose qu'il y a un saut d'ordre 0, un saut d'ordre 1, ou une rupture complète (c'est-à-dire saut d'ordre 0 et 1 en un même point) dans la fonction vraie. A partir de données bruitées de la fonction, on veut localiser la position et estimer l'amplitude de ce saut.

Remarquons d'abord que la localisation précise dans un intervalle $[t_i, t_{i+1}]$ est impossible, par exemple: si les données sont $z_1=0, \dots, z_i=0, z_{i+1}=1, \dots, z_n=1$, toutes les positions α dans l'intervalle $[t_i, t_{i+1}]$ sont évidemment possibles pour le saut d_α^0 . Si un saut est détecté dans cet intervalle, on prendra donc la convention de le placer au centre.

On souhaite estimer en même temps ρ et l'indice de la position i . Nous proposons ici de réaliser cette recherche de ρ et de i par GCV. Ceci a été fait pour les problèmes de lissage quadratique (Girard [88a], Girard et Laurent [88]).

Notons $V_1^*(\rho, i) = V_1^*(\rho, P)$ et $V_2^*(\rho, i) = V_2^*(\rho, P)$ les fonctions de GCV associées à un saut d'ordre 0 ou/et 1 à $\alpha = t_i$. On minimise $V_1^*(\rho, i)$ et $V_2^*(\rho, i)$ par rapport à ρ et l'indice i . Plusieurs observations

notées à partir des expériences numériques sont données dans la section suivante.

4.5 Expériences numériques

De nombreuses expérimentations numériques ont été réalisées, on présente ici quelques exemples d'applications. Dans tous les exemples de ce paragraphe, on considère que les points t_i , $i=1,\dots,n$ sont équidistants avec $n=201$, et

$$z_i \sim \text{Poisson}(g(t_i)), \quad i=1,\dots,n,$$

où $g \in \{g_1, g_2, g_3\}$ est une fonction définie sur $[0,1]$. Nous avons pris

$$g_1(t) = \begin{cases} 1+10 \times \exp\left(\frac{-(t-0.5)^2}{2 \times 0.4^2}\right), & 0 \leq t \leq 0.57, \\ 11+10 \times \exp\left(\frac{-(t-0.5)^2}{2 \times 0.4^2}\right), & 0.57 < t \leq 1, \end{cases}$$

qui a une discontinuité d'ordre 0,

$$g_2(t) = \begin{cases} 1+25 \times \sin(2\pi t), & 0 \leq t \leq 0.5, \\ 1+25 \times \sin(2\pi(t-1)), & 0.5 \leq t \leq 1, \end{cases}$$

qui a une discontinuité d'ordre 1, et

$$g_3(t) = \begin{cases} 1+10 \times \sin(2\pi t), & 0 \leq t \leq 0.4, \\ 6+3 \times \cos(4\pi t), & 0.4 < t \leq 1, \end{cases}$$

qui a une discontinuité complète.

Les résultats obtenus pour ces trois exemples sont donnés dans les figures 4.1 à 4.6. Notons ρ_1 et ρ_2 les paramètres de lissage qui minimisent respectivement $V_1(\rho, i_{\text{exact}})$ et $V_2(\rho, i_{\text{exact}})$, où i_{exact} note la vraie position de la discontinuité. Pour ces trois exemples, on a observé que les splines de lissages, obtenues par la méthode itérative décrite dans la section 4.2, en supposant que la vraie position de la discontinuité est donnée et avec le paramètre de lissage ρ_1 et ρ_2 sont visuellement les mêmes. Sur les figures 4.1, 4.3 et 4.5 on représente les données (par +), la fonction g (par la ligne continue), et la spline de lissage obtenue

en supposant que la vraie position de la discontinuité soit donnée et avec le paramètre de lissage p_1 respectivement pour les fonctions g_1 , g_2 et g_3 .

Pour le choix de α par GCV, on a étudié les fonctions $V_1^*(p_1, i)$ et $V_2^*(p_1, i)$. Sur les figures 4.2, 4.4 et 4.6 on représente ces deux fonctions $V_1^*(p_1, i)$ (par -----) et $V_2^*(p_1, i)$ (par -----), montrent que les minimums globaux (disons i_{opt}) sont tout près de la vraie position de la discontinuité.

En fait, avec ces valeurs optimales (p_{opt}, i_{opt}) on a obtenu les splines de lissage (non présentées ici) aussi bonnes que celles obtenues avec la vraie position de la discontinuité (représentées figures 4.1, 4.3 et 4.5). Notons que les indices cherchés valent respectivement 116, 100, et 80.

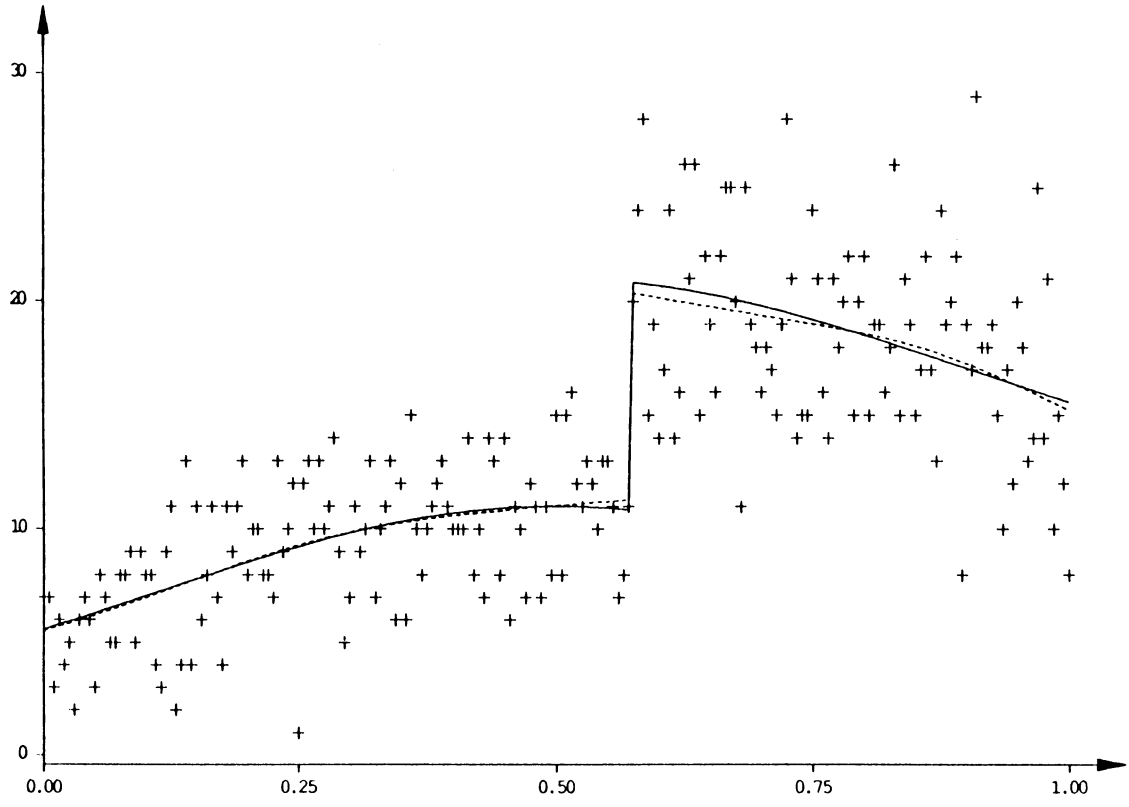


Figure 4.1 g_1

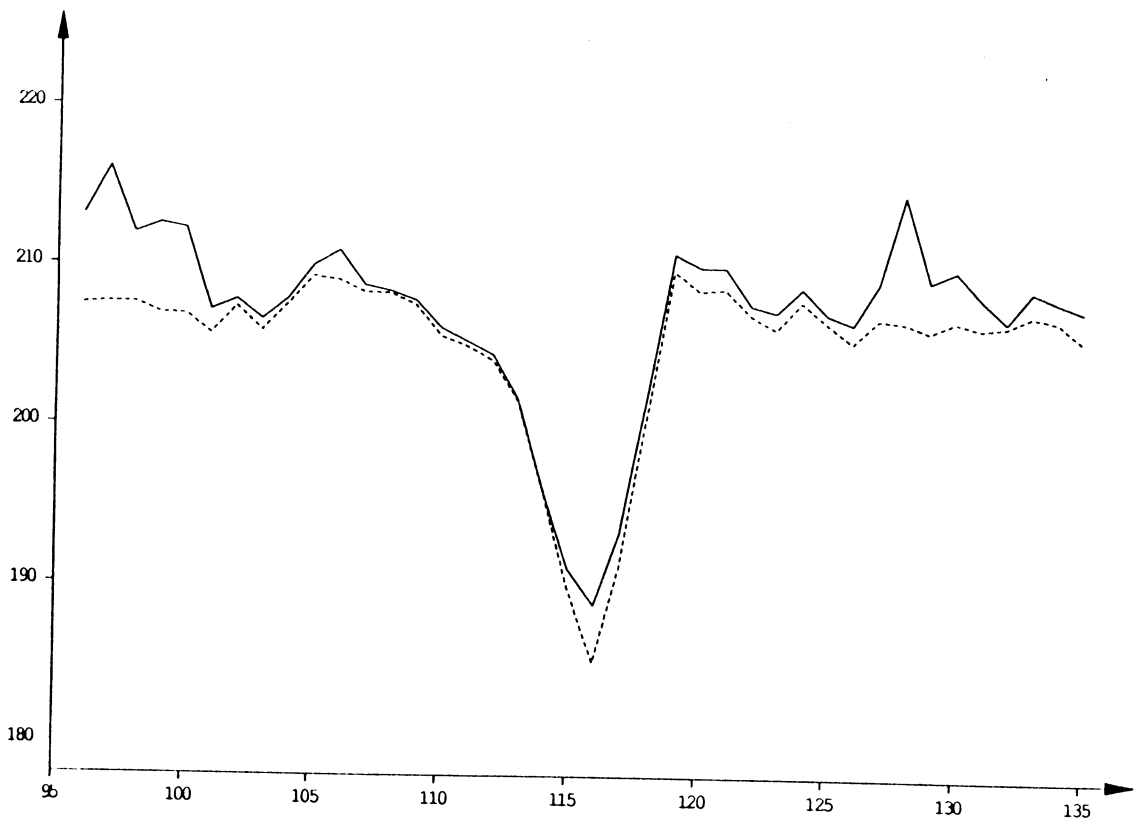
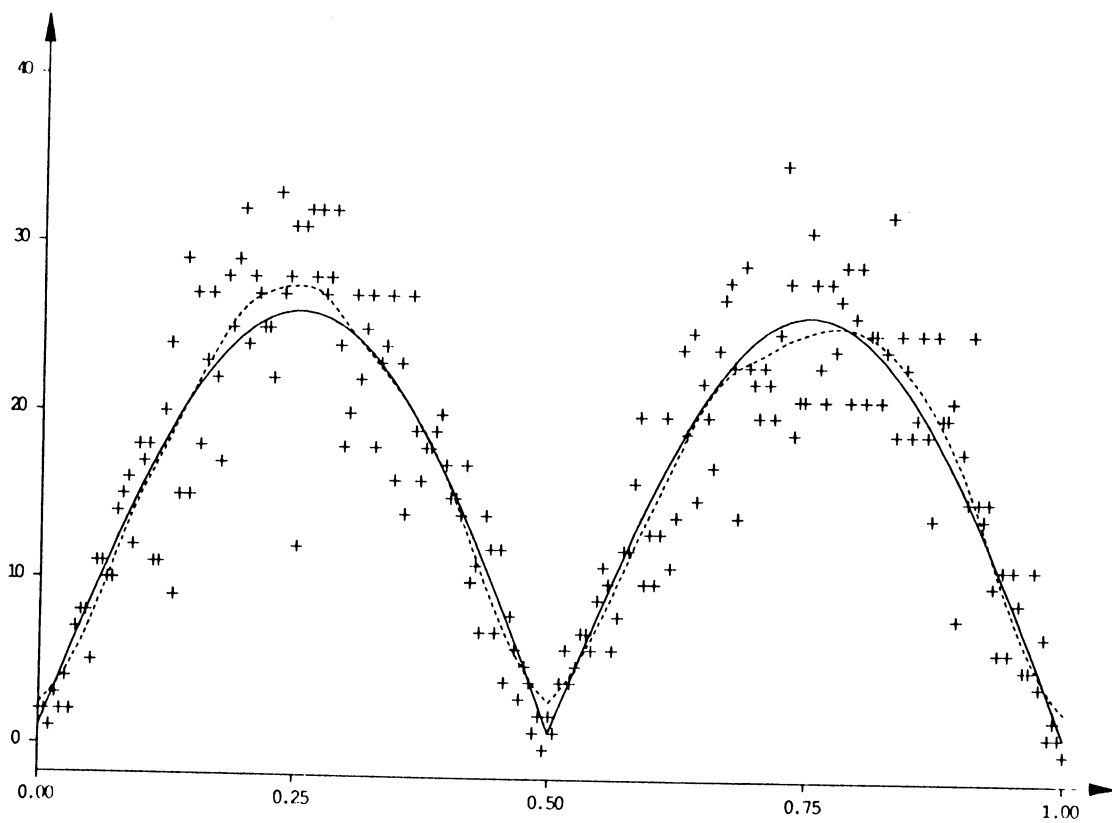
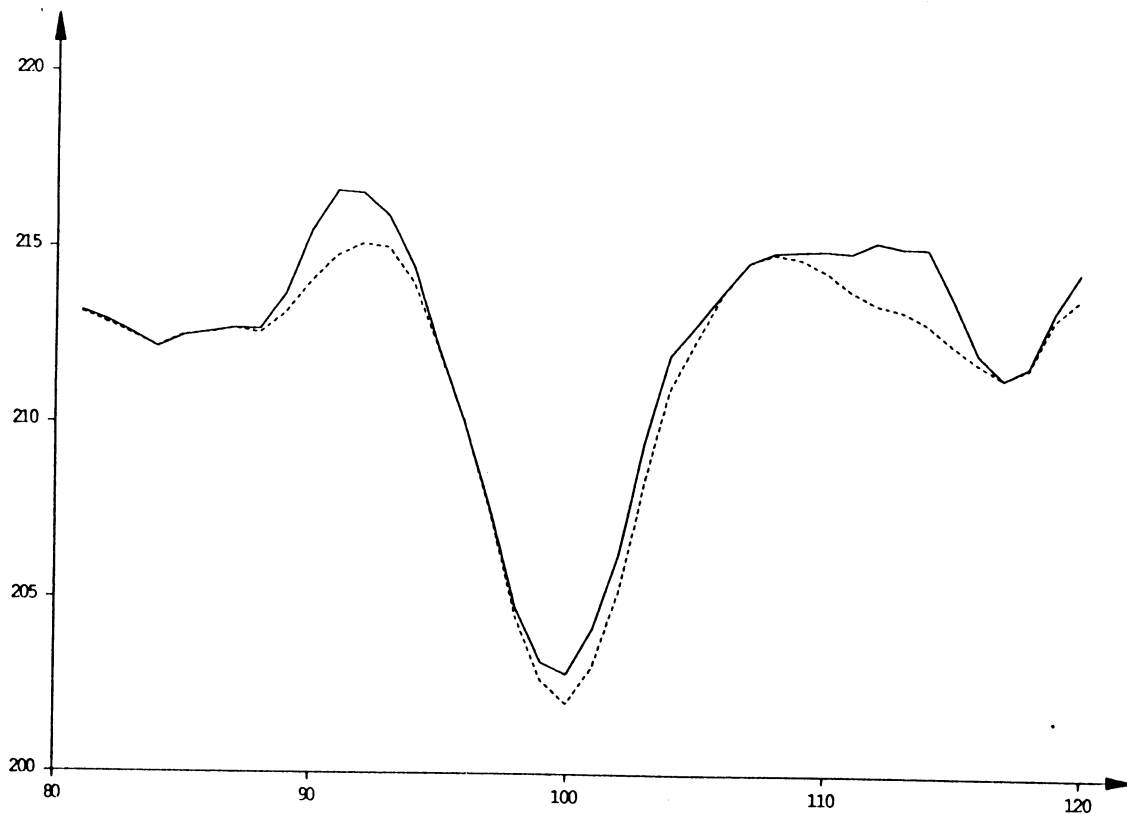


Figure 4.2 g_1

----- : $V_1(\rho_1,i)$, — : $V_2(\rho_1,i)$

Figure 4.3 g_2 Figure 4.4 g_2

----- : $V_1^*(p_1, i)$, — : $V_2^*(p_1, i)$

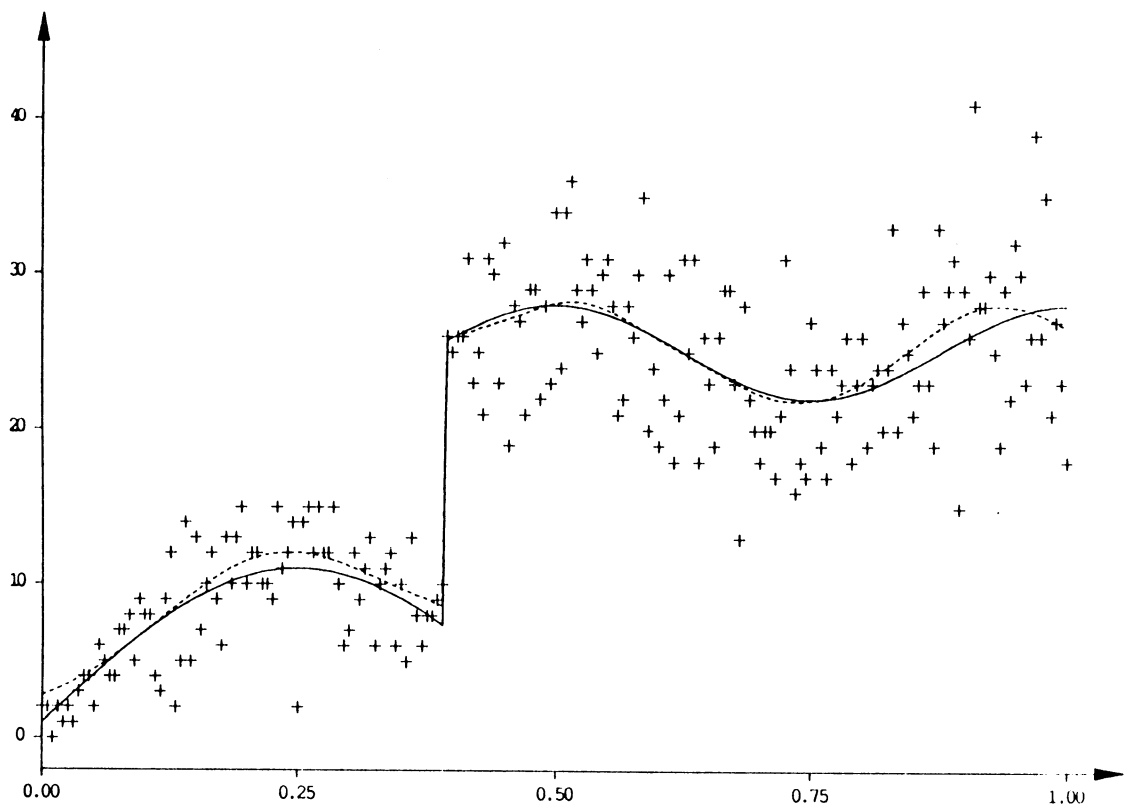


Figure 4.5 g_3

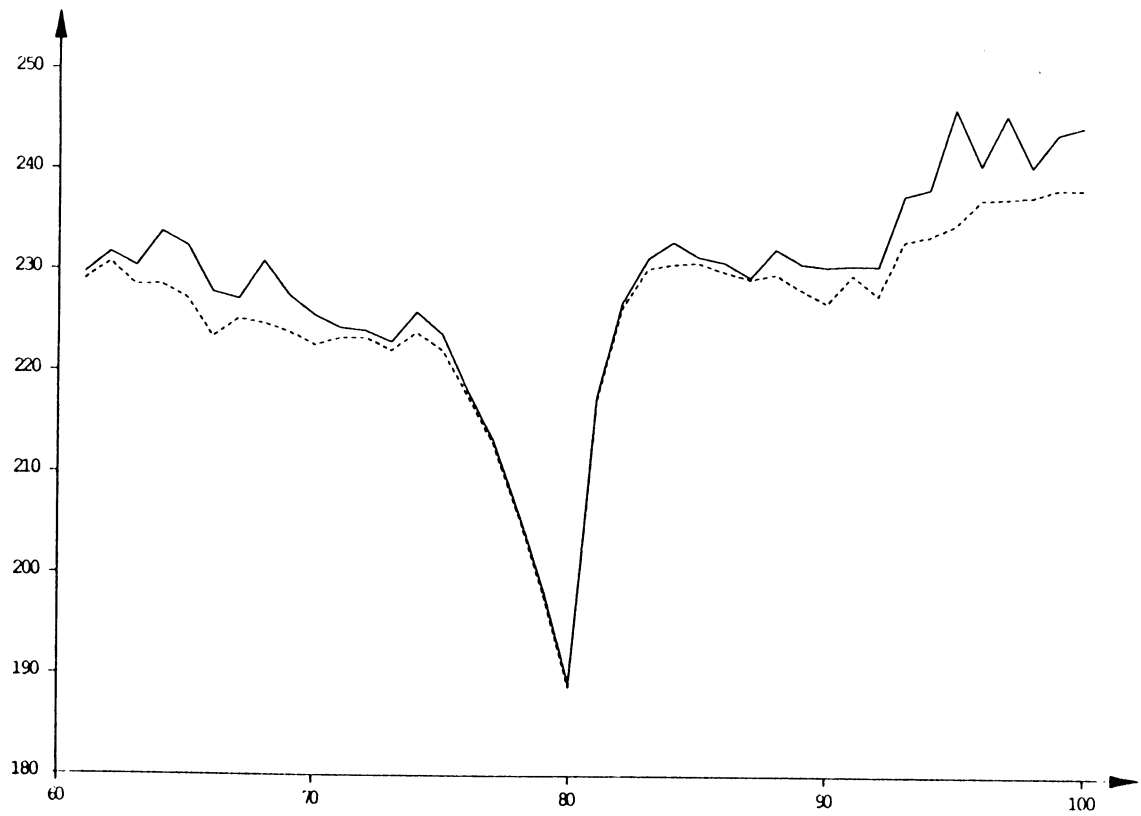


Figure 4.6 g_3

----- : $V_1^*(p_1,i)$, ——— : $V_2^*(p_1,i)$

Remarque 4.5.1: En pratique, il a été observé que pour un ρ fixé (raisonnablement), les minima de $V_1^*(\rho, i)$ et $V_2^*(\rho, i)$ par rapport à l'indice i sont très bien marqués quand il y a réellement une discontinuité.

Remarque 4.5.2: On a remarqué que, dans une large plage de valeurs de ρ autour de sa valeur optimale, l'indice i minimisant $V_1^*(\rho, i)$ et $V_2^*(\rho, i)$ est très souvent indépendant de ρ . L'exploitation de cette propriété serait utile pour réduire le coût global pour les problèmes de très grande taille.

Le critère pour arrêter les itérations était

$$\left(\frac{\|y^k - y^{k-1}\|^2}{\|y^k\|^2} < 10^{-10} \right).$$

Le nombre d'itérations nécessaires pour la convergence était en moyenne 6. Dans les exemples précédents, le coût de chaque itération était au plus $m+1$ splines (avec $m=2$). Pour les résultats numériques présentés ici nous avons pris $h=10^{-2}$.

Chapitre V

Choix du paramètre de lissage pour un problème d'enveloppe

5.1 Introduction

Dans ce chapitre nous considérons le problème de l'estimation d'une fonction g (lisse et continue) à partir de données bruitées $z_1, \dots, z_n$ du type:

$$(5.1) \quad z_i = z_i^0 + \epsilon_i, \quad i=1, \dots, n,$$

où z_i^0 est la réalisation d'une variable aléatoire Z_i^0 de loi uniforme sur $[0, g(t_i)]$ et ϵ_i est la réalisation d'une variable aléatoire E_i de loi normale centrée et de variance v^2 . C'est-à-dire, si la fonction g représente une courbe ouverte, alors les données seront uniformément réparties entre l'axe des x et la courbe $g(t)$ mais avec une perturbation gaussienne. Par exemple sur la figure 5.1a on représente la fonction g (par la ligne continue) et l'ensemble des données (par +). Un exemple de données pour une courbe fermée g est représentée sur la figure 5.1b.

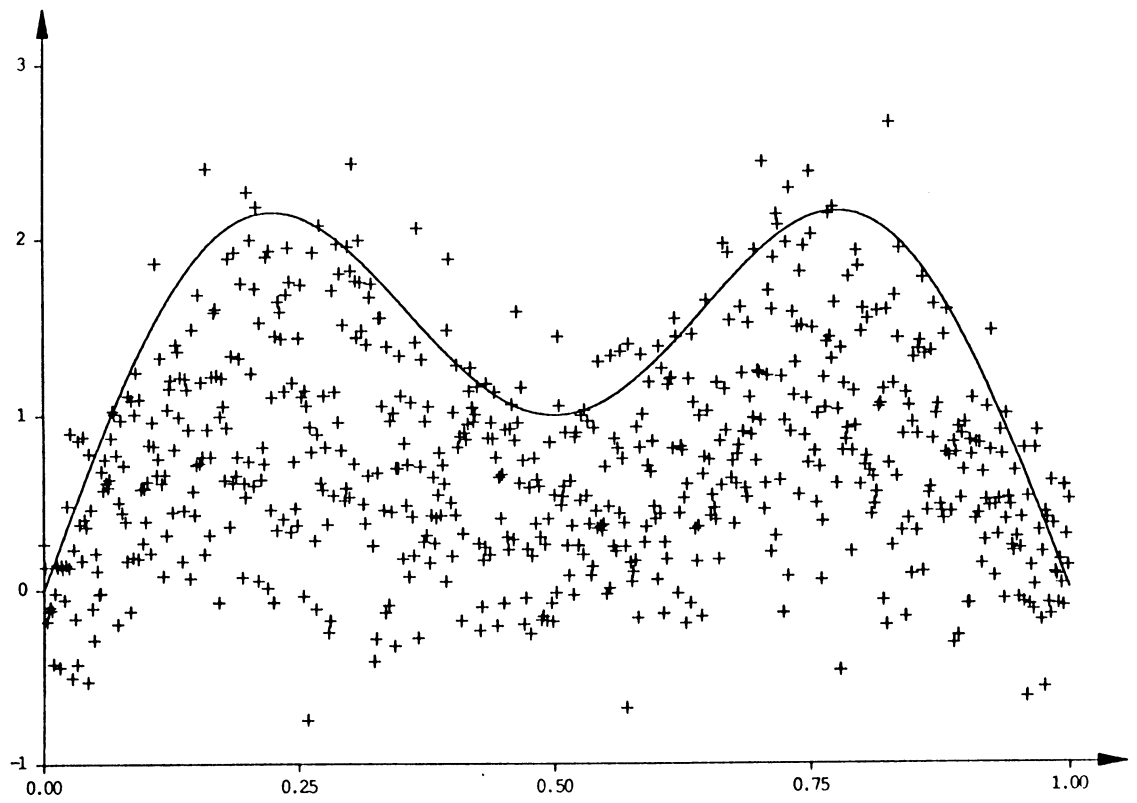


Figure 5.1a

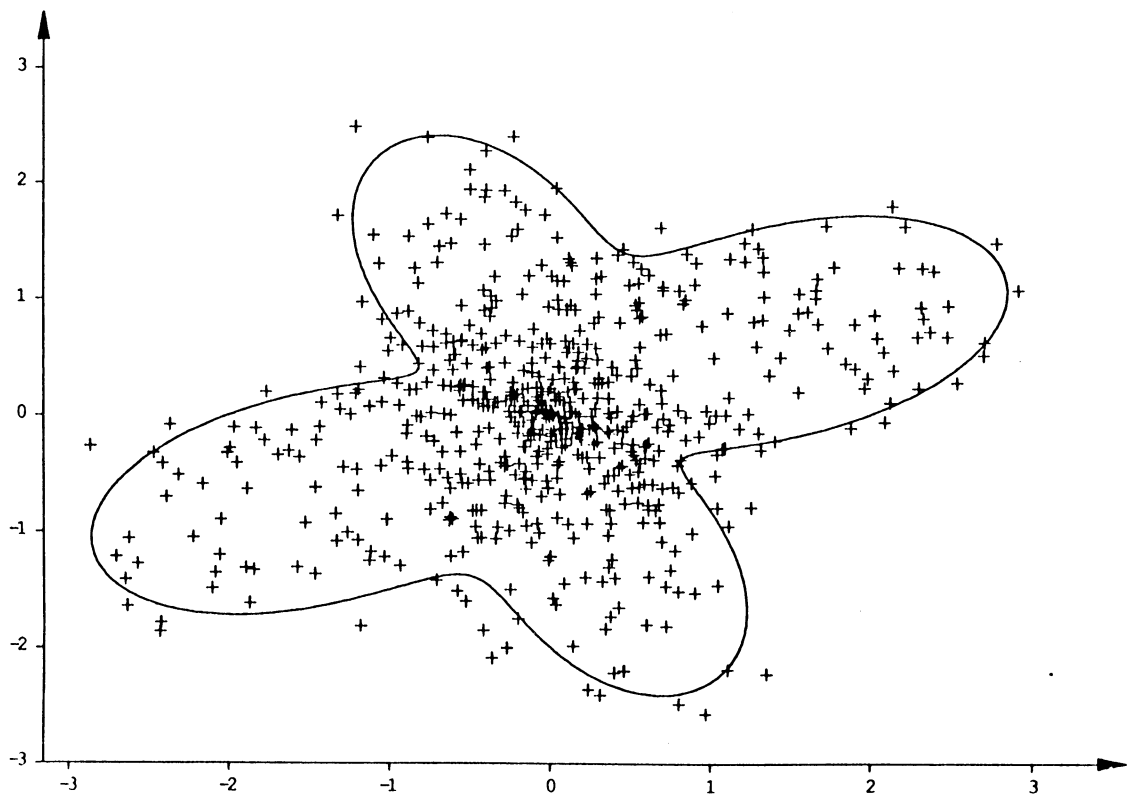


Figure 5.1b

Pour résoudre ce problème nous appliquons le principe de maximum de vraisemblance, comme dans le chapitre III, qui nous conduit à résoudre un problème de lissage non linéaire qui nécessite un bon choix du paramètre de lissage.

D'abord dans la section 5.2 nous formulons le problème en rappelons les fonctions de vraisemblances associées à ce type de données. La section 5.3 est consacrée à l'estimation de g pour une valeur fixé p du paramètre de lissage. Comme dans tout problème de lissage, le choix de ce paramètre est extrêmement important. Ceci est expliqué dans la section 5.3 avec des exemples numériques. On présente dans la dernière section 5.4 une méthode de GCV pour le choix de p ; ainsi que les résultats des expériences numériques effectuées.

5.2 Formulation du problème

Soit $z=(z_1, \dots, z_n)$ le vecteur des données du type (5.1). La variance du bruit v^2 qui est inconnue, peut être estimée. Ici on suppose que l'on connaît la valeur v^2 , et que la fonction g est positive et définie sur $[0,1]$.

Maintenant, on considère la fonction de vraisemblance pour ce type de données. La densité de probabilité de Z_i^0 est donnée par:

$$\begin{aligned} d_{Z_i^0}(u) &= \frac{1}{g(t_i)} \Pi_{[0, g(t_i)]}(u) \\ &= \begin{cases} \frac{1}{g(t_i)} & , u \in [0, g(t_i)] \\ 0 & , \text{sinon} \end{cases} \end{aligned}$$

Et la densité de probabilité de E_i est donnée par:

$$d_{E_i}(u) = \frac{1}{v\sqrt{2\pi}} \exp\left(-\frac{u^2}{2v^2}\right).$$

Alors, la densité de probabilité de la variable observée, que on notera par $d(z_i)$, est obtenue par convolution de $d_{Z_i^0}$ et d_{E_i} . C'est-à-dire

$$d(z_i) = (d_{Z_i^0} * d_{E_i})(z_i) = \frac{1}{g(t_i)} \int_0^{g(t_i)} \frac{1}{v\sqrt{2\pi}} \exp\left(-\frac{(z_i-u)^2}{2v^2}\right) du.$$

D'où la fonction de log-vraisemblance correspondant au modèle de données est:

$$L(\mu, z) = \ln \left(\prod_{i=1}^n d(z_i) \right)$$

$$= \sum_{i=1}^n \left[\ln \left(\int_0^{\mu_i} \frac{1}{v\sqrt{2\pi}} \exp \left(-\frac{(z_i-u)^2}{2v^2} \right) du \right) - \ln(\mu_i) \right]$$

où $\mu = (f(t_1), \dots, f(t_n))$ et f est l'approximation cherchée de g . Etant donné que la fonction g est positive, nous introduisons le terme de régularisation $R(\mu)$ suivant:

$$R(\mu) = (\ln \mu)^T \Omega (\ln \mu)$$

avec $\ln \mu = (\ln(\mu_1), \dots, \ln(\mu_n))$ pour des raisons classiques (à cause du changement de variable $f = \exp(\theta)$).

Notons:

$$\Phi_{\rho, z}(\mu) = R(\mu) - \rho L(\mu, z).$$

Alors, la minimisation de $\Phi_{\rho, z}(\mu)$ qui doit se faire sous la contrainte de positivité est équivalente à minimiser $\Psi_{\rho, z}(y) = \Phi_{\rho, z}(\psi(y_1), \dots, \psi(y_n))$ où $\psi(.) = \exp(.)$, par rapport à y . En conséquence, l'estimateur de g , $g_{\rho, z}$, est l'exponentielle de la spline cubique naturelle qui interpole $y_{\rho, z, i}$, $i=1, \dots, n$ en $t_1, \dots, t_n$ avec $y_{\rho, z}$ solution de:

$$(5.2) \quad \min_y \left\{ y^T \Omega y + \rho \sum_{i=1}^n \phi(\psi(y_i), z_i) \right\}$$

pour une valeur ρ du paramètre de lissage. Dans l'expression ci-dessus on a:

$$(5.3) \quad \phi(\psi(y_i), z_i) = \ln \left(\int_0^{\psi(y_i)} \frac{1}{v\sqrt{2\pi}} \exp \left(-\frac{(z_i-u)^2}{2v^2} \right) du \right) - y_i.$$

5.3 Estimateurs pour différentes valeurs de ρ

Pour résoudre le problème de lissage non quadratique (5.2), on doit utiliser une méthode itérative. Au début on a appliqué la méthode de Gauss-Siedel sur une grille de m points pour $m \ll n$. Mais pour les exemples considérés on a observé que cette méthode converge très lentement. Alors on a dû l'abandonner.

Après on a utilisé les itérations de Newton relaxée rappelées dans la section 1.6.1. En moyenne la convergence est obtenue en dix itérations.

Le calcul exact de la fonction $\phi(\psi(y_i), z_i)$ est compliqué puisqu'il nécessite le calcul de la fonction de répartition de Gauss. Pour tous les exemples numériques donnés dans ce chapitre, on approche la Gaussienne par des splines cubiques normalisées. Pour que cette approximation soit valable on impose la condition suivante sur les y_i pour bien définir la fonction $\phi(\psi(y_i), z_i)$,

$$\exp(y_i) > z_i - \alpha/2 \text{ où } \alpha = 3\sqrt{2} \, v.$$

Heureusement ça ne complique pas le calcul.

On envisage aussi l'introduction des contraintes de périodicité qui permettent surtout l'estimation de courbes fermées.

Pour les exemples numériques de cette section on a pris $n=650$. Sur les figures 5.2 à 5.5 on représente les données (par +), la fonction g_1^0 précisée dans la section 5.4 (par la ligne continue) et les splines de lissage obtenues pour quatre valeurs de ρ . De même, sur les figures 5.6 à 5.9, on présente les résultats pour la fonction g_1^f périodique. La première valeur de ρ (voir figures 5.2 et 5.6) conduit à un lissage beaucoup trop fort, la deuxième (voir figures 5.3 et 5.7) conduit à un lissage assez fort, mais elle permet déjà d'obtenir l'allure de la courbe. La quatrième valeur (voir figures 5.5 et 5.9) est trop grand et on remarque les oscillations. Enfin la troisième valeur (voir figures 5.4 et 5.8) est la meilleur des quatres; elle permet de reconstituer de façon précise la forme de la courbe initiale.

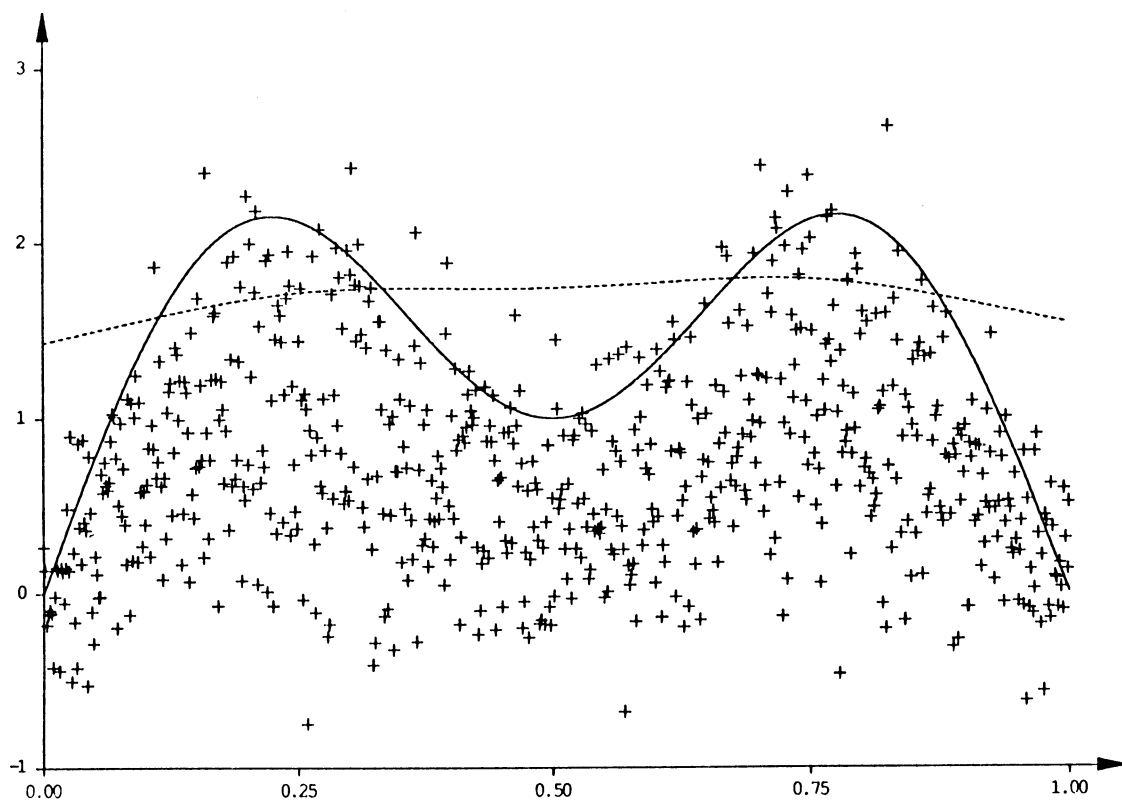


Figure 5.2 $g_1^0, \rho=1e0$

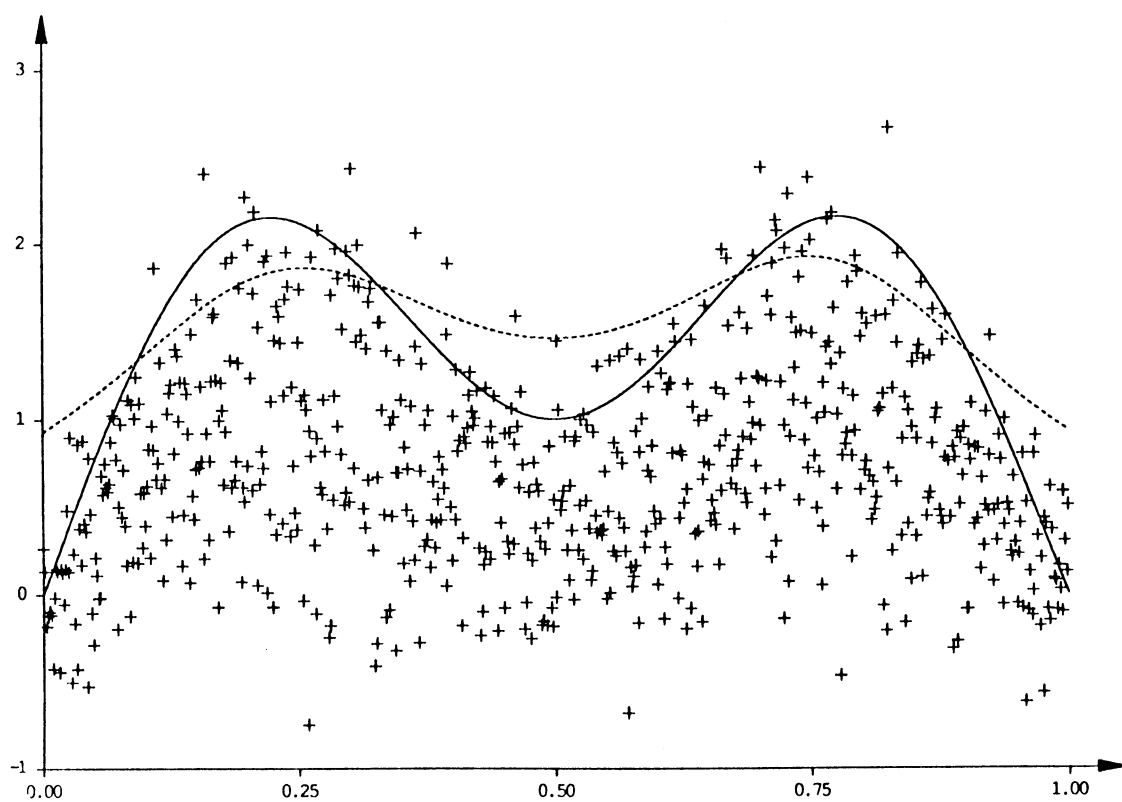


Figure 5.3 $g_1^0, \rho=1e1$

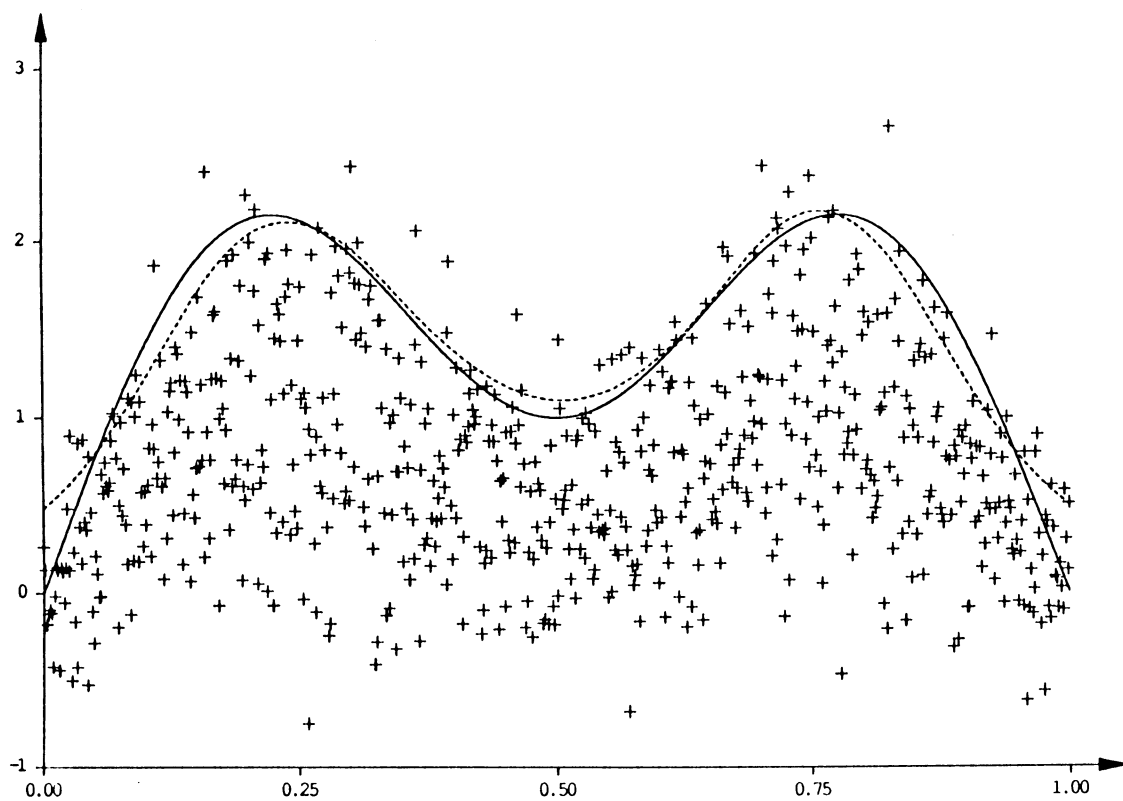


Figure 5.4 $g_1^0, \rho=1e2$

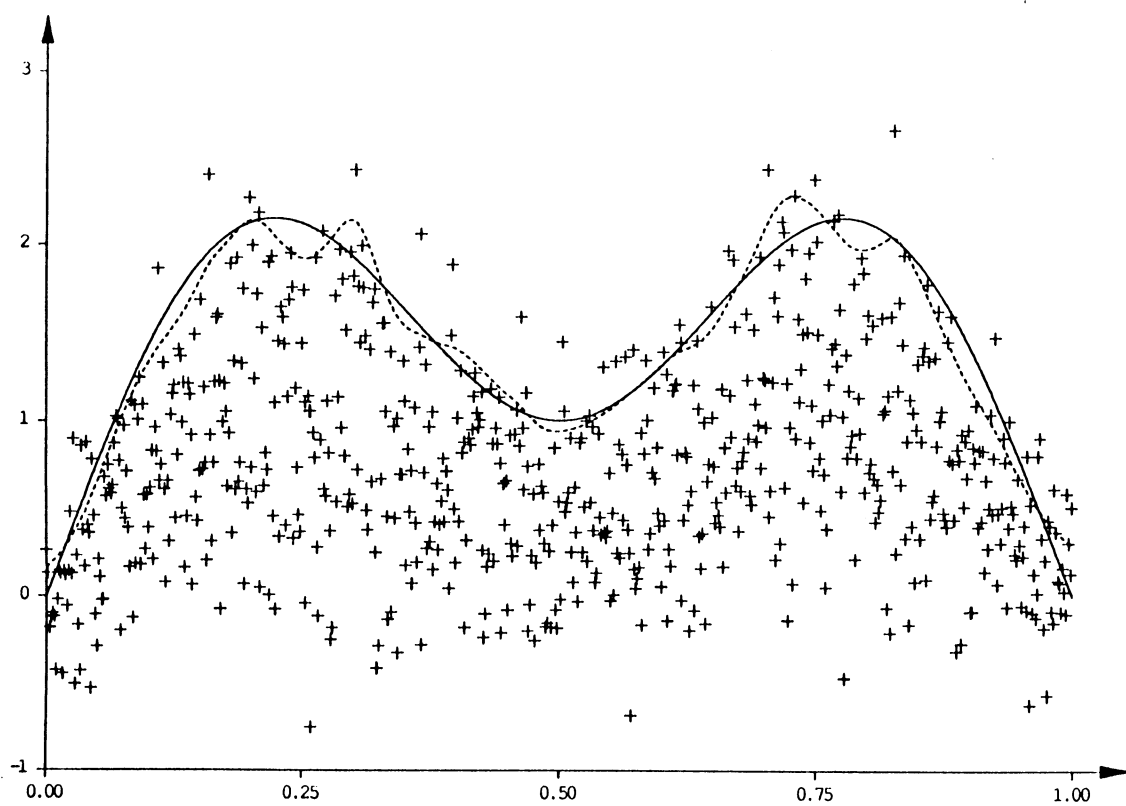
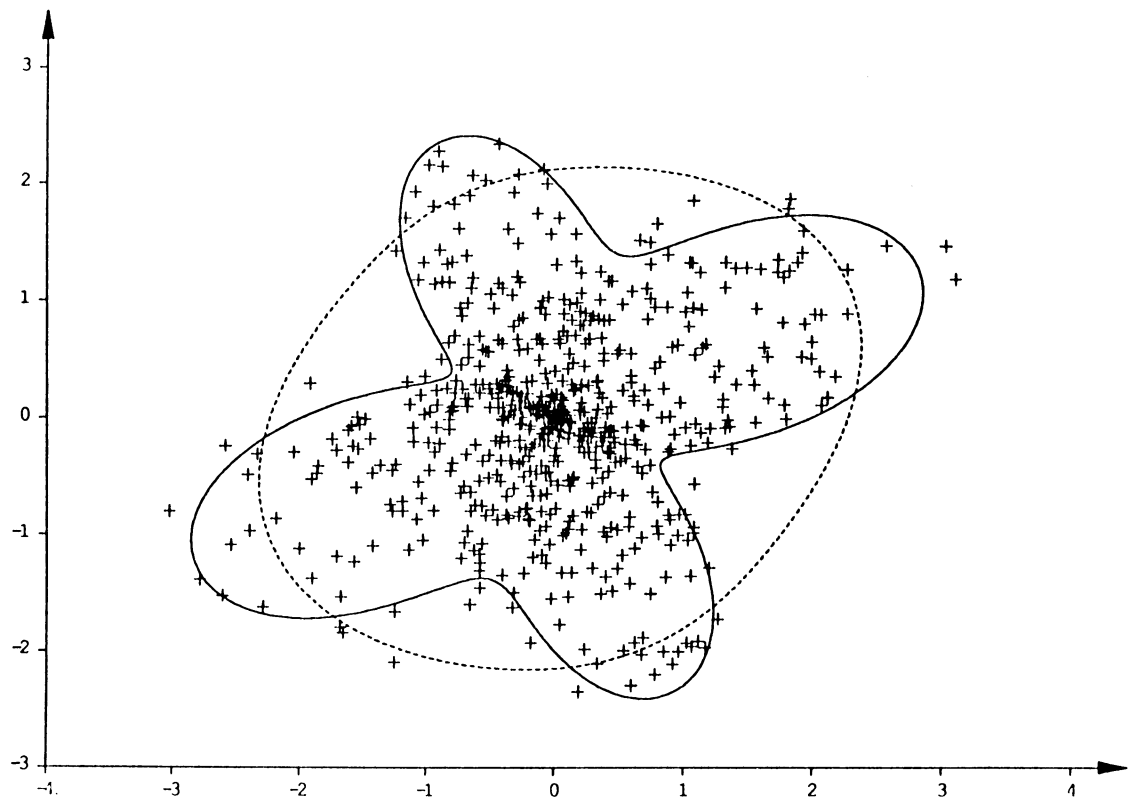
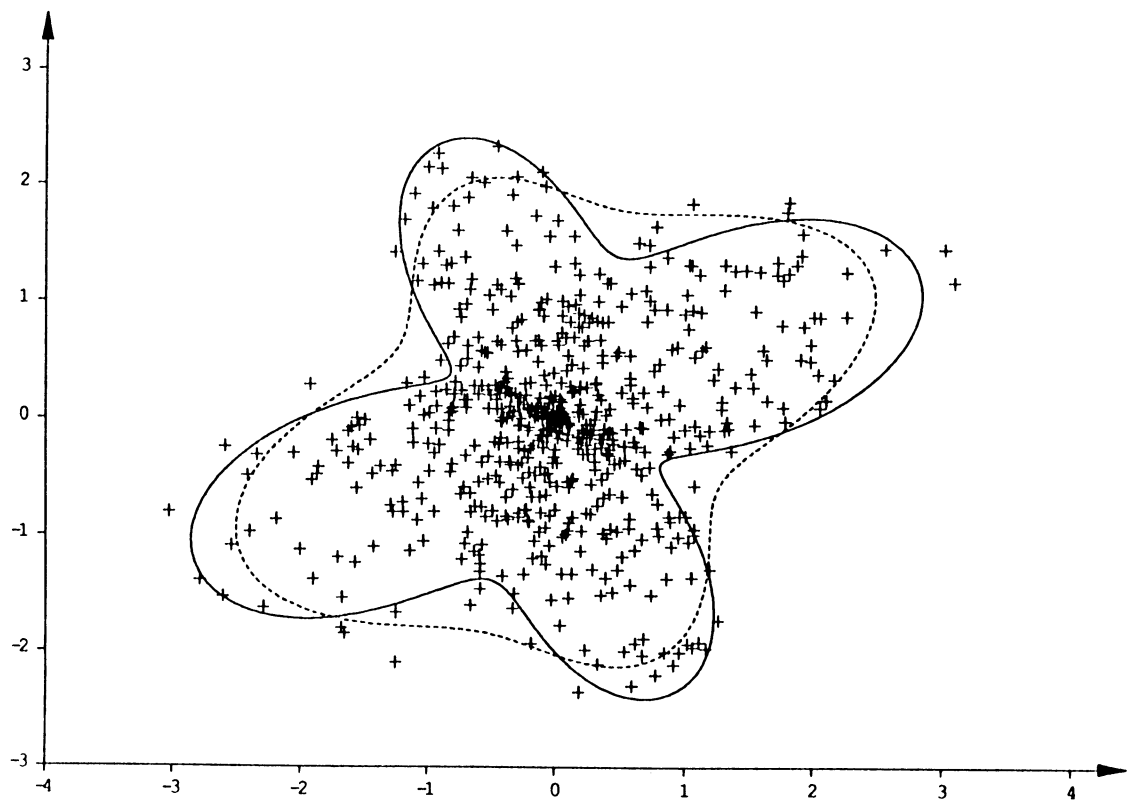


Figure 5.5 $g_1^0, \rho=1e4$

Figure 5.6 $g_1^f, \rho = 1e-1$ Figure 5.7 $g_1^f, \rho = 1e0$

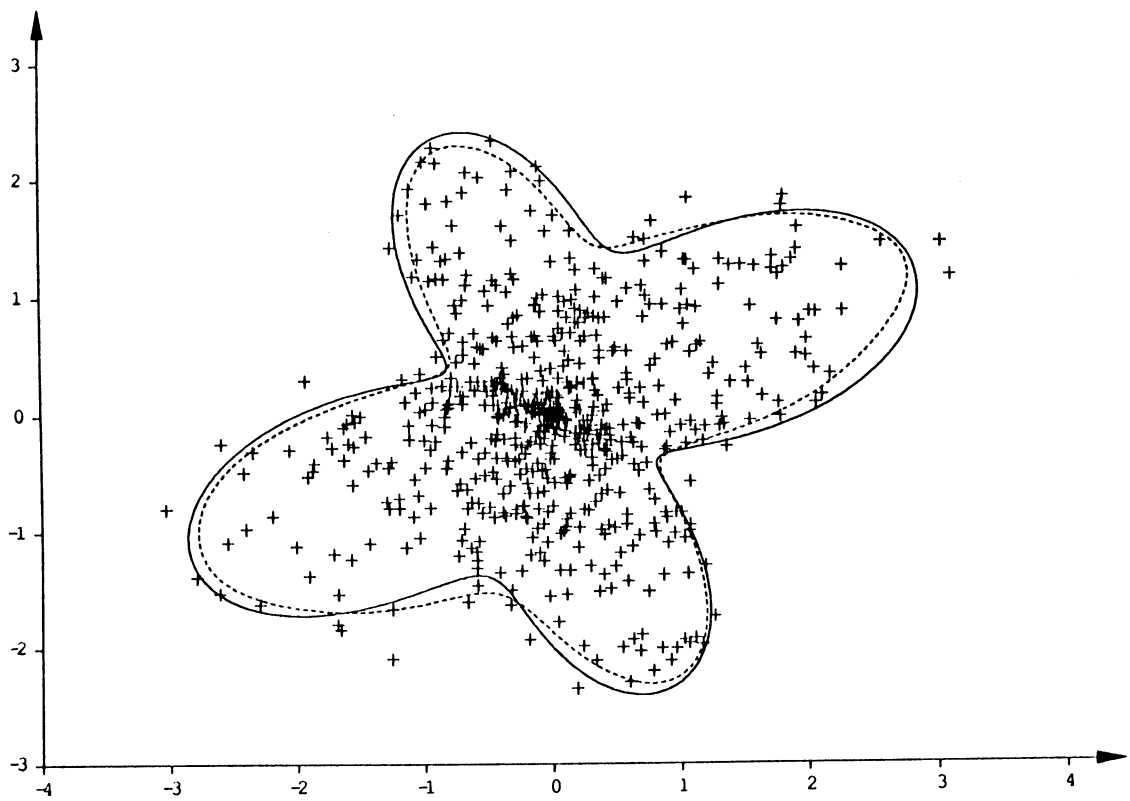


Figure 5.8 $g_1^f, \rho=1e1$

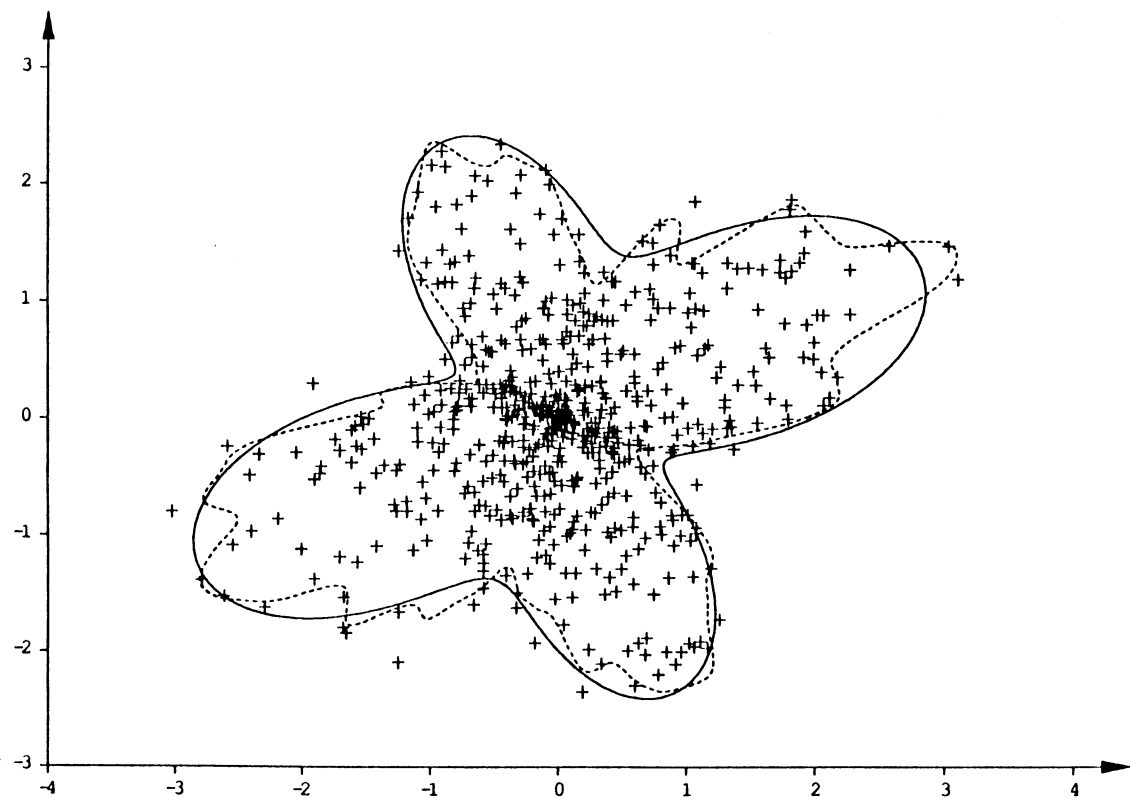


Figure 5.9 $g_1^f, \rho=1e3$

5.4 Choix de ρ

Pour le problème de lissage classique la fonction de validation croisée généralisée est définie par:

$$V(\rho) = \frac{\frac{1}{n} \sum_{k=1}^n [y_{\rho,z,k} - z_k]^2}{\left[\frac{1}{n} \text{Tr}(I - A_\rho) \right]^2}$$

où A_ρ est la matrice de lissage classique. Ici on propose de généraliser la fonction $V(\rho)$, pour le problème de lissage considéré dans ce chapitre, en utilisant l'expression suivante:

$$(5.4) \quad V(\rho) = \frac{\frac{1}{n} \sum_{k=1}^n \phi(\psi(y_{\rho,z,k}), z_k)}{\left[\frac{1}{n} \text{Tr}(I - A_\rho^{NM}) \right]^2}$$

où A_ρ^{NM} est la matrice de lissage obtenue à la fin des itérations. Nous définissons l'erreur quadratique par l'expression:

$$R(\rho) = \frac{1}{n} \sum_{k=1}^n [\psi(y_{\rho,z,k}) - z_k]^2.$$

Par analogie avec le cas classique on espère que la valeur de ρ qui minimise la fonction GCV est proche de celle qui minimise l'erreur quadratique $R(\rho)$. On définit l'inefficacité de GCV par:

$$\text{inefficacité} = \frac{R(\rho_V)}{R(\hat{\rho})}$$

où $\hat{\rho}$ est la valeur qui minimise l'erreur $R(\rho)$ définie ci-dessus et ρ_V minimise $V(\rho)$.

Remarque 5.4.1

Ici pour définir GCV, on ne peut pas appliquer le traitement du chapitre III pour les raisons suivantes. Dans la section 3.2.1, on a supposé que la fonction de log-vraisemblance $\phi(\mu_i, z_i)$ est convexe par rapport à μ_i et elle réalise la minimum à $\mu_i = z_i$. Ces deux suppositions ne sont pas valables ici.

Dans la section 3.2.3, nous avons supposé que $w_i \neq 0$, $i=1, \dots, n$ où

$w_i = \frac{1}{2} \frac{\partial^2 F}{\partial y_i^2}(y_{z,i}, z_i)$ pour $F(y_i, z_i) = \phi(\psi(y_i), z_i)$. Mais, ici les fonctions

$F(y_i, z_i)$ ne sont pas strictement convexe et la condition sur w_i n'est pas toujours satisfaite.

Les résultats numériques suivants montre que la généralisation de GCV donnée par (5.4) produit un bon choix de ρ .

Remarque 5.4.2

On a fait des expériences pour $n=200$. Il a été observé que les fonctions $R(\rho)$ et $V(\rho)$ sont convexes. Mais la valeur de ρ qui minimise $V(\rho)$ n'est pas un bon choix, c'est-à-dire que l'inefficacité est plus grande que 3.5. Par exemple dans les figures 5.10a et 5.10b, on a tracé les données (par +), la fonction exacte g_1^0 (par la ligne continue) et respectivement les splines de lissage avec le paramètre ρ_V qui minimise $V(\rho)$ et $\hat{\rho}$ qui minimise $R(\rho)$.

Ceci est dû (au moins c'est ce qu'on pense) au fait que pour $n=200$, il n'y pas assez des données dans le voisinage de la courbe exacte.

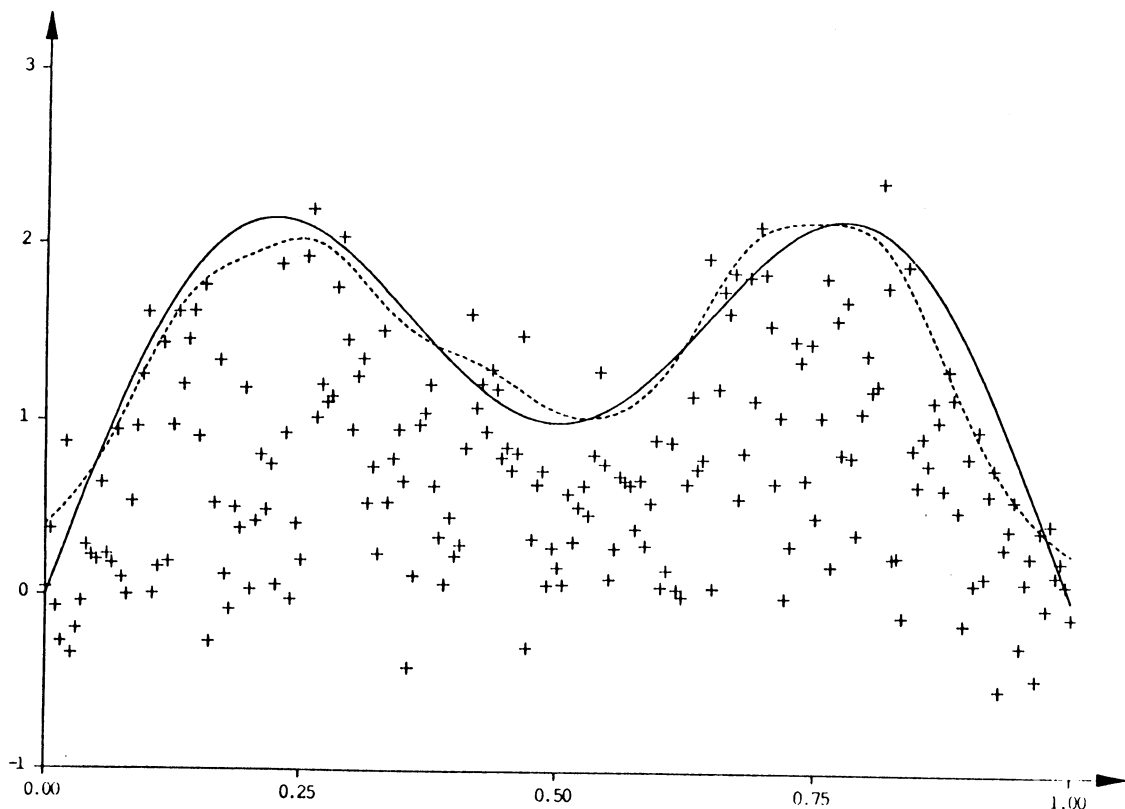


Figure 5.10a g_1^0 , $n=200$, spline de lissage avec $\hat{\rho}$

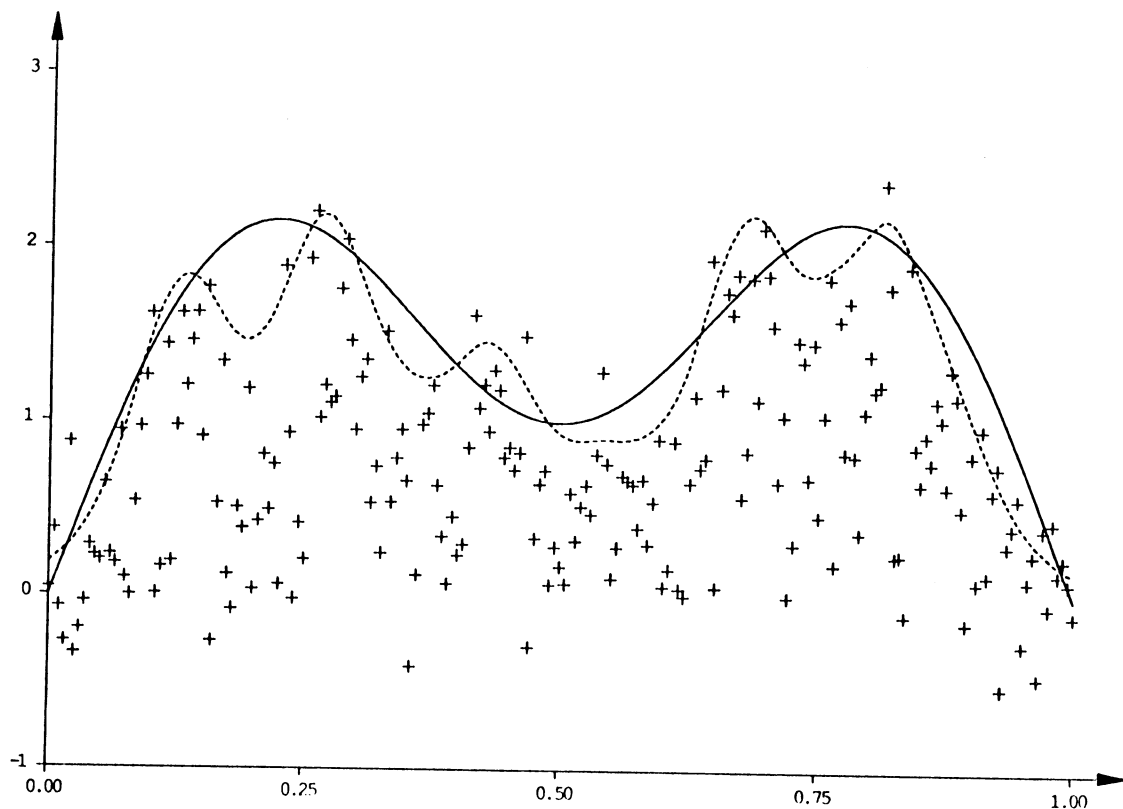


Figure 5.10b g_1^0 , $n=200$, spline de lissage avec ρ_V

Exemples d'utilisations

D'abord on illustre le comportement de la GCV avec quelques exemples qui sont les suivants: les données fictives sont générées par les trois fonctions suivantes pour $n=650$. On considère deux courbes ouvertes et une courbe fermée.

$$g_1^0(t) = 2\sin(\pi t) + \sin(3\pi t), \quad t \in [0, 1],$$

$$g_2^0(t) = 1 + 15.6 t^3 [2584t^8 + 105(1-t)^4] (1-t)^6, \quad t \in [0, 1],$$

$$g_1^f(t) = 2 + 0.4\sin(2t) + 0.8\sin(4t), \quad t \in [0, 2\pi],$$

Sur les figures 5.11 à 5.13, on représente les données (par +), les fonctions exactes (par la ligne continue) et les fonctions splines obtenues avec le paramètre de lissage qui minimise $V(\rho)$. La trace est approchée par l'estimateur de Monte-Carlo.

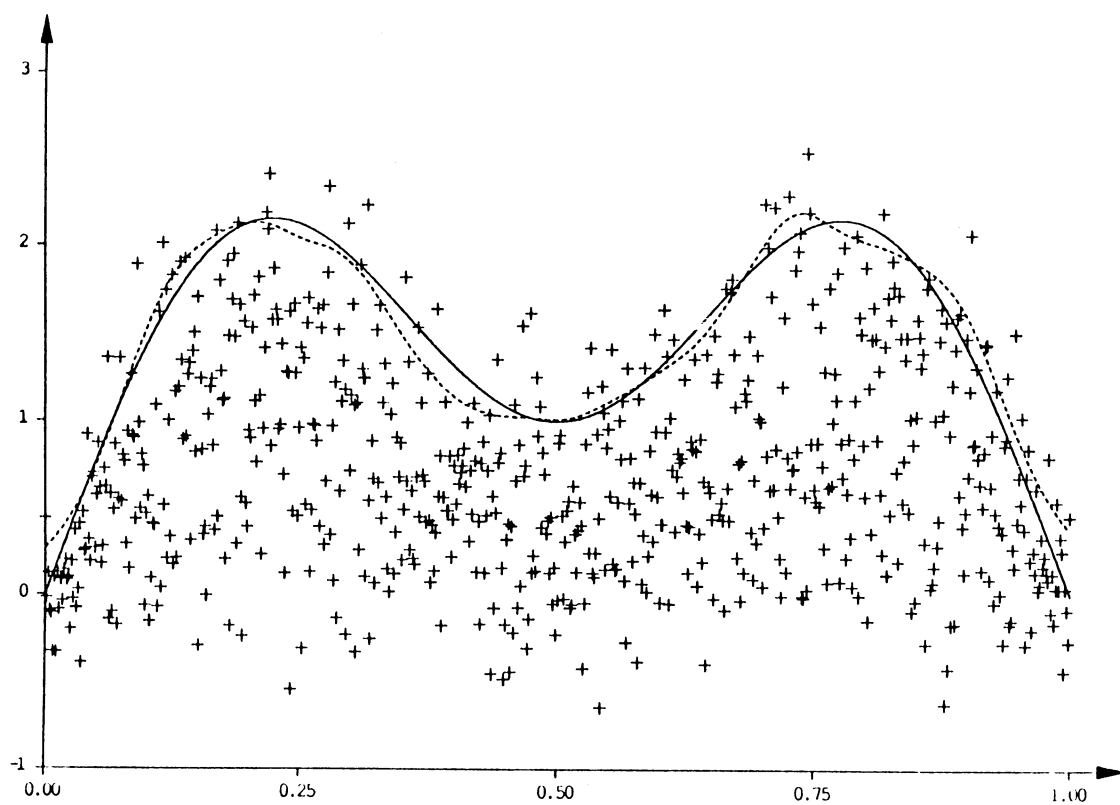


Figure 5.11 $g_1^0, v=0.3$

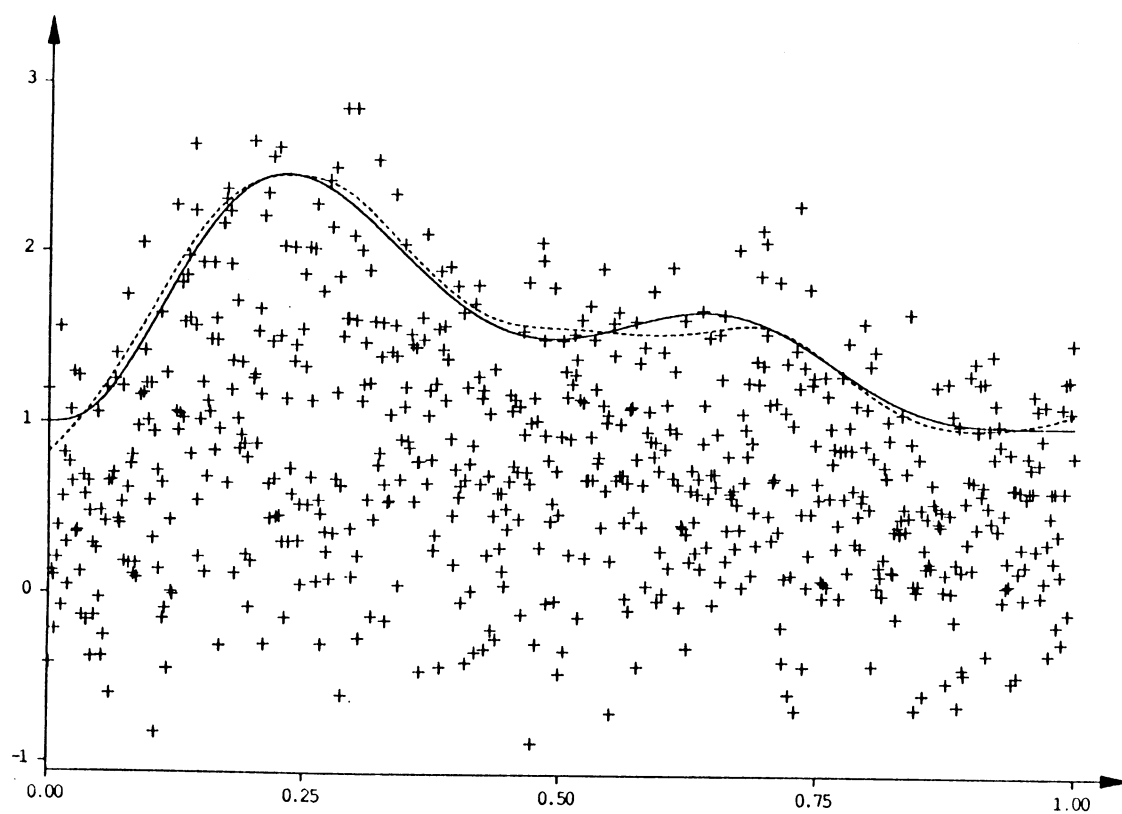


Figure 5.12 $g_2^0, v=0.4$

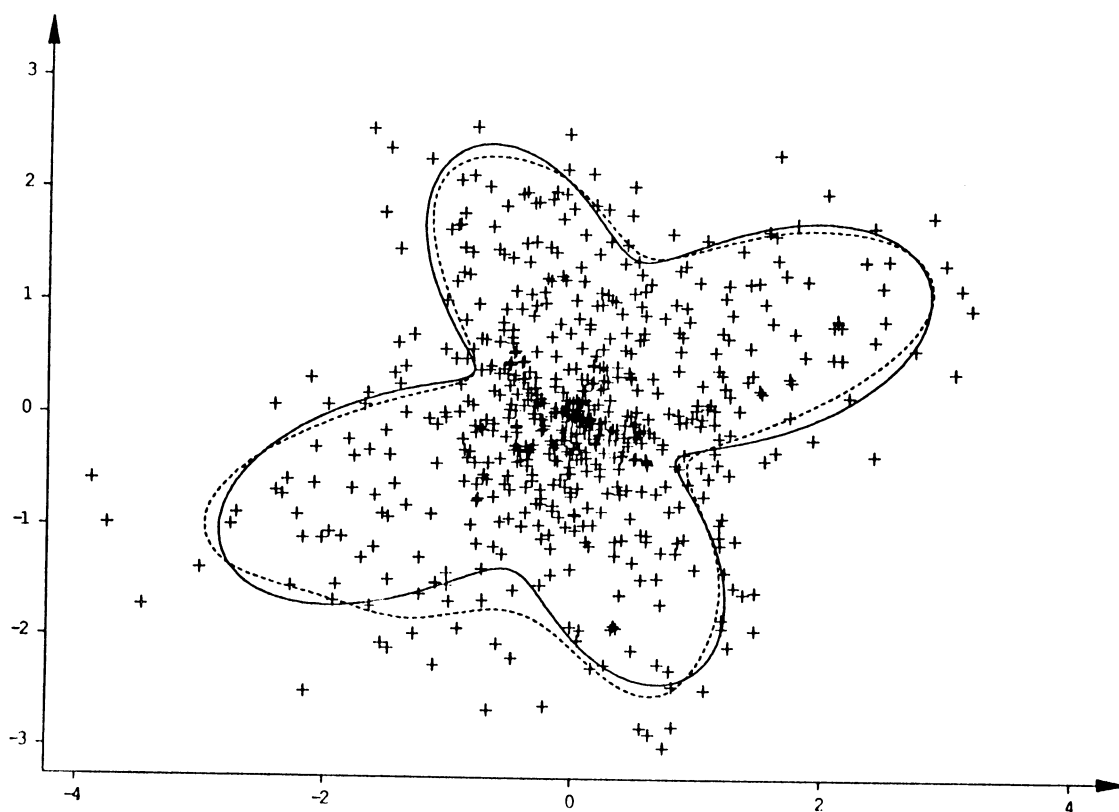


Figure 5.13 $g_1^f, v=0.6$

Pendant l'exécution on a noté des inefficacités qui sont présentées dans le tableau 5.1.

Tableau 5.1

Exemple	v	inefficacité	$\min_{\rho} R(\rho)$
g_1^0	0.3	1.03	6.85
g_2^0	0.4	1.05	1.79
g_1^f	0.6	1.01	1.39e+1

Pour une validation expérimentale plus complète, on a simulé 10 fois un exemple de données pour chaque valeur de $v \in \{0.2, 0.3, 0.4\}$ et la fonction g déjà explicitée. La distribution des inefficacités obtenues sur 90 cas est indiquée dans le tableau 5.2 suivant.

Tableau 5.2

inefficacité	(2.50,∞)	(1.66,2.50]	(1.25,1.66]	(1.00,1.25]
nombre des cas	1	11	22	56

Les résultats du tableau ci-dessus permettent de dire que la méthode de GCV proposée ici conduit à un bon choix de ρ .

Le nombre d'itération nécessaires pour la convergence était en moyenne 6, c'est-à-dire le coût de calculer la solution pour une valeur de ρ était 6 splines avec poids.

Annexe

Expression analytique pour $(B'_\rho)_z$

Nous considérons le problème de minimisation général:

$$(A.1) \quad \min_{y \in \mathbb{R}^n} \Psi_{\rho,z}(y)$$

où $\Psi_{\rho,z}(y) = y^T \Omega y + \rho \sum_{i=1}^n \phi(\eta_i(y), z_i)$, dont la solution est notée par y_z .

Soit $B_\rho(z) = (\eta_1(y_z), \dots, \eta_n(y_z))^T$ l'opérateur de lissage associé au problème (A.1). Il faut noter que B_ρ est non linéaire.

Dans ce paragraphe nous donnons l'expression explicite pour la dérivée de $B_\rho(z)$ par rapport à z , notée par $(B'_\rho)_z$.

Notons $f_i(y, z_i) = \phi(\eta_i(y), z_i)$, pour $i=1, \dots, n$ et pour chaque i , soit $J_i(y, z_i)$ le vecteur Jacobien de $f_i(y, z_i)$ par rapport à y i.e.

$$J_i(y, z_i) \equiv \nabla_y f_i(y, z_i) = \left(\frac{\partial f_i}{\partial y_1}, \dots, \frac{\partial f_i}{\partial y_n} \right)^T.$$

Le vecteur Jacobien de $\Psi_{\rho,z}(y)$ par rapport à y est alors défini par:

$$\Lambda(y, z) = 2 \left((\Omega y)_1, \dots, (\Omega y)_n \right)^T + \rho \sum_{i=1}^n J_i(y, z_i).$$

Puisque y_z minimise (A.1), nous avons pour tout z , $\Lambda(y_z, z) \equiv 0$ et donc la matrice Jacobienne de l'opérateur $z \mapsto \Lambda(y_z, z)$ est nulle, i.e.

$$\nabla_z \Lambda(y_z, z) + \nabla_y \Lambda(y_z, z) \cdot \nabla_z y_z = 0$$

où

$$\nabla_z \Lambda(y, z) = \rho \sum_{i=1}^n \nabla_z J_i(y, z_i), \quad \nabla_y \Lambda(y, z) = 2\Omega + \rho \sum_{i=1}^n \nabla_y J_i(y, z_i)$$

avec

$$\nabla_z J_i(y, z_i) = \begin{bmatrix} 0 \dots 0 \frac{\partial^2 f_i}{\partial z_i \partial y_1} 0 \dots 0 \\ \vdots \vdots \vdots \\ 0 \dots 0 \frac{\partial^2 f_i}{\partial z_i \partial y_n} 0 \dots 0 \end{bmatrix},$$

$$\nabla_y J_i(y, z_i) = \begin{bmatrix} \frac{\partial^2 f_i}{\partial y_1^2} \dots \frac{\partial^2 f_i}{\partial y_n \partial y_1} \\ \vdots \vdots \\ \frac{\partial^2 f_i}{\partial y_1 \partial y_n} \dots \frac{\partial^2 f_i}{\partial y_n^2} \end{bmatrix}$$

et

$$\nabla_z y_z = \begin{bmatrix} \frac{\partial y_{z,1}}{\partial z_1} \dots \frac{\partial y_{z,1}}{\partial z_n} \\ \vdots \vdots \\ \frac{\partial y_{z,n}}{\partial z_1} \dots \frac{\partial y_{z,n}}{\partial z_n} \end{bmatrix}.$$

Soit $W = \frac{1}{2} \sum_{i=1}^n \nabla_y J_i(y_z, z_i)$, $V = \frac{1}{2} \sum_{i=1}^n \nabla_z J_i(y_z, z_i)$ et $X = \nabla_z y_z$ à calculer.

On peut ainsi déterminer X par:

$$X = -(\Omega + \rho W)^{-1} \rho V.$$

Mais on a $B_\rho(z) = (\eta_1(y_z), \dots, \eta_n(y_z))^T$, donc

$$(A.2) \quad (B_\rho)_z = \Theta X = -\Theta(\Omega + \rho W)^{-1} \rho V$$

où

$$\Theta = \begin{bmatrix} \frac{\partial \eta_1}{\partial y_1}(y_z) \dots \frac{\partial \eta_1}{\partial y_n}(y_z) \\ \vdots \vdots \\ \frac{\partial \eta_n}{\partial y_1}(y_z) \dots \frac{\partial \eta_n}{\partial y_n}(y_z) \end{bmatrix}.$$

Bibliographie

(Abramowitz et Stegun [72])

M. Abramowitz et I. A. Stegun [1972]

Handbook of mathematical functions, Dover Publications, Inc., NY.

(Craven et Wahba [79])

P. Craven et G. Wahba [1979]

Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by generalized cross-validation, Num. Math. 31, 377-403.

(Demmler et Reinsch [75])

A. Demmler et C. Reinsch [1975]

Oscillation matrices with spline smoothing, Num. Math. 24, 375-382.

(Deshpande et Girard [89])

L. N. Deshpande et D. Girard [1989]

Cross-validation for regularized non-linear least square problems, RT 54, TIM3-IMAG, Université de Grenoble.

(Deshpande et Girard [90])

L. N. Deshpande et D. Girard [1990]

Fast computation of cross-validated robust splines and other non-linear smoothing splines, à paraître dans les proceedings de congrès 'Approximation of Curves and Surfaces', Chamonix, France, Juin 1990.

(Erisman et Tinney [75])

A. M. Erisman et W. F. Tinney [1975]

On computing certain elements of the inverse of a sparse matrix, Commun. ACM 18, 177-179.

(Ekblom et Madsen [89])

H. Ekblom et K. Madsen [1989]

Algorithms for non-linear Huber estimation, BIT 29, 60-76.

(Girard [86])

D. Girard [1986]

Practical optimal regularization of large linear systems, M²AN, RAIRO, Anal. Num., 20, 75-87.

(Girard [87a])

D. Girard [1987a]

Un algorithme rapide pour le calcul de la trace de l'inverse d'une grande matrice, RR 665-M, TIM3-IMAG, Université de Grenoble.

(Girard [87b])

D. Girard [1987b]

Optimal regularized reconstruction in computerized tomography, SIAM J. Sci. Stat. Comput. vol. 8, 934-950.

(Girard [87c])

D. Girard [1987c]

Un algorithme simple et rapide pour la validation corisée généralisée sur problèmes de grande taille, RR 669-M, TIM3-IMAG, Université de Grenoble.

(Girard [87d])

D. Girard [1987d]

A fast 'Monte-Carlo cross-validation' procedure for large least squares problems with noisy data, RR 687-M, TIM3-IMAG, Université de Grenoble; Num. Math. 56, 1-23 [1989].

(Girard [88a])

D. Girard [1988a]

Détection de discontinuités dans un signal (ou une image) par Inf-Convolution spline et validation croisée: un algorithme rapide non paramétré, RR 702-I-M, TIM3-IMAG, Université de Grenoble.

(Girard [88b])

D. Girard [1988b]

Sur les calculs d'incertitudes dans les méthodes de régression par fonctions splines, RT 47, TIM3-IMAG, Université de Grenoble.

(Girard [89])

D. Girard [1989]

Asymptotic optimality of the fast randomized versions of GCV and C_L in ridge regression and regularization, RR 793-M, TIM3-IMAG, Université de Grenoble.

(Girard [90])

D. Girard [1990]

From template matching to optimal approximation by piecewise smooth curves, to appear in SPIE Proceedings on "Curves and surfaces in computer vision and graphics", Santa Clara, CA, Feb. 11-16.

(Girard et Laurent [88])

D. Girard et P. J. Laurent [1988]

Splines and estimation of nonlinear parameters, RR 750-M, TIM3-IMAG, Université de Grenoble; Mathematical Methods in CAGD (Lyche, Schumaker, eds.) Academic Press 273-298 [1989].

(Golub et al [79])

G. H. Golub, M. Heath et G. Wahba [1979]

Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics, 21, 215-224.

(Green [90])

P. J. Green [1990]

Bayesian reconstructions from emission tomography data using a modified EM algorithm, IEEE Trans. on Medical imaging 9, 84-93.

(Haken et Madsen [89])

E. Haken et K. Madsen [1989]

Algorithms for non-linear Huber estimation, BIT 29, 60-76.

(Harouche et al [85])

E. Harouche, S. Peleg et L. S. Davis [1985]

Noisy image restoration by cost function minimization, Pattern Recognition Letters 3, 65-69.

(Huber [73])

P. Huber [1973]

Robust regression: asymptotics, conjectures and Monte-Carlo, The ann. of Stat. 1, 799-821.

(Huber [81])

P. Huber [1981]

Robust Statistics, John Wiley, New York.

(Hutchinson et DeHoog [85])

M. F. Hutchinson et F. R. DeHoog [1985]

Smoothing noisy data with spline functions, Numer. Math., 47, 99-106.

(Laurent [72])

P. J. Laurent [1972]

Approximation et optimisation, Hermann- Paris.

(Laurent [81])

P. J. Laurent [1981]

Inf-convolution spline pour l'approximation de données discontinues, RR 270, TIM3-IMAG, Université de Grenoble; M²AN, RAIRO, Anal. Num. 20, 89-111 [1986].

(Laurent [88])

P. J. Laurent [1988]

Inf-convolution splines, RR 749M, TIM3-IMAG, Université de Grenoble; à paraître dans "Constructive Approximation".

(Laurent et Utreras[86])

P. J. Laurent et F. Utreras [1986]

Optimal smoothing of noisy broken data, J. of Approx. Theory and its Appl. 2, 71-94.

(Li [85])

K. C. Li [1985]

From Stein's unbiased risk estimates to the method of generalized cross-validation, Ann. Statist. 13, 1352-1377.

(Li [86])

K. C. Li [1986]

Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing, Ann. Statist. 14, 1101-1112.

(Mallows [73])

C. L. Mallows [1973]

Some concepts on C_p , Technometrics 15, 661-675.

(Ortega et Rheinboldt [70])

J. M. Ortega et W. C. Rheinboldt [1970]

Iterative solution of non-linear equations in several variables, Academic Press, New York.

(Paihua [78])

L. Paihua [1978]

Quelques méthodes numériques pour les fonctions spline à une et plusieurs variables, Thèse, Grenoble.

(Reinsch [67])

C. Reinsch [1967]

Smoothing by spline functions, Num. Math. 10, 177-183.

(Reinsch [71])

C. Reinsch [1971]

Smoothing by spline functions II, Num. Math. 16, 451-454.

(Schumaker et Utreras [88])

L. L. Schumaker et F. Utreras [1988]

Asymptotic properties of complete smoothing splines and applications, SIAM J. Sci. Stat. Comput. vol. 9, 24-38.

(Shiau et al [86])

J. J. Shiau, G. Wahba et D. R. Johnson [1986]

Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise smooth two and three dimensional objective analysis, J. Atmospheric and Oceanic Technology 3, 714-725.

(Silverman [84])

B. W. Silverman [1984]

A fast and efficient cross-validation method for smoothing parameter choice in spline regression, J. Amer. Statist. Ass., 79, 584-589.

(Silverman et al [90])

B. W. Silverman, M. C. Jones, J. D. Wilson et D. W. Nychka [1990]

A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography, J. R. Statist. Soc. B 52, 271-324.

(Snyder et Miller [85])

D. L. Snyder et M. I. Miller [1985]

The use of sieves to stabilize images produced with the EM algorithm for emission tomography, IEEE Trans. Nuc. Sci. 32, 3864-3872.

(Speckman [85])

P. Speckman [1985]

Spline smoothing and optimal rates of convergence in nonparametric regression models, Ann. Statist. 13, 970-983.

(Stone [74])

M. Stone [1974]

Cross-validatory choice and assesement of statistical prediction, J. R. S. S. Ser B, 36(2), 111-147.

(O'Sullivan [83])

F. O'Sullivan [1983]

On the analysis of some penalised likelihood schemes, Tech. Rpt. No. 726, Statistics Dept., University of Wisconsin, Madison, WI.

(O'Sullivan [86])

F. O'Sullivan [1986]

A statistical perspective on ill-posed inverse problems, Stats. Sci. 1, 502-527.

(O'Sullivan [88a])

F. O'Sullivan [1988a]

Fast computation of fully automated log-density and log-hazard estimators, SIAM J. Sci. Stat. Comput. vol. 9, 363-379.

(O'Sullivan [88b])

F. O'Sullivan [1988b]

Nonparametric estimation of relative risk using splines and cross-validation, SIAM J. Sci. Stat. Comput. vol. 9, 531-542.

(O'Sullivan et Wahba [84])

F. O'Sullivan et G. Wahba [1984]

A cross-validated Bayesian retrieval algorithm for non-linear remote sensing experiments, Tech. Rep. no. 747. Dept. of Statistics, University of Wisconsin.

(O'Sullivan et al [86])

F. O'Sullivan, B. S. Yandell et J. Raynor [1986]

Automatic smoothing of regression functions in generalised linear models, J. Amer. Stat. Assos., 81, 96-103.

(Utreras [80])

F. Utreras [1980]

Sur le choix du paramètre d'ajustement dans le lissage par fonctions spline, Num. Math. 34, 15-28

(Utreras [81])

F. Utreras [1981]

Optimal smoothing of noisy data using spline functions, SIAM J. Sci. Stat. Comp. 2, 349-362.

(Utreras [83])

F. Utreras [1983]

Natural spline functions, their associated eigenvalue problem, Numer. Math. 42, 107-117.

(Wahba [75])

G. Wahba [1975]

Smoothing noisy data with spline functions, Num. Math. 24, 383-393.

(Wahba [77a])

G. Wahba [1977a]

Practical approximate solution of linear operator equations when the data are noisy, SIAM J. Numer. Anal., 14, 651-667.

(Wahba [77b])

G. Wahba [1977b]

A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem, The Annals of Stat. vol. 13, 1378-1402.

(Wahba [86])

G. Wahba [1986]

Partial and interaction spline models for the semiparametric estimation of functions of several variables, Computer Science and Statistics, T. J. Boardman ed., Proc. of the 18th Symposium on the Interface, Fort Collins.

(Wahba et Wold [75])

G. Wahba et S. Wold [1975]

A completely automatic French curve: Fitting spline functions by cross-validation, Comm. Statist. 4,1-7.

(Wold [74])

S. Wold [1974]

Spline functions in data analysis, Technometrics, 16, 1, 1-11.