

Maxiset, sélection de modèles et estimation d'images par bandelettes

E. Le Pennec / LPMA / Université Denis Diderot - Paris VII
F. Autin, J.-M. Loubes, V. Rivoirard,
S. Mallat, Ch. Dossal et G. Peyré

02 Octobre 2006

Maxiset, sélection de modèles, bandelettes

Maxiset, sélection de modèles, bandelettes

- Estimation dans un modèle de bruit blanc gaussien :

$$dY = f(t)dt + \epsilon dW \quad .$$

Maxiset, sélection de modèles, bandelettes

- Estimation dans un modèle de bruit blanc gaussien :

$$dY = f(t)dt + \epsilon dW \quad .$$

- Estimation de f dans des bases : seuillage et sélection de modèles.

Maxiset, sélection de modèles, bandelettes

- Estimation dans un modèle de bruit blanc gaussien :

$$dY = f(t)dt + \epsilon dW \quad .$$

- Estimation de f dans des bases : seuillage et sélection de modèles.
- Maxiset : espace adaptée à un estimateur.

Maxiset, sélection de modèles, bandelettes

- Estimation dans un modèle de bruit blanc gaussien :

$$dY = f(t)dt + \epsilon dW \quad .$$

- Estimation de f dans des bases : seuillage et sélection de modèles.
- Maxiset : espace adaptée à un estimateur.
- Maxiset pour la sélection de modèle : espace d'approximation de la théorie de l'approximation.

Maxiset, sélection de modèles, bandelettes

- Estimation dans un modèle de bruit blanc gaussien :

$$dY = f(t)dt + \epsilon dW \quad .$$

- Estimation de f dans des bases : seuillage et sélection de modèles.
- Maxiset : espace adaptée à un estimateur.
- Maxiset pour la sélection de modèle : espace d'approximation de la théorie de l'approximation.
- Nécessité d'avoir une représentation creuse (approximation).

Maxiset, sélection de modèles, bandelettes

- Estimation dans un modèle de bruit blanc gaussien :

$$dY = f(t)dt + \epsilon dW \quad .$$

- Estimation de f dans des bases : seuillage et sélection de modèles.
- Maxiset : espace adaptée à un estimateur.
- Maxiset pour la sélection de modèle : espace d'approximation de la théorie de l'approximation.
- Nécessité d'avoir une représentation creuse (approximation).
- Images, représentations géométriques et estimation en bandelettes.

Plan

Plan

- Estimation par projection.

Plan

- Estimation par projection.
- Estimation oracle et approximation.

Plan

- Estimation par projection.
- Estimation oracle et approximation.
- Estimateur par seuillage et estimateur par sélection de modèles.

Plan

- Estimation par projection.
- Estimation oracle et approximation.
- Estimateur par seuillage et estimateur par sélection de modèles.
- Maxiset et espace d'approximations.

Plan

- Estimation par projection.
- Estimation oracle et approximation.
- Estimateur par seuillage et estimateur par sélection de modèles.
- Maxiset et espace d'approximations.
- Importance du choix des modèles.

Plan

- Estimation par projection.
- Estimation oracle et approximation.
- Estimateur par seuillage et estimateur par sélection de modèles.
- Maxiset et espace d'approximations.
- Importance du choix des modèles.
- 1D : Signaux , Fourier et ondelettes.

Plan

- Estimation par projection.
- Estimation oracle et approximation.
- Estimateur par seuillage et estimateur par sélection de modèles.
- Maxiset et espace d'approximations.
- Importance du choix des modèles.
- 1D : Signaux , Fourier et ondelettes.
- 2D : Images , ondelettes et représentations géométriques.

Plan

- Estimation par projection.
- Estimation oracle et approximation.
- Estimateur par seuillage et estimateur par sélection de modèles.
- Maxiset et espace d'approximations.
- Importance du choix des modèles.
- 1D : Signaux , Fourier et ondelettes.
- 2D : Images , ondelettes et représentations géométriques.
- Estimation d'images par bandelettes.

Estimation par projection

Estimation par projection

- Modèle de bruit blanc :

$$dY = f(t)dt + \epsilon dW \quad .$$

Estimation par projection

- Modèle de bruit blanc :

$$dY = f(t)dt + \epsilon dW \quad .$$

- Dans la suite, projection sur un espace de dimension N et calibrage $\epsilon = \frac{1}{\sqrt{N}}$ (lien avec la régression) :

$$Y = f + \frac{1}{\sqrt{N}}W \quad .$$

Estimation par projection

- Modèle de bruit blanc :

$$dY = f(t)dt + \epsilon dW \quad .$$

- Dans la suite, projection sur un espace de dimension N et calibrage $\epsilon = \frac{1}{\sqrt{N}}$ (lien avec la régression) :

$$Y = f + \frac{1}{\sqrt{N}}W \quad .$$

- Propriétés de f : propriétés dans le domaine continu.

Estimation par projection

- Modèle de bruit blanc :

$$dY = f(t)dt + \epsilon dW \quad .$$

- Dans la suite, projection sur un espace de dimension N et calibrage $\epsilon = \frac{1}{\sqrt{N}}$ (lien avec la régression) :

$$Y = f + \frac{1}{\sqrt{N}}W \quad .$$

- Propriétés de f : propriétés dans le domaine continu.
- Estimateur par projection : $F = P_{\hat{m}}Y$ avec \hat{m} s.e.v. (modèle) à choisir.

Estimation par projection

- Modèle de bruit blanc :

$$dY = f(t)dt + \epsilon dW \quad .$$

- Dans la suite, projection sur un espace de dimension N et calibrage $\epsilon = \frac{1}{\sqrt{N}}$ (lien avec la régression) :

$$Y = f + \frac{1}{\sqrt{N}}W \quad .$$

- Propriétés de f : propriétés dans le domaine continu.
- Estimateur par projection : $F = P_{\hat{m}}Y$ avec \hat{m} s.e.v. (modèle) à choisir.
- Choix de la collection \mathcal{M} de modèles m .

Estimation par projection

- Modèle de bruit blanc :

$$dY = f(t)dt + \epsilon dW \quad .$$

- Dans la suite, projection sur un espace de dimension N et calibrage $\epsilon = \frac{1}{\sqrt{N}}$ (lien avec la régression) :

$$Y = f + \frac{1}{\sqrt{N}}W \quad .$$

- Propriétés de f : propriétés dans le domaine continu.
- Estimateur par projection : $F = P_{\hat{m}}Y$ avec \hat{m} s.e.v. (modèle) à choisir.
- Choix de la collection \mathcal{M} de modèles m .
- Choix du modèle \hat{m} utilisé dans l'estimateur.

Estimation par projection

- Modèle de bruit blanc :

$$dY = f(t)dt + \epsilon dW \quad .$$

- Dans la suite, projection sur un espace de dimension N et calibrage $\epsilon = \frac{1}{\sqrt{N}}$ (lien avec la régression) :

$$Y = f + \frac{1}{\sqrt{N}}W \quad .$$

- Propriétés de f : propriétés dans le domaine continu.
- Estimateur par projection : $F = P_{\hat{m}}Y$ avec \hat{m} s.e.v. (modèle) à choisir.
- Choix de la collection \mathcal{M} de modèles m .
- Choix du modèle \hat{m} utilisé dans l'estimateur.
- Critère : risque quadratique

$$E(\|f - F\|^2)$$

Estimation oracle dans une base

Estimation oracle dans une base

- Base o.n. $\{b_n\}_n$ et $m = \text{vect}\{b_n\}_{n \in \Gamma}$

Estimation oracle dans une base

- Base o.n. $\{b_n\}_n$ et $m = \text{vect}\{b_n\}_{n \in \Gamma}$
- Décomposition de $Y = f + \frac{1}{\sqrt{N}}W$ dans une base orthonormée

$$Y = \sum_{b_n} \langle Y, b_n \rangle b_n = \sum_{b_n} \left(\langle f, b_n \rangle + \frac{1}{\sqrt{N}} \langle W, b_n \rangle \right) b_n \quad .$$

Estimation oracle dans une base

- Base o.n. $\{b_n\}_n$ et $m = \text{vect}\{b_n\}_{n \in \Gamma}$
- Décomposition de $Y = f + \frac{1}{\sqrt{N}}W$ dans une base orthonormée

$$Y = \sum_{b_n} \langle Y, b_n \rangle b_n = \sum_{b_n} \left(\langle f, b_n \rangle + \frac{1}{\sqrt{N}} \langle W, b_n \rangle \right) b_n \quad .$$

- Estimateur F par projection (conservation/élimination de coordonnées) :

$$F = P_m Y = Y_\Gamma = \sum_{n \in \Gamma} \langle Y, b_n \rangle b_n \quad .$$

Estimation oracle dans une base

- Base o.n. $\{b_n\}_n$ et $m = \text{vect}\{b_n\}_{n \in \Gamma}$
- Décomposition de $Y = f + \frac{1}{\sqrt{N}}W$ dans une base orthonormée

$$Y = \sum_{b_n} \langle Y, b_n \rangle b_n = \sum_{b_n} \left(\langle f, b_n \rangle + \frac{1}{\sqrt{N}} \langle W, b_n \rangle \right) b_n \quad .$$

- Estimateur F par projection (conservation/élimination de coordonnées) :

$$F = P_m Y = Y_\Gamma = \sum_{n \in \Gamma} \langle Y, b_n \rangle b_n \quad .$$

- Minimisation du risque quadratique :

$$E(\|f - F\|^2) = \sum_{n \notin \Gamma} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma} \frac{1}{N} \quad .$$

Estimation oracle dans une base

- Base o.n. $\{b_n\}_n$ et $m = \text{vect}\{b_n\}_{n \in \Gamma}$
- Décomposition de $Y = f + \frac{1}{\sqrt{N}}W$ dans une base orthonormée

$$Y = \sum_{b_n} \langle Y, b_n \rangle b_n = \sum_{b_n} \left(\langle f, b_n \rangle + \frac{1}{\sqrt{N}} \langle W, b_n \rangle \right) b_n \quad .$$

- Estimateur F par projection (conservation/élimination de coordonnées) :

$$F = P_m Y = Y_\Gamma = \sum_{n \in \Gamma} \langle Y, b_n \rangle b_n \quad .$$

- Minimisation du risque quadratique :

$$E(\|f - F\|^2) = \sum_{n \notin \Gamma} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma} \frac{1}{N} \quad .$$

- Solution : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \frac{1}{\sqrt{N}}\}$ et $F_O = Y_{\Gamma_O}$.

Estimation oracle dans une base

- Base o.n. $\{b_n\}_n$ et $m = \text{vect}\{b_n\}_{n \in \Gamma}$
- Décomposition de $Y = f + \frac{1}{\sqrt{N}}W$ dans une base orthonormée

$$Y = \sum_{b_n} \langle Y, b_n \rangle b_n = \sum_{b_n} \left(\langle f, b_n \rangle + \frac{1}{\sqrt{N}} \langle W, b_n \rangle \right) b_n \quad .$$

- Estimateur F par projection (conservation/élimination de coordonnées) :

$$F = P_m Y = Y_\Gamma = \sum_{n \in \Gamma} \langle Y, b_n \rangle b_n \quad .$$

- Minimisation du risque quadratique :

$$E(\|f - F\|^2) = \sum_{n \notin \Gamma} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma} \frac{1}{N} \quad .$$

- Solution : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \frac{1}{\sqrt{N}}\}$ et $F_O = Y_{\Gamma_O}$.
- Problème : demande la connaissance de f ! (Oracle)

Oracle, risque et approximation

Oracle, risque et approximation

- Risque quadratique de l'estimateur oracle F_O :

$$E(\|f - F_O\|^2) = \sum_{n \notin \Gamma_O} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma_O} \frac{1}{N}$$

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \frac{1}{N} |\Gamma_O| \quad .$$

Oracle, risque et approximation

- Risque quadratique de l'estimateur oracle F_O :

$$E(\|f - F_O\|^2) = \sum_{n \notin \Gamma_O} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma_O} \frac{1}{N}$$

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \frac{1}{N} |\Gamma_O| \quad .$$

- Compromis entre erreur d'approximation et nombre de termes.

Oracle, risque et approximation

- Risque quadratique de l'estimateur oracle F_O :

$$E(\|f - F_O\|^2) = \sum_{n \notin \Gamma_O} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma_O} \frac{1}{N}$$

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \frac{1}{N} |\Gamma_O| \quad .$$

- Compromis entre erreur d'approximation et nombre de termes.
- Théorie de l'approximation :

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \frac{1}{N} |\Gamma_O| \leq C \left(\frac{1}{N} \right)^{\frac{\beta}{\beta+1}}$$
$$\Leftrightarrow \min_{\dim(m) \leq M} \|f - P_m f\|^2 \leq C M^{-\beta} \Leftrightarrow f \in \mathcal{A}^\beta \quad .$$

Oracle, risque et approximation

- Risque quadratique de l'estimateur oracle F_O :

$$E(\|f - F_O\|^2) = \sum_{n \notin \Gamma_O} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma_O} \frac{1}{N}$$

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \frac{1}{N} |\Gamma_O| \quad .$$

- Compromis entre erreur d'approximation et nombre de termes.
- Théorie de l'approximation :

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \frac{1}{N} |\Gamma_O| \leq C \left(\frac{1}{N} \right)^{\frac{\beta}{\beta+1}}$$
$$\Leftrightarrow \min_{\dim(m) \leq M} \|f - P_m f\|^2 \leq C M^{-\beta} \Leftrightarrow f \in \mathcal{A}^\beta \quad .$$

- Minimax : pour Θ , classe de fonctions, recherche d'une base telle que $\Theta \subset \mathcal{A}^\beta$ avec β optimal ($(\frac{1}{N})^{\frac{\beta}{\beta+1}}$ vitesse minimax).

Oracle, risque et approximation

- Risque quadratique de l'estimateur oracle F_O :

$$E(\|f - F_O\|^2) = \sum_{n \notin \Gamma_O} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma_O} \frac{1}{N}$$

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \frac{1}{N} |\Gamma_O| \quad .$$

- Compromis entre erreur d'approximation et nombre de termes.

- Théorie de l'approximation :

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \frac{1}{N} |\Gamma_O| \leq C \left(\frac{1}{N} \right)^{\frac{\beta}{\beta+1}}$$
$$\Leftrightarrow \min_{\dim(m) \leq M} \|f - P_m f\|^2 \leq C M^{-\beta} \Leftrightarrow f \in \mathcal{A}^\beta \quad .$$

- Minimax : pour Θ , classe de fonctions, recherche d'une base telle que

$\Theta \subset \mathcal{A}^\beta$ avec β optimal ($(\frac{1}{N})^{\frac{\beta}{\beta+1}}$ vitesse minimax).

- Maxiset : pour une base fixée, ensemble des fonctions estimées avec une vitesse $(\frac{1}{N})^{\frac{\beta}{\beta+1}} = \mathcal{A}^\beta$.

Estimateur par seuillage

Estimateur par seuillage

- Oracle : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \frac{1}{\sqrt{N}}\}$ et $F_O = Y_{\Gamma_O}$.

Estimateur par seuillage

- Oracle : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \frac{1}{\sqrt{N}}\}$ et $F_O = Y_{\Gamma_O}$.
- Stratégie : garder les grands coefficients.

Estimateur par seuillage

- Oracle : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \frac{1}{\sqrt{N}}\}$ et $F_O = Y_{\Gamma_O}$.
- Stratégie : garder les grands coefficients.
- Seuillage : $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq T \left(\frac{1}{\sqrt{N}} \right)\}$ et $F_S = Y_{\Gamma_S}$.

Estimateur par seuillage

- Oracle : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \frac{1}{\sqrt{N}}\}$ et $F_O = Y_{\Gamma_O}$.
- Stratégie : garder les grands coefficients.
- Seuillage : $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq T \left(\frac{1}{\sqrt{N}} \right)\}$ et $F_S = Y_{\Gamma_S}$.
- **Théorème (Donoho, Johnstone)** : Si $T \left(\frac{1}{\sqrt{N}} \right) = \lambda \sqrt{\frac{\log N}{N}}$, alors

$$E(\|f - F_S\|^2) \leq C(\log N)E(\|f - F_O\|^2)$$

$$E(\|f - F_S\|^2) \leq C \min_{\Gamma} \|f - f_{\Gamma}\|^2 + \lambda^2 \frac{\log N}{N} |\Gamma| \quad \text{plus fin.}$$

Estimateur par seuillage

- Oracle : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \frac{1}{\sqrt{N}}\}$ et $F_O = Y_{\Gamma_O}$.
- Stratégie : garder les grands coefficients.
- Seuillage : $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq T \left(\frac{1}{\sqrt{N}}\right)\}$ et $F_S = Y_{\Gamma_S}$.
- **Théorème (Donoho, Johnstone)** : Si $T \left(\frac{1}{\sqrt{N}}\right) = \lambda \sqrt{\frac{\log N}{N}}$, alors

$$E(\|f - F_S\|^2) \leq C(\log N)E(\|f - F_O\|^2)$$

$$E(\|f - F_S\|^2) \leq C \min_{\Gamma} \|f - f_{\Gamma}\|^2 + \lambda^2 \frac{\log N}{N} |\Gamma| \quad \text{plus fin.}$$

- **Théorème (Maxiset) (Cohen, DeVore, Kerkyacharian, Picard)** :

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N}\right)^{\frac{\beta}{\beta+1}} \Leftrightarrow f \in V_{\frac{2\beta}{\beta+1}}^*$$

$$\Leftrightarrow \min_{\Gamma} \|f - f_{\Gamma}\|^2 + \lambda^2 T^2 |\Gamma| \leq CT^2 \frac{\beta}{\beta+1}$$

$$\Leftrightarrow \min_{\dim(m) \leq M} \|f - P_m f\|^2 \leq CM^{-\beta} \Leftrightarrow f \in \mathcal{A}^{\beta} \quad .$$

Seuillage et sélection de modèles

Seuillage et sélection de modèles

- Risque oracle : $\|f - f_{\Gamma_O}\|^2 + \frac{1}{N}|\Gamma_O|$.

Seuillage et sélection de modèles

- Risque oracle : $\|f - f_{\Gamma_O}\|^2 + \frac{1}{N}|\Gamma_O|$.
- Analogue empirique : $\|Y - Y_{\Gamma}\|^2 + \frac{\lambda_N}{N}|\Gamma|$.

Seuillage et sélection de modèles

- Risque oracle : $\|f - f_{\Gamma_O}\|^2 + \frac{1}{N}|\Gamma_O|$.
- Analogue empirique : $\|Y - Y_{\Gamma}\|^2 + \frac{\lambda_N}{N}|\Gamma|$.
- Minimisation : $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq \sqrt{\frac{\lambda_N}{N}}\}$ (seuillage) et $F_S = Y_{\Gamma_S}$.

Seuillage et sélection de modèles

- Risque oracle : $\|f - f_{\Gamma_O}\|^2 + \frac{1}{N}|\Gamma_O|$.
- Analogue empirique : $\|Y - Y_{\Gamma}\|^2 + \frac{\lambda_N}{N}|\Gamma|$.
- Minimisation : $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq \sqrt{\frac{\lambda_N}{N}}\}$ (seuillage) et $F_S = Y_{\Gamma_S}$.
- Cadre de la sélection de modèles avec $\text{pen}(m) = \frac{\lambda_N}{N} \dim(m)$:

$$F_S = \underset{P_m Y, m \in \mathcal{M}_N}{\text{argmin}} \|Y - P_m Y\|^2 + \text{pen}(m) \quad .$$

Seuillage et sélection de modèles

- Risque oracle : $\|f - f_{\Gamma_O}\|^2 + \frac{1}{N}|\Gamma_O|$.
- Analogue empirique : $\|Y - Y_{\Gamma}\|^2 + \frac{\lambda_N}{N}|\Gamma|$.
- Minimisation : $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq \sqrt{\frac{\lambda_N}{N}}\}$ (seuillage) et $F_S = Y_{\Gamma_S}$.
- Cadre de la sélection de modèles avec $\text{pen}(m) = \frac{\lambda_N}{N} \dim(m)$:

$$F_S = \underset{P_m Y, m \in \mathcal{M}_N}{\text{argmin}} \|Y - P_m Y\|^2 + \text{pen}(m) \quad .$$

- L'ensemble \mathcal{M}_N des modèles m parcourent l'ensemble des sous-espaces engendrés par les N vecteurs de bases.

Seuillage et sélection de modèles

- Risque oracle : $\|f - f_{\Gamma_O}\|^2 + \frac{1}{N}|\Gamma_O|$.
- Analogue empirique : $\|Y - Y_{\Gamma}\|^2 + \frac{\lambda_N}{N}|\Gamma|$.
- Minimisation : $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq \sqrt{\frac{\lambda_N}{N}}\}$ (seuillage) et $F_S = Y_{\Gamma_S}$.
- Cadre de la sélection de modèles avec $\text{pen}(m) = \frac{\lambda_N}{N} \dim(m)$:

$$F_S = \underset{P_m Y, m \in \mathcal{M}_N}{\text{argmin}} \|Y - P_m Y\|^2 + \text{pen}(m) \quad .$$

- L'ensemble \mathcal{M}_N des modèles m parcourent l'ensemble des sous-espaces engendrés par les N vecteurs de bases.
- Inégalité de Kraft satisfaite pour $\lambda_N = \lambda \sqrt{\log N}$:

$$\sum_{m \in \mathcal{M}_N} e^{-\lambda_N \dim(m)} < +\infty \quad .$$

Seuillage et sélection de modèles

- Risque oracle : $\|f - f_{\Gamma_O}\|^2 + \frac{1}{N}|\Gamma_O|$.
- Analogue empirique : $\|Y - Y_{\Gamma}\|^2 + \frac{\lambda_N}{N}|\Gamma|$.
- Minimisation : $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq \sqrt{\frac{\lambda_N}{N}}\}$ (seuillage) et $F_S = Y_{\Gamma_S}$.
- Cadre de la sélection de modèles avec $\text{pen}(m) = \frac{\lambda_N}{N} \dim(m)$:

$$F_S = \underset{P_m Y, m \in \mathcal{M}_N}{\text{argmin}} \|Y - P_m Y\|^2 + \text{pen}(m) \quad .$$

- L'ensemble \mathcal{M}_N des modèles m parcourent l'ensemble des sous-espaces engendrés par les N vecteurs de bases.
- Inégalité de Kraft satisfaite pour $\lambda_N = \lambda \sqrt{\log N}$:

$$\sum_{m \in \mathcal{M}_N} e^{-\lambda_N \dim(m)} < +\infty \quad .$$

- **Théorème (Barron, Birgé, Massart)** : Pour λ assez grand,

$$E(\|f - F_S\|^2) \leq C \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + \lambda^2 \frac{\log N}{N} \dim(m) \quad .$$

Seuillage et sélection de modèles

- Risque oracle : $\|f - f_{\Gamma_O}\|^2 + \frac{1}{N}|\Gamma_O|$.
- Analogue empirique : $\|Y - Y_{\Gamma}\|^2 + \frac{\lambda_N}{N}|\Gamma|$.
- Minimisation : $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq \sqrt{\frac{\lambda_N}{N}}\}$ (seuillage) et $F_S = Y_{\Gamma_S}$.
- Cadre de la sélection de modèles avec $\text{pen}(m) = \frac{\lambda_N}{N} \dim(m)$:

$$F_S = \underset{P_m Y, m \in \mathcal{M}_N}{\text{argmin}} \|Y - P_m Y\|^2 + \text{pen}(m) \quad .$$

- L'ensemble \mathcal{M}_N des modèles m parcourent l'ensemble des sous-espaces engendrés par les N vecteurs de bases.
- Inégalité de Kraft satisfaite pour $\lambda_N = \lambda \sqrt{\log N}$:

$$\sum_{m \in \mathcal{M}_N} e^{-\lambda_N \dim(m)} < +\infty \quad .$$

- **Théorème (Barron, Birgé, Massart)** : Pour λ assez grand,

$$E(\|f - F_S\|^2) \leq C \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + \lambda^2 \frac{\log N}{N} \dim(m) \quad .$$

- Cadre permettant de travailler dans plusieurs bases à la fois...

Sélection de modèles

Sélection de modèles

- **Théorème (Barron, Birgé, Massart)** : Si la collection \mathcal{M}_N de modèles m satisfait une inégalité de Kraft pour des coefficients $\lambda_{N,m}$ ($\sum_{m \in \mathcal{M}_N} e^{-\lambda_{N,m} \dim(m)} < +\infty$) alors pour $\text{pen}(m) = (C_1 + C_2 \lambda_{N,m}) \frac{\dim(m)}{N}$

$$F_S = \underset{P_m Y, m \in \mathcal{M}_N}{\text{argmin}} \|Y - P_m Y\|^2 + \text{pen}(m)$$

satisfait

$$E(\|f - F_S\|^2) \leq C \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + (C_1 + C_2 \lambda_{N,m}) \frac{\dim(m)}{N} .$$

Sélection de modèles

- **Théorème (Barron, Birgé, Massart)** : Si la collection \mathcal{M}_N de modèles m satisfait une inégalité de Kraft pour des coefficients $\lambda_{N,m}$ ($\sum_{m \in \mathcal{M}_N} e^{-\lambda_{N,m} \dim(m)} < +\infty$) alors pour $\text{pen}(m) = (C_1 + C_2 \lambda_{N,m}) \frac{\dim(m)}{N}$

$$F_S = \underset{P_m Y, m \in \mathcal{M}_N}{\text{argmin}} \|Y - P_m Y\|^2 + \text{pen}(m)$$

satisfait

$$E(\|f - F_S\|^2) \leq C \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + (C_1 + C_2 \lambda_{N,m}) \frac{\dim(m)}{N} .$$

- **Théorème (Maxiset)** : Si $\mathcal{M}_N \subset \mathcal{M}_{N+1}$, $\lambda_{N,m} = \lambda_N$ et $1 \leq \frac{\lambda_N}{\lambda_{N/2}} \leq (2 - 2\epsilon)$ alors

$$E(\|f - F_S\|^2) \leq C \left(\frac{\lambda_N}{N} \right)^{\frac{\beta}{\beta+1}}$$

$$\Leftrightarrow \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + (C_1 + C_2 \lambda_N) \frac{\dim(m)}{N} \leq C \left(\frac{\lambda_N}{N} \right)^{\frac{\beta}{\beta+1}}$$

$$\Leftrightarrow \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + T^2 \dim(m) \leq C (T^2)^{\frac{\beta}{\beta+1}}$$

$$\Leftrightarrow \min_{\dim(m) \leq M} \|f - P_m f\|^2 \leq C M^{-\beta} \Leftrightarrow f \in \mathcal{A}^\beta .$$

Esquisse de preuve – 1

Esquisse de preuve – 1

• $f \in \mathcal{A}^\beta \implies E(\|f - F_S\|^2) \leq C \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}}$: sélection de modèles.

Esquisse de preuve – 1

- $f \in \mathcal{A}^\beta \implies E(\|f - F_S\|^2) \leq C \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}}$: sélection de modèles.
- Sens inverse :

$$E(\|f - F_S\|^2) \geq C \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + \lambda_N \frac{\dim(m)}{N} \quad ???$$

Esquisse de preuve – 1

- $f \in \mathcal{A}^\beta \implies E(\|f - F_S\|^2) \leq C \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}}$: sélection de modèles.
- Sens inverse :

$$E(\|f - F_S\|^2) \geq C \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + \lambda_N \frac{\dim(m)}{N} \quad ???$$

- Cas du seuillage :

Esquisse de preuve – 1

- $f \in \mathcal{A}^\beta \implies E(\|f - F_S\|^2) \leq C \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}}$: sélection de modèles.
- Sens inverse :

$$E(\|f - F_S\|^2) \geq C \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + \lambda_N \frac{\dim(m)}{N} \quad ???$$

- Cas du seuillage :
 - On montre

$$\|f - F_S\|^2 \geq \|f - f_{T/2}\|^2 \quad .$$

Esquisse de preuve – 1

- $f \in \mathcal{A}^\beta \implies E(\|f - F_S\|^2) \leq C \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}}$: sélection de modèles.
- Sens inverse :

$$E(\|f - F_S\|^2) \geq C \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + \lambda_N \frac{\dim(m)}{N} \quad ???$$

- Cas du seuillage :
- On montre

$$\|f - F_S\|^2 \geq \|f - f_{T/2}\|^2 \quad .$$

- Ensuite on obtient

$$\|f - f_{T/2}\|^2 \leq CT^{\frac{2\alpha}{\alpha+1}} \implies \|f - f_{T/2}\|^2 + T^2 M \leq C_\alpha T^{\frac{2\alpha}{\alpha+1}} \quad .$$

Esquisse de preuve – 2

- \hat{m} : modèle sélectionné qui minimise

$$\|Y - P_m Y\|^2 + \lambda_N \frac{\dim(m)}{N} .$$

Esquisse de preuve – 2

- \hat{m} : modèle sélectionné qui minimise

$$\|Y - P_m Y\|^2 + \lambda_N \frac{\dim(m)}{N} \quad .$$

- m_O : modèle oracle qui minimise pour K suffisamment grand

$$\|f - P_m f\|^2 + \lambda_N \frac{\dim(m)}{KN} \quad .$$

Esquisse de preuve – 2

- \hat{m} : modèle sélectionné qui minimise

$$\|Y - P_m Y\|^2 + \lambda_N \frac{\dim(m)}{N} .$$

- m_O : modèle oracle qui minimise pour K suffisamment grand

$$\|f - P_m f\|^2 + \lambda_N \frac{\dim(m)}{KN} .$$

- En général, on a pas

$$\|f - P_{\hat{m}} Y\|^2 \geq \|f - P_{m_O} f\|^2 .$$

Esquisse de preuve – 2

- \hat{m} : modèle sélectionné qui minimise

$$\|Y - P_m Y\|^2 + \lambda_N \frac{\dim(m)}{N} \quad .$$

- m_O : modèle oracle qui minimise pour K suffisamment grand

$$\|f - P_m f\|^2 + \lambda_N \frac{\dim(m)}{KN} \quad .$$

- En général, on a pas

$$\|f - P_{\hat{m}} Y\|^2 \geq \|f - P_{m_O} f\|^2 \quad .$$

- On va montrer que

$$E(\|f - P_{\hat{m}} Y\|^2) \leq C \left(\frac{\lambda_N}{N} \right)^{\frac{\beta}{\beta+1}} \implies \|f - P_{m_O} f\|^2 + \lambda_N \frac{\dim(m_O)}{KN} \leq C \left(\frac{\lambda_N}{N} \right)^{\frac{\beta}{\beta+1}}$$

Esquisse de preuve – 3

● Par définition :

$$\|Y - P_{\hat{m}}Y\|^2 + \lambda_N \frac{\dim \hat{m}}{N} \leq \|Y - P_{m_O}Y\|^2 + \lambda_N \frac{\dim m_O}{N}$$

Esquisse de preuve – 3

- Par définition :

$$\|Y - P_{\hat{m}}Y\|^2 + \lambda_N \frac{\dim \hat{m}}{N} \leq \|Y - P_{m_O}Y\|^2 + \lambda_N \frac{\dim m_O}{N}$$

- En passant à l'espérance, on obtient

$$E\left(\lambda_N \frac{\dim \hat{m}}{N}\right) \leq \|f - P_{m_O}f\|^2 + \frac{\dim m_O}{N} + \lambda_N \frac{\dim m_O}{N}$$

Esquisse de preuve – 3

- Par définition :

$$\|Y - P_{\hat{m}}Y\|^2 + \lambda_N \frac{\dim \hat{m}}{N} \leq \|Y - P_{m_O}Y\|^2 + \lambda_N \frac{\dim m_O}{N}$$

- En passant à l'espérance, on obtient

$$E\left(\lambda_N \frac{\dim \hat{m}}{N}\right) \leq \|f - P_{m_O}f\|^2 + \frac{\dim m_O}{N} + \lambda_N \frac{\dim m_O}{N}$$

- Par ailleurs,

$$\|f - P_{\hat{m}}f\|^2 + \lambda_N \frac{\dim \hat{m}}{KN} \geq \|f - P_{m_O}f\|^2 + \lambda_N \frac{\dim m_O}{KN}$$

Esquisse de preuve – 3

- Par définition :

$$\|Y - P_{\hat{m}}Y\|^2 + \lambda_N \frac{\dim \hat{m}}{N} \leq \|Y - P_{m_O}Y\|^2 + \lambda_N \frac{\dim m_O}{N}$$

- En passant à l'espérance, on obtient

$$E\left(\lambda_N \frac{\dim \hat{m}}{N}\right) \leq \|f - P_{m_O}f\|^2 + \frac{\dim m_O}{N} + \lambda_N \frac{\dim m_O}{N}$$

- Par ailleurs,

$$\|f - P_{\hat{m}}f\|^2 + \lambda_N \frac{\dim \hat{m}}{KN} \geq \|f - P_{m_O}f\|^2 + \lambda_N \frac{\dim m_O}{KN}$$

- En passant à l'espérance, on en déduit

$$E(\|f - P_{\hat{m}}f\|^2) \geq \frac{K-1}{K} \left(\|f - P_{m_O}f\|^2 - \frac{\dim(m_O)}{(K-1)N} \right)$$

Esquisse de preuve – 4

- On introduit alors la dépendance en N pour montrer par récurrence que si

$$E(\|f - P_{\hat{m}}Y\|^2) \leq C \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}}$$

alors, sous une hypothèse de croissance pour λ_N et de structure des modèles, pour K et C_2 assez grand

$$\frac{\dim(m_O(N))}{KN} \leq C_2 \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}} .$$

Esquisse de preuve – 4

- On introduit alors la dépendance en N pour montrer par récurrence que si

$$E(\|f - P_{\hat{m}}Y\|^2) \leq C \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}}$$

alors, sous une hypothèse de croissance pour λ_N et de structure des modèles, pour K et C_2 assez grand

$$\frac{\dim(m_O(N))}{KN} \leq C_2 \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}} .$$

- Ceci implique

$$\|f - P_{m_O(N)}f\|^2 \leq \left(\frac{K}{K-1}C + C_2\right) \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}} .$$

Esquisse de preuve – 4

- On introduit alors la dépendance en N pour montrer par récurrence que si

$$E(\|f - P_{\hat{m}}Y\|^2) \leq C \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}}$$

alors, sous une hypothèse de croissance pour λ_N et de structure des modèles, pour K et C_2 assez grand

$$\frac{\dim(m_O(N))}{KN} \leq C_2 \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}} .$$

- Ceci implique

$$\|f - P_{m_O(N)}f\|^2 \leq \left(\frac{K}{K-1}C + C_2\right) \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}} .$$

- Enfin sous l'hypothèse, $\lambda_N \leq 2(1 - \epsilon)\lambda_{N/2}$, on démontre

$$\|f - P_{m_O}f\|^2 + \lambda_N \frac{\dim(m_O)}{KN} \leq \left(\frac{K}{K-1}C + C_2 + C_3\right) \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}} .$$

Choix de la collection de modèles

Choix de la collection de modèles

- Trois critères :
 - Espace d'approximation \mathcal{A}^β grand,
 - Collection pas trop grande (λ_N au plus logarithmique),
 - Algorithmique pour la minimisation.

Choix de la collection de modèles

- Trois critères :
 - Espace d'approximation \mathcal{A}^β grand,
 - Collection pas trop grande (λ_N au plus logarithmique),
 - Algorithmique pour la minimisation.
- \mathcal{M} non linéaire : m sous-espaces engendrés par des vecteurs d'une base o.n.
Espaces d'approximation non linéaire, $\lambda_N \simeq \log N$ et seuillage.

Choix de la collection de modèles

- Trois critères :
 - Espace d'approximation \mathcal{A}^β grand,
 - Collection pas trop grande (λ_N au plus logarithmique),
 - Algorithmique pour la minimisation.
- \mathcal{M} non linéaire : m sous-espaces engendrés par des vecteurs d'une base o.n.
Espaces d'approximation non linéaire, $\lambda_N \simeq \log N$ et seuillage.
- \mathcal{M} linéaire : m sous-espaces croissants engendrés par des vecteurs d'une base o.n. dans l'ordre.
Espaces d'approximation linéaire, $\lambda_N \simeq C$ et seuillage.

Choix de la collection de modèles

- Trois critères :
 - Espace d'approximation \mathcal{A}^β grand,
 - Collection pas trop grande (λ_N au plus logarithmique),
 - Algorithmique pour la minimisation.
- \mathcal{M} non linéaire : m sous-espaces engendrés par des vecteurs d'une base o.n.
Espaces d'approximation non linéaire, $\lambda_N \simeq \log N$ et seuillage.
- \mathcal{M} linéaire : m sous-espaces croissants engendrés par des vecteurs d'une base o.n. dans l'ordre.
Espaces d'approximation linéaire, $\lambda_N \simeq C$ et seuillage.
- \mathcal{M} hautement non linéaire : m sous-espaces engendrés par des vecteurs d'une base choisie parmi un dictionnaire de bases.
Espaces d'approximations hautement non linéaires...
 $\lambda_N \simeq C \log N$ si nb total de vecteurs de base = $\mathcal{O}(N^C)$.
Seuillage et algorithme de recherche de meilleure base si structure...

Choix de la collection de modèles

- Trois critères :
 - Espace d'approximation \mathcal{A}^β grand,
 - Collection pas trop grande (λ_N au plus logarithmique),
 - Algorithmique pour la minimisation.
- \mathcal{M} non linéaire : m sous-espaces engendrés par des vecteurs d'une base o.n.
Espaces d'approximation non linéaire, $\lambda_N \simeq \log N$ et seuillage.
- \mathcal{M} linéaire : m sous-espaces croissants engendrés par des vecteurs d'une base o.n. dans l'ordre.
Espaces d'approximation linéaire, $\lambda_N \simeq C$ et seuillage.
- \mathcal{M} hautement non linéaire : m sous-espaces engendrés par des vecteurs d'une base choisie parmi un dictionnaire de bases.
Espaces d'approximations hautement non linéaires...
 $\lambda_N \simeq C \log N$ si nb total de vecteurs de base = $\mathcal{O}(N^C)$.
Seuillage et algorithme de recherche de meilleure base si structure...
- Exemples autour des fonctions C^α .

Fonctions C^α et Fourier

Fonctions C^α et Fourier

- Fonctions C^α : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).

Fonctions C^α et Fourier

- Fonctions C^α : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).
- Seuillage dans la base de Fourier :

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} .$$

Fonctions C^α et Fourier

- Fonctions C^α : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).
- Seuillage dans la base de Fourier :

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} .$$

- Disparition du facteur $\log N$ en sélectionnant les M premiers coefficients au lieu des M plus grands (approximation linéaire).

Fonctions C^α et Fourier

- Fonctions C^α : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).
- Seuillage dans la base de Fourier :

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} .$$

- Disparition du facteur $\log N$ en sélectionnant les M premiers coefficients au lieu des M plus grands (approximation linéaire).
- Maxiset redonne le résultat : $C^\alpha \subset H^\alpha \subset \mathcal{A}^{2\alpha}(\text{Lin}) \subset \mathcal{A}^{2\alpha}(\text{NonLin})$ puisque $\mathcal{A}^{2\alpha}(\text{Lin})$ est une version faible de H^α .

Fonctions C^α et Fourier

- Fonctions C^α : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).
- Seuillage dans la base de Fourier :

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} .$$

- Disparition du facteur $\log N$ en sélectionnant les M premiers coefficients au lieu des M plus grands (approximation linéaire).
- Maxiset redonne le résultat : $C^\alpha \subset H^\alpha \subset \mathcal{A}^{2\alpha}(\text{Lin}) \subset \mathcal{A}^{2\alpha}(\text{NonLin})$ puisque $\mathcal{A}^{2\alpha}(\text{Lin})$ est une version faible de H^α .
- Pour les fonctions C^α par morceaux : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).

Fonctions C^α et Fourier

- Fonctions C^α : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).
- Seuillage dans la base de Fourier :

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} .$$

- Disparition du facteur $\log N$ en sélectionnant les M premiers coefficients au lieu des M plus grands (approximation linéaire).
- Maxiset redonne le résultat : $C^\alpha \subset H^\alpha \subset \mathcal{A}^{2\alpha}(\text{Lin}) \subset \mathcal{A}^{2\alpha}(\text{NonLin})$ puisque $\mathcal{A}^{2\alpha}(\text{Lin})$ est une version faible de H^α .
- Pour les fonctions C^α par morceaux : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).
- Pour $\alpha > 1$, C^α par morceaux $\not\subset \mathcal{A}^{2\alpha}(\text{NonLin})$.

Fonctions C^α et Fourier

- Fonctions C^α : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).
- Seuillage dans la base de Fourier :

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} .$$

- Disparition du facteur $\log N$ en sélectionnant les M premiers coefficients au lieu des M plus grands (approximation linéaire).
- Maxiset redonne le résultat : $C^\alpha \subset H^\alpha \subset \mathcal{A}^{2\alpha}(\text{Lin}) \subset \mathcal{A}^{2\alpha}(\text{NonLin})$ puisque $\mathcal{A}^{2\alpha}(\text{Lin})$ est une version faible de H^α .
- Pour les fonctions C^α par morceaux : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).
- Pour $\alpha > 1$, C^α par morceaux $\not\subset \mathcal{A}^{2\alpha}(\text{NonLin})$.
- On a uniquement C^α par morceaux $\subset \mathcal{A}^2(\text{Lin})$!

Fonctions C^α et Fourier

- Fonctions C^α : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).
- Seuillage dans la base de Fourier :

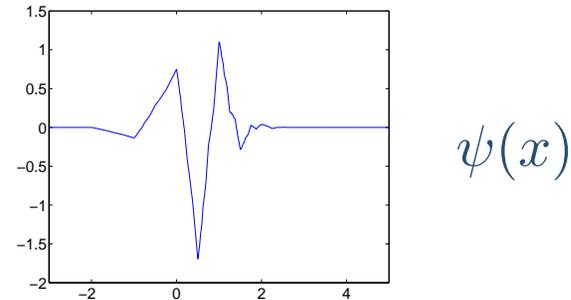
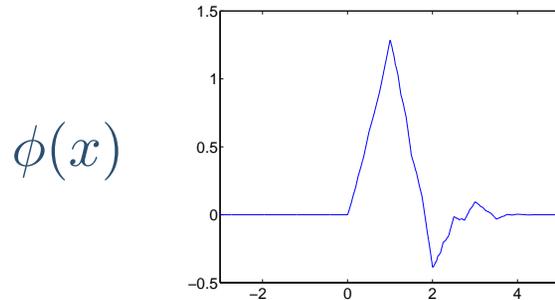
$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} .$$

- Disparition du facteur $\log N$ en sélectionnant les M premiers coefficients au lieu des M plus grands (approximation linéaire).
- Maxiset redonne le résultat : $C^\alpha \subset H^\alpha \subset \mathcal{A}^{2\alpha}(\text{Lin}) \subset \mathcal{A}^{2\alpha}(\text{NonLin})$ puisque $\mathcal{A}^{2\alpha}(\text{Lin})$ est une version faible de H^α .
- Pour les fonctions C^α par morceaux : vitesse minimax $(\frac{1}{N})^{\frac{2\alpha}{2\alpha+1}}$ ($\beta = 2\alpha$).
- Pour $\alpha > 1$, C^α par morceaux $\not\subset \mathcal{A}^{2\alpha}(\text{NonLin})$.
- On a uniquement C^α par morceaux $\subset \mathcal{A}^2(\text{Lin})$!
- Besoin d'autres bases...

Base d'ondelettes 1D de $L^2[0, 1]$

Base d'ondelettes 1D de $L^2[0, 1]$

- Construite à partir d'une fonction d'échelle $\phi(x)$ et d'une ondelette mère $\psi(x)$



qui sont dilatées par 2^j et translatées de $2^j n$

$$\phi_{j,n}(x) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{x - 2^j n}{2^j}\right) \quad , \quad \psi_{j,n}(x) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{x - 2^j n}{2^j}\right) \quad .$$

Base d'ondelettes 1D de $L^2[0, 1]$

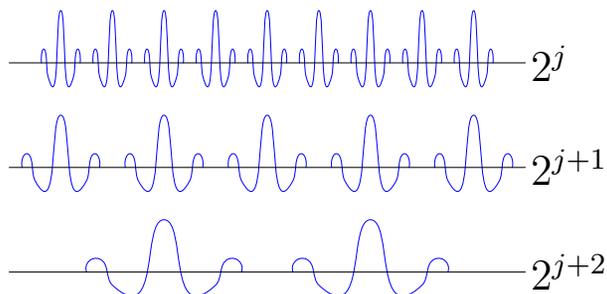
- Construite à partir d'une fonction d'échelle $\phi(x)$ et d'une ondelette mère $\psi(x)$



qui sont dilatées par 2^j et translatées de $2^j n$

$$\phi_{j,n}(x) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{x - 2^j n}{2^j}\right) \quad , \quad \psi_{j,n}(x) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{x - 2^j n}{2^j}\right) \quad .$$

- $\mathbf{B} = \left\{ \psi_{j,n} \right\}_{j \in \mathbb{N}, 2^j n \in [0,1]}$ est une base orthonormale de $L^2[0, 1]$.



Fonctions C^α et ondelettes

Fonctions C^α et ondelettes

- Maxisets bien étudiés dans ce cadre (C , DV , K , P , *Autin*, *Rivoirard*).

Fonctions C^α et ondelettes

- Maxisets bien étudiés dans ce cadre (C , DV , K , P , *Autin*, *Rivoirard*).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^{\beta/2}$.

Fonctions C^α et ondelettes

- Maxisets bien étudiés dans ce cadre ($C, DV, K, P, Autin, Rivoirard$).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^{\beta/2}$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{W}B_{2/(\beta+1),2/(\beta+1)}^{\beta/2}$ avec $\mathcal{W}B_{p,q}^s$ version faible de $B_{p,q}^s$.

Fonctions C^α et ondelettes

- Maxisets bien étudiés dans ce cadre (C , DV , K , P , *Autin*, *Rivoirard*).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^{\beta/2}$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{W}B_{2/(\beta+1),2/(\beta+1)}^{\beta/2}$ avec $\mathcal{W}B_{p,q}^s$ version faible de $B_{p,q}^s$.
- $C^\alpha \subsetneq \mathcal{A}^{2\alpha}(\text{Lin}) = B_{2,\infty}^\alpha \subsetneq C^\alpha$ par morceaux $\subsetneq \mathcal{A}^{2\alpha}(\text{NonLin}) = \mathcal{W}B_{p,p}^\alpha$ avec $p = 2/(2\alpha + 1)$.

Fonctions C^α et ondelettes

- Maxisets bien étudiés dans ce cadre (C , DV , K , P , *Autin*, *Rivoirard*).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^{\beta/2}$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{W}B_{2/(\beta+1),2/(\beta+1)}^{\beta/2}$ avec $\mathcal{W}B_{p,q}^s$ version faible de $B_{p,q}^s$.
- $C^\alpha \subsetneq \mathcal{A}^{2\alpha}(\text{Lin}) = B_{2,\infty}^\alpha \subsetneq C^\alpha$ par morceaux $\subsetneq \mathcal{A}^{2\alpha}(\text{NonLin}) = \mathcal{W}B_{p,p}^\alpha$ avec $p = 2/(2\alpha + 1)$.
- Estimation par seuillage dans une base d'ondelettes pour $f \in C^\alpha$ par morceaux :

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} .$$

Fonctions C^α et ondelettes

- Maxisets bien étudiés dans ce cadre (C , DV , K , P , *Autin*, *Rivoirard*).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^{\beta/2}$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{W}B_{2/(\beta+1),2/(\beta+1)}^{\beta/2}$ avec $\mathcal{W}B_{p,q}^s$ version faible de $B_{p,q}^s$.
- $C^\alpha \subsetneq \mathcal{A}^{2\alpha}(\text{Lin}) = B_{2,\infty}^\alpha \subsetneq C^\alpha$ par morceaux $\subsetneq \mathcal{A}^{2\alpha}(\text{NonLin}) = \mathcal{W}B_{p,p}^\alpha$ avec $p = 2/(2\alpha + 1)$.
- Estimation par seuillage dans une base d'ondelettes pour $f \in C^\alpha$ par morceaux :
$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} .$$
- Disparition du facteur $\log N$ si $f \in C^\alpha$ et utilisation de \mathcal{M} linéaire.

Fonctions C^α et ondelettes

- Maxisets bien étudiés dans ce cadre (C , DV , K , P , *Autin*, *Rivoirard*).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^{\beta/2}$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{W}B_{2/(\beta+1),2/(\beta+1)}^{\beta/2}$ avec $\mathcal{W}B_{p,q}^s$ version faible de $B_{p,q}^s$.
- $C^\alpha \subsetneq \mathcal{A}^{2\alpha}(\text{Lin}) = B_{2,\infty}^{2\alpha} \subsetneq C^\alpha$ par morceaux $\subsetneq \mathcal{A}^{2\alpha}(\text{NonLin}) = \mathcal{W}B_{p,p}^{2\alpha}$ avec $p = 2/(2\alpha + 1)$.
- Estimation par seuillage dans une base d'ondelettes pour $f \in C^\alpha$ par morceaux :

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} .$$

- Disparition du facteur $\log N$ si $f \in C^\alpha$ et utilisation de \mathcal{M} linéaire.
- Clé : approximation

$$\min_{\dim(m) \leq M} \|f - P_m f\|^2 \leq CM^{-2\alpha}$$

$$\|f - f_M\|^2 \leq CM^{-2\alpha} .$$

Base d'ondelettes 2D séparables

Base d'ondelettes 2D séparables

- La famille

$$\left\{ \begin{array}{l} \phi_{j,n_1}(x_1) \psi_{j,n_2}(x_2) \quad , \quad \psi_{j,n_1}(x_1) \phi_{j,n_2}(x_2) \\ \psi_{j,n_1}(x_1) \psi_{j,n_2}(x_2) \end{array} \right\}_{(j,n_1,n_2) \in \mathbb{Z}^3}$$

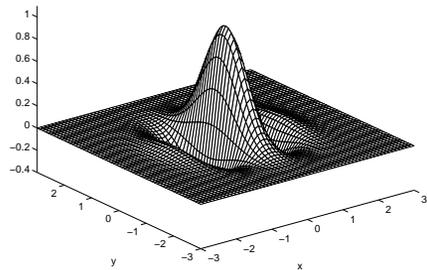
est une base orthonormée de $L^2[0, 1]^2$.

Base d'ondelettes 2D séparables

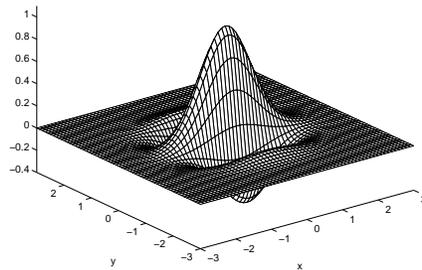
● La famille

$$\left\{ \begin{array}{l} \phi_{j,n_1}(x_1) \psi_{j,n_2}(x_2) \quad , \quad \psi_{j,n_1}(x_1) \phi_{j,n_2}(x_2) \\ \psi_{j,n_1}(x_1) \psi_{j,n_2}(x_2) \end{array} \right\}_{(j,n_1,n_2) \in \mathbb{Z}^3}$$

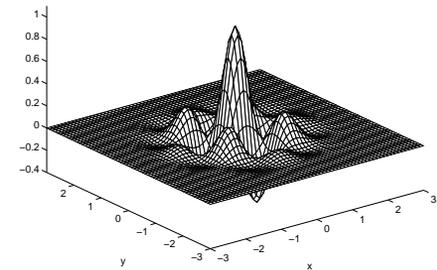
est une base orthonormée de $L^2[0, 1]^2$.



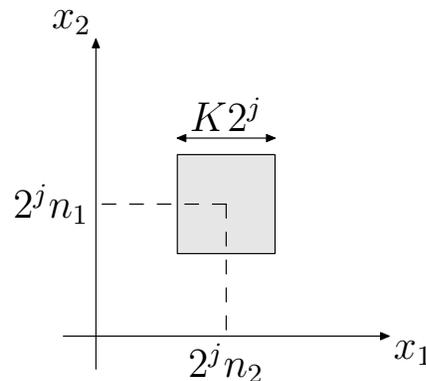
$$\phi_{j,n_1}(x_1) \psi_{j,n_2}(x_2)$$



$$\psi_{j,n_1}(x_1) \phi_{j,n_2}(x_2)$$



$$\psi_{j,n_1}(x_1) \psi_{j,n_2}(x_2)$$



Support
des ondelettes
isotropes

Fonctions C^α et ondelettes 2D

Fonctions C^α et ondelettes 2D

- Pour $f \in C^\alpha - C^\alpha$ (C^α en dehors de contours C^α) (Korostelev, Tsybakov) : vitesse minimax $(\frac{1}{N})^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).

Fonctions C^α et ondelettes 2D

- Pour $f \in C^\alpha - C^\alpha$ (C^α en dehors de contours C^α) (Korostelev, Tsybakov) : vitesse minimax $(\frac{1}{N})^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^\beta$.

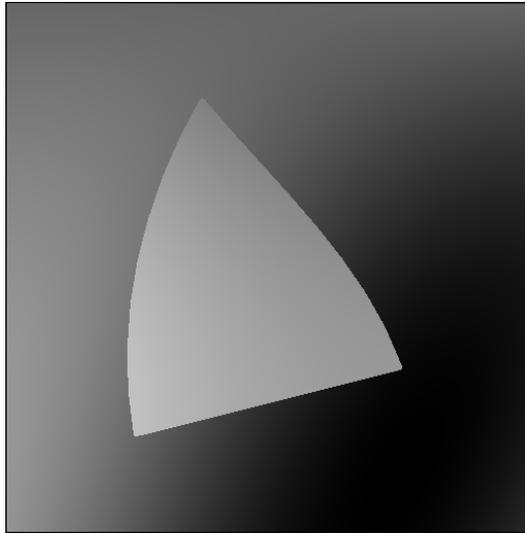
Fonctions C^α et ondelettes 2D

- Pour $f \in C^\alpha - C^\alpha$ (C^α en dehors de contours C^α) (Korostelev, Tsybakov) : vitesse minimax $(\frac{1}{N})^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^\beta$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{W}B_{2/(2\beta+1),2/(2\beta+1)}^\beta$ avec $\mathcal{W}B_{p,q}^s$ version faible de $B_{p,q}^s$.

Fonctions C^α et ondelettes 2D

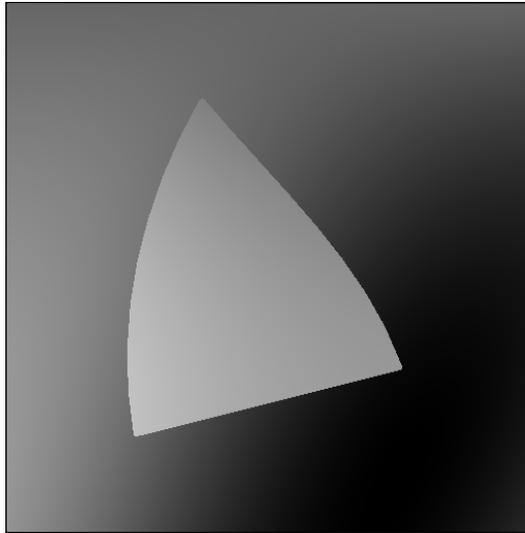
- Pour $f \in C^\alpha - C^\alpha$ (C^α en dehors de contours C^α) (Korostelev, Tsybakov) : vitesse minimax $(\frac{1}{N})^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^\beta$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{W}B_{2/(2\beta+1), 2/(2\beta+1)}^\beta$ avec $\mathcal{W}B_{p,q}^s$ version faible de $B_{p,q}^s$.
- $C^\alpha \subsetneq \mathcal{A}^\alpha(\text{Lin}) = B_{2,\infty}^\alpha \subsetneq \mathcal{A}^\alpha(\text{NonLin}) = \mathcal{W}B_{p,p}^\alpha$ avec $p = 2/(2\alpha + 1)$.

Fonctions C^α et ondelettes 2D



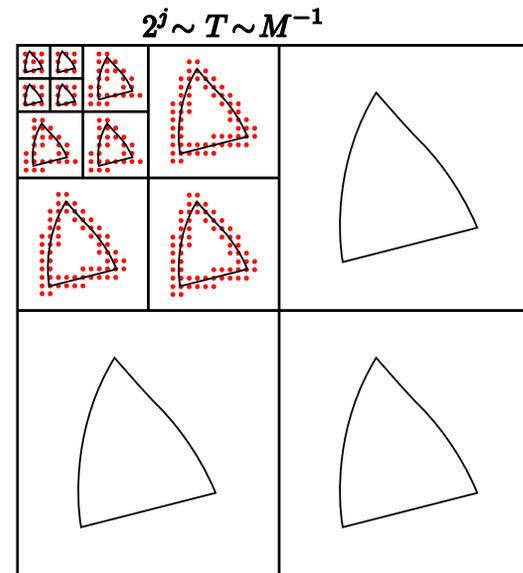
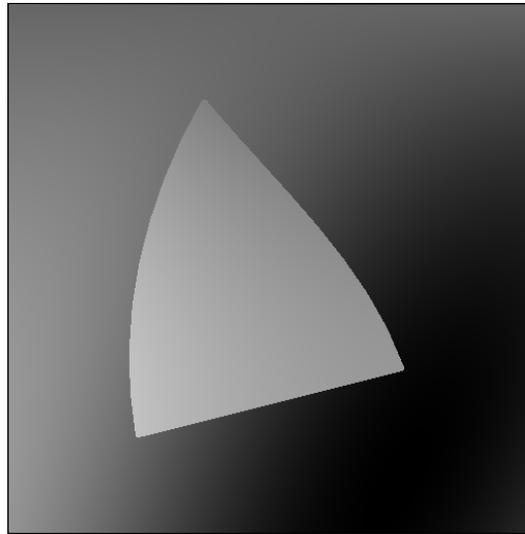
- Pour $f \in C^\alpha - C^\alpha$ (C^α en dehors de contours C^α) (Korostelev, Tsybakov) : vitesse minimax $(\frac{1}{N})^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^\beta$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{W}B_{2/(2\beta+1), 2/(2\beta+1)}^\beta$ avec $\mathcal{W}B_{p,q}^s$ version faible de $B_{p,q}^s$.
- $C^\alpha \subsetneq \mathcal{A}^\alpha(\text{Lin}) = B_{2,\infty}^\alpha \subsetneq \mathcal{A}^\alpha(\text{NonLin}) = \mathcal{W}B_{p,p}^\alpha$ avec $p = 2/(2\alpha + 1)$.
- Pour $\alpha > 1$, $C^\alpha - C^\alpha \not\subset \mathcal{A}^\alpha(\text{NonLin})$.

Fonctions C^α et ondelettes 2D



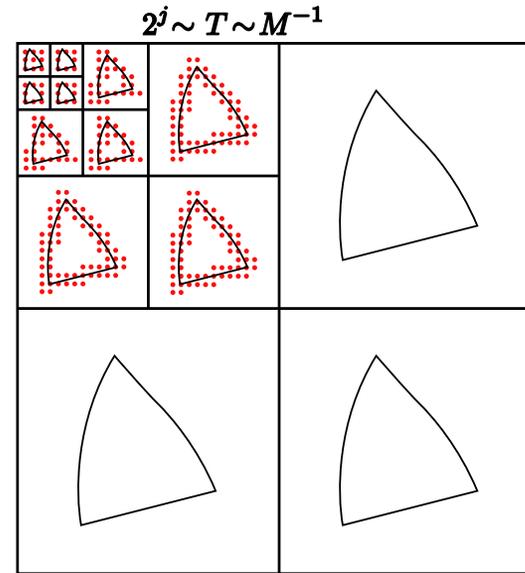
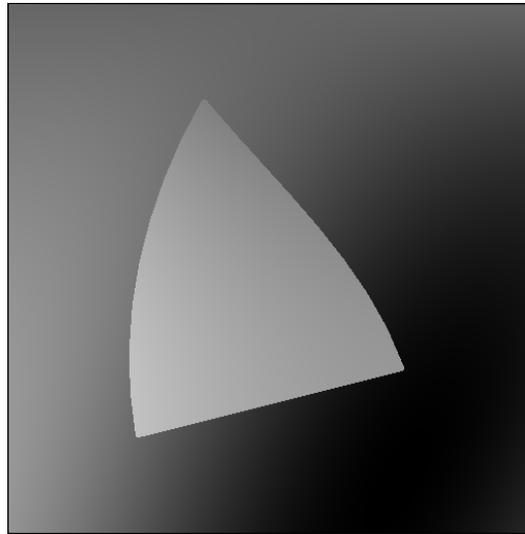
- Pour $f \in C^\alpha - C^\alpha$ (C^α en dehors de contours C^α) (Korostelev, Tsybakov) : vitesse minimax $(\frac{1}{N})^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^\beta$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{W}B_{2/(2\beta+1), 2/(2\beta+1)}^\beta$ avec $\mathcal{W}B_{p,q}^s$ version faible de $B_{p,q}^s$.
- $C^\alpha \subsetneq \mathcal{A}^\alpha(\text{Lin}) = B_{2,\infty}^\alpha \subsetneq \mathcal{A}^\alpha(\text{NonLin}) = \mathcal{W}B_{p,p}^\alpha$ avec $p = 2/(2\alpha + 1)$.
- Pour $\alpha > 1$, $C^\alpha - C^\alpha \not\subset \mathcal{A}^\alpha(\text{NonLin})$.
- Pour $\alpha > 1$, $C^\alpha - C^\alpha \subset \mathcal{A}^1(\text{NonLin})$.

Fonctions C^α et ondelettes 2D



- Pour $f \in C^\alpha - C^\alpha$ (C^α en dehors de contours C^α) (Korostelev, Tsybakov) : vitesse minimax $(\frac{1}{N})^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^\beta$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{WB}_{2/(2\beta+1), 2/(2\beta+1)}^\beta$ avec $\mathcal{WB}_{p,q}^s$ version faible de $B_{p,q}^s$.
- $C^\alpha \subsetneq \mathcal{A}^\alpha(\text{Lin}) = B_{2,\infty}^\alpha \subsetneq \mathcal{A}^\alpha(\text{NonLin}) = \mathcal{WB}_{p,p}^\alpha$ avec $p = 2/(2\alpha + 1)$.
- Pour $\alpha > 1$, $C^\alpha - C^\alpha \not\subset \mathcal{A}^\alpha(\text{NonLin})$.
- Pour $\alpha > 1$, $C^\alpha - C^\alpha \subset \mathcal{A}^1(\text{NonLin})$.
- Avec M ondelettes : $\|f - f_M\|^2 \leq C M^{-1}$.

Fonctions C^α et ondelettes 2D

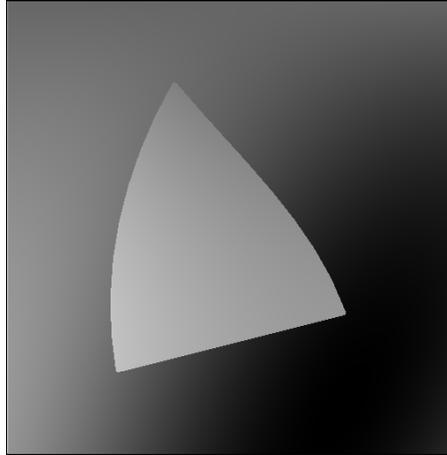


- Pour $f \in C^\alpha - C^\alpha$ (C^α en dehors de contours C^α) (Korostelev, Tsybakov) : vitesse minimax $(\frac{1}{N})^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^\beta$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{WB}_{2/(2\beta+1), 2/(2\beta+1)}^\beta$ avec $\mathcal{WB}_{p,q}^s$ version faible de $B_{p,q}^s$.
- $C^\alpha \subsetneq \mathcal{A}^\alpha(\text{Lin}) = B_{2,\infty}^\alpha \subsetneq \mathcal{A}^\alpha(\text{NonLin}) = \mathcal{WB}_{p,p}^\alpha$ avec $p = 2/(2\alpha + 1)$.
- Pour $\alpha > 1$, $C^\alpha - C^\alpha \not\subset \mathcal{A}^\alpha(\text{NonLin})$.
- Pour $\alpha > 1$, $C^\alpha - C^\alpha \subset \mathcal{A}^1(\text{NonLin})$.
- Avec M ondelettes : $\|f - f_M\|^2 \leq C M^{-1}$.
- Besoin de $\|f - f_M\|^2 \leq C M^{-\alpha}$ pour le risque minimax.

Éléments géométriques pour les contours

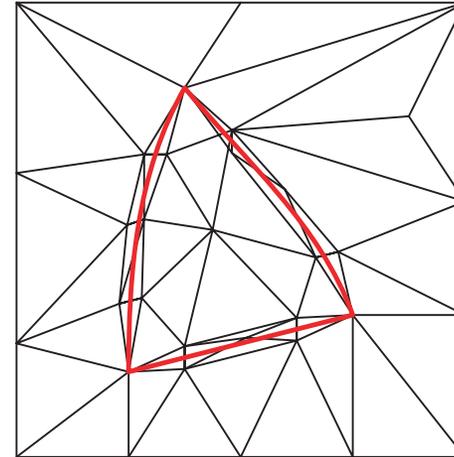
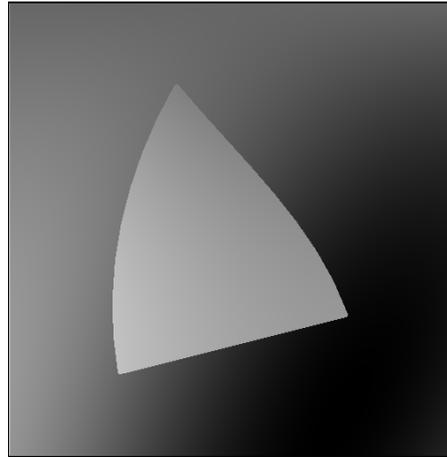
Éléments géométriques pour les contours

- Approximation de $f \in C^\alpha - C^\alpha$:

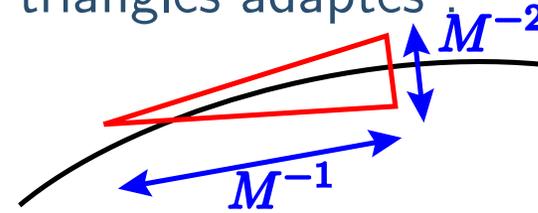


Éléments géométriques pour les contours

- Approximation de $f \in C^\alpha - C^\alpha$:

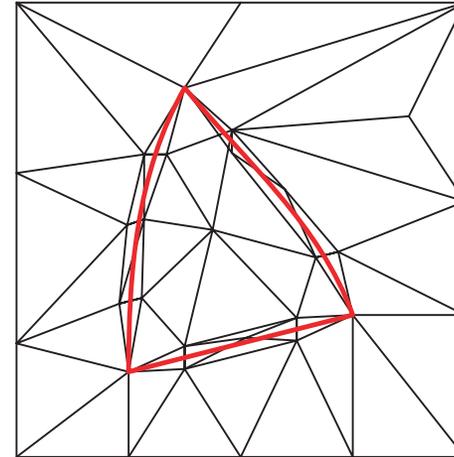
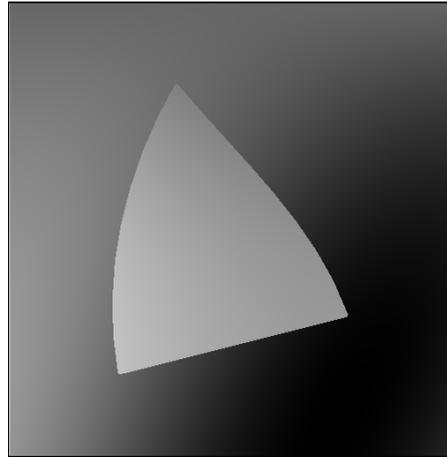


- Approximation linéaire par morceau sur M triangles adaptés :
si $\alpha \geq 2$ alors $\|f - f_M\|^2 \leq C M^{-2}$.

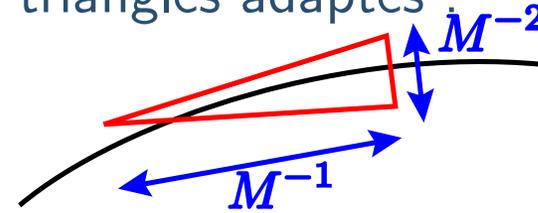


Éléments géométriques pour les contours

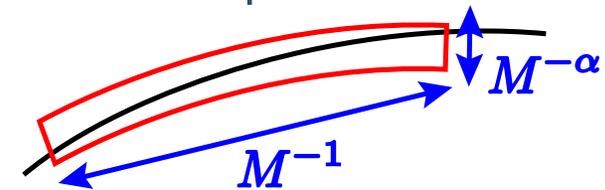
- Approximation de $f \in C^\alpha - C^\alpha$:



- Approximation linéaire par morceau sur M triangles adaptés :
si $\alpha \geq 2$ alors $\|f - f_M\|^2 \leq C M^{-2}$.

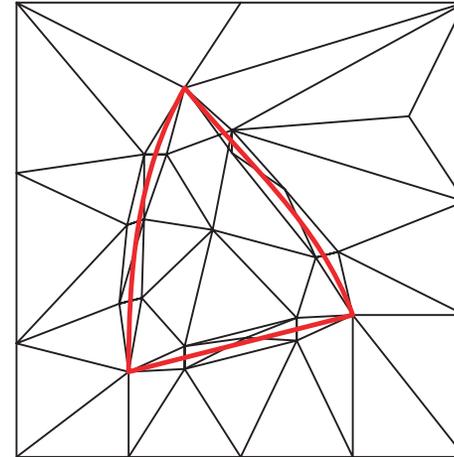
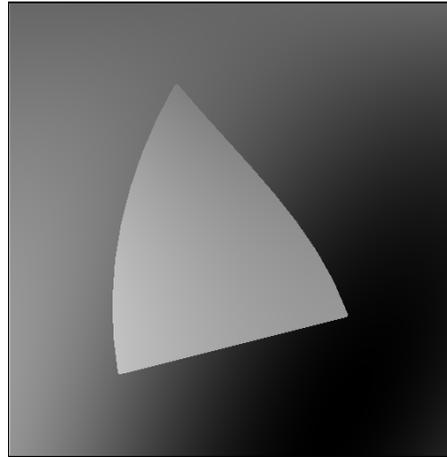


- Approximation d'ordre élevé avec M "éléments" adaptés :
 $\|f - f_M\|^2 \leq C M^{-\alpha}$.

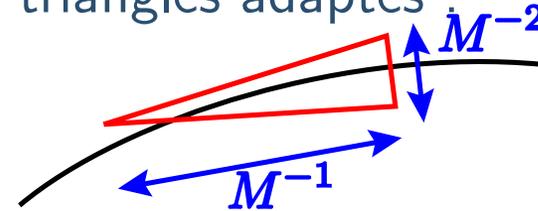


Éléments géométriques pour les contours

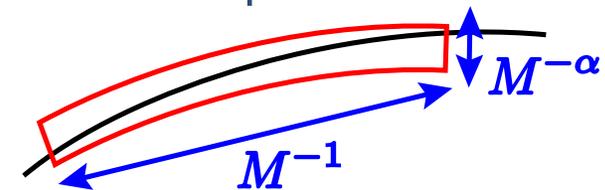
- Approximation de $f \in C^\alpha - C^\alpha$:



- Approximation linéaire par morceau sur M triangles adaptés :
si $\alpha \geq 2$ alors $\|f - f_M\|^2 \leq C M^{-2}$.



- Approximation d'ordre élevé avec M "éléments" adaptés :
 $\|f - f_M\|^2 \leq C M^{-\alpha}$.



- Pas de bases et optimisation difficile.

Curvelets

Curvelets

- Les curvelets définissent un “tight frame” de $L^2[0, 1]^2$ avec des éléments allongés et orientés (Candes, Donoho) : $\{c_j(R_\theta x - \eta)\}_{j,\theta,\eta}$.



Curvelets

- Les curvelets définissent un “tight frame” de $L^2[0, 1]^2$ avec des éléments allongés et orientés (Candes, Donoho) : $\{c_j(R_\theta x - \eta)\}_{j,\theta,\eta}$.



- Si f est $C^\alpha - C^\alpha$ alors avec M curvelets :

$$\|f - f_M\|^2 \leq C (\log M)^3 M^{-2} \quad \text{si } \alpha \geq 2.$$

Curvelets

- Les curvelets définissent un “tight frame” de $L^2[0, 1]^2$ avec des éléments allongés et orientés (Candes, Donoho) : $\{c_j(R_\theta x - \eta)\}_{j,\theta,\eta}$.



- Si f est $C^\alpha - C^\alpha$ alors avec M curvelets :

$$\|f - f_M\|^2 \leq C (\log M)^3 M^{-2} \quad \text{si } \alpha \geq 2.$$

- Quasi optimal pour $\alpha = 2$.

Curvelets

- Les curvelets définissent un “tight frame” de $L^2[0, 1]^2$ avec des éléments allongés et orientés (Candes, Donoho) : $\{c_j(R_\theta x - \eta)\}_{j,\theta,\eta}$.



- Si f est $C^\alpha - C^\alpha$ alors avec M curvelets :

$$\|f - f_M\|^2 \leq C (\log M)^3 M^{-2} \quad \text{si } \alpha \geq 2.$$

- Quasi optimal pour $\alpha = 2$.
- En pratique, estimation dans l'espace des coefficients.

Curvelets

- Les curvelets définissent un “tight frame” de $L^2[0, 1]^2$ avec des éléments allongés et orientés (Candes, Donoho) : $\{c_j(R_\theta x - \eta)\}_{j,\theta,\eta}$.



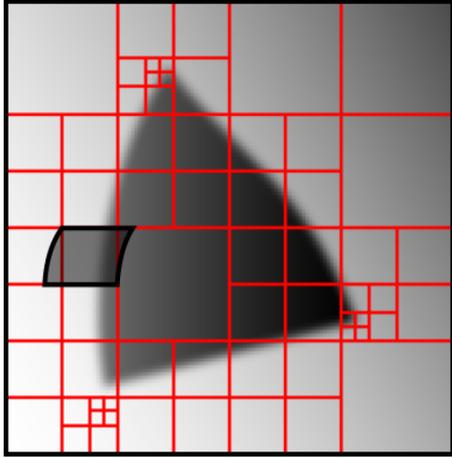
- Si f est $C^\alpha - C^\alpha$ alors avec M curvelets :

$$\|f - f_M\|^2 \leq C (\log M)^3 M^{-2} \quad \text{si } \alpha \geq 2.$$

- Quasi optimal pour $\alpha = 2$.
- En pratique, estimation dans l'espace des coefficients.
- Discrétisation complexe et difficultés pour obtenir des bases orthogonales ou des bases de Riesz.

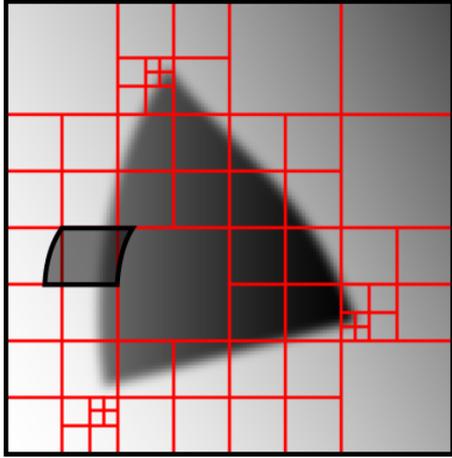
Bandelettes

Bandelettes



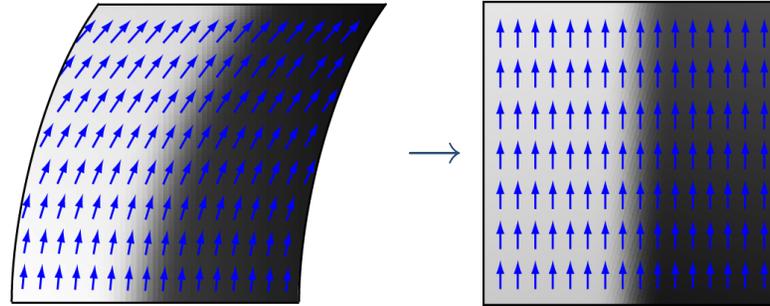
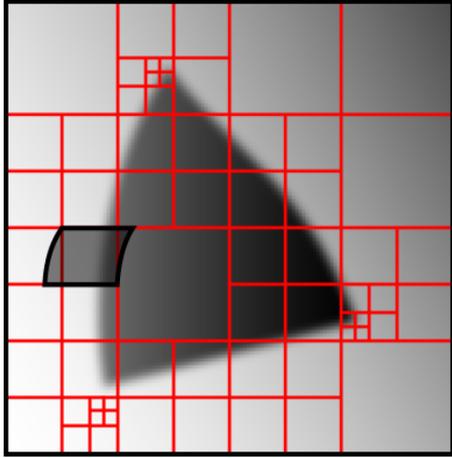
- Image $C^\alpha - C^\alpha$ simple

Bandelettes



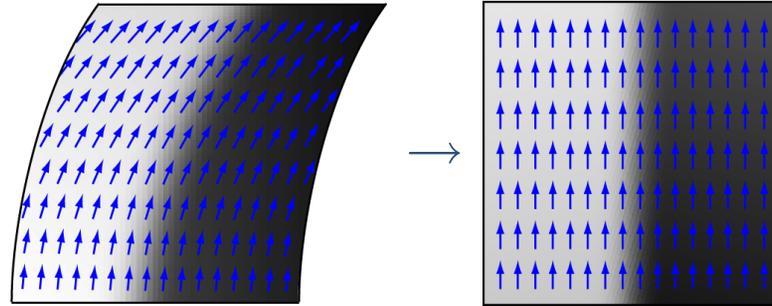
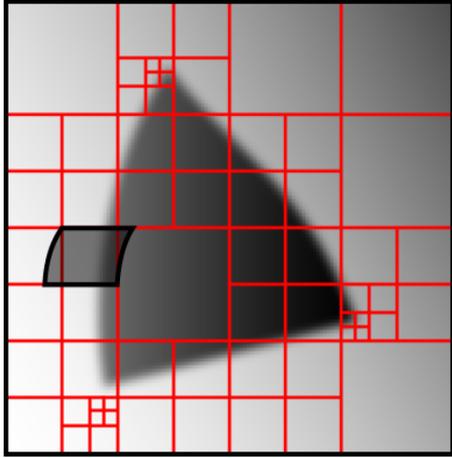
- Image $C^\alpha - C^\alpha$ simple par morceaux.

Bandelettes



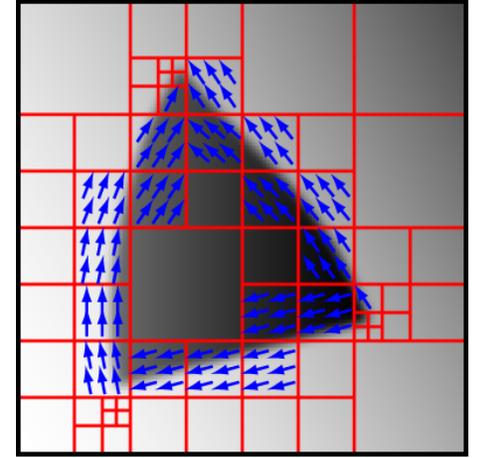
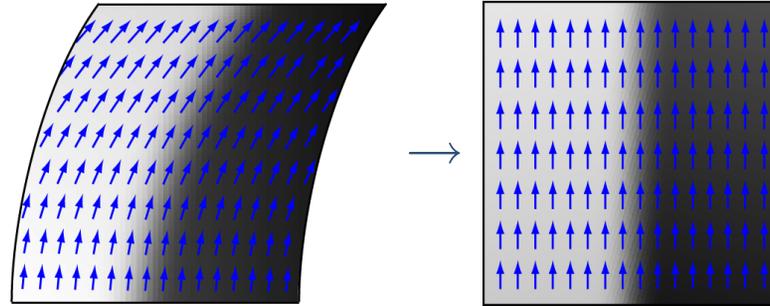
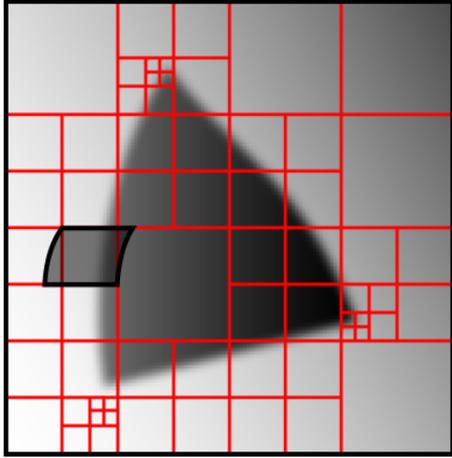
- Image $C^\alpha - C^\alpha$ simple par morceaux.
- Déformation locale \implies singularité verticale/horizontale .

Bandelettes



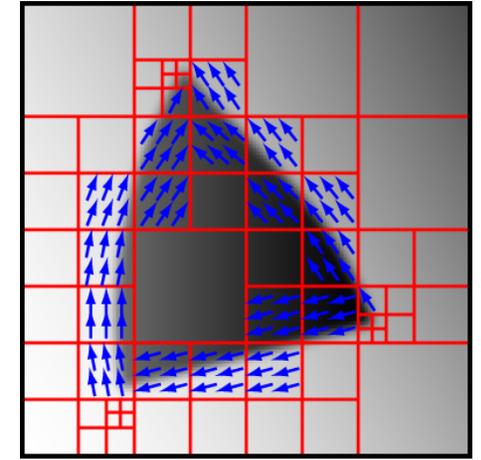
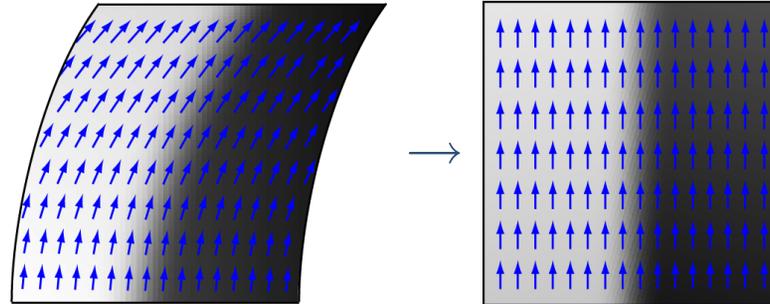
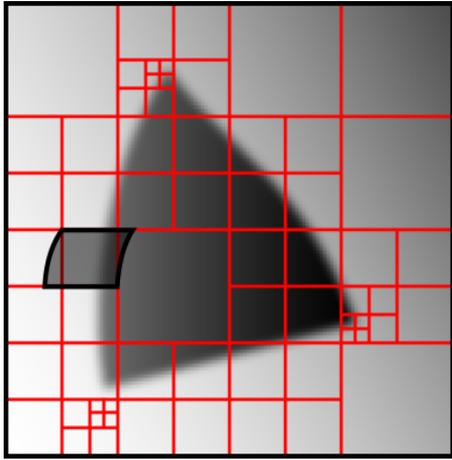
- Image $C^\alpha - C^\alpha$ simple par morceaux.
- Déformation locale \implies singularité verticale/horizontale .
- Bandelettes locales : préimage d'une base adaptée.

Bandelettes



- Image $C^\alpha - C^\alpha$ simple par morceaux.
- Déformation locale \implies singularité verticale/horizontale .
- Bandelettes locales : préimage d'une base adaptée.
- Base de bandelettes définie par :
 - une segmentation dyadique et
 - une géométrie dans chaque carré.

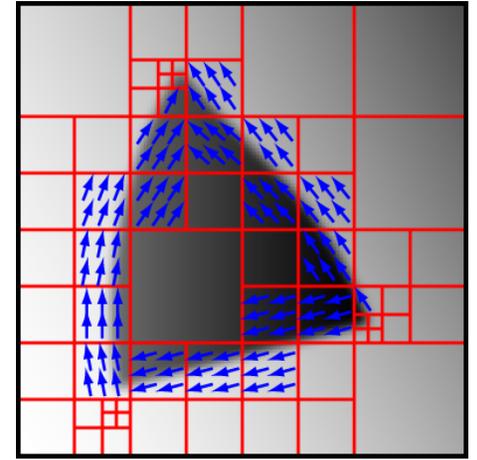
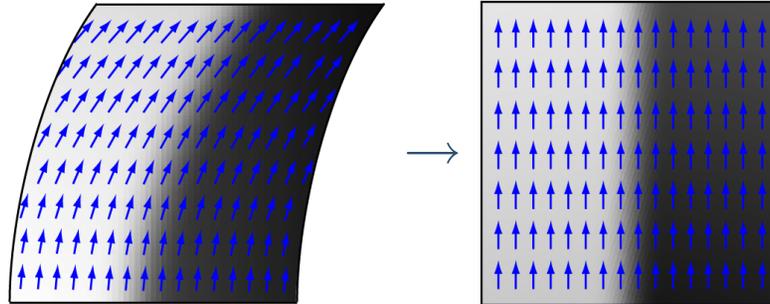
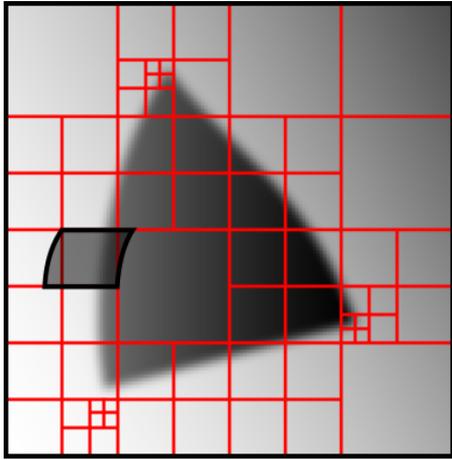
Bandelettes



- Image $C^\alpha - C^\alpha$ simple par morceaux.
- Déformation locale \implies singularité verticale/horizontale .
- Bandelettes locales : préimage d'une base adaptée.
- Base de bandelettes définie par :
 - une segmentation dyadique et
 - une géométrie dans chaque carré.
- **Théorème** : Si f est $C^\alpha - C^\alpha$, alors, dans la meilleure base,

$$\|f - f_M\|^2 \leq C(\log M)M^{-\alpha} .$$

Bandelettes



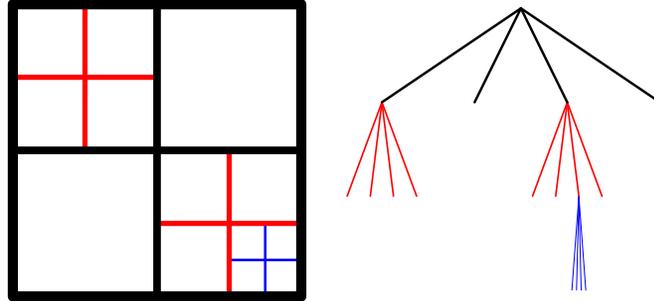
- Image $C^\alpha - C^\alpha$ simple par morceaux.
- Déformation locale \implies singularité verticale/horizontale .
- Bandelettes locales : préimage d'une base adaptée.
- Base de bandelettes définie par :
 - une segmentation dyadique et
 - une géométrie dans chaque carré.
- **Théorème** : Si f est $C^\alpha - C^\alpha$, alors, dans la meilleure base,

$$\|f - f_M\|^2 \leq C(\log M)M^{-\alpha} .$$

- Discrétisation de la géométrie donne une famille de bases avec beaucoup de vecteurs communs et un algorithme de recherche de meilleure base.

Sélection de modèles en bandelettes

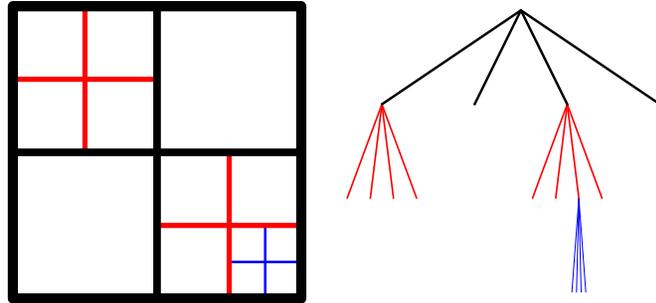
Sélection de modèles en bandelettes



- m sous-espace engendrées par des vecteurs d'une même base de bandelettes prise parmi les bases précédentes,

$$F_S = \operatorname{argmin}_{m \in \mathcal{M}} \|Y - P_m Y\|^2 + \lambda \log N \frac{\dim(m)}{N} .$$

Sélection de modèles en bandelettes

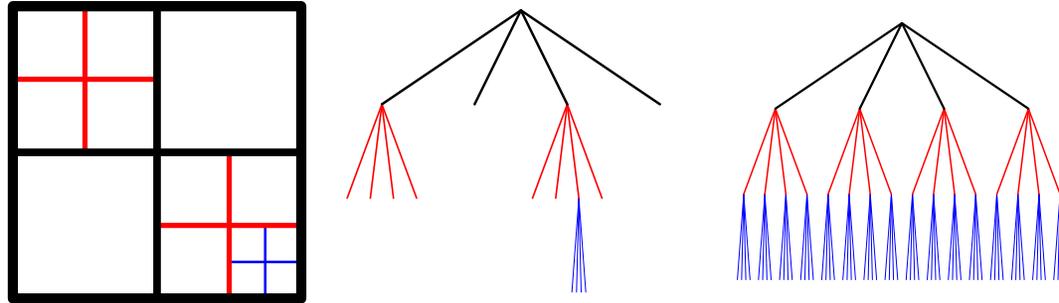


- m sous-espace engendrées par des vecteurs d'une même base de bandelettes prise parmi les bases précédentes,

$$F_S = \operatorname{argmin}_{m \in \mathcal{M}} \|Y - P_m Y\|^2 + \lambda \log N \frac{\dim(m)}{N} .$$

- Minimisation à 2 étages :
 - à base fixée, seuillage à $\sqrt{\lambda \frac{\log N}{N}}$ (facile),
 - recherche de meilleure base (difficile).

Sélection de modèles en bandelettes

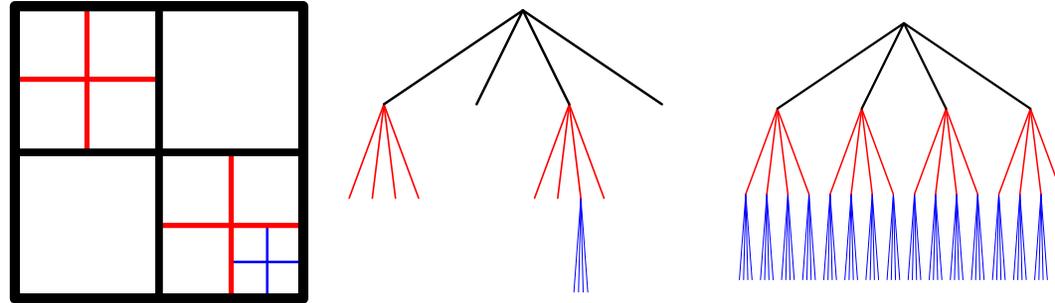


- m sous-espace engendrées par des vecteurs d'une même base de bandelettes prise parmi les bases précédentes,

$$F_S = \operatorname{argmin}_{m \in \mathcal{M}} \|Y - P_m Y\|^2 + \lambda \log N \frac{\dim(m)}{N} .$$

- Minimisation à 2 étages :
 - à base fixée, seuillage à $\sqrt{\lambda \frac{\log N}{N}}$ (facile),
 - recherche de meilleure base (difficile).
- Structure hiérarchique de la partition et additivité de la fonction à minimiser : utilisation de l'algorithme de meilleure base (*Coifman, Wickerhauser*) (CART).

Sélection de modèles en bandelettes

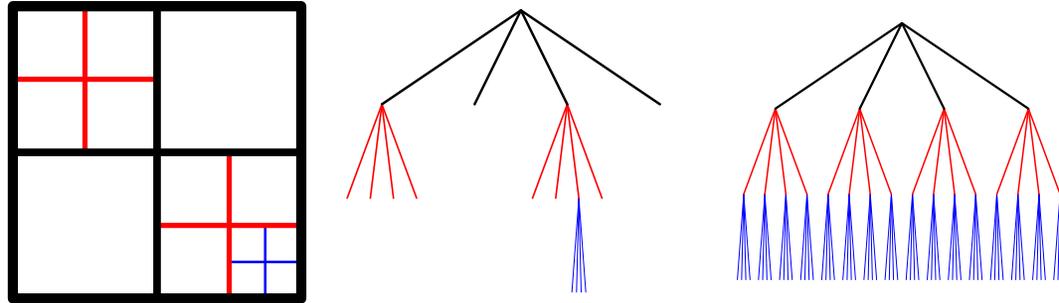


- m sous-espace engendrées par des vecteurs d'une même base de bandelettes prise parmi les bases précédentes,

$$F_S = \operatorname{argmin}_{m \in \mathcal{M}} \|Y - P_m Y\|^2 + \lambda \log N \frac{\dim(m)}{N} .$$

- Minimisation à 2 étages :
 - à base fixée, seuillage à $\sqrt{\lambda \frac{\log N}{N}}$ (facile),
 - recherche de meilleure base (difficile).
- Structure hiérarchique de la partition et additivité de la fonction à minimiser : utilisation de l'algorithme de meilleure base (*Coifman, Wickerhauser*) (CART).
- Exploration exhaustive des géométries dans chaque carré.

Sélection de modèles en bandelettes



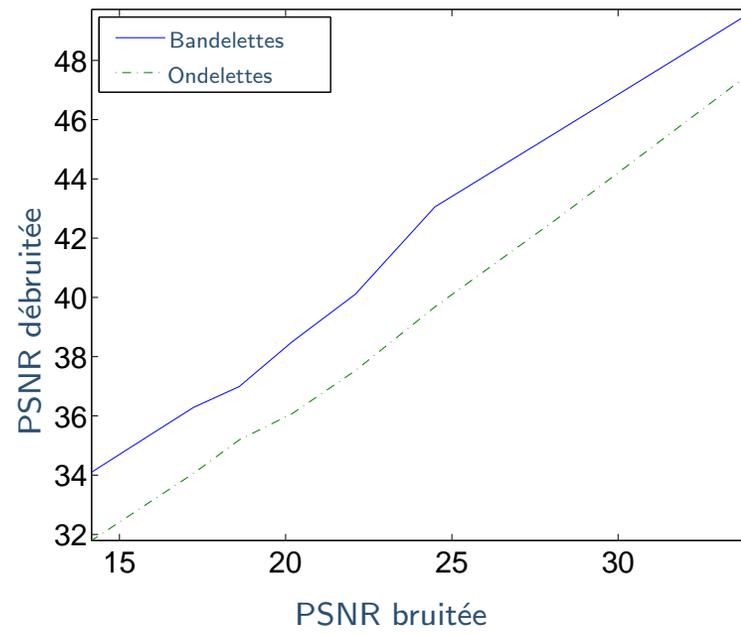
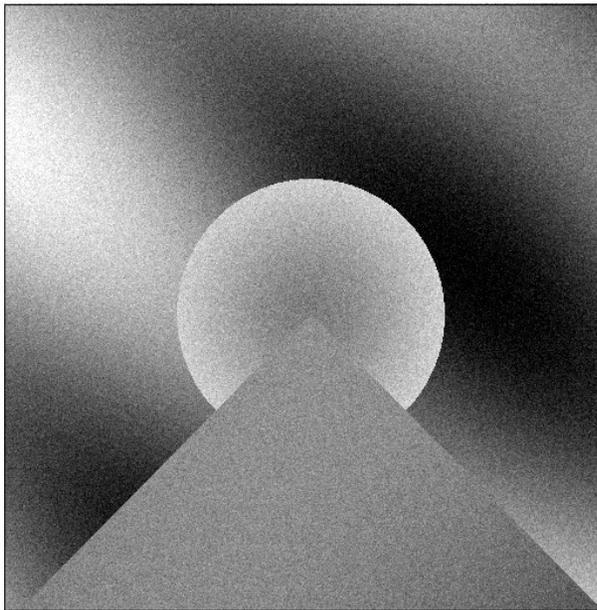
- m sous-espace engendrées par des vecteurs d'une même base de bandelettes prise parmi les bases précédentes,

$$F_S = \operatorname{argmin}_{m \in \mathcal{M}} \|Y - P_m Y\|^2 + \lambda \log N \frac{\dim(m)}{N} .$$

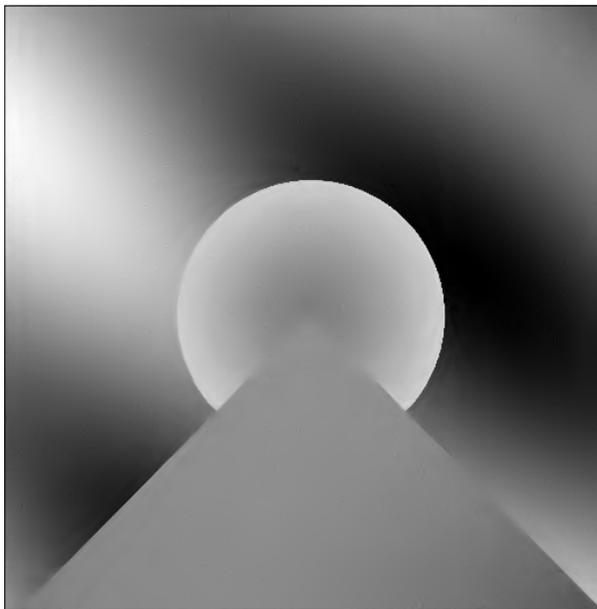
- Minimisation à 2 étages :
 - à base fixée, seuillage à $\sqrt{\lambda \frac{\log N}{N}}$ (facile),
 - recherche de meilleure base (difficile).
- Structure hiérarchique de la partition et additivité de la fonction à minimiser : utilisation de l'algorithme de meilleure base (*Coifman, Wickerhauser*) (CART).
- Exploration exhaustive des géométries dans chaque carré.
- Algorithme polynomial permettant d'atteindre le risque quasi optimal :

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{\alpha}{\alpha+1}} .$$

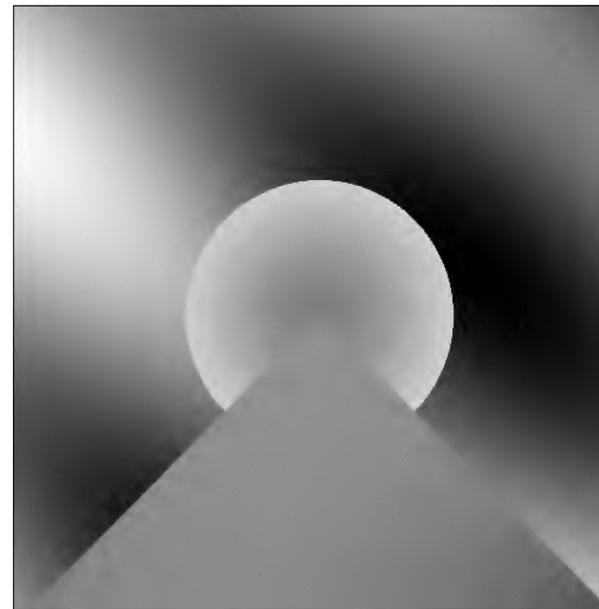
Bruitée (20,19 dB)



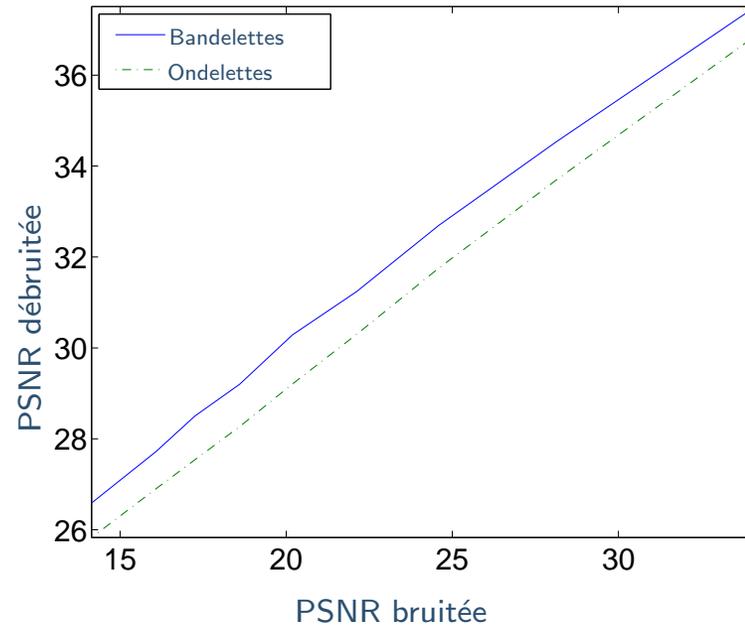
Bandelettes (30,29 dB)



Ondelettes (28,21 dB)



Bruitée (20,19 dB)



Bandelettes (30,29 dB)



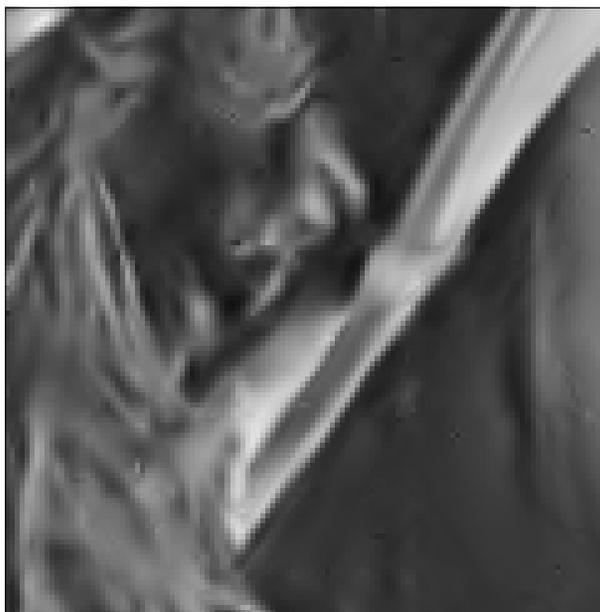
Ondelettes (28,21 dB)



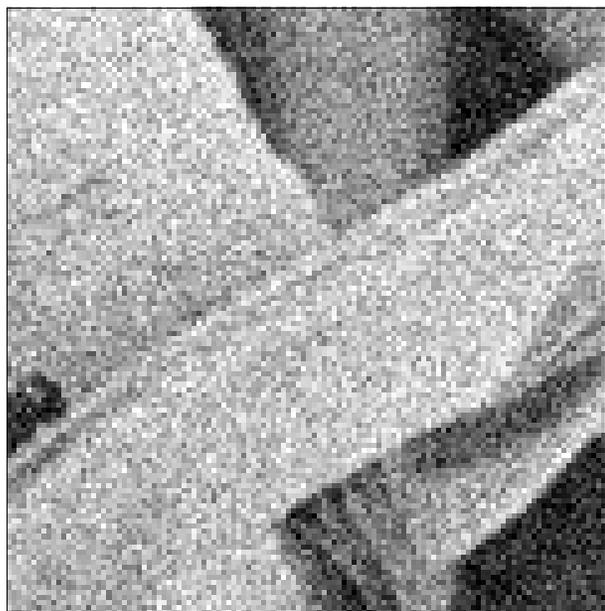
Bruitée



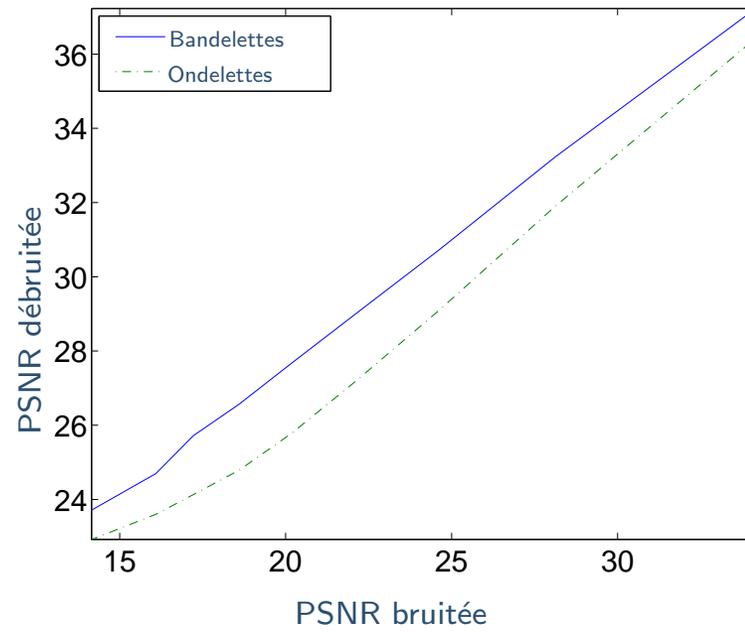
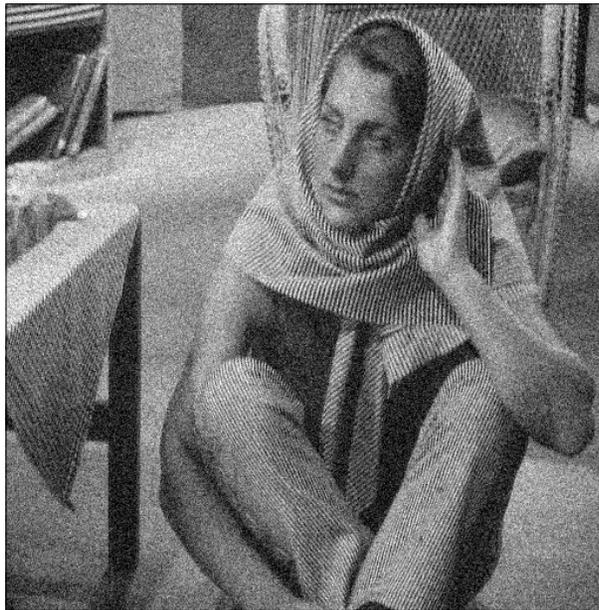
Bandelettes



Ondelettes



Bruitée (20,19 dB)



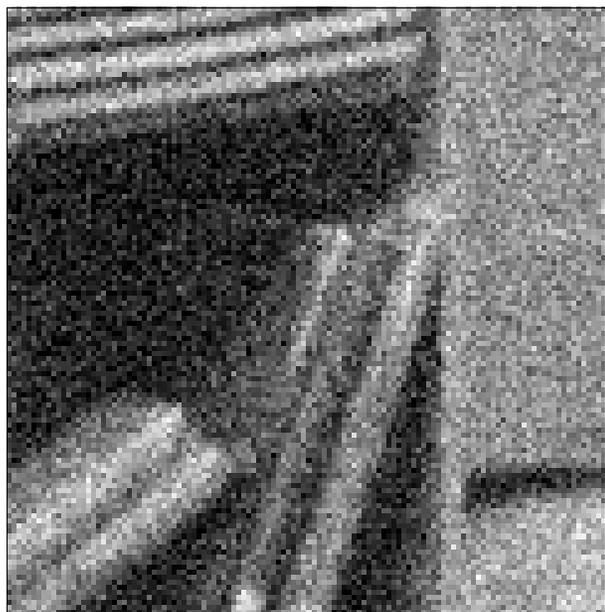
Bandelettes (27,68 dB)



Ondelettes (25,79 dB)



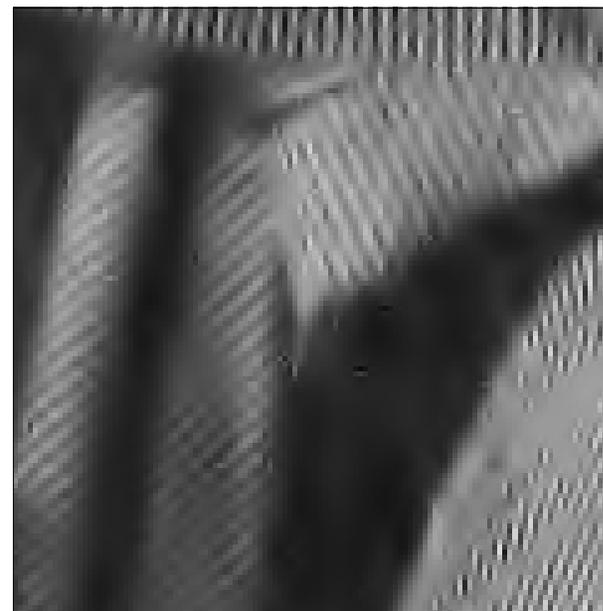
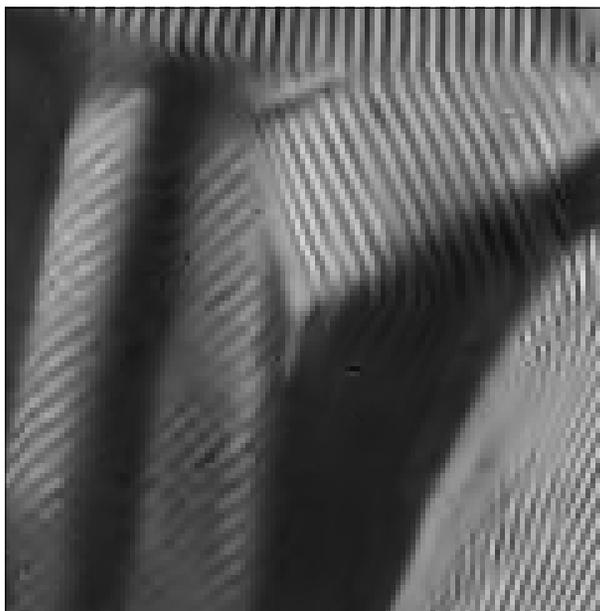
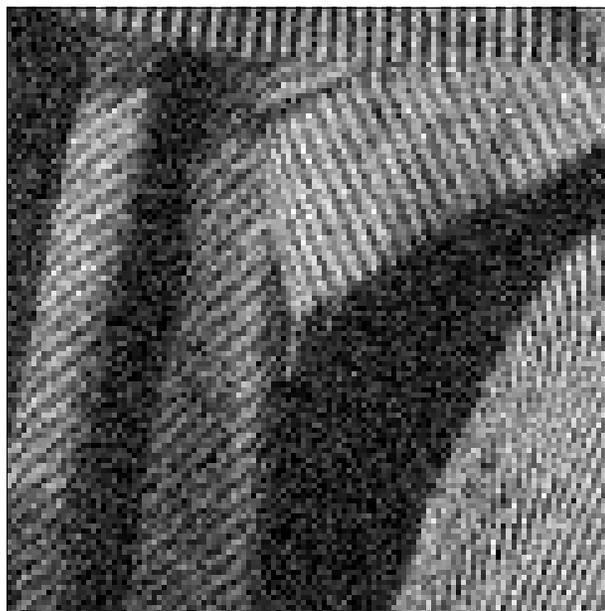
Bruitée



Bandelettes



Ondelettes



Conclusion

Conclusion

- Pour la sélection de modèles, $\text{maxiset} = \text{espace d'approximation}$.

Conclusion

- Pour la sélection de modèles, $\text{maxiset} = \text{espace d'approximation}$.
- Importance d'avoir une représentation adaptée.

Conclusion

- Pour la sélection de modèles, $\text{maxiset} = \text{espace d'approximation}$.
- Importance d'avoir une représentation adaptée.
- Bandelettes : une représentation adaptée à la géométrie des images $C^\alpha - C^\alpha$.

Conclusion

- Pour la sélection de modèles, $\text{maxiset} = \text{espace d'approximation}$.
- Importance d'avoir une représentation adaptée.
- Bandelettes : une représentation adaptée à la géométrie des images $C^\alpha - C^\alpha$.
- Passage au modèle d'échantillonnage se fait bien.

Conclusion

- Pour la sélection de modèles, $\text{maxiset} = \text{espace d'approximation}$.
- Importance d'avoir une représentation adaptée.
- Bandelettes : une représentation adaptée à la géométrie des images $C^\alpha - C^\alpha$.
- Passage au modèle d'échantillonnage se fait bien.
- Problèmes ouverts :

Conclusion

- Pour la sélection de modèles, $\text{maxiset} = \text{espace d'approximation}$.
- Importance d'avoir une représentation adaptée.
- Bandelettes : une représentation adaptée à la géométrie des images $C^\alpha - C^\alpha$.
- Passage au modèle d'échantillonnage se fait bien.
- Problèmes ouverts :
 - Caractérisation *fonctionnelle* du maxiset de l'estimateur en bandelettes ?

Conclusion

- Pour la sélection de modèles, maxiset = espace d'approximation.
- Importance d'avoir une représentation adaptée.
- Bandelettes : une représentation adaptée à la géométrie des images $C^\alpha - C^\alpha$.
- Passage au modèle d'échantillonnage se fait bien.
- Problèmes ouverts :
 - Caractérisation *fonctionnelle* du maxiset de l'estimateur en bandelettes ?
 - Maxiset pour des pénalisations différentes (l^1, \dots)

Conclusion

- Pour la sélection de modèles, maxiset = espace d'approximation.
- Importance d'avoir une représentation adaptée.
- Bandelettes : une représentation adaptée à la géométrie des images $C^\alpha - C^\alpha$.
- Passage au modèle d'échantillonnage se fait bien.
- Problèmes ouverts :
 - Caractérisation *fonctionnelle* du maxiset de l'estimateur en bandelettes ?
 - Maxiset pour des pénalisations différentes (l^1, \dots)
 - Quelle représentation pour quel problème ou quel problème pour quelle représentation ?