

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

E. Le Pennec / LPMA / Université Paris Diderot (Paris 7)

F. Autin, J.-M. Loubes, V. Rivoirard,
S. Mallat, Ch. Dossal et G. Peyré

8 Janvier 2007

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

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

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

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

- 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, quelle base donne $\Theta \subset \mathcal{A}^\beta$ avec β optimal ($(\frac{1}{N})^{\frac{\beta}{\beta+1}}$ vitesse minimax).
- Maxiset : pour une base fixée, quel est l'ensemble des fonctions estimées avec une vitesse $(\frac{1}{N})^{\frac{\beta}{\beta+1}}$? Ici \mathcal{A}^β .

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

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

$$E(\|f - F_S\|^2) \leq C \min_{\Gamma} \|f - f_{\Gamma}\|^2 + \lambda^2 \frac{\log N}{N} |\Gamma| + \frac{1}{N} \quad \text{plus fin.}$$

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

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

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

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

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 $\lambda_N = \lambda \log N$:

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

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

$$E(\|f - F_S\|^2) \leq C \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + \lambda \frac{\log N}{N} \dim(m) + \frac{1}{N} \quad .$$

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

$$E(\|f - F_S\|^2) \leq C \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + (C_1 + C_2 \lambda_{N,m}) \frac{\dim(m)}{N} + \frac{1}{N} .$$

- **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

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

● Sens inverse :

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

● Non, mais

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

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

$$\implies f \in \mathcal{A}^\beta$$

● 3 cas (par ordre de complexité) :

- Seuillage (*Cohen, DeVore, Picard, Kerkycharian*),
- Modèles emboîtés,
- Cas général...

Le cas du seuillage – 1

● Clé : $\operatorname{argmin}_{P_m f, m \in \mathcal{M}_N} \|f - P_m f\|^2 + \lambda_N \frac{\dim(m)}{N} = f_T$ avec $T^2 = \frac{\lambda_N}{N}$.

●
$$\|f - F_S\|^2 = \sum_{\substack{|\langle Y, b_k \rangle| \geq T \\ |\langle f, b_k \rangle| \geq T/2}} \left| \left\langle \frac{1}{\sqrt{N}} W, b_k \right\rangle \right|^2 + \sum_{\substack{|\langle Y, b_k \rangle| \geq T \\ |\langle f, b_k \rangle| < T/2}} \left| \left\langle \frac{1}{\sqrt{N}} W, b_k \right\rangle \right|^2$$

$$+ \sum_{\substack{|\langle Y, b_k \rangle| < T \\ |\langle f, b_k \rangle| \geq T/2}} |\langle f, b_k \rangle|^2 + \sum_{\substack{|\langle Y, b_k \rangle| < T \\ |\langle f, b_k \rangle| < T/2}} |\langle f, b_k \rangle|^2$$

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

● D'où

$$\|f - f_{T/2}\|^2 \leq E(\|f - F_S\|^2) \leq CT^{\frac{2\beta}{\beta+1}} .$$

● Erreur d'approximation \neq compromis erreur d'approximation et nombre de coefficients... :

$$\|f - f_{T/2}\|^2 + T^2 M \geq \|f - f_{T/2}\|^2 .$$

Le cas du seuillage – 2

● Mais

$$\|f - f_{T/2}\|^2 \leq CT^{\frac{2\beta}{\beta+1}} \Rightarrow \|f - f_{T/2}\|^2 + T^2 M \leq C_\beta T^{\frac{2\beta}{\beta+1}} \quad !$$

● Explication :

$$\|f - f_{T/2}\|^2 \geq \text{Card} \{ \text{coeff} \in]T/2, T] \} T^2 \quad .$$

● D'où :

$$\|f - f_{T/2}\|^2 \leq CT^{\frac{2\beta}{\beta+1}} \implies \text{Card} \{ |\text{coeff}| \in]T/2, T] \} \leq CT^{\frac{-2}{\beta+1}} \quad .$$

● On en déduit :

$$\text{Card} \{ |\text{coeff}| \in]0, T] \} \leq C_\beta T^{\frac{-2}{\beta+1}} \quad .$$

● Ou encore :

$$T^2 M \leq C_\beta T^{\frac{2\beta}{\beta+1}} \quad \dots$$

Cas des modèles emboîtés – 1

- \hat{m} : modèle sélectionné qui minimise

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

- m_O : modèle oracle qui minimise

$$\|f - P_m f\|^2 + \lambda_N \frac{\dim(m)}{4N} .$$

- On va montrer que

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

- Première étape :

$$\begin{aligned} \|f - P_{\hat{m}} Y\|^2 &= \|f - P_{\hat{m}} f\|^2 + \|P_{\hat{m}} f - P_{\hat{m}} Y\|^2 \\ &= \|f - P_{m_O} f\|^2 \\ &\quad + (\|P_{\hat{m}}(f - Y)\|^2 + \|f - P_{\hat{m}} f\|^2 - \|f - P_{m_O} f\|^2) . \end{aligned}$$

- Reste à prouver que la parenthèse est positive...

Cas des modèles emboîtés – 2

● $\|P_{\hat{m}}(f - Y)\|^2 \geq \|f - P_{m_O}f\|^2 - \|f - P_{\hat{m}}f\|^2 \quad ?$

● Si $\hat{m} \subset m_O$:

$$\|f - P_{m_O}f\|^2 - \|f - P_{\hat{m}}f\|^2 \leq 0 \quad .$$

● Sinon $\hat{m} \supset m_O$ et on est ramené à montrer que

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

● Or par définition de \hat{m} et m_O

$$\|P_{\hat{m} \setminus m_O}f\|^2 \leq \frac{1}{4} \frac{\lambda_N}{N} (\dim(\hat{m}) - \dim(m_O)) \quad \text{et} \quad \|P_{\hat{m} \setminus m_O}Y\|^2 \geq \frac{\lambda_N}{N} (\dim(\hat{m}) - \dim(m_O)) \quad .$$

● Par inégalité triangulaire :

$$\|P_{\hat{m} \setminus m_O}(Y - f)\| \geq \|P_{\hat{m} \setminus m_O}Y\| - \|P_{\hat{m} \setminus m_O}f\| \geq \frac{1}{2} \sqrt{\frac{\lambda_N}{N} (\dim(\hat{m}) - \dim(m_O))}$$

$$\|P_{\hat{m} \setminus m_O}(Y - f)\| \geq \|P_{\hat{m} \setminus m_O}f\| \quad .$$

Cas des modèles emboîtés – 3

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

et donc

$$\|f - P_{\hat{m}}Y\|^2 \leq CT^{\frac{2\beta}{\beta+1}} \implies \|f - P_{m_O}f\|^2 \leq CT^{\frac{2\beta}{\beta+1}} .$$

● Il reste à montrer que

$$\|f - P_{m_O}f\|^2 + \frac{1}{4}T^2 \dim(m_O) \leq C_\beta T^{\frac{2\beta}{\beta+1}} \dots$$

● m_O dépend de T : $m_O(T/2)$ n'est pas optimal pour T mais pour $T/2$.

$$\begin{aligned} \|f - P_{m_O(T)}f\|^2 + \lambda_N \frac{\dim(m_O(T))}{4N} &\leq \|f - P_{m_O(T/2)}f\|^2 + \lambda_N \frac{\dim(m_O(T/2))}{4N} \\ \dim(m_O(T)) - \dim(m_O(T/2)) &\leq \frac{4N}{\lambda_N} (\|f - P_{m_O(T/2)}f\|^2 - \|f - P_{m_O(T)}f\|^2) \\ \dim(m_O(T)) - \dim(m_O(T/2)) &\leq \frac{4N}{\lambda_N} C_\beta (T/2)^{\frac{2\beta}{\beta+1}} \end{aligned}$$

● En sommant sur les $T/2^k$, on conclut.

● Rq : le seuillage est un cas particulier (somme directe de modèles emboîtés).

Cas général – 1

- \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

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

Cas général – 2

- 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,

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

- On en déduit

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

- Rq : $E(\|f - P_{\hat{m}}f\|^2) \leq E(\|f - P_{\hat{m}}Y\|^2)$.

Cas général – 3

- On montre alors que

$$E(\|f - P_{\hat{m}} Y\|^2) \leq C \left(\frac{\lambda_N}{N} \right)^{\frac{\beta}{\beta+1}} \implies \frac{\dim(m_O)}{KN} \leq C_2 \left(\frac{\lambda_N}{N} \right)^{\frac{\beta}{\beta+1}} .$$

- Introduction de la dépendance en N du modèle oracle + une hypothèse de croissance pour λ_N et de structure des modèles donne le résultat , pour K et C_2 assez grand !
- 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

- 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

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

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

- Seuillage dans la base de Fourier :

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

- Vitesse quasi optimale ! (Rq : disparition du log pour des modèles emboîtés)
- Approche maxiset pour les vitesses $(\frac{\log N}{N})^{\frac{2\alpha}{2\alpha+1}}$:

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} \Leftrightarrow f \in \mathcal{A}^{2\alpha} = \mathcal{W}H^\alpha$$

avec $\mathcal{W}H^\alpha$ version faible de H^α .

- Minimax-Maxiset :

$$\forall f \in C^\alpha, E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} \Leftrightarrow C^\alpha \subset \mathcal{W}H^\alpha .$$

Fonctions C^α par morceaux et Fourier

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

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

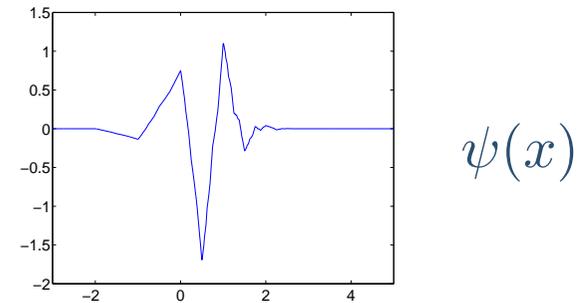
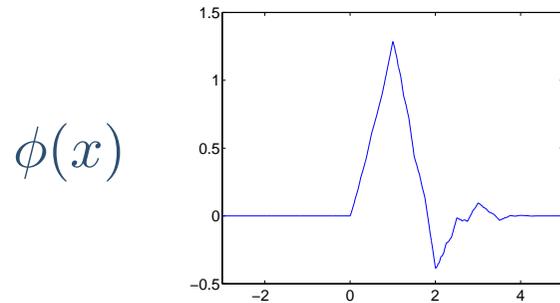
- Seuillage dans la base de Fourier ($\alpha > 1$) :

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

- Maxiset : C^α par morceaux $\subset \mathcal{A}^2 = \mathcal{W}H^1$.
- mais C^α par morceaux $\not\subset \mathcal{A}^{2\alpha} = \mathcal{W}H^\alpha$.
- Pour obtenir la vitesse minimax, C^α par morceaux $\subset \mathcal{A}^{2\alpha}$.
- Besoin d'autres bases pour atteindre la vitesse minimax !

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

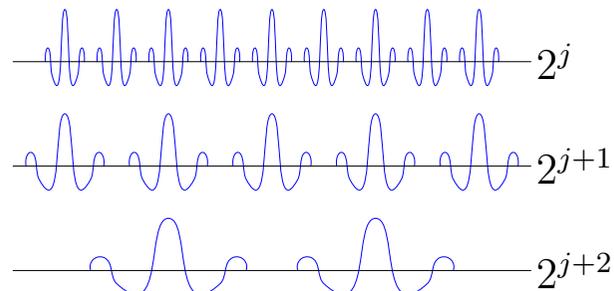
- Construite à partir d'une fonction d'échelle $\phi(x)$ et d'une ondelette mère $\psi(x)$



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

- 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 $B_{p,q}^s$.
- $C^\alpha \subsetneq \mathcal{A}^{2\alpha}(\text{Lin}) = B_{2,\infty}^\alpha \subsetneq C^\alpha \text{ par morceaux} \subsetneq \mathcal{A}^{2\alpha}(\text{NonLin}) = \mathcal{W}B_{p,p}^\alpha$ 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.
- Clé : approximation

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

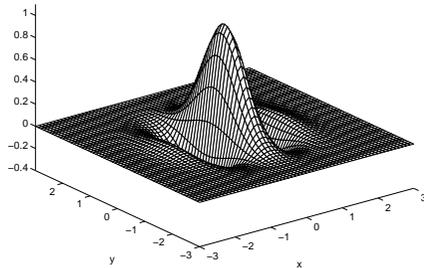
$$\|f - f_M\|^2 \leq CM^{-2\alpha} .$$

Base d'ondelettes 2D séparables

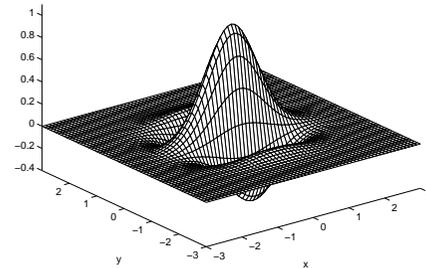
● La famille

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

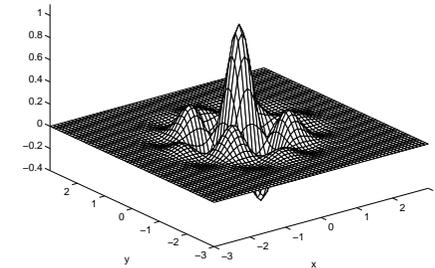
est une base orthonormée de $L^2[0, 1]^2$.



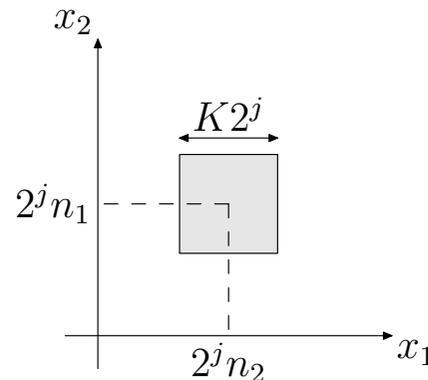
$$\phi_{j,n_1}(x_1) \psi_{j,n_2}(x_2)$$



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

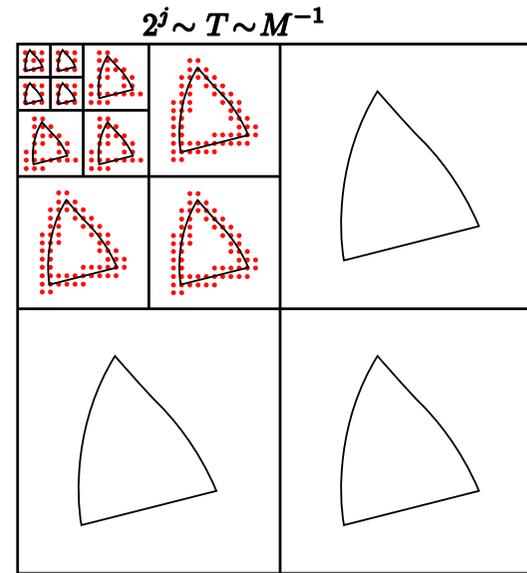
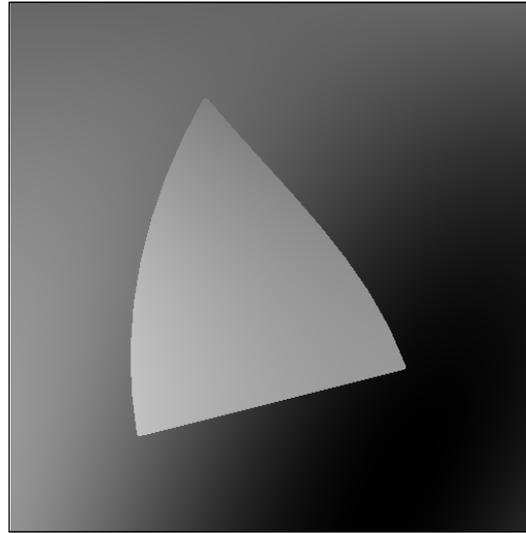


$$\psi_{j,n_1}(x_1) \psi_{j,n_2}(x_2)$$



Support
des ondelettes
isotropes

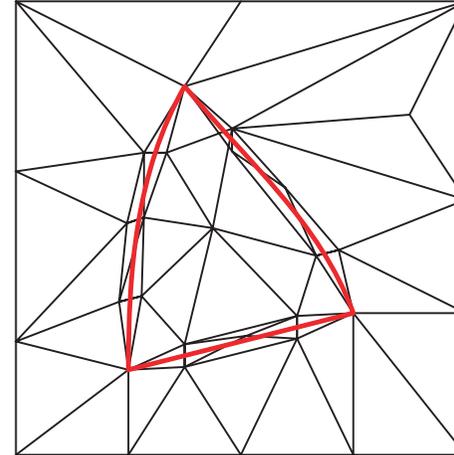
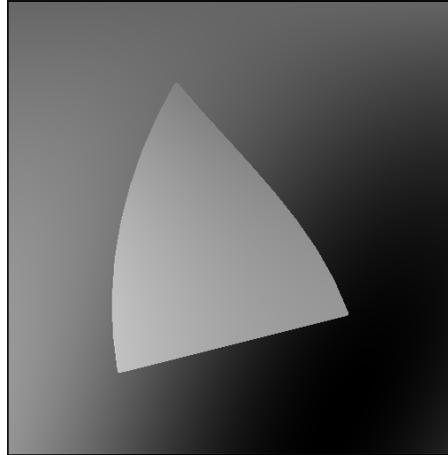
Fonctions C^α et ondelettes 2D



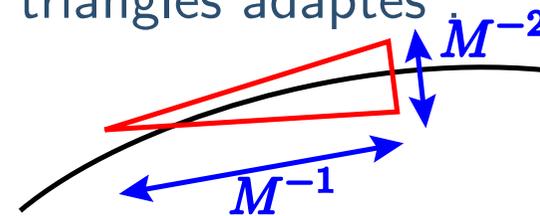
- Pour $f \in C^\alpha - C^\alpha$ (C^α en dehors de contours C^α) (Korostelev, Tsybakov) :
vitesse minimax $(\frac{1}{N})^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- $\mathcal{A}^\beta(\text{Lin}) = B_{2,\infty}^\beta$.
- $\mathcal{A}^\beta(\text{NonLin}) = \mathcal{W}B_{2/(2\beta+1), 2/(2\beta+1)}^\beta$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

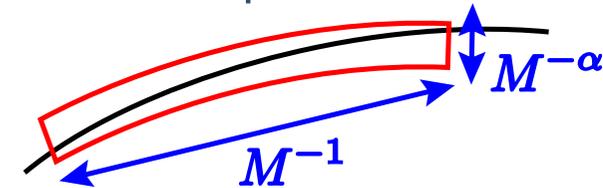
- Approximation de $f \in C^\alpha - C^\alpha$:



- 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

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

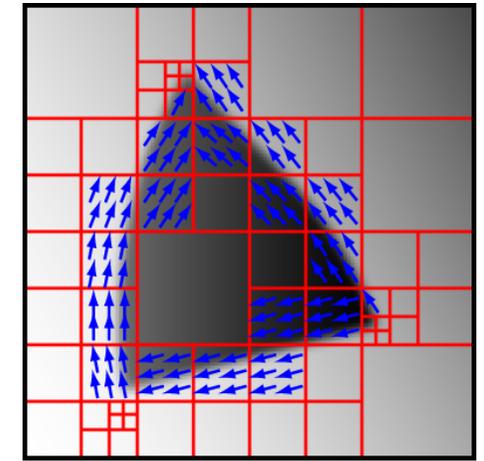
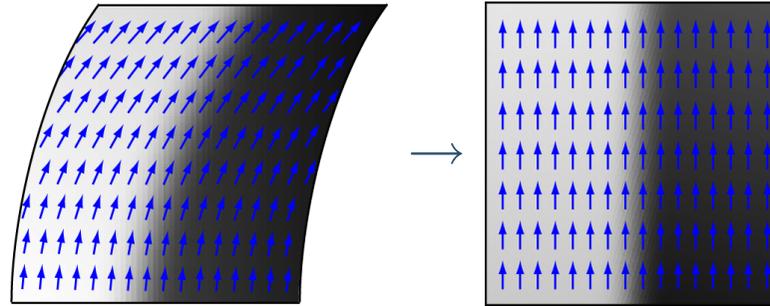
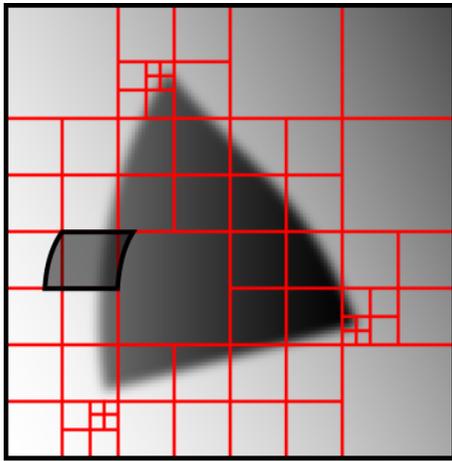


- Si f est $C^\alpha - C^\alpha$ alors avec M curvelets :

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

- 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



- 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.

Estimation géométrique

- Contrôle polynomial en fonction de N sur le nombre total de bandelettes.
- Sélection de bandelettes :

$$F_S = \operatorname{argmin} \|Y - P_{\mathcal{M}}Y\|^2 + \lambda \frac{\log N}{N} \dim(\mathcal{M})$$

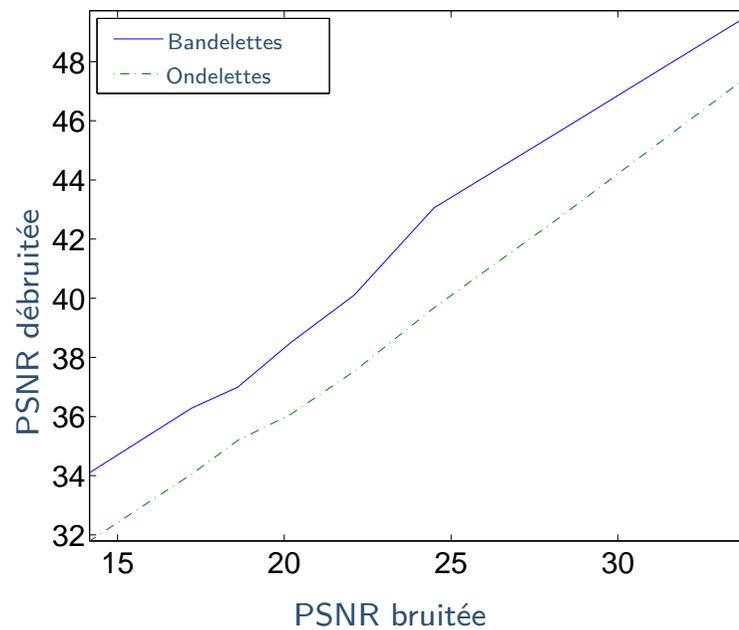
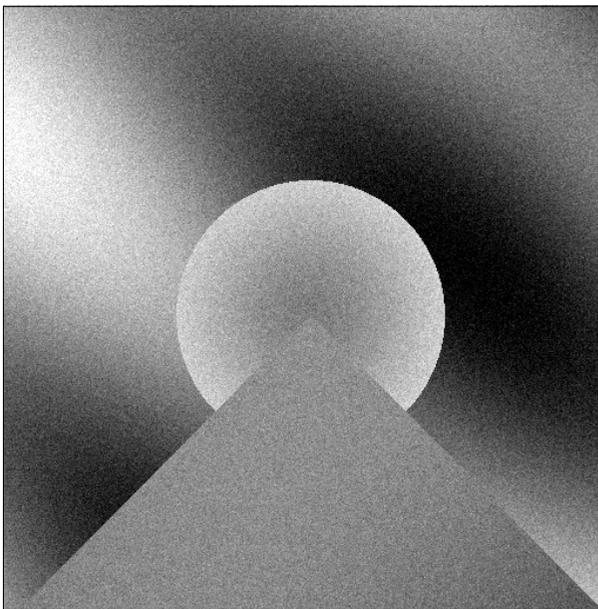
avec \mathcal{M} qui contient les sous-espaces des bases de bandelettes.

- Minimisation à 2 étages :
 - à base fixée, seuillage (facile),
 - recherche de meilleure base (difficile).
- Structure hiérarchique de la partition et additivité de la fonction à minimiser :
algorithme de meilleure base de Wickerhauser (CART).
- Exploration exhaustive des géométries dans chaque carré.
- Quasi optimalité : si $f \in \mathbf{C}^\alpha - \mathbf{C}^\alpha$ alors

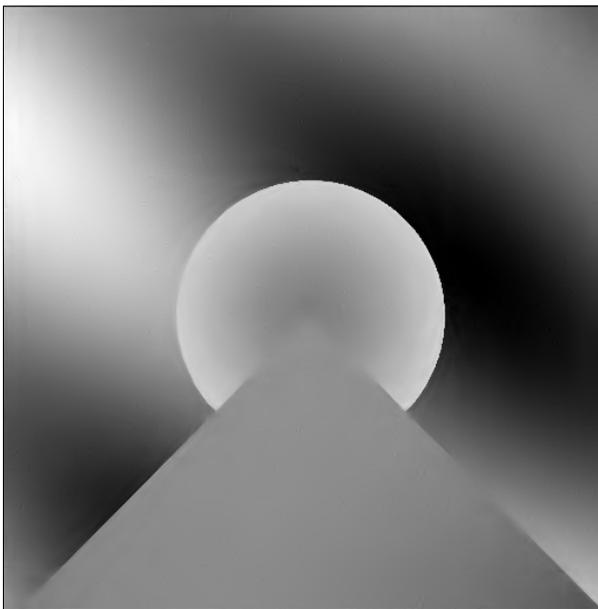
$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{\alpha}{\alpha+1}} .$$

- Maxiset :
$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{\alpha}{\alpha+1}} \Leftrightarrow f \in \mathcal{A}^\alpha \Leftrightarrow \forall M, \exists \mathcal{B}, \|f - f_M\|^2 \leq CM^{-\alpha} .$$

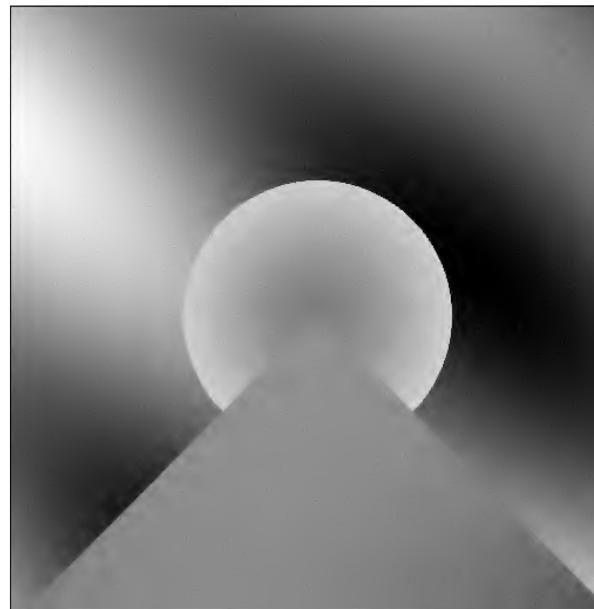
Bruitée (20,19 dB)



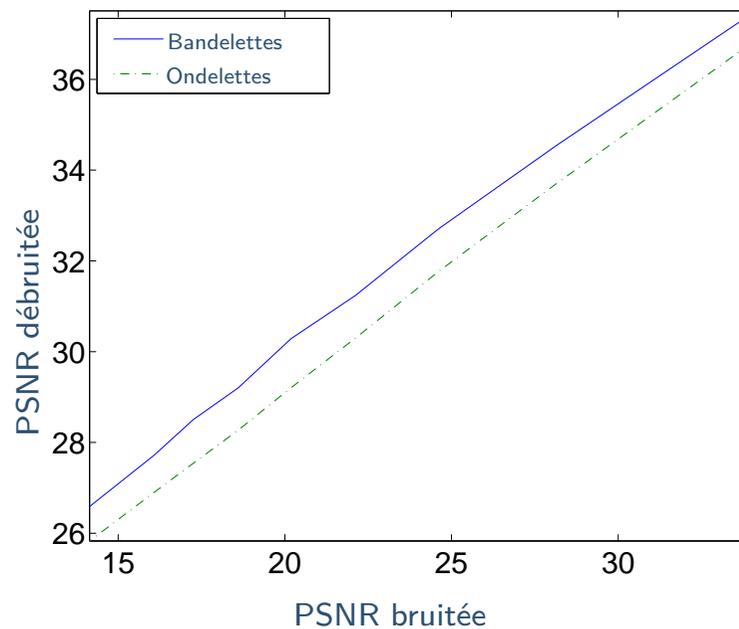
Bandelettes (30,29 dB)



Ondelettes (28,21 dB)



Bruitée (20,19 dB)



Bandelettes (30,29 dB)



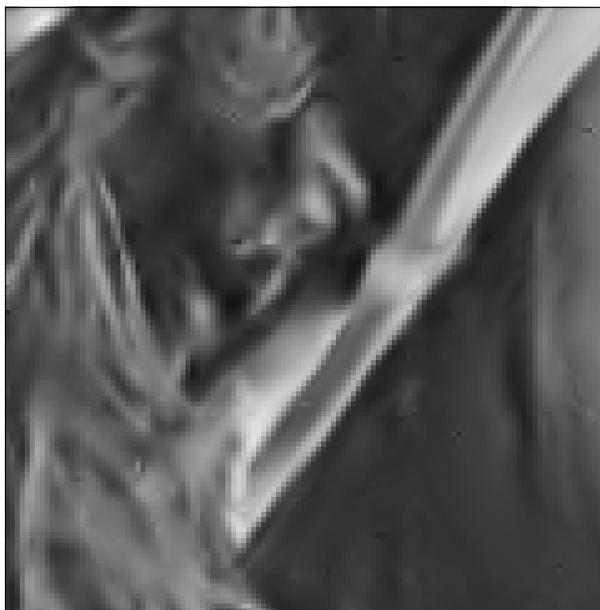
Ondelettes (28,21 dB)



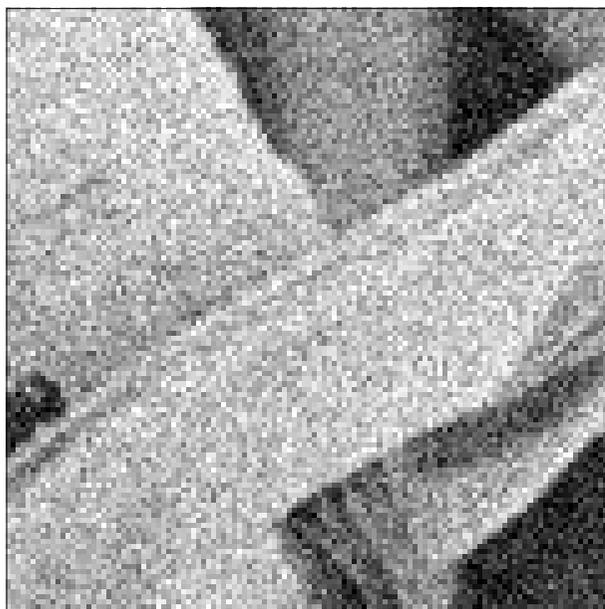
Bruitée



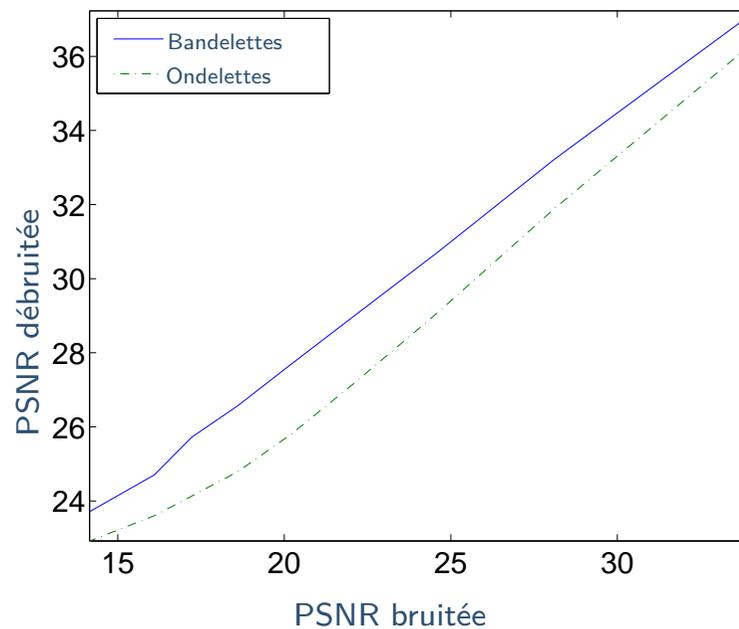
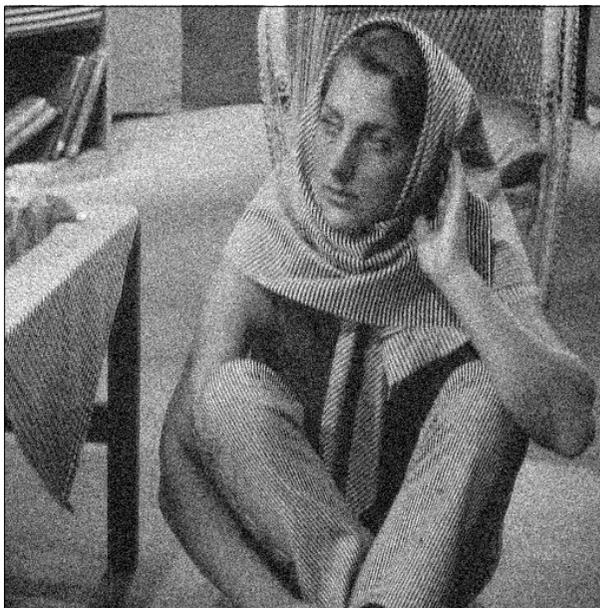
Bandelettes



Ondelettes



Bruitée (20,19 dB)



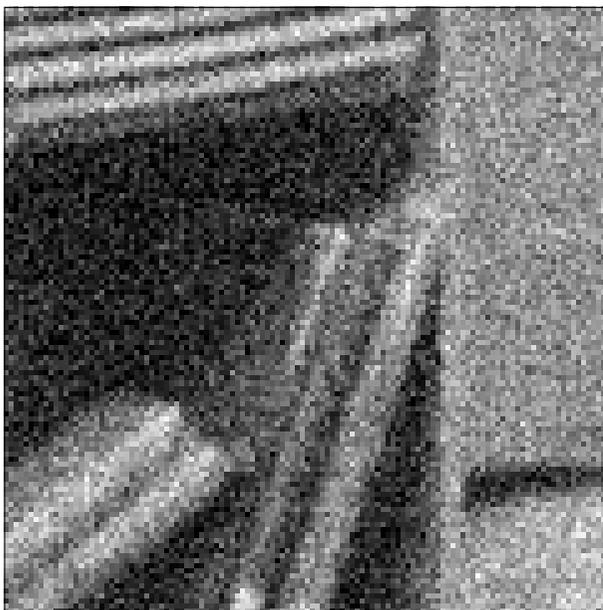
Bandelettes (27,68 dB)



Ondelettes (25,79 dB)



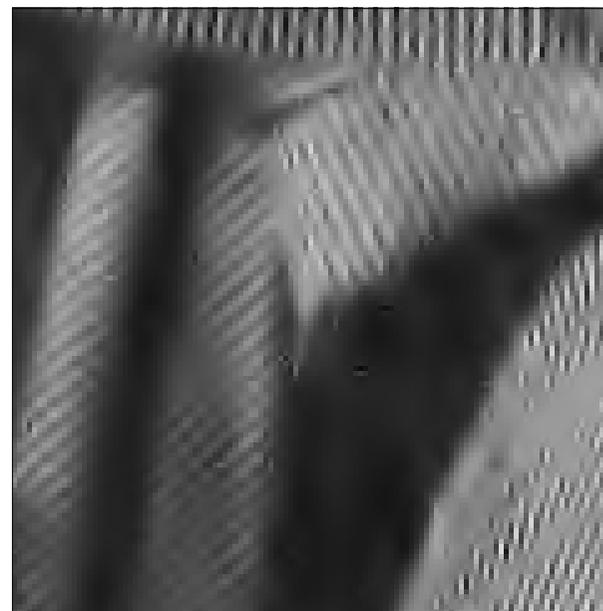
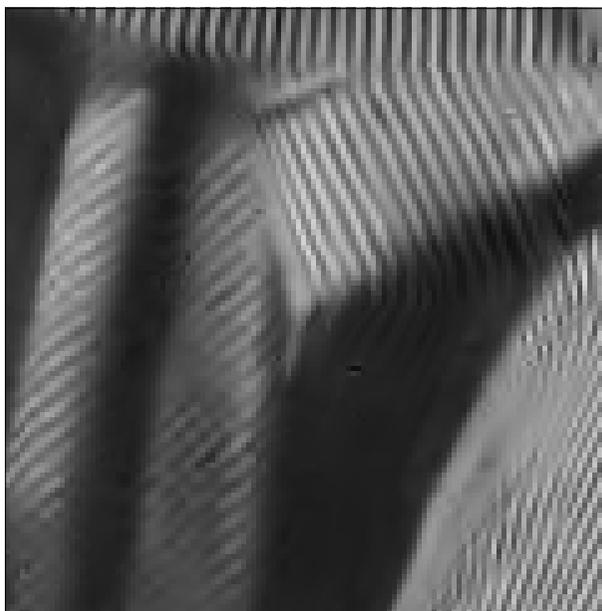
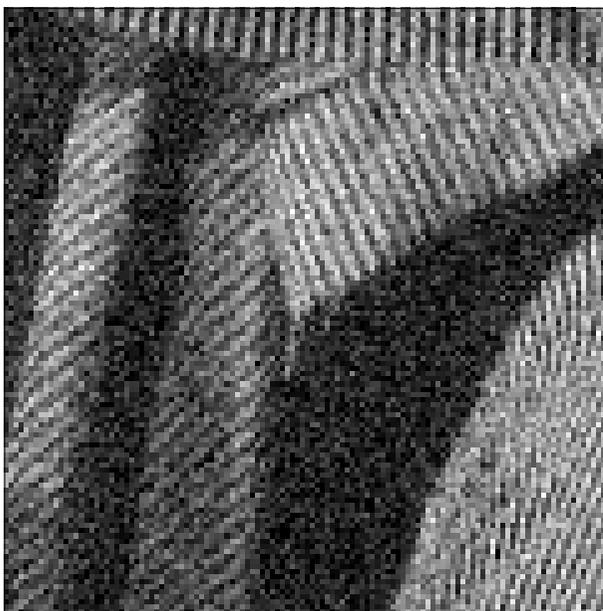
Bruitée



Bandelettes



Ondelettes



Un peu plus d'image ?

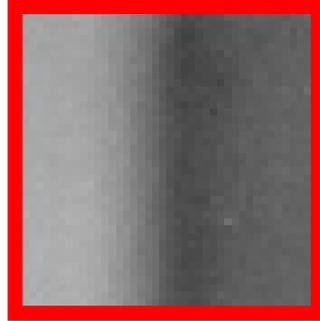
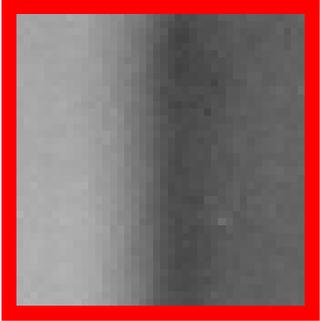
OUI ou NON

Conclusion

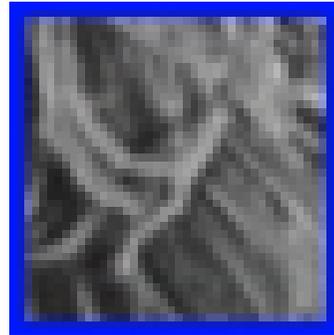
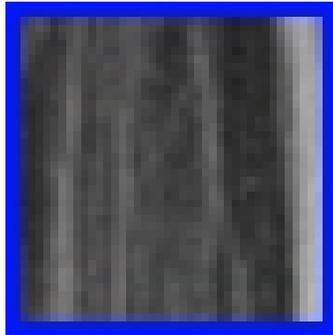
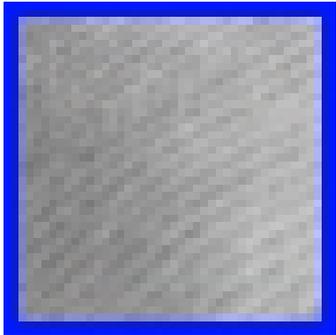
- Pour la sélection de modèles, maxiset = espace d'approximation.
- Importance d'avoir une représentation adaptée.
- Bandelettes : une représentation adaptée à la géométrie des images $\mathbf{C}^\alpha - \mathbf{C}^\alpha$.
- Passage au modèle d'échantillonnage/norme empirique se fait bien.
- Problèmes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

Vers des images plus nature

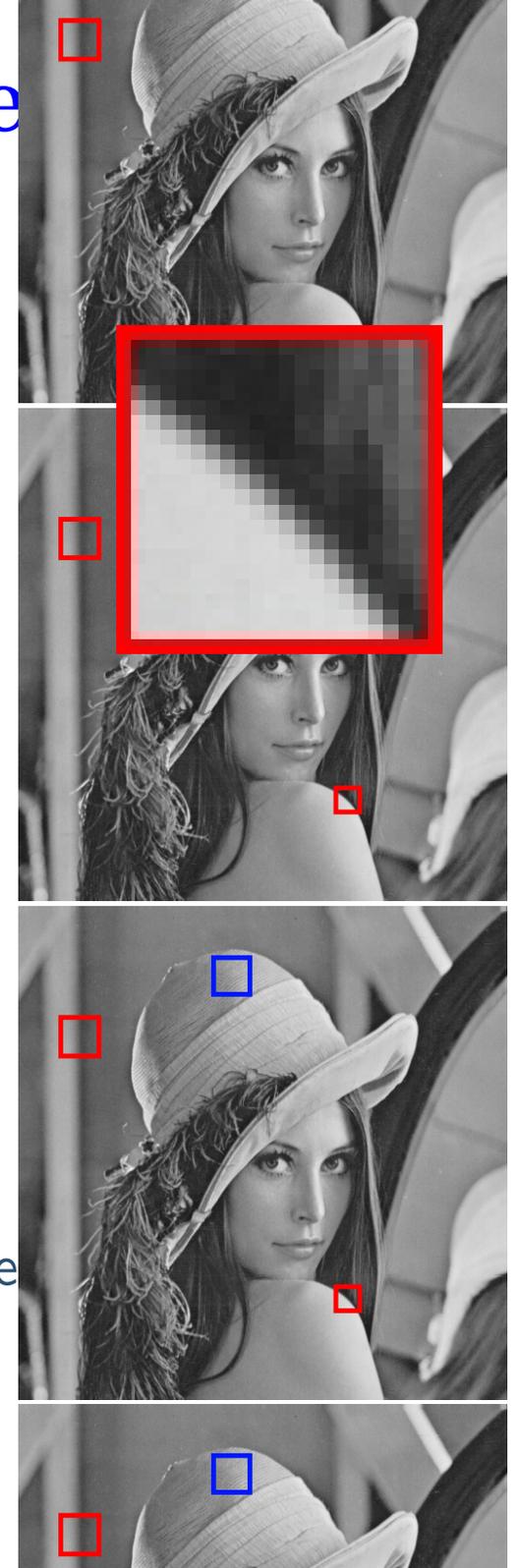
- Modèle $C^\alpha - C^\alpha$ simpliste.
- Les contours sont flous :



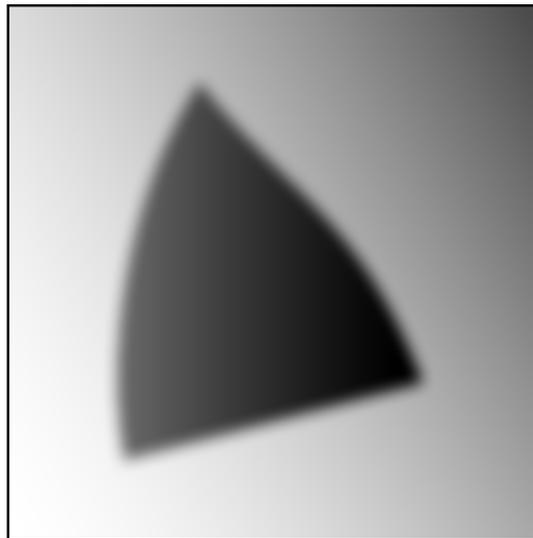
- Pas un problème pour les bandelettes.
- La géométrie vie à plusieurs échelles :



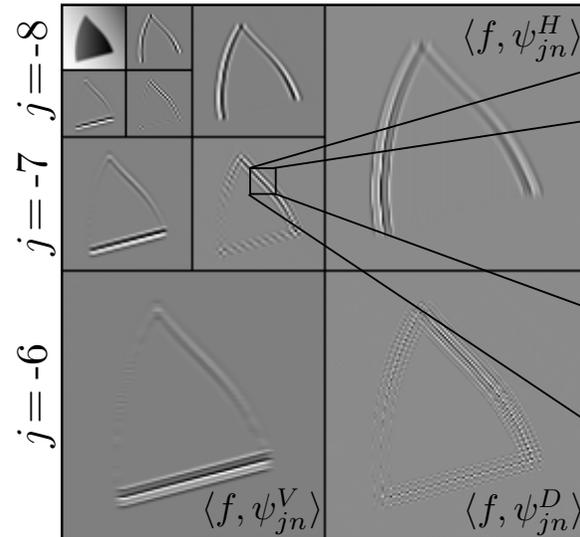
- Comment incorporer une géométrie multiéchelle ?
- Comment éviter les effets de bords sans perdre l'orthogonalité ?



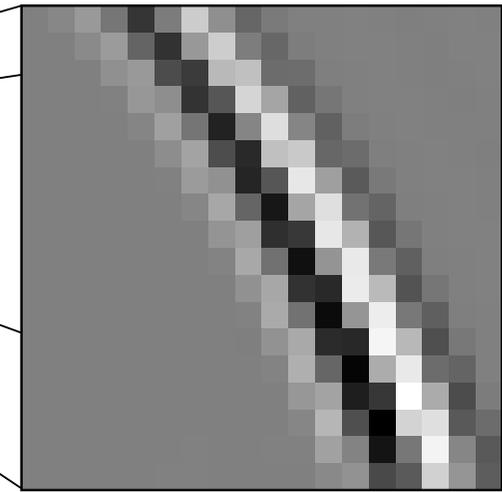
Retour vers les ondelettes



(a) Fonction originale



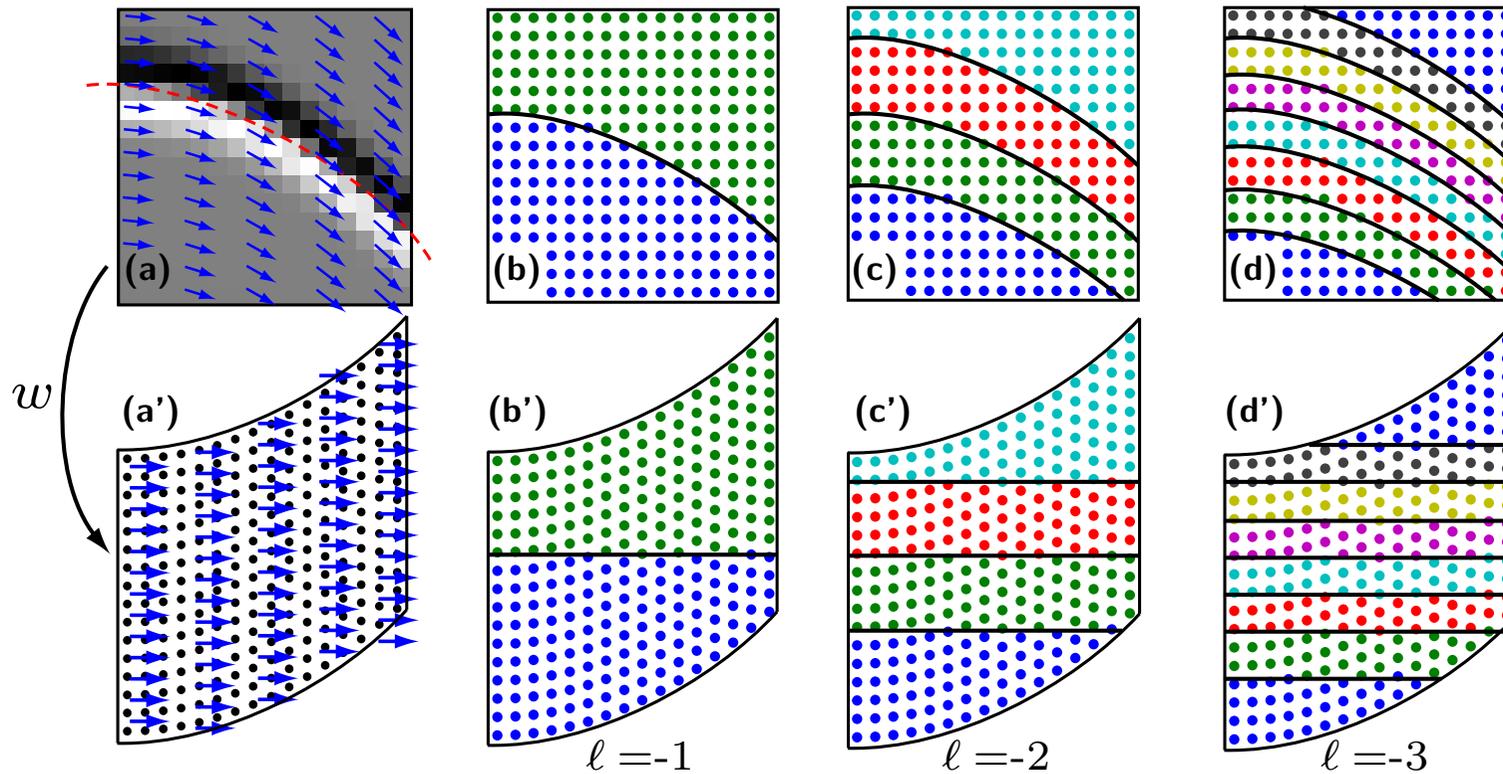
(b) Transformée en ondelettes



(c) Zoom

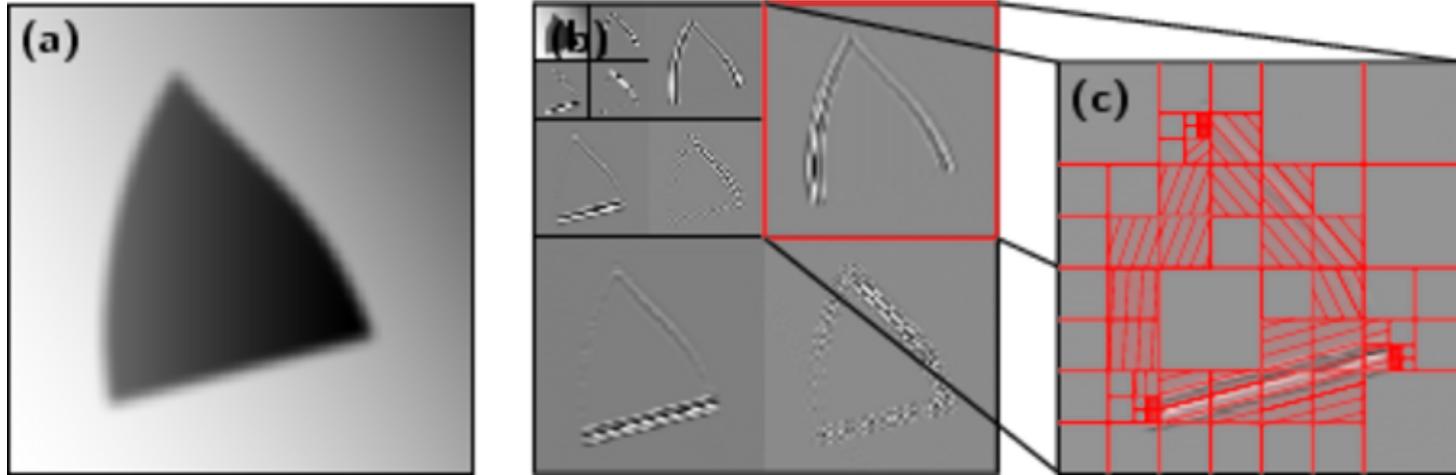
- Ondelettes : bonne représentation multiéchelles.
- Analogie avec le système visuel.
- Existence de régularité pour les coefficients d'ondelettes.
- Comment l'exploiter ?
- Utilisation du contexte dans JPEG 2000.
- Edgeprint (*Vetterli, Dragotti, Baraniuk*) : modélisation explicite dans le contexte du codage.
- Bandelettes sur les coefficients ?

Base locale de bandelettes



- Bandelettes 2G (*Peyré*) : Changement de base orthogonale adapté à la géométrie sur les coefficients d'ondelette.
- Multirésolution d'espaces d'approximations polynomiales par morceaux.
- Base des compléments orthogonaux de ces espaces (*Alpert*).
- Image des ondelettes par ce changement de bases : bandelettes 2G.

Base de bandelettes 2G



- Base de bandelettes :
 - segmentation dyadique des sous-bandes,
 - géométrie dans chaque carré.
- Algorithme d'optimisation par programmation dynamique (CART) de

$$\|f - f_M\|^2 + T^2 M \quad .$$

- **Théorème** : Si f est $C^\alpha - C^\alpha$, alors, dans la meilleure base,

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

Estimation en bandelettes 2G

- Contrôle polynomial en fonction de N sur le nombre total de bandelettes.
- Estimation par sélection de modèle :

$$F_S = \operatorname{argmin} \|Y - P_{\mathcal{M}}Y\|^2 + \lambda \frac{\log N}{N} \dim(\mathcal{M})$$

avec \mathcal{M} qui parcourt les sous-espaces d'une famille de bases de bandelettes 2G.

- Quasi optimalité : si $f \in \mathbf{C}^\alpha - \mathbf{C}^\alpha$ alors

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

- Maxiset :

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N} \right)^{\frac{\alpha}{\alpha+1}} \Leftrightarrow f \in \mathcal{A}^\alpha \Leftrightarrow \forall M, \exists \mathcal{B}, \|f - f_M\|^2 \leq CM^{-\alpha} .$$

- Conjecture : $f \in \mathcal{A}^\beta \Leftrightarrow \exists \mathcal{B}, \forall M, \|f - f_M\|^2 \leq CM^{-\beta}$!
- Expérimentation numérique en cours...

Conclusion

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