Noyaux autoreproduisants à base d'ondelettes et débruitage de signaux sur plans d'expérience aléatoires

Réunion de rentrée MAFIA - Nissan les Enserune, 2007

A. Antoniadis*, U. Amato et M. Pensky

* Laboratoire LJK

Université Joseph Fourier

Grenoble, France

Introduction

A partir d'observations bruitées d'un fonction f échantillonnée sur un plan $\{t_1, \ldots, t_n\}$ estimer la fonction f. Modèle additif gaussien :

$$Y_i = f(t_i) + \epsilon_i, \quad i = 1, \ldots, n.$$

On suppose que f appartient à une classe suffisamment flexible d'un espace fonctionnel infini-dimensionnel.

Introduction

A partir d'observations bruitées d'un fonction f échantillonnée sur un plan $\{t_1, \dots, t_n\}$ estimer la fonction f. Modèle additif gaussien :

$$Y_i = f(t_i) + \epsilon_i, \quad i = 1, \dots, n.$$

On suppose que f appartient à une classe suffisamment flexible d'un espace fonctionnel infini-dimensionnel.

Une méthode populaire pour estimer la fonction de *regression* à partir des observations est la *régularisation* : minimiser une fonction objective qui établit un équilibre entre le degré de régularité de la solution et sa cohérence aux données.

Résumé

- Formulation du problème

Résumé

- Formulation du problème
- Espaces hilbertiens à noyau autoreproduisant

Résumé

- Formulation du problème
- Espaces hilbertiens à noyau autoreproduisant
- Ondelettes et espaces autoreproduisants

Résumé

- Formulation du problème
- Espaces hilbertiens à noyau autoreproduisant
- Ondelettes et espaces autoreproduisants
- Estimation et optimisation

Résumé

- Formulation du problème
- Espaces hilbertiens à noyau autoreproduisant
- Ondelettes et espaces autoreproduisants
- Estimation et optimisation
- Propriétés asymptotiques

Résumé

- Formulation du problème
- Espaces hilbertiens à noyau autoreproduisant
- Ondelettes et espaces autoreproduisants
- Estimation et optimisation
- Propriétés asymptotiques
- Algorithme de calcul et exemples.

Formulation

Soit $\mathcal{F}$ un espace fonctionnel et considérons

$$\boxed{\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \{(y_i - f(t_i))\}^2 + \lambda J(f)}$$

- Comment imposer à l'estimation de f un rétrecissement ?

Formulation

Soit $\mathcal{F}$ un espace fonctionnel et considérons

$$\boxed{\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \{(y_i - f(t_i))\}^2 + \lambda J(f)}$$

- Comment imposer à l'estimation de f un rétrecissement ?
- Quel type de pénalité $J(\cdot)$ choisir ?

Espaces auto-reproduisants (RKHS)

Soit $\mathcal{H}$ un espace de Hilbert fonctionnel séparable constitué de fonctions définies sur un domaine $\mathcal{T}$.

- Si, $\forall t \in \mathcal{T}$ la fonctionnelle d'évaluation $\delta_t f = f(t)$ est définie et continue sur $\mathcal{H}$, alors $\mathcal{H}$ est un *espace auto-reproduisant*.

Espaces auto-reproduisants (RKHS)

Soit $\mathcal{H}$ un espace de Hilbert fonctionnel séparable constitué de fonctions définies sur un domaine $\mathcal{T}$.

- Si, $\forall t \in \mathcal{T}$ la fonctionnelle d'évaluation $\delta_t f = f(t)$ est définie et continue sur $\mathcal{H}$, alors $\mathcal{H}$ est un *espace auto-reproduisant*.
- Pour un RKHS, il existe un *noyau auto-reproduisant* $K(\cdot,\cdot)$ tel que $\forall f \in \mathcal{H}, \forall x \in \mathcal{T}, \quad \langle K(t,\cdot), f\rangle = f(t)$.

Espaces auto-reproduisants (RKHS)

Soit $\mathcal{H}$ un espace de Hilbert fonctionnel séparable constitué de fonctions définies sur un domaine $\mathcal{T}$.

- Si, $\forall t \in \mathcal{T}$ la fonctionnelle d'évaluation $\delta_t f = f(t)$ est définie et continue sur $\mathcal{H}$, alors $\mathcal{H}$ est un *espace auto-reproduisant*.
- Pour un RKHS, il existe un *noyau auto-reproduisant* $K(\cdot, \cdot)$ tel que $\forall f \in \mathcal{H}, \forall x \in \mathcal{T}, \quad \langle K(t, \cdot), f \rangle = f(t)$.
- Il existe une correspondance bijective entre RKHS et un noyau semi-défini positif (Aronszjan (1951)).

Espaces auto-reproduisants (RKHS)

Soit $\mathcal{H}$ un espace de Hilbert fonctionnel séparable constitué de fonctions définies sur un domaine $\mathcal{T}$.

- Si, $\forall t \in \mathcal{T}$ la fonctionnelle d'évaluation $\delta_t f = f(t)$ est définie et continue sur $\mathcal{H}$, alors $\mathcal{H}$ est un *espace auto-reproduisant*.
- Pour un RKHS, il existe un *noyau auto-reproduisant* $K(\cdot,\cdot)$ tel que $\forall f \in \mathcal{H}, \forall x \in \mathcal{T}, \quad \langle K(t,\cdot), f\rangle = f(t)$.
- Il existe une correspondance bijective entre RKHS et un noyau semi-défini positif (Aronszjan (1951)).
- $\mathcal{H}_K$: RKHS généré par le noyau K.

Espaces auto-reproduisants (RKHS)

Soit $\mathcal{H}$ un espace de Hilbert fonctionnel séparable constitué de fonctions définies sur un domaine $\mathcal{T}$.

- Si, $\forall t \in \mathcal{T}$ la fonctionnelle d'évaluation $\delta_t f = f(t)$ est définie et continue sur $\mathcal{H}$, alors $\mathcal{H}$ est un *espace auto-reproduisant*.
- Pour un RKHS, il existe un *noyau auto-reproduisant* $K(\cdot,\cdot)$ tel que $\forall f \in \mathcal{H}, \forall x \in \mathcal{T}, \quad \langle K(t,\cdot), f\rangle = f(t)$.
- Il existe une correspondance bijective entre RKHS et un noyau semi-défini positif (Aronszjan (1951)).
- $\mathcal{H}_K$: RKHS généré par le noyau K.

Ondelettes et RKHS

Objectif : Utiliser une base d'ondelettes dans laquelle $\mathbf{f}$ sera développée en utilisant de manière explicite le plan d'expérience et la *régularité* de f. Schématiquement, la méthode est constituée de étapes suivantes :

Ondelettes et RKHS

Objectif : Utiliser une base d'ondelettes dans laquelle $\mathbf{f}$ sera développée en utilisant de manière explicite le plan d'expérience et la *régularité* de f. Schématiquement, la méthode est constituée de étapes suivantes :

- Choix de la sous-famille de la base d'ondelettes fondée sur le plan d'expérience.
- Estimation par pénalisation appropriée de la décomposition de f sur cette sous-famille.

Ondelettes et RKHS (suite)

On désire estimer f sur un inervalle compact D. Sans perdre de généralité, nous supposerons $D = [0,1]$.
Notons $G_{-1} = \{-1\} \times \{0\}$, $G_0 = \{0\} \times \{0,1\}$ et pour tout entier $L \geq 1$ soit
$G_L = \{L\} \times \{k \in \{0,\dots,2^L-1\};\ k/2 \notin \mathbb{Z}\}$.
On construit une base d'ondelettes orthonormée de $L_2([0,1))$ en périodisant une base orthonormée de $L_2(\mathbb{R})$ générée par les dilatations et les translations d'une fonction échelle à support compact, $\phi(t)$, et de l'ondelette mère correspondante, $\psi(t)$, découlant d'une AMR r-regulière de $L^2(\mathbb{R})$.

(suite)

On obtient ainsi

$$L_2([0,1)) = V_0 \oplus W_0 \oplus W_1 \oplus \cdots ,$$

où V_0 est l'espace des constantes (engendré par $\phi_{0,0} = \psi_{-1,0}$) et W_j est l'espace 2^j-dimensionnel engendré par les ondelettes indexées par G_j.

Pour $L \geq 1$ définissons $V_L = V_0 \oplus \oplus_{j=0}^{L-1} W_j$. Pour tout $f \in L^2[0,1]$ et tout entier $j_0 \geq 0$, on note $u_{j_0 k} = \langle f, \phi_{j_0 k} \rangle$ ($k = 0, 1, \ldots, 2^{j_0} - 1$) et $w_{jk} = \langle f, \psi_{jk} \rangle$ ($j \geq j_0$; $k = 0, 1, \ldots, 2^j - 1$);

(suite)

On a

$$f(t) = \sum_{k=0}^{2^{j_0}-1} u_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} w_{jk}\psi_{jk}(t), \quad t \in [0,1].$$

La totalité des indices (j,k) décrivant les ondelettes de cette base sera noté $G = \cup_{j\geq -1} G_j$. Ainsi on a

$$f = \sum_{g\in G} f_g \psi_g$$

où ψ_g est l'ondelette indexée par $g \in G$ et f_g est le coefficient correspondant.

RKHS et ondelettes

Pour toute fonction

$$\Gamma : G \longrightarrow [0, \infty)$$

on définit l'espace de Hilbert

$$\mathcal{H}_\Gamma = \{ f \in L_2([0,1)) : \sum_{g \in G} \Gamma(g) |f_g|^2 < \infty \},$$

avec

$$\langle f, h \rangle_\Gamma = \sum_{g \in G} f_g h_g \Gamma(g),$$

et norme associé $\| \cdot \|_\Gamma$.

Comme G_L est un sous-ensemble fini de G, on a

$$V_L \subset \mathcal{H}_\Gamma$$

pour chaque $L \geq 0$.
De plus, pour tout $f \in \mathcal{H}_\Gamma$,

$$\lim_{L \to \infty} \|f - \pi_L(f)\|_\Gamma = 0, \tag{1}$$

où π_L est le projecteur orthogonal de $L_2([0,1))$ sur V_L.

Soit maintenant $\Gamma : G \longrightarrow [0, \infty)$ telle que

$$\sum_{j \geq 0} 2^{j/2} \Gamma_j^{-1/2} = B_1 < \infty$$

avec

$$\Gamma_j = \min_{g \in G_{j+1}} |\Gamma(g)|.$$

Alors

$$\boxed{\text{L'espace } \mathcal{H}_\Gamma \text{ est un espace auto-reproduisant.}}$$

Pour montrer que $\mathcal{H}_\Gamma$ est un RKHS choisissons $M > 0$ et $B_2 > 0$ tels que : $\forall x \in [0,1)$ et $\forall j \geq 0$

- il y ait au plus M ondelettes de base indéxées par des indices de G_{j+1} qui ne sont pas nulles en x, et
- le module maximum des fonctions de base indéxées par G_{j+1} soit $\leq B_2 2^{j/2}$.

Ceci est possible car les ondelettes considérées sont compactes et périodiques.

Posons $\gamma := MB_2\sqrt{B_1}$. Soit $f \in \mathcal{H}_\Gamma$ et notons

$$f_j := \max_{g \in G_{j+1}} |f_g|.$$

L'inégalité de Schwartz implique que

$$|f(x)| \leq MB_2 \sum_{j \geq 0} f_j 2^{j/2} \leq \gamma \|f\|_\Gamma.$$

Donc $\mathcal{H}_\Gamma$ est un RKHS and comme il admet un sous-espace dense de fonctions continues et que γ ne dépend pas de x, toute fonction de $\mathcal{H}_\Gamma$ est continue.

Les noyaux autoreproduisants sont

$$K(x,\cdot) = \sum_{g \in G} \frac{\psi_g(x)}{\Gamma(g)} \psi_g, \quad x \in [0,1).$$

On peut encore écrire

$$K_j(s,t) = \sum_{k=0}^{2^j-1} \frac{\psi_{j,k}(s)}{\Gamma((j,k))} \psi_{j,k}(t)$$

On a donc

$$\mathcal{H}_\Gamma = V_0 \oplus \oplus_{j \geq 0} \mathcal{W}_{j,\Gamma}, \tag{2}$$

où chaque espace de détails est le RKHS associé au noyau K_j, i.e. le RKHS engendré par un ensemble d'ondelettes à l'échelle j. Notons aussi que si Γ est uniquement fonction de j dans G_j, alors K est une somme pondérée, Γ_j^{-1}, des noyaux

$$K_j(s,t) = \sum_{k=0}^{2^j-1} \psi_{j,k}(s)\psi_{j,k}(t).$$

Théorème de représentation

Il s'agit d'un lemme général d'optimisation dans des RKHS (Kimeldorf & Wahba (1971)) pour résoudre le problème suivant :
Etant donnés $\{y_i, t_i;\ i = 1, 2, \ldots, n\}$, $y_i \in \mathbb{R}$ et $t_i \in \mathcal{T}$, K et (éventuellement) M fonctions particulières $\{\Phi_1, \ldots \Phi_M\}$ sur $\mathcal{T}$, déterminer f de la forme $f(s) = \sum_{d=1}^{M} a_d \Phi_d(s) + h(s)$ avec $h \in \mathcal{H}_K$ minimisant

$$\mathcal{I}(f, \mathbf{y}) := \sum_{i=1}^{n} \mathcal{C}(y_i, f(t_i)) + \lambda^2 \|h\|^2_{\mathcal{H}_K}, \qquad (3)$$

où $\mathcal{C}$ est une fonction convexe de f.

Si l'on suppose que le minimiseur de $\mathcal{C}(y_i, f(t_i))$ dans l'espace engendré par les Φ_d'est unique, alors l'argmin de $\mathcal{I}(f, \mathbf{y})$ est de la forme :

$$f(s) = \sum_{d=1}^{M} a_d \Phi_d(s) + \sum_{i=1}^{n} c_i K(t_i, s). \tag{4}$$

Les vecteurs $\mathbf{a} = (a_1, \ldots, a_M)^T$ et $\mathbf{c} = (c_1, \ldots, c_n)^T$ sont déterminés numériquement en substituant (4) dans le premier terme de (3).

D'après le lemme précédent, la solution au problème de minimisation $f_\lambda = \mathrm{argmin}_{h \in \mathcal{H}_\Gamma} \mathcal{I}(h, \mathbf{y})$ où

$$\mathcal{I}(h, \mathbf{y}) = \sum_{i=1}^{n} (h(t_i) - y_i)^2 + \lambda^2 \|h\|^2_{\mathcal{H}_\Gamma},$$

s'ecrit en fonction des $K(t_i, \cdot)$, i .e.
$f_\lambda(x) = \sum_{i=1}^{n} u_i K(t_i, x)$, avec

$$\mathbf{u} = (K + \lambda^2 I_n)^{-1}(y),$$

et K désigne, par abus de notation, la matrice de Gram $(K(t_i, t_j))$.

Remarque

L'idée est simple. Au lieu d'utiliser la décomposition infinie $\oplus_{j=0}^{\infty}\mathcal{W}_{j,\Gamma}$, nous allons tronquer les développements à une résolution maximale J. Il s'agit ainsi d'approcher f_λ par le choix de J et la restriction de $\mathcal{I}(h,\mathbf{y})$ à $\mathcal{H}_{J,\Gamma}$. Plus précisément, pour J assez grand, on résout la minimisation dans $\mathcal{H}_{J,\Gamma}$ défini par le noyau tronqué

$$K_J(x,y) = \sum_{g\in\cup_{0\leq j\leq J}G_j} \frac{\psi_g(x)}{\Gamma(g)}\psi_g(y), \quad x,y\in[0,1).$$

Si K_J est la matrice de Gram et $\mathbf{u}_J$ la solution correspondante, on a

$$\mathbf{u} - \mathbf{u}_J = (K_J + \lambda^2 I_n)^{-1}(K_J - K)(K + \lambda^2 I_n)^{-1} f|_X,$$

ce qui donne

$$\max |u_i - u_{i,J}| \leq \|(K + \lambda^2 I_n)^{-1}\|_\infty \|(K_J + \lambda^2 I_n)^{-1}\|_\infty \|K - K_J\|_\infty$$

Or $\mu_{\max}(K_J) \leq \mu_{\max}(K_{J+1}) \cdots \to \mu_{\max}(K)$, et $\mu_{\min}(K_J) \geq \mu_{\min}(K_{J+1}) \cdots \to \mu_{\min}(K) > 0$. Donc $\|(K_J + \lambda^2 I_n)^{-1}\|_\infty$ sont uniformément bornés en J.

De plus,

$$\lim_{J\to\infty} \|K - K_J\|_\infty = 0.$$

Par conséquent

$$\lim_{J\to\infty} \max_i |u_i - u_{i,J}| = 0 \tag{5}$$

et

$$f_{J,\lambda} \to f_\lambda,$$

quand $J \to \infty$.

Pénalisation par ondelettes

Pour simplifier, on suppose que Γ ne dépend que de j dans G_j, et donc K est une Γ_j^{-1}-pondération, des noyaux

$$K_j(s,t) = \sum_{k=0}^{2^j-1} \psi_{j,k}(s)\psi_{j,k}(t).$$

Dans ce cas on a

$$\mathcal{H}_\Gamma = \{1\} \oplus \oplus_{j\geq 0}\Gamma_j^{-1}\mathcal{W}_j. \tag{6}$$

Par orthogonalité des ondelettes les $\mathcal{W}_j$ sont des sous-espaces orthogonaux de$\mathcal{H}_\Gamma$.

Par analogie avec le lissage par fonctions spline, une façon d'estimer f serait de trouver $f \in \mathcal{H}_\Gamma$ minimisant

$$\frac{1}{n}\sum_{i=1}^{n}\{y_i - f(t_i)\}^2 + \lambda^2 \sum_{j\geq 0} \theta_j^{-1} \|P_j f\|_\Gamma^2, \tag{7}$$

où $\theta_j \geq 0$. Si $\theta_j = 0$, il convient de prendre $\|P_j f\|_\Gamma^2 = 0$, avec la convention $0/0 = 0$.

Notons que pour $J \geq \log_2(n)$ le minimum est atteint dans $\mathcal{H}_{J,\Gamma}$ et le problème de minimisation peut être enoncé comme : trouver $f \in \mathcal{H}_{J,\Gamma}$ pour minimiser

$$\frac{1}{n}\sum_{i=1}^{n}\{y_i - f(t_i)\}^2 + \lambda^2 \sum_{0\leq j\leq J} \theta_j^{-1}\|P_j f\|_{J,\Gamma}^2. \tag{8}$$

Une telle procédure d'estimation est controlée par une pénalité quadratique et telle quelle elle produira des estimateurs linéaires dont les taux asymptotiques seront optimaux pour des classes de fonctions suffisamment régulières.

Rétrecissement (shrinkage)

On propose de minimiser plutôt

$$\frac{1}{n}\sum_{i=1}^{n}\{y_i - f(x_i)\}^2 + \lambda^2 R_J(f), avec R_J(f) = \sum_{j=0}^{J}\|P_j f\|_{J,\Gamma}. \tag{9}$$

Notons que $R_J(f)$ n'est pas une norme sur $\mathcal{H}_{J,\Gamma}$ mais une pseudo-norme : $\forall f, h \in \mathcal{H}_{J,\Gamma}$, $R_J(f) \geq 0$, $R_J(cf) = |c|R_J(f)$, $R_J(f+h) \leq R_J(f) + R_J(h)$, et , pour tout $f \in \mathcal{H}_{J,\Gamma}$ non constant, $R_J(f) > 0$. De plus

$$\sum_{j=0}^{J}\|P_j f\|^2_{\mathcal{H}_{J,\Gamma}} \leq R_J(f)^2 \leq J\sum_{j=0}^{J}\|P_j f\|^2_{\mathcal{H}_{J,\Gamma}}. \tag{10}$$

Remarque

Si $n = 2^J$, la procédure classique de rétrécissement par ondelettes est obtenue par minimisation de

$$\|\mathbf{y} - W\boldsymbol{\beta}\|_n^2 + \lambda^2 \sum_{j=0}^{J} \sum_{k=0}^{2^j-1} 2^{j/2} |\beta_{j,k}|.$$

Pour chaque $(j,k) \in G_j$, le produit tensoriel $\psi_{j,k}(s)\psi_{j,k}(t)$ définit un noyau ondelette et si $\mathcal{W}_{j,k}$ est le RKHS correspondant, on voit que la pelnalisation L_1 des coefficients équivaut à (10) avec $R_J(f) = \sum_{j=0}^{J} \sum_{k=0}^{2^j-1} \|P_{j,k} f\|_{\mathcal{W}_{j,k}}$, interprétant ainsi la pénalité comme la somme des normes des composantes.

Solution et propriétés asymptotiques

Considérons la décomposition de $\mathcal{H}_{J,\Gamma}$ en

$$\mathcal{H}_{J,\Gamma} = V_0 \oplus \oplus_{j=0}^{J} \mathcal{W}_{j,\Gamma}.$$

Alors il existe une solution de (9) dans $\mathcal{H}_{J,\Gamma}$.

Dans les problèmes d'estimation non paramétrique par ondelettes, on considère souvent les espaces $B^s_{p,q}([0,1))$. Pour des ondelettes suffisamment régulières d'une fonction f dans la boule unité de $B^s_{p,q}([0,1))$ on a

$$\left(\sum_{j=0}^{J} 2^{j\left((2s+1)\frac{p}{2}-1\right)\frac{q}{p}} \left\{ \sum_{k=0}^{2^j-1} |d_{j,k}|^p \right\}^{\frac{q}{p}} \right)^{\frac{1}{q}} \leq 1. \qquad (11)$$

Soit $s \geq 0$, $p, q \geq 1$ avec $\rho = 2/(2s+1) \leq \min\{p, q\}$ et $J < \infty$. Supposons aussi que $\mathbf{d}_J = \{d_g, g \in \cup_{j=0}^{J} G_j\}$ satisfait à (11). Prenant $\Gamma_j = 2^{-j((2s+1)\frac{p}{2}-1)}$ on a $f \in \mathcal{H}_{\Gamma,J}$. On peut alors énoncer le théorème suivant :

Sous les conditions ci-dessus

- si f_0 n'est pas constante, et si $\lambda_n^{-1} = O_P(n^{(2-\rho)/4}) R_J^{(1-\rho)/2}(f_0)$, on a $\frac{1}{n}\|\mathbf{f} - \mathbf{f}_0\|_n = O_P(\lambda_n) R_J^{1/2}(f_0)$;
- si f_0 est constante alors $\frac{1}{n}\|\hat{\mathbf{f}} - \mathbf{f}_0\|_n = O_P(\max\{(n\lambda_n)^{-2/3}, n^{-1/2}\})$.

En choisissant de manière appropriée le paramètre λ le taux asymptotique de l'estimateur est $n^{-(2-\rho)/4}$. C'est le taux minimax sans terme en $\log n$. Le résultat précédent repose sur un théorème général de Sara van de Geer (Théorème 10.2) à condition de contrôler la δ-entropie de la boule unité des espaces $B^s_{p,q}([0,1])$. Un résultat récent de Picard et Kerkyacharian (Esaim, 2003) permet d'affirmer que

$$c\delta^{-\frac{1}{s}} \leq H(\delta, \mathcal{B}^s_{p,q}) \leq C\delta^{-\frac{1}{s}}, \quad \delta > 0. \tag{12}$$

ce qui permet d'etablir le résultat.

Algorithme de calcul

Une formulation équivalente plus facile est calculer est de considérer $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_J)^T$ et $f \in \mathcal{H}_{J,\Gamma}$ minimisant

$$\frac{1}{n}\sum_{i=1}^{n}\{y_i - f(x_i)\}^2 + \lambda_0 \sum_{0\leq j\leq J} \theta_j^{-1}\|P_j f\|_{J,\Gamma}^2 + \nu\sum_{j=0}^{J}\theta_j, \quad (13)$$

sous la contrainte $\theta_j \geq 0,\ j = 0, \ldots, J$, où λ_0 est une positive arbitraire et $\nu = \nu_n$ un paramètre de lissage.

Soit $\nu = \lambda^4/(4\lambda_0)$. D'une part, si $\hat{f}$ minimise (9), en posant $\hat{\theta}_j = \lambda_0^{1/2}\lambda^{-1/2}\|P_j\hat{f}\|_{J,\Gamma}$, le couple $(\hat{\boldsymbol{\theta}}, \hat{f})$ minimise (13). D'autre part si $(\hat{\boldsymbol{\theta}}, \hat{f})$ est un couple qui minimise (13), alors $\hat{f}$ minimise (9).

La pénalité additionnelle sur les θ dans (13) permet d'avoir certaines composantes de $\boldsymbol{\theta}$ nulles, produisant ainsi l'annulation de certains détails et conduisant à du seuillage des coefficients.

La solution f est de la forme

$$f(x) = b + \sum_{i=1}^{n} c_i \sum_{j=0}^{J} K_j(t_i, x),$$

avec $\mathbf{c} = (c_1, \ldots, c_n)^T \in \mathbb{R}^n$, $b \in \mathbb{R}$ et K_j le noyau de $\mathcal{W}_{j,\Gamma}$. Notons encore K_j la matrice $\{K_j(t_i, t_j)\}, i = 1, \ldots, n, j = 1, \ldots, n$ et $K_{\boldsymbol{\theta}}$ la matrice $\sum_{j=0}^{J} \theta_j K_j$, et soit 1_r le vecteur colonne à r composantes égales à 1.

Alors on a $\mathbf{f} = K_{\boldsymbol{\theta}}\mathbf{c} + b1_n$, et (13) s'exprime comme :

$$\frac{1}{n}\|\mathbf{y} - \sum_{j=0}^{J} \theta_j K_j \mathbf{c} - b1_n\|_n^2 + \lambda_0 \sum_{j=0}^{J} \mathbf{c}^T K_j \mathbf{c} + \nu \sum_{j=0}^{J} \theta_j, \quad (14)$$

avec $\theta_j \geq 0, j = 0, \ldots, J$.

Si les composantes de $\boldsymbol{\theta}$ sont fixées, alors (14) s'écrit

$$\min_{\mathbf{c},b} \|\mathbf{y} - K_{\boldsymbol{\theta}}\mathbf{c} - b1_n\|_n^2 + n\lambda_0 \mathbf{c}^T K_{\boldsymbol{\theta}}\mathbf{c}, \tag{15}$$

qui peut être résolu par des méthodes linéaires. D'un autre coté, si $\mathbf{c}$ et b sont fixés, notons $\mathbf{d}_j = K_j\mathbf{c}$ et D la matrice d'ordre $n \times J$ dont la $j$ème colonne est $\mathbf{d}_j$.

On montre alors que $\boldsymbol{\theta}$ minimisant (14) est solution de

$$\min_{\boldsymbol{\theta}} \|\mathbf{z} - D\boldsymbol{\theta}\|_n^2 + n\nu \sum_{j=0}^{J} \theta_j, \tag{16}$$

où $\mathbf{z} = \mathbf{y} - (1/2)n\lambda_0\mathbf{c} - b1_n$.

On peut alors itérer entre (15) et (16). A chaque itération (14) décroit.

L'expression(16) équivaut à

$$\min_{\boldsymbol{\theta}} \|\mathbf{z} - D\boldsymbol{\theta}\|_n^2 \text{ sous } \theta_j \geq 0, \ \sum_{j=0}^{J} \theta_j \leq M, \tag{17}$$

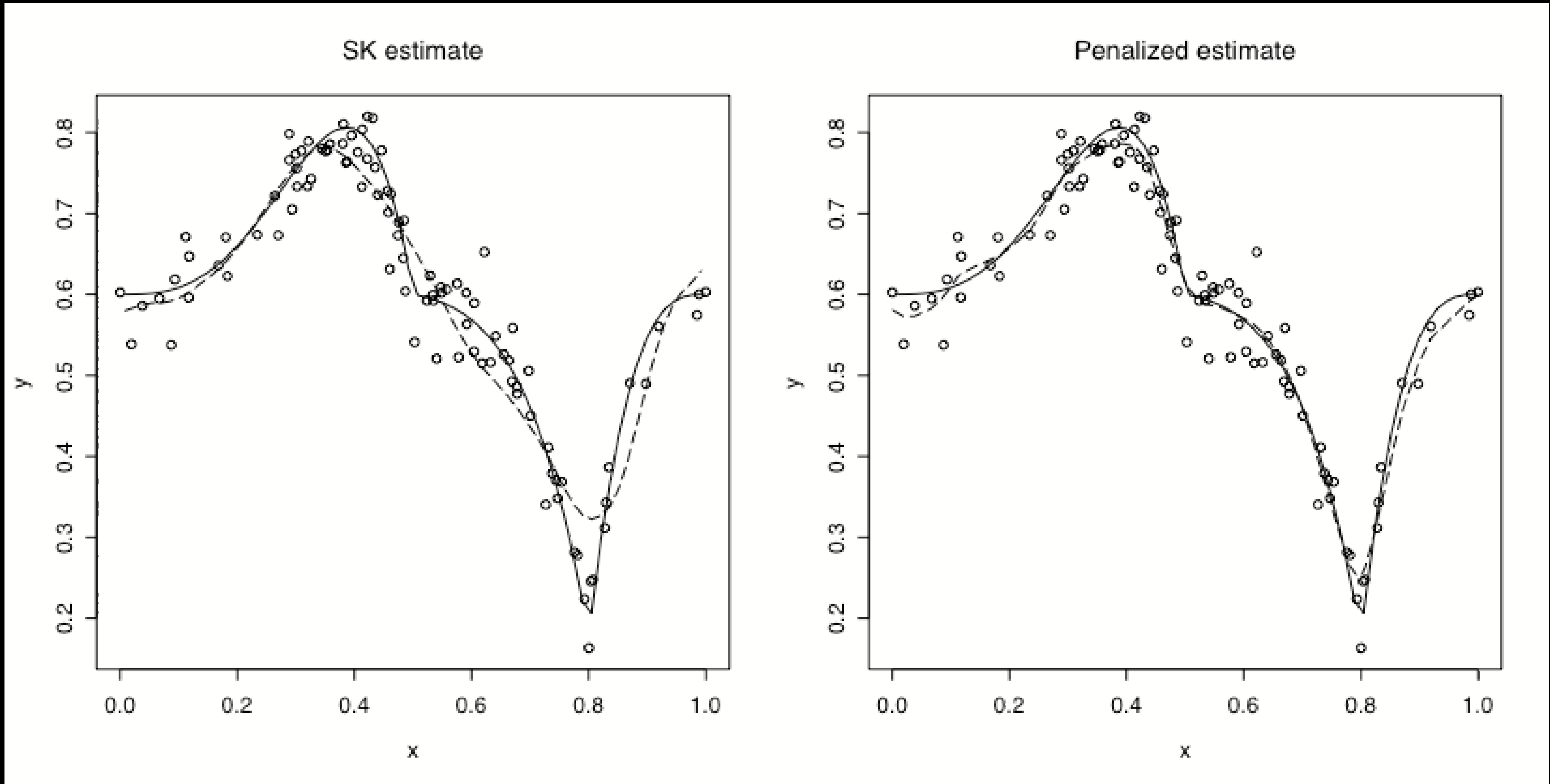
pour un certain $M \geq 0$. Si l'algorithme itérant entre (15) et (16) converge, alors la solution est également un point fixe de l'algorithme qui itère entre (15) et (17) pour M fixé.

Algorithme

1. Initialisation: Fixer $\theta_j = 1, j = 0, \ldots, J$.
2. Résoudre en $\mathbf{c}$ et b avec (15).
3. Pour $\mathbf{c}$ et b obtenus au pas 2, résoudre en $\boldsymbol{\theta}$ avec (17).
4. Avec le nouveau $\boldsymbol{\theta}$, résoudre en $\mathbf{c}$ et b avec (15).

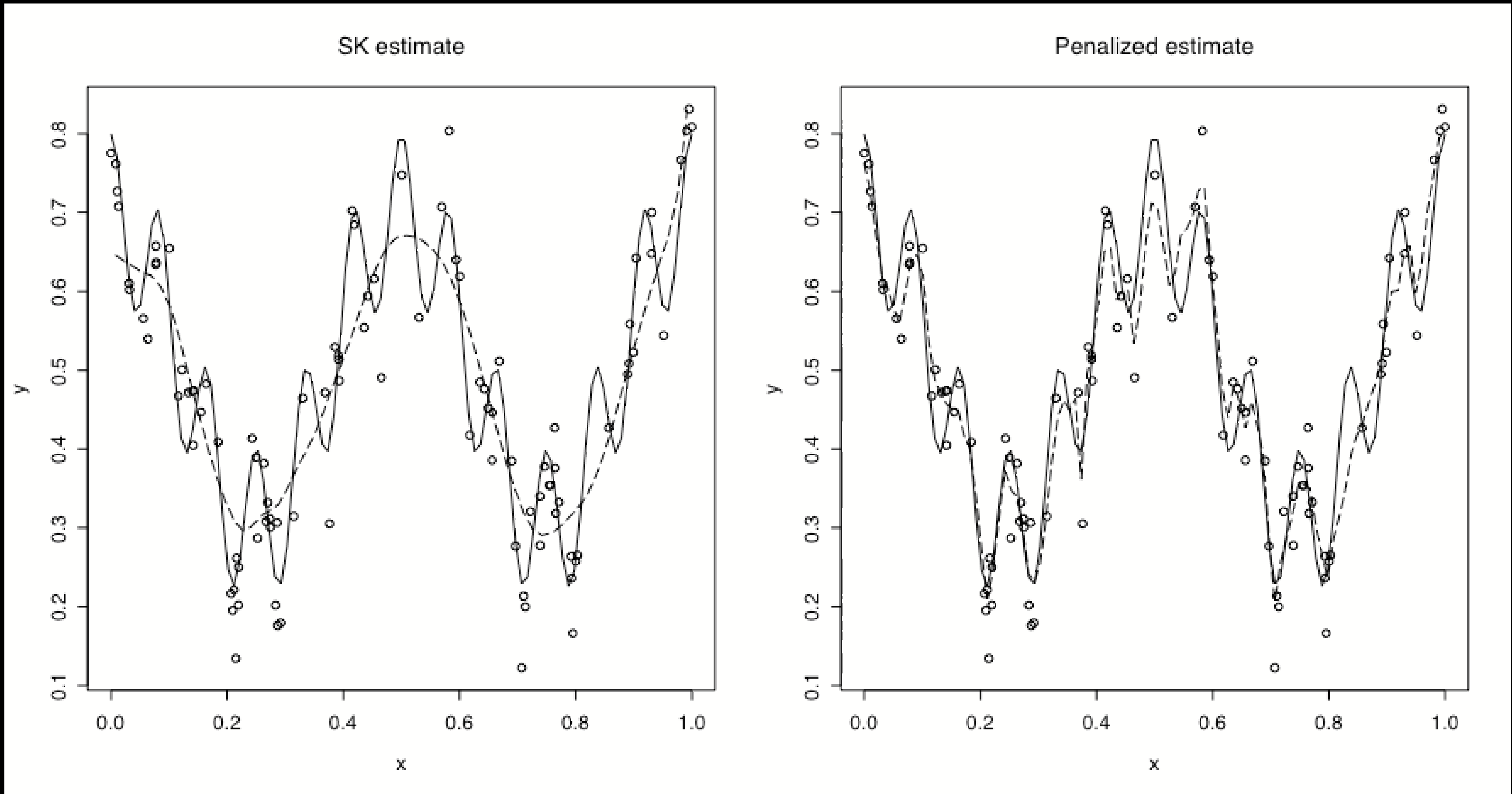
Cette procédure à un pas global est aussi efficace que celle procédant à toutes les itérations (voir Antoniadis et Fan (2001)).

Exemple



Simulation.

Exemple



Simulation.