Agrégation d'estimateurs affines pour la régression

Joseph Salmon
(en collaboration avec **Arnak Dalalyan**)

Séminaire de probabilités et statistiques
Angers
Janvier 2011

Plan

Introduction

Motivations

- Théorique : inégalités d'oracles (grande dimension, sparsité), adaptation
- Applications variées : **traitement d'images**, génétique, problèmes inverses (estimation de dérivée, déconvolution avec un noyau connu, tomographie, etc.)

Heuristique sous-jacente

- L'agrégation/mélange d'estimateurs est plus stable que la sélection d'un seul estimateur

Motivation principale : débruitage d'images

Modèle du bruit additif

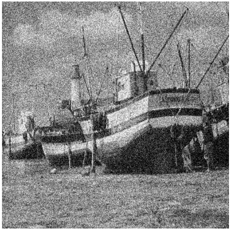


Image observée

Motivation principale : débruitage d'images

Modèle du bruit additif

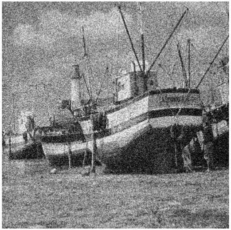


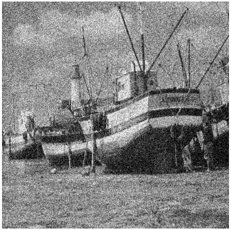
Image observée

$=$

Image idéale

Motivation principale : débruitage d'images

Modèle du bruit additif



$=$

$+$

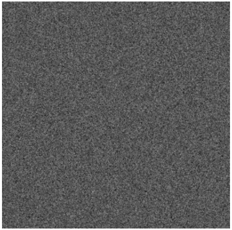


Image observée

Image idéale

Bruit

Notation : $\boxed{Y = f + \varepsilon}$

Méthodes à patchs et redondance

Introduction aux Non-Local Means (NLM)

Introduction des patchs pour le débruitage Buades Coll et Morel (2005)
État de l'art : Dabov et al. (2007), Mairal et al. (2009)

- Patch à débruiter
- Patchs similaires : poids importants
- Patchs peu similaires : poids faibles
- Patchs très différents : poids quasi nuls

Méthodes à patchs et redondance

Introduction aux Non-Local Means (NLM)

Introduction des patchs pour le débruitage Buades Coll et Morel (2005)
État de l'art : Dabov et al. (2007), Mairal et al. (2009)

- Patch à débruiter
- Patchs similaires : poids importants
- Patchs peu similaires : poids faibles
- Patchs très différents : poids quasi nuls

Mélanger lissage classique et NL-Means

- Y est le vecteur contenant les intensité des pixels d'un patch bruité.
- L'approche par filtrage estime le vrai patch f par AY, A est une matrice de convolution.
  - Inégalité oracle exacte pour l'agrégation d'estimateur AY pour des matrices de projections Leung et Barron (2006)
- L'approche NL-Means estime f en combinant des patchs voisins b qui sont essentiellement indépendant de Y.
  - Inégalité oracle exacte pour l'agrégation d'estimateur construit sur un échantillon indépendant Dalalyan et Tsybakov (2007,2008)
- On veut étendre ces résultats à l'agrégation d'éléments de la forme $AY + b$ avec A et b indépendants de Y

Notations et modèle

Modèle hétéroscédastique gaussien

$$Y_i = f_i + \sigma_i \varepsilon_i, \quad i = 1, \cdots, n \qquad (\star)$$

ε_i i.i.d $\mathcal{N}(0,1)$ et $\Sigma = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_n^2)$ (Σ connue)

- Rem 1 : grille d'observation fixe (pixels) $f_i = f(x_i)$, $(x_i)_{i=1,\cdots,n}$
- Rem 2 : $\Sigma = \sigma^2 Id$, modèle homoscédastique

But : estimer f par $\hat{f}$, avec un faible risque

$$r = \mathbb{E}\left(\left\|f - \hat{f}_n\right\|_n^2\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(f_i - \hat{f}_i)^2\right)$$

Lien problème inverse/hétéroscédasticité

T : opérateur ***connu*** sur un Hilbert $(\mathcal{H}, \langle\cdot|\cdot\rangle_{\mathcal{H}})$
Y : processus aléatoire , pour un $h \in \mathcal{H}$

$$Y = Th + \varepsilon\xi \quad \Longleftrightarrow \quad Y(g) = \langle Th|g\rangle_{\mathcal{H}} + \varepsilon\xi(g), \quad \forall g \in \mathcal{H},$$

T^* : l'adjoint de T ; si $T^*\,T$ est compact, par SVD

$$T\phi_k = b_k\psi_k, \quad T^*\psi_k = b_k\phi_k, \qquad k \in \mathbb{N},$$

b_k : valeurs singulières,$\{\phi_k\}$: base orthonormale de $\mathcal{H}$, $\{\psi_k\}$: base orthonormale de $\mathrm{Im}(T) \subset \mathcal{H}$. Le modèle se ré-écrit

$$Y(\psi_k) = \langle h|\phi_k\rangle_{\mathcal{H}}\, b_k + \varepsilon\xi(\psi_k), \qquad k \in \mathbb{N}.$$

Si $b_k \neq 0$ le modèle est équivalent à $(\star)$, avec $f_i = \langle h|\phi_i\rangle_{\mathcal{H}}$ et $\sigma_i = \varepsilon b_i^{-1}$

Cadre de l'agrégation et inégalité d'oracle

On dispose d'une famille de « pré-estimateur » :
$\mathcal{F}_\Lambda = \{\hat{f}_\lambda \in \mathbb{R}^n, \lambda \in \Lambda\}$ avec $\Lambda \subset \mathbb{R}^M$
But : fournir un inégalité d'oracle pour un estimateur $\hat{f}_{agr}$

Inégalité d'oracle Nermirovski (2000)

$$\mathbb{E}\|\hat{f}_{agr} - f\|_n^2 \leq C_n \inf_{\lambda \in \Lambda} \mathbb{E}\|\hat{f}_\lambda - f\|_n^2 + R_{n,\Lambda}$$

- l'**oracle** est l'élément $\hat{f}^{\lambda^*}$ qui minimise $\mathbb{E}\|\hat{f}_\lambda - f\|_n^2$ sur $\mathcal{F}_\lambda$
- $C_n \geq 1$. Si $C_n = 1$: on parle d'inégalité d'oracle **exacte**
- $R_{n,\Lambda} \overset{n\to\infty}{\longrightarrow} 0$: prix à payer car on ne connait pas l'oracle, dépend de la complexité de Λ, du bruit, ...

Rem 1 : on n'a pas besoin d'évaluer le terme d'approximation
Rem 2 : optimalité (borne inférieure) pour certain Λ

Estimation Non-Paramétrique contre Agrégation

	Disponible	Non Disponible	Cible
Estimation NP	$\boldsymbol{Y}$	f	trouver le meilleur estimateur
Agrégation	$\boldsymbol{Y}, \mathcal{F}_\Lambda$	f	trouver un estimateur (presque) aussi bon que le meilleur d'une certaine famille

Avantage : pas besoin d'évaluer le terme d'erreur d'approximation

Méthode par pénalisation

Famille de « pré-estimateurs » : $\mathcal{F}_\Lambda = (\hat{f}_\lambda)_{\lambda\in\Lambda} \in \mathbb{R}^n$ avec $\Lambda \subset \mathbb{R}^M$.
$\hat{r}_\lambda$: estimateur sans biais du risque, $\mathbb{E}(\hat{r}_\lambda) = \mathbb{E}(\|f - \hat{f}_\lambda\|_n^2) = r_\lambda$
On suppose que $\hat{f}_\lambda = f_\lambda = \sum_{j=1}^M \lambda_j \varphi_j$

Méthode par pénalisation

$$\hat{f}^{\mathrm{Pen}} = f_{\hat{\lambda}}, \quad \text{où} \quad \hat{\lambda} = \arg\min_{\lambda\in\Lambda} \Big(\underbrace{\hat{r}_\lambda}_{\text{adéquation}} + \underbrace{\mathrm{Pen}(\lambda)}_{\text{régularisation}} \Big)$$

- $\mathrm{Pen}(\lambda) = \beta\|\lambda\|_2^2$: Ridge Tikhonov [43]
- $\mathrm{Pen}(\lambda) = \beta\|\lambda\|_0$: AIC,BIC Akaike [74], Schwarz [78]
- $\mathrm{Pen}(\lambda) = \beta\|\lambda\|_1$: LASSO Tibshirani [96]

Rem 1 : pour les deux derniers des inégalités d'oracles existent
Rem 2 : β paramètre de lissage

Versions par blocs possibles...

Agrégation à poids exponentiels (EWA)

- Extension : élargir l'espace de recherche, changer de pénalité
- Espace de recherche :
  $\mathcal{P}_\Lambda = \{p$: probabilité sur Λ tel que $\mathbb{E} \int_\Lambda \|\hat{f}_\lambda\|_n^2 p(d\lambda) < \infty\}$
- Pénalisation étendue : $\hat{f}^{\mathrm{Pen}} = \int_\Lambda \hat{f}_\lambda \hat{\pi}^{\mathrm{Pen}}(d\lambda)$ avec

$$\hat{\pi}^{\mathrm{Pen}} = \arg\min_{p \in \mathcal{P}_\Lambda} \left(\int_\Lambda \hat{r}_\lambda p(d\lambda) + \int_\Lambda \mathrm{Pen}(\lambda) p(d\lambda) \right)$$

- Pénalisation KL : π *a priori* sur Λ, l'EWA est

$$\hat{\pi}^{\mathsf{EWA}} = \arg\min_{p \in \mathcal{P}_\Lambda} \left(\int_\Lambda \hat{r}_\lambda p(d\lambda) + \frac{\beta}{n} \mathcal{K}(p, \pi) \right)$$

$$\hat{f}^{\mathsf{EWA}} = \int_\Lambda \hat{f}_\lambda \hat{\pi}^{\mathsf{EWA}}(d\lambda)$$

avec $\mathcal{K}(p, \pi)$ la divergence de Kullback-Leibler entre deux mesures de probabilité $p, \pi \in \mathcal{P}_\Lambda$.

$$\mathcal{K}(p, \pi) = \begin{cases} \int_\Lambda \log\left(\frac{dp}{d\pi}(\lambda)\right) p(d\lambda) & \text{si } p \ll \pi, \\ +\infty & \text{sinon.} \end{cases}$$

Agrégation à poids exponentiels (EWA)

- Extension : élargir l'espace de recherche, changer de pénalité
- Espace de recherche :
  $\mathcal{P}_\Lambda = \{p : \text{probabilité sur } \Lambda \text{ tel que } \mathbb{E} \int_\Lambda \|\hat{f}_\lambda\|_n^2 p(d\lambda) < \infty\}$
- Pénalisation étendue : $\hat{f}^{\mathrm{Pen}} = \int_\Lambda \hat{f}_\lambda \hat{\pi}^{\mathrm{Pen}}(d\lambda)$ avec

$$\hat{\pi}^{\mathrm{Pen}} = \arg\min_{p \in \mathcal{P}_\Lambda} \left(\int_\Lambda \hat{r}_\lambda p(d\lambda) + \int_\Lambda \mathrm{Pen}(\lambda) p(d\lambda) \right)$$

- Pénalisation KL : π *a priori* sur Λ, l'EWA est

$$\hat{\pi}^{\mathsf{EWA}} = \arg\min_{p \in \mathcal{P}_\Lambda} \left(\int_\Lambda \hat{r}_\lambda p(d\lambda) + \frac{\beta}{n} \mathcal{K}(p, \pi) \right)$$

$$\hat{f}^{\mathsf{EWA}} = \int_\Lambda \hat{f}_\lambda \hat{\pi}^{\mathsf{EWA}}(d\lambda)$$

avec $\mathcal{K}(p, \pi)$ la divergence de Kullback-Leibler entre deux mesures de probabilité $p, \pi \in \mathcal{P}_\Lambda$.

$$\mathcal{K}(p, \pi) = \begin{cases} \int_\Lambda \log\left(\frac{dp}{d\pi}(\lambda)\right) p(d\lambda) & \text{si } p \ll \pi, \\ +\infty & \text{sinon.} \end{cases}$$

Agrégation à poids exponentiels (EWA)

On choisit un *a priori* $\pi \in \mathcal{P}_\Lambda$ (mesure de probabilité sur Λ)

L'agrégat à poids exponentiels (EWA)

Mesure de Gibbs : mesure *a posteriori* sur Λ

$$\boxed{\hat{\pi}^{\mathsf{EWA}}(d\lambda) \propto \exp(-n\hat{r}_\lambda/\beta)\pi(d\lambda)}$$

- β : un paramètre de lissage (température)
- $\hat{r}_\lambda$: estimateur sans biais du risque $\mathbb{E}\|\hat{f}_\lambda - f\|_n^2$

L'agrégat à poids exponentiel est l'espérance *a posteriori* :

$$\boxed{\hat{f}^{\mathsf{EWA}} = \int_\Lambda \hat{f}_\lambda \hat{\pi}^{\mathsf{EWA}}(d\lambda)}$$

Rem : Si $\beta \to 0$, $\hat{f}^{\mathsf{EWA}} \to \hat{f}_{\lambda^*}$ avec $\lambda^* = \arg\min_\lambda \hat{r}_\lambda$
Si $\beta \to \infty$, $\hat{f}^{\mathsf{EWA}} \to \int_\Lambda \hat{f}_\lambda \pi(d\lambda)$

Agrégation d'estimateurs affines

Cadre de l'agrégation

- Point de départ : famille de « pré-estimateurs » : $\mathcal{F}_\Lambda = (\hat{f}_\lambda)_{\lambda \in \Lambda} \in \mathbb{R}^n$ avec $\Lambda \subset \mathbb{R}^M$
- Objectif : approcher f aussi bien que le ferait le meilleur élément de cette famille.

Rem : les pré-estimateurs doivent dépendre des données (!)

Pré-estimateurs affines

$$\boxed{\hat{f}_\lambda = A_\lambda \boldsymbol{Y} + b_\lambda}$$

- A_λ est une matrice déterministe $n \times n$
- b_λ est un vecteur déterministe de $\mathbb{R}^n$
- A_λ et b_λ sont indépendants de $\boldsymbol{Y}$

Agrégation d'estimateurs affines

Cadre de l'agrégation

- Point de départ : famille de « pré-estimateurs » : $\mathcal{F}_\Lambda = (\hat{f}_\lambda)_{\lambda \in \Lambda} \in \mathbb{R}^n$ avec $\Lambda \subset \mathbb{R}^M$
- Objectif : approcher f aussi bien que le ferait le meilleur élément de cette famille.

Rem : les pré-estimateurs doivent dépendre des données (!)

Pré-estimateurs affines

$$\boxed{\hat{f}_\lambda = A_\lambda \boldsymbol{Y} + b_\lambda}$$

- A_λ est une matrice déterministe $n \times n$
- b_λ est un vecteur déterministe de $\mathbb{R}^n$
- A_λ et b_λ sont indépendants de $\boldsymbol{Y}$

Cas constant : $A_\lambda = 0$

Pré-estimateurs déterministes : $\hat{f}_\lambda = f_\lambda (= b_\lambda)$

Exemples de familles de pré-estimateurs

- $\mathcal{F}_\Lambda = \{\varphi_1, \cdots, \varphi_M\}$ est un « dictionnaire » fixé et fini
- $\mathcal{F}_\Lambda = \mathrm{conv}(\varphi_1, \cdots, \varphi_M)$ combinaisons convexes d'éléments d'un « dictionnaire » : $f_\lambda = \sum_{j=1}^M \lambda_j \varphi_j$, où les φ_j sont fixées
- $\mathcal{F}_\Lambda = \mathrm{vect}(\varphi_1, \cdots, \varphi_M)$ combinaisons linéaires d'éléments d'un « dictionnaire » : $f_\lambda = \sum_{j=1}^M \lambda_j \varphi_j$, où les φ_j sont fixées
- $\mathcal{F}_\Lambda = \mathrm{vect}_S(\varphi_1, \cdots, \varphi_M)$ combinaisons "creuses" (sparse) d'éléments d'un « dictionnaire » :
  $f_\lambda = \sum_{j=1}^M \lambda_j \varphi_j$, avec au plus S coefficients λ_j non nuls

Bornes inférieures disponibles : Tsybakov (2003), Bunea Tsybakov Wegkamp (2007)

Cas linéaire : $\hat{f}_\lambda = A_\lambda \boldsymbol{Y} \quad (b_\lambda = 0)$

Moindres carrés ordinaires

$\{\mathcal{S}_\lambda : \lambda \in \Lambda\}$ ensemble de sous-espaces de $\mathbb{R}^n$
A_λ : projecteurs orthogonaux sur $\mathcal{S}_\lambda$ Leung et Barron [06], Alquier et Lounici [à paraître], Rigollet et Tsybakov [à paraître]

Cas des matrices diagonales

$$A_\lambda = \mathrm{diag}(a_1, \ldots, a_n)$$

- Projections ordonnées : $a_k = \mathbb{1}_{(k \leq \lambda)}$ pour λ entier, ie. $\Lambda = \{1, \ldots, n\}$
- Filtre de Pinsker : $a_k = \left(1 - \frac{k^\alpha}{w}\right)_+$, où $x_+ = \max(x, 0)$ et $w, \alpha > 0$, i.e., $\Lambda = (\mathbb{R}_+^*)^2$

Autres cas possibles : Kernel ridge regression, traitement par blocs, etc.

Cas linéaire : $\hat{f}_\lambda = A_\lambda \boldsymbol{Y} \quad (b_\lambda = 0)$

Moindres carrés ordinaires

$\{\mathcal{S}_\lambda : \lambda \in \Lambda\}$ ensemble de sous-espaces de $\mathbb{R}^n$
A_λ : projecteurs orthogonaux sur $\mathcal{S}_\lambda$ Leung et Barron [06], Alquier et Lounici [à paraître], Rigollet et Tsybakov [à paraître]

Cas des matrices diagonales

$$A_\lambda = \mathrm{diag}(a_1, \ldots, a_n)$$

- Projections ordonnées : $a_k = \mathbb{1}_{(k \leq \lambda)}$ pour λ entier, ie. $\Lambda = \{1, \ldots, n\}$
- Filtre de Pinsker : $a_k = \left(1 - \frac{k^\alpha}{w}\right)_+$, où $x_+ = \max(x, 0)$ et $w, \alpha > 0$, i.e., $\Lambda = (\mathbb{R}_+^*)^2$

Autres cas possibles : Kernel ridge regression, traitement par blocs, etc.

Pré-estimateurs affines et estimation du risque

Formule de Stein [81]

Si $\hat{f}$ est un estimateur différentiable presque partout en Y et que $\partial_{Y_i}\hat{f}_i$ est intégrable alors

$$\hat{r} = \| \boldsymbol{Y} - \hat{f} \|_n^2 + \frac{2}{n}\sum_{i=1}^{n} \partial_{Y_i}\hat{f}_i \sigma_i^2 - \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$$

est un estimateur sans biais du risque $\mathbb{E}(\hat{r}_n) = r$

Cas affine : $\hat{f}_\lambda = A_\lambda \boldsymbol{Y} + b_\lambda$

Conclusion : $\boxed{\hat{r}_\lambda = \| \boldsymbol{Y} - \hat{f}_\lambda \|_n^2 + \frac{2}{n}\mathrm{Tr}(\Sigma A_\lambda) - \frac{1}{n}\mathrm{Tr}(\Sigma)}$

est un estimateur sans biais du risque r_λ

Conditions du théorème principal

Condition C_1

- Matrices A_λ : projections orthogonales ($A_\lambda^2 = A_\lambda^\top = A_\lambda$)
- Vecteurs b_λ : $A_\lambda b_\lambda = 0$

Exemple : A_λ projections sur des sous espaces de $\mathbb{R}^n$ Leung et Barron [06]

Condition C_2

- Matrices A_λ : symétriques, semi-définies positives
- $A_\lambda A_{\lambda'} = A_{\lambda'} A_\lambda, \forall \lambda, \lambda' \in \Lambda$
- Vecteurs b_λ : $A_{\lambda'} b_\lambda = 0, \forall \lambda, \lambda' \in \Lambda$

Exemple : les A_λ sont des estimateurs par seuillage de type James-Stein en deux blocs Leung [04]

Conditions du théorème principal

Condition C_1

- Matrices A_λ : projections orthogonales ($A_\lambda^2 = A_\lambda^\top = A_\lambda$)
- Vecteurs b_λ : $A_\lambda b_\lambda = 0$

Exemple : A_λ projections sur des sous espaces de $\mathbb{R}^n$ Leung et Barron [06]

Condition C_2

- Matrices A_λ : symétriques, semi-définies positives
- $A_\lambda A_{\lambda'} = A_{\lambda'} A_\lambda, \forall \lambda, \lambda' \in \Lambda$
- Vecteurs b_λ : $A_{\lambda'} b_\lambda = 0, \forall \lambda, \lambda' \in \Lambda$

Exemple : les A_λ sont des estimateurs par seuillage de type James-Stein en deux blocs Leung [04]

Énoncé du théorème principal

Borne PAC-bayésienne Dalalyan et S. [Soumis]

Si $\mathbf{C}_1$ ou $\mathbf{C}_2$ est vérifiée, alors $\hat{f}^{\text{EWA}}$ vérifie pour tout *a priori* π :

$$\mathbb{E}(\|\hat{f}^{\text{EWA}} - f\|_n^2) \leq \inf_{p \in \mathcal{P}_\Lambda} \left(\int_\Lambda \mathbb{E}\|\hat{f}_\lambda - f\|_n^2 \, p(d\lambda) + \frac{\beta}{n} \mathcal{K}(p, \pi) \right)$$

où $\mathcal{K}(p, \pi)$ est la divergence de Kullback-Leibler entre p et π, et $\beta \geq 4 \max_{i=1,\dots,n} \sigma_i^2$ sous $\mathbf{C}_1$ et $\beta \geq 8 \max_{i=1,\dots,n} \sigma_i^2$ sous $\mathbf{C}_2$, avec $\mathcal{K}(p, \pi)$ la divergence de Kullback-Leibler entre deux mesures de probabilité $p, \pi \in \mathcal{P}_\Lambda$.

$$\mathcal{K}(p, \pi) = \begin{cases} \int_\Lambda \log\left(\frac{dp}{d\pi}(\lambda)\right) p(d\lambda) & \text{si } p \ll \pi, \\ +\infty & \text{sinon.} \end{cases}$$

Corollaire : cas discret

Inégalité Oracle : $\Lambda = [\![1, M]\!]$, π uniforme

Si $\mathbf{C}_1$ ou $\mathbf{C}_2$ est vérifiée, et si π est uniforme sur $[\![1, M]\!]$, alors

$$\mathbb{E}(\|\hat{f}^{\mathsf{EWA}} - f\|_n^2) \leq \inf_{\lambda \in [\![1,M]\!]} \left(\mathbb{E}\|\hat{f}_\lambda - f\|_n^2 + \frac{\beta \log(M)}{n} \right)$$

pour $\beta \geq \alpha \max_{i=1,\dots,n} \sigma_i^2$. $\alpha = 4$ sous $\mathbf{C}_1$ et $\alpha = 8$ sous $\mathbf{C}_2$.

- Pour $b_\lambda = 0$, résultat de Leung et Barron (2006)
- Pour $A_\lambda = 0$, et si pour tout i, $\sigma_i = \sigma$: l'inégalité est optimale Tsybakov (2003)

Inégalité Oracle Sparse

Scénario sparse : il existe un vecteur sparse $\lambda^* \in \Lambda = \mathbb{R}^M$ tel que $\hat{f}_{\lambda^*} \approx f$. Choix d'un *a priori* favorisant la sparsité

$$\pi(d\lambda) \propto \prod_{j=1}^{M} \frac{1}{(1 + |\lambda_j/\tau|^2)^2} \, \mathbb{1}_\Lambda(\lambda),$$

$\tau > 0$: paramètre de concentration.

Inégalité Oracle

Prenons π définit ci-dessus, supposons que $\lambda \mapsto r_\lambda$ est $\mathcal{C}^1$, et qu'il existe une matrice $\mathcal{M}$ de taille $M \times M$ telle que :

$$r_\lambda - r_{\lambda'} - \nabla r_{\lambda'}^\top(\lambda - \lambda') \leq (\lambda - \lambda')^\top \mathcal{M}(\lambda - \lambda'), \quad \forall \lambda, \lambda' \in \Lambda.$$

Si $\mathbf{C}_1$ ou $\mathbf{C}_2$ est vérifiée

$$\mathbb{E}(\|\hat{f}^{\mathsf{EWA}} - f\|_n^2) \leq \inf_{\lambda \in \mathbb{R}^M} \left\{ \mathbb{E}\|\hat{f}_\lambda - f\|_n^2 + \frac{4\beta}{n} \sum_{j=1}^{M} \log\left(1 + \frac{|\lambda_j|}{\tau}\right) \right\} + \mathrm{Tr}(\mathcal{M})\tau^2$$

pour $\beta \geq \alpha \max_{i=1,\dots,n} \sigma_i^2$. $\alpha = 4$ sous $\mathbf{C}_1$ et $\alpha = 8$ sous $\mathbf{C}_2$.

Point de vue minimax : cas homoscédastique, $\Sigma = \sigma^2 Id$

$\theta_k(f) = \langle f | \varphi_k \rangle_n$: coefficients de la transformée (orthogonale) de Fourier discrète de f, notée $\mathcal{D}f$

Ellipsoïde de Sobolev : $\mathcal{F}(\alpha, R) = \{f \in \mathbb{R}^n : \sum_{k=1}^n k^{2\alpha} \theta_k(f)^2 \leq R\}$

Théorème de Pinsker

$$\begin{aligned}\inf_{\hat{f}} \sup_{f \in \mathcal{F}(\alpha,R)} \mathbb{E}(\|\hat{f} - f\|_n^2) &\sim \inf_{A} \sup_{f \in \mathcal{F}(\alpha,R)} \mathbb{E}(\|A\,\boldsymbol{Y} - f\|_n^2) \\ &\sim \inf_{w>0} \sup_{f \in \mathcal{F}(\alpha,R)} \mathbb{E}(\|A_{\alpha,w}\,\boldsymbol{Y} - f\|_n^2)\end{aligned}$$

$\inf$ est sur tous les estimateurs $\hat{f}$ possibles et
$A_{\alpha,w} = \mathcal{D}^\top \mathrm{diag}\left((1 - k^\alpha / w)_+ ; k = 1, \ldots, n\right) \mathcal{D}$: Filtre de Pinsker

Morale : Estimateurs linéaires minimax sur les ellipsoïdes

EWA, adaptation et simulations

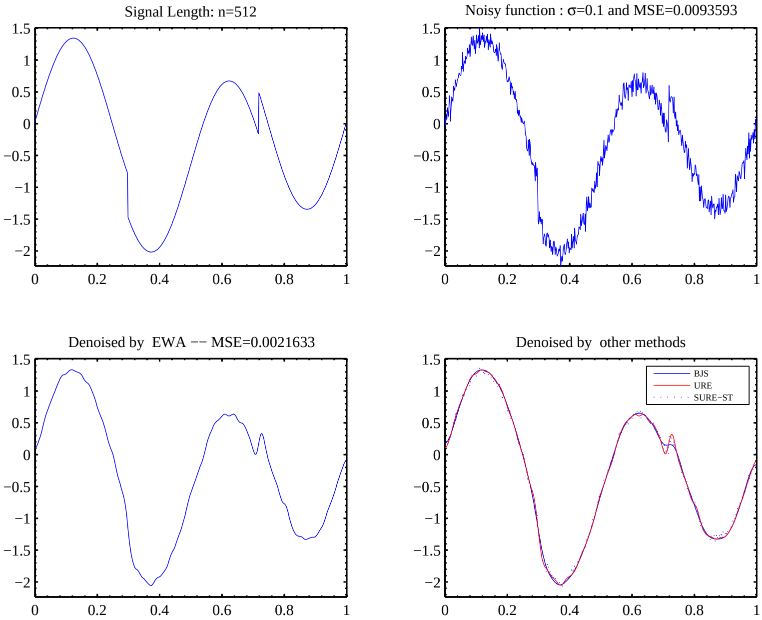
EWA sur des filtres de Pinsker : $\hat{f}_{\alpha,w} = \mathcal{D}^\top A_{\alpha,w} \mathcal{D}\, \boldsymbol{Y}$ ($\mathcal{D}$: DCT),
avec $A_{\alpha,w} = \mathrm{diag}((1 - \frac{k^\alpha}{w})_+, k = 1, \cdots, n)$
Choix *a priori* π :

- Tirer α selon une lois exponentielle de paramètre 1
- Sachant α, tirer w selon $\frac{2 n_\sigma^{-\alpha/(2\alpha+1)}}{\left(1+n_\sigma^{-\alpha/(2\alpha+1)} w\right)^3}$ avec $n_\sigma = n/\sigma^2$

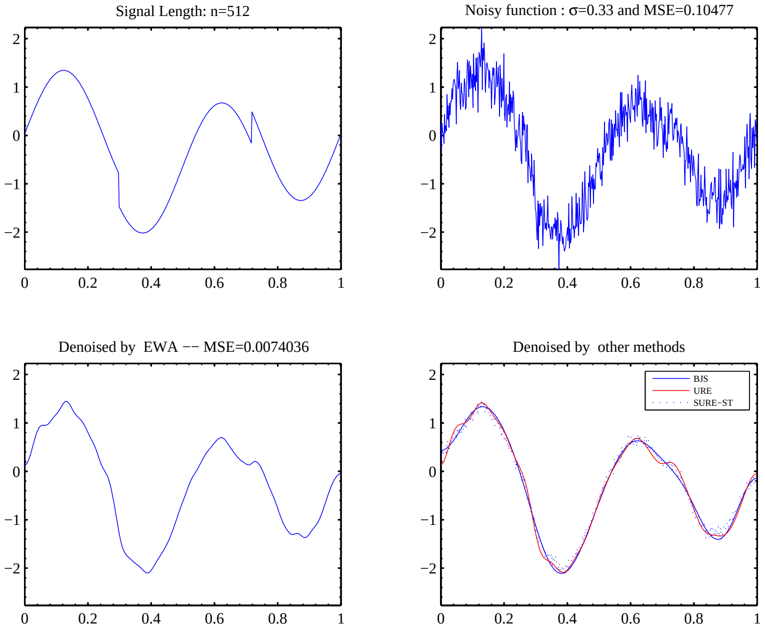
Performance

- Théorique : adaptatif au sens minimax exacte sur les ellipsoïdes $\{\mathcal{F}(\alpha, R) : \alpha > 0, R > 0\}$
- Pratique : performances au moins aussi bonnes que celle d'autres estimateurs adaptatifs classiques (Sure-Shrink Donoho et Johnstone [95] , James-Stein par blocs Cai [99] , minimiseur du risque empirique Cavalier et al. [02])

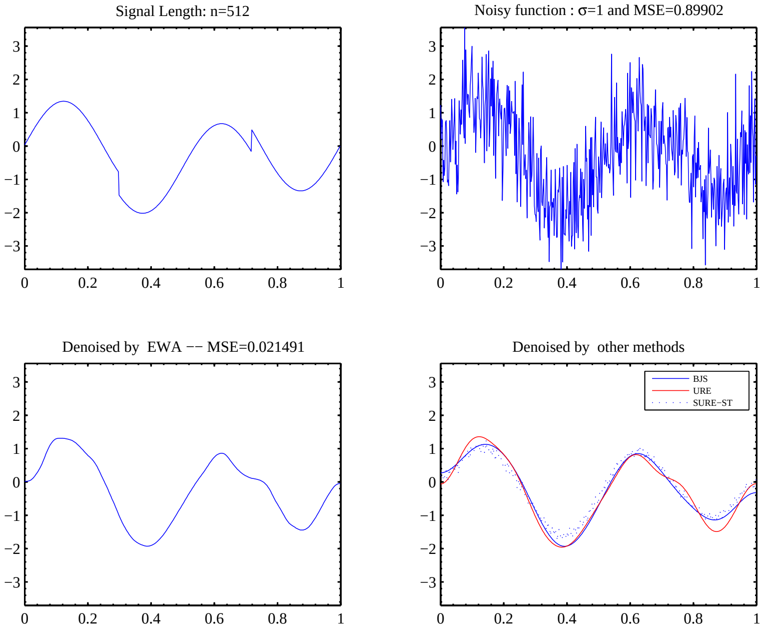
Application à des signaux 1D



Application à des signaux 1D



Application à des signaux 1D



Conclusion

Contributions

- Inégalité d'oracle exacte pour certains estimateurs affines
- Résultat sur l'adaptation en fonction de la régularité des signaux

Ouvertures possibles et travaux en cours

- Liens avec les travaux sur l'apprentissage pour le traitement par patchs
- Autres modèles de bruits : modèle de Poisson notamment

http ://people.math.jussieu.fr/ $\sim$ salmon/

Conclusion

Contributions

- Inégalité d'oracle exacte pour certains estimateurs affines
- Résultat sur l'adaptation en fonction de la régularité des signaux

Ouvertures possibles et travaux en cours

- Liens avec les travaux sur l'apprentissage pour le traitement par patchs
- Autres modèles de bruits : modèle de Poisson notamment

http ://people.math.jussieu.fr/ $\sim$ salmon/

Références I

- H. Akaike.
  A new look at the statistical model identification.
  IEEE Trans. Automatic Control, AC-19 :716–723, 1974.
  System identification and time-series analysis.
- P. Alquier and K. Lounici.
  Pac-bayesian bounds for sparse regression estimation with exponential weights.
  hal-00465801, submitted, 2010.
- A. Buades, B. Coll, and J-M. Morel.
  A review of image denoising algorithms, with a new one.
  Multiscale Model. Simul., 4(2) :490–530, 2005.
- F. Bunea, A. B. Tsybakov, and M. H. Wegkamp.
  Aggregation for Gaussian regression.
  Ann. Statist., 35(4) :1674–1697, 2007.

Références II

- T. T. Cai.
  Adaptive wavelet estimation : a block thresholding and oracle inequality approach.
  Ann. Statist., 27(3) :898–924, 1999.

- L. Cavalier, G. K. Golubev, D. Picard, and A. B. Tsybakov.
  Oracle inequalities for inverse problems.
  Ann. Statist., 30(3) :843–874, 2002.

- K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian.
  Image denoising by sparse 3-D transform-domain collaborative filtering.
  IEEE Trans. Image Process., 16(8) :2080–2095, 2007.

- D. L. Donoho and I. M. Johnstone.
  Adapting to unknown smoothness via wavelet shrinkage.
  J. Amer. Statist. Assoc., 90(432) :1200–1224, 1995.

- A. S. Dalalyan and A. B. Tsybakov.
  Aggregation by exponential weighting, sharp oracle inequalities and sparsity.
  In *COLT*, pages 97–111, 2007.

Références III

- A. S. Dalalyan and A. B. Tsybakov.
  Aggregation by exponential weighting, sharp pac-bayesian bounds and sparsity.
  Mach. Learn., 72(1-2) :39–61, 2008.
- A. S. Dalalyan and A. B. Tsybakov.
  Sparse regression learning by aggregation and Langevin Monte-Carlo.
  In *COLT*, 2009.
- G. Leung and A. R. Barron.
  Information theory and mixing least-squares regressions.
  IEEE Trans. Inf. Theory, 52(8) :3396–3410, 2006.
- G. Leung.
  Information Theory and Mixing Least Squares Regression.
  PhD thesis, Yale University, 2004.
- J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman.
  Non-local sparse models for image restoration.
  ICCV, 2009.

Références IV

- A. S. Nemirovski.
  Topics in non-parametric statistics, volume 1738 of *Lecture Notes in Math.*
  Springer, Berlin, 2000.
- Ph. Rigollet and A. B. Tsybakov.
  Exponential screening and optimal rates of sparse estimation.
  arXiv :1003.2654, submitted, 2010.
- G. Schwarz.
  Estimating the dimension of a model.
  Ann. Statist., 6(2) :461–464, 1978.
- C. M. Stein.
  Estimation of the mean of a multivariate normal distribution.
  Ann. Statist., 9(6) :1135–1151, 1981.
- R. Tibshirani.
  Regression shrinkage and selection via the lasso.
  J. Roy. Statist. Soc. Ser. B, 58(1) :267–288, 1996.

Références V

- A. N. Tikhonov.

  On the stability of inverse problems.

  C. R. (Doklady) Acad. Sci. URSS (N.S.), 39 :176–179, 1943.

- A. B. Tsybakov.

  Optimal rates of aggregation.

  In *COLT*, pages 303–313, 2003.