Maxisets pour la sélection de modèles

E. Le Pennec / LPMA / Université Paris Diderot
F. Autin, J.-M. Loubes et V. Rivoirard

18 Juin 2009

Sélection de modèles, Minimax et Maxiset

- 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.
- Mesure de performance:
  - Minimax: recherche d'un estimateur adaptée à un espace.
  - Maxiset : recherche des espaces bien adaptés à un estimateur.
- Maxisets pour la sélection de modèle $\Leftrightarrow$ espaces d'approximation de la théorie de l'approximation.
- Nécessité d'avoir une représentation creuse (approximation).
- Illustrations avec des ondelettes.

Message pour la plage

- On estime bien les fonctions qu'on approche bien et uniquement celles-ci...

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.
- Estimation par ondelettes.

Estimation par projection

- Modèle de bruit blanc :

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

- Dans la suite, projection sur un espace de dimension finie 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 : $F = P_{\widehat{m}} Y$ avec $\widehat{m}$ s.e.v. (modèle) à choisir.
- Choix de la collection $\mathcal{M}$ de modèles m .
- Choix du modèle $\widehat{m}$ utilisé dans l'estimateur.
- Critère : risque quadratique

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

Estimation oracle dans une base

- Base o.n. $\{b_n\}_n$ et modèles $m = \mathrm{vect}\{b_n\}_{n\in\Gamma}$.
- 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 $F_O = Y_{\Gamma_0}$.
- 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 ($\left(\frac{1}{N}\right)^{\frac{\beta}{\beta+1}}$ vitesse minimax).
- Maxiset : pour une base fixée, quel est l'ensemble des fonctions estimées avec une vitesse $\left(\frac{1}{N}\right)^{\frac{\beta}{\beta+1}}$? Ici $\mathcal{A}^\beta$.

Estimateur par seuillage

- Oracle : $\Gamma_O = \{n, |\langle f, b_n\rangle| \geq \frac{1}{\sqrt{N}}\}$ et $F_O = Y_{\Gamma_0}$.
- Stratégie : garder les grands coefficients parmi les $N/\log N$ premiers.
- Seuillage : $\Gamma_S = \{n \leq N, |\langle Y, b_n\rangle| \geq T\left(\frac{1}{\sqrt{N}}\right)\}$ et $F_S = Y_{\Gamma_S}$.
- **Théorème (*Donoho, Johnstone*) :** Si $T\left(\frac{1}{\sqrt{N}}\right) = \lambda\sqrt{\frac{\log N}{N}}$, alors

$$E(\|f - F_S\|^2) \leq C \min_{\Gamma\subset[1,N/\log N]} \|f - f_\Gamma\|^2 + \lambda^2 \frac{\log N}{N}|\Gamma| + \frac{1}{N}$$

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

$$E(\|f - F_S\|^2) \leq C \left(\frac{\log N}{N}\right)^{\frac{\beta}{\beta+1}} = C\rho_{\beta,N} \quad \left(\Leftrightarrow f \in V^*_{\frac{2\beta}{\beta+1}}\right)$$

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

$$\text{et } \|f - P_{V_N} f\|^2 \leq C\left(\frac{\log N}{N}\right)^{\frac{\beta}{\beta+1}}$$

$$\Leftrightarrow \min_{\dim(m)\leq M} \|f - P_m f\|^2 \leq C M^{-\beta} \text{ et } \|f - P_{V_N} f\|^2 \leq C\rho_{\beta,N}$$

$$\Leftrightarrow f \in \mathcal{A}^\beta \cap \mathcal{L}^\beta \quad .$$

Seuillage et sélection de modèles

- Risque oracle : $\|f - f_{\Gamma_O}\|^2 + \frac{1}{N}|\Gamma_O| \quad .$
- Analogue empirique : $\|Y - Y_\Gamma\|^2 + \frac{\lambda_N}{N}|\Gamma| \quad .$
- 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 $\mathrm{pen}(m) = \frac{\lambda_N}{N}\dim(m)$:

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

- 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 $\mathrm{pen}(m) = (C_1 + C_2\lambda_{N,m})\frac{\dim(m)}{N}$

$$F_S = \operatorname*{argmin}_{P_mY,\ m\in\mathcal{M}_N} \|Y - P_mY\|^2 + \mathrm{pen}(m)$$

satisfait

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

- **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

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

$$\Leftrightarrow \min_{m\in\mathcal{M}_N} \|f - P_mf\|^2 + \lambda_N\frac{\dim(m)}{KN} \leq C\left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}} = C\rho_{\beta,N}$$

$$\Leftrightarrow \min_{\dim(m)\leq M} \|f - P_mf\|^2 \leq CM^{-\beta} \text{ et } \|f - P_{V_N}f\|^2 \leq C\rho_{\beta,N}$$

$$\Leftrightarrow f \in \mathcal{A}^{\beta} \cap \mathcal{L}^{\beta} \quad .$$

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

$$\begin{aligned} E(\|f - F_S\|^2) &\leq C \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}} \\ &\implies \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + \lambda_N \frac{\dim(m)}{KN} \leq C' \left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}} \\ &\implies f \in \mathcal{A}^{\beta} \end{aligned}$$

- 3 cas (par ordre de complexité):
  - Seuillage (*Cohen, DeVore, Picard, Kerkyacharian*),
  - Modèles emboîtés,
  - Cas général...

Le cas du seuillage – 1

- Clé : $\underset{P_m f, m \in \mathcal{M}_N}{\operatorname{argmin}} \|f - P_m f\|^2 + \lambda_N \frac{\dim(m)}{N} = f_T$ avec $T^2 = \frac{\lambda_N}{N}$.
- $$\|f - F_S\|^2 = \sum_{\substack{|\langle Y, b_k\rangle| \geq T \\ |\langle f, b_k\rangle \geq T/2}} |\langle \frac{1}{\sqrt{N}} W, b_k\rangle|^2 + \sum_{\substack{|\langle Y, b_k\rangle| \geq T \\ |\langle f, b_k\rangle < T/2}} |\langle \frac{1}{\sqrt{N}} W, b_k\rangle|^2$$
$$+ \sum_{\substack{|\langle Y, b_k\rangle| < T \\ |\langle f, b_k\rangle| \geq T/2}} |\langle f, b_k\rangle|^2 + \sum_{\substack{|\langle Y, b_k\rangle| < T \\ |\langle f, b_k\rangle| < T/2}} |\langle f, b_k\rangle|^2$$
$$\|f - F_S\|^2 \geq \|f - f_{T/2}\|^2 \quad .$$
- D'où
$$\|f - f_{T/2}\|^2 \leq E(\|f - F_S\|^2) \leq C T^{\frac{2\beta}{\beta+1}} \quad .$$
- Erreur d'approximation $\neq$ compromis erreur d'approximation et nombre de coefficients...:
$$\|f - f_{T/2}\|^2 + T^2 M \geq \|f - f_{T/2}\|^2 \quad .$$

Le cas du seuillage – 2

- Mais
$$\|f - f_{T/2}\|^2 \leq C T^{\frac{2\beta}{\beta+1}} \Rightarrow \|f - f_{T/2}\|^2 + T^2 M \leq C_\beta T^{\frac{2\beta}{\beta+1}} \quad !$$
- Explication :
$$\|f - f_{T/2}\|^2 \geq \mathrm{Card}\left\{\text{coeff} \in]T/4, T/2]\right\} T^2/16 \quad .$$
- D'où :
$$\|f - f_{T/2}\|^2 \leq C T^{\frac{2\beta}{\beta+1}} \implies \mathrm{Card}\left\{|\text{coeff}| \in]T/4, T/2]\right\} \leq C' T^{\frac{-2}{\beta+1}} \quad .$$
- On en déduit :
$$\mathrm{Card}\left\{|\text{coeff}| \in]0, T]\right\} \leq C_\beta T^{\frac{-2}{\beta+1}} \quad .$$
- Ou encore :
$$T^2 M \leq C_\beta T^{\frac{2\beta}{\beta+1}} \quad \ldots$$

Cas des modèles emboîtés – 1

- $\widehat{m}$: modèle sélectionné qui minimise

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

- m_O : modèle oracle qui minimise

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

- On va montrer que

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

- Première étape :

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

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

Cas des modèles emboîtés – 2

- $$\|P_{\widehat{m}}(f-Y)\|^2 \geq \|f - P_{m_O} f\|^2 - \|f - P_{\widehat{m}} f\|^2 \quad ?$$
- Si $\widehat{m} \subset m_O$:
$$\|f - P_{m_O} f\|^2 - \|f - P_{\widehat{m}} f\|^2 \leq 0 \quad .$$
- Sinon $\widehat{m} \supset m_O$ et on est ramené à montrer que
$$\|P_{\widehat{m}}(f-Y)\|^2 \geq \|P_{\widehat{m}\backslash m_O} f\|^2 \quad .$$
- Or par définition de $\widehat{m}$ et m_O

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

- Par inégalité triangulaire :

$$\|P_{\widehat{m}\backslash m_O}(Y-f)\| \geq \|P_{\widehat{m}\backslash m_O} Y\| - \|P_{\widehat{m}\backslash m_O} f\| \geq \frac{1}{2}\sqrt{\frac{\lambda_N}{N}(\dim(\widehat{m}) - \dim(m_O))}$$
$$\|P_{\widehat{m}\backslash m_O}(Y-f)\| \geq \|P_{\widehat{m}\backslash m_O} f\| \quad .$$

Cas des modèles emboîtés – 3

- On a obtenu $$\|f - P_{\widehat{m}} Y\|^2 \geq \|f - P_{m_O} f\|^2 \quad .$$
  et donc $$\|f - P_{\widehat{m}} Y\|^2 \leq C T^{\frac{2\beta}{\beta+1}} \implies \|f - P_{m_O} f\|^2 \leq C T^{\frac{2\beta}{\beta+1}} \quad .$$
- Il reste à montrer que $$\|f - P_{m_O} f\|^2 + \frac{1}{4} T^2 \dim(m_O) \leq C_\beta T^{\frac{2\beta}{\beta+1}} \quad \ldots$$
- m_O dépend de T : $m_O(T/2)$ n'est pas optimal pour T mais pour $T/2$.

$$\|f - P_{m_O(T)} f\|^2 + \lambda_N \frac{\dim(m_O(T))}{4N} \leq \|f - P_{m_O(T/2)} f\|^2 + \lambda_N \frac{\dim(m_O(T/2))}{4N}$$

$$\dim(m_O(T)) - \dim(m_O(T/2)) \leq \frac{4N}{\lambda_N} (\|f - P_{m_O(T/2)} f\|^2 - \|f - P_{m_O(T)} f\|^2)$$

$$\dim(m_O(T)) - \dim(m_O(T/2)) \leq \frac{4N}{\lambda_N} C_\beta (T/2)^{\frac{2\beta}{\beta+1}}$$

- En sommant sur les $T/2^k$, on conclut.
- Rq: le seuillage est un cas particulier (somme directe de modèles emboîtés).

Cas général – 1

- $\widehat{m}$: modèle sélectionné qui minimise

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

- m_O : modèle oracle qui minimise pour K suffisament grand

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

- En général, on a pas

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

- On va montrer que

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

Cas général – 2

- Définition de $\hat{m}$:

$$\frac{\lambda_N}{N}(\dim(\hat{m}) - \dim(m_0)) \leq \|P_{m_0}Y - f\|^2 - \|P_{\hat{m}}Y - f\|^2 + \frac{2}{\sqrt{N}}\langle W_{\hat{m}+m_0}, P_{\hat{m}}Y - P_{m_0}Y\rangle.$$

- Événement:

$$A_N = \{\sup_{m,m'} \left((\dim(m) + \dim(m'))^{-1}\|W'_{m+m}\|^2\right) \leq \lambda_N\}$$

- Kraft + Lemme de Cirelson $\quad P(A_N) \geq p > 0$
- Contrôle sous A_N de $(\dim(\hat{m}) - \dim(m_0))$:

$$\mathbb{1}_{A_N}\frac{\lambda_N}{KN}(\dim(\hat{m}) - \dim(m_0)) \leq \mathbb{1}_{A_N}\Big(\frac{\frac{2}{\gamma}+1}{1-\gamma}\|P_{\hat{m}}Y - f\|^2 + \frac{\frac{2}{\gamma}-1}{1-\gamma}\|P_{m_0}Y - f\|^2 + \frac{2\gamma}{1-\gamma}\frac{\lambda_N}{N}\dim(m_0)\Big)$$

Cas général – 3

- Définition de m_0
$$\|P_{m_0}f - f\|^2 \leq \|P_{\hat{m}}f - f\|^2 + \frac{\lambda_N}{KN}(\dim(\hat{m}) - \dim(m_0))$$
- En combinant avec le contrôle sous A_N:
$$\left(1 - \frac{\frac{2}{\gamma}+1}{K(1-\gamma)}\right)\mathbb{1}_{A_N}\|P_{m_0}f - f\|^2$$
$$\leq \left(1 + \frac{\frac{2}{\gamma}-1}{K(1-\gamma)}\right)\|P_{\hat{m}}Y - f\|^2$$
$$+ \frac{\frac{2}{\gamma}+1}{K(1-\gamma)}\frac{1}{N}\|W_{m_0}\|^2 + \frac{2\gamma}{K(1-\gamma)}\frac{\lambda_N}{N}\mathbb{1}_{A_N}\dim(m_0).$$
- Passage à l'espérance possible:
$$\|P_{m_0} - f\|^2 \leq \frac{\tilde{K} + \frac{4}{\gamma}}{\tilde{K}P\{A_N\}}E\left[\|P_{\hat{m}}Y - f\|^2\right]$$
$$+ \left(\frac{K(\frac{2}{\gamma}+1)}{\tilde{K}\lambda_N P\{A_N\}} + \frac{2K\gamma}{\tilde{K}}\right)\frac{\lambda_N}{KN}\dim(m_0)$$

avec $\tilde{K} = K(1-\gamma) - \frac{2}{\gamma} - 1.$

Cas général – 4

- Récurrence sur N:
$$\|P_{m_0(N)}f - f\|^2 + \frac{\lambda_N}{KN}\dim(m_0(N)) \leq C'\left(\frac{\lambda_N}{N}\right)^{\frac{\beta}{\beta+1}}$$

- Définition de $m_0(N)$
$$\|P_{m_0(N)}f - f\|^2 + \frac{\lambda_N}{KN}\dim(m_0(N)) \leq \|P_{m_0(N/2)}f - f\|^2 + \frac{\lambda_N}{KN}\dim(m_0(N/2))$$

- Existence d'un β_N (suffisament petit) tel que:
$$\begin{aligned}\|P_{m_0(N)}f - f\|^2 &+ \frac{\lambda_N}{KN}\dim(m_0(N)) \\ &\leq (1-\beta_N)\frac{\tilde{K}+\frac{4}{\gamma}}{\tilde{K}P\{A_{N/2}\}}E\left[\|P_{\hat{m}(N/2)}Y - f\|^2\right] \\ &\quad + \beta_N\left(\|s_{m_0(N/2)} - s\|^2 + \frac{2\lambda_{N/2}}{KN}\dim(m_0(N/2))\right).\end{aligned}$$

- Conditions sur λ_N (assez grand) et β assurent la récurrence...

Choix de la collection

- Compromis entre
  - une collection riche ayant de bonne propriétés d'approximations,
  - une collection pauvre permettant de choisir une pénalité petite...
- Exemples à partir d'une base d'ondelettes $(\psi_{j,k})$sur $[0,1]$.
- Modèles: espaces m engendrés par $\psi_{j,k}$ pour $(j,k) \in I_m$.
- Collection de modèles $=$ collection de sous-ensembles d'indices.
- Exemples:
  - Linéaire: $$I_{m_J} = \{(j,k), j \le J, k \in [0, 2^j - 1]\}$$
    pour tout J dans $\mathbb{N}$,
  - Seuillage: $$I_{m_I} = \{(j,k) \in I\} (= I)$$
    pour tout I dans $\mathcal{P}(\{(j,k), j \le J\})$,
  - Collection à taille contrôlée:
    $$I_m = \{(j,k) \in I, \quad I \cap \{(j',k)\}_k = \{(j',k)\}_k, \quad \forall j' \le J,$$
    $$\text{et } |\{(j',k)\}_k| = \frac{2^J}{(j-J)^\theta}, \quad \forall j' > J\}$$
    pour tout J dans $\mathbb{N}$.
- Étude des maxisets associés et de la calculabilité.

Estimateur linéaire

- Estimateur classique linéaire de projection sur les espaces d'approximations en ondelettes rendu adaptatif par sélection de modèles.
- Pénalisation (avec abus de notation...)

$$\hat{m} = \underset{m_J}{\operatorname{argmin}} \|Y - P_{m_J} Y\|^2 + \frac{\lambda_N}{N} \dim(m_J)$$

- Collection "pauvre" (modèles emboités), λ_N constant suffit.
- Maxiset:

$$E(\|f - P_{\hat{m}} Y|^2) \leq C \left(\frac{\lambda_N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} \Leftrightarrow f \in \mathcal{B}^{\alpha}_{2,\infty}$$

  où $\mathcal{B}^s_{p,q}$ est l'espace de Besov "classique".
- Résultat déjà connu...
- Estimateur rendu calculable en restreignant J: $2^J \leq 2^{J_0(N)} \simeq \frac{N}{\lambda_N}$.
- Maxiset non modifié.

Seuillage

- Collection de tous les sous-ensembles d'échelle inférieure à J.
- Collection riche qui nécessite une pénalité en $\log 2^J = J$.
- Classiquement, on prend $2^J = 2^{J_0(N)} \simeq \frac{N}{\lambda_N}$.
- Dans ce cas, sélection de modèles = seuillage ⇒ pas de problème de calculabilité!
- Maxiset:

$$E(\|f - P_{\hat{m}} Y|^2) \leq C \left(\frac{\lambda_N}{N} \right)^{\frac{2\alpha}{2\alpha+1}} \Leftrightarrow f \in \mathcal{W}^{2/(1+2\alpha)} \cap \mathcal{B}_{2,\infty}^{\alpha/(1+2\alpha)}$$

  où $\mathcal{W}^{2/(1+2\alpha)}$ est un espace de Besov faible.
- Besov faible caractérisé par l'approximation:

$$f \in \mathcal{W}^{2/(1+2\alpha)} \Leftrightarrow \inf_{\Gamma, |\Gamma| \leq M} \|f - f_M|^2 \leq C M^{-2\alpha} \quad .$$

- Résultat déjà connu...
- Jeu possible sur le Besov fort.

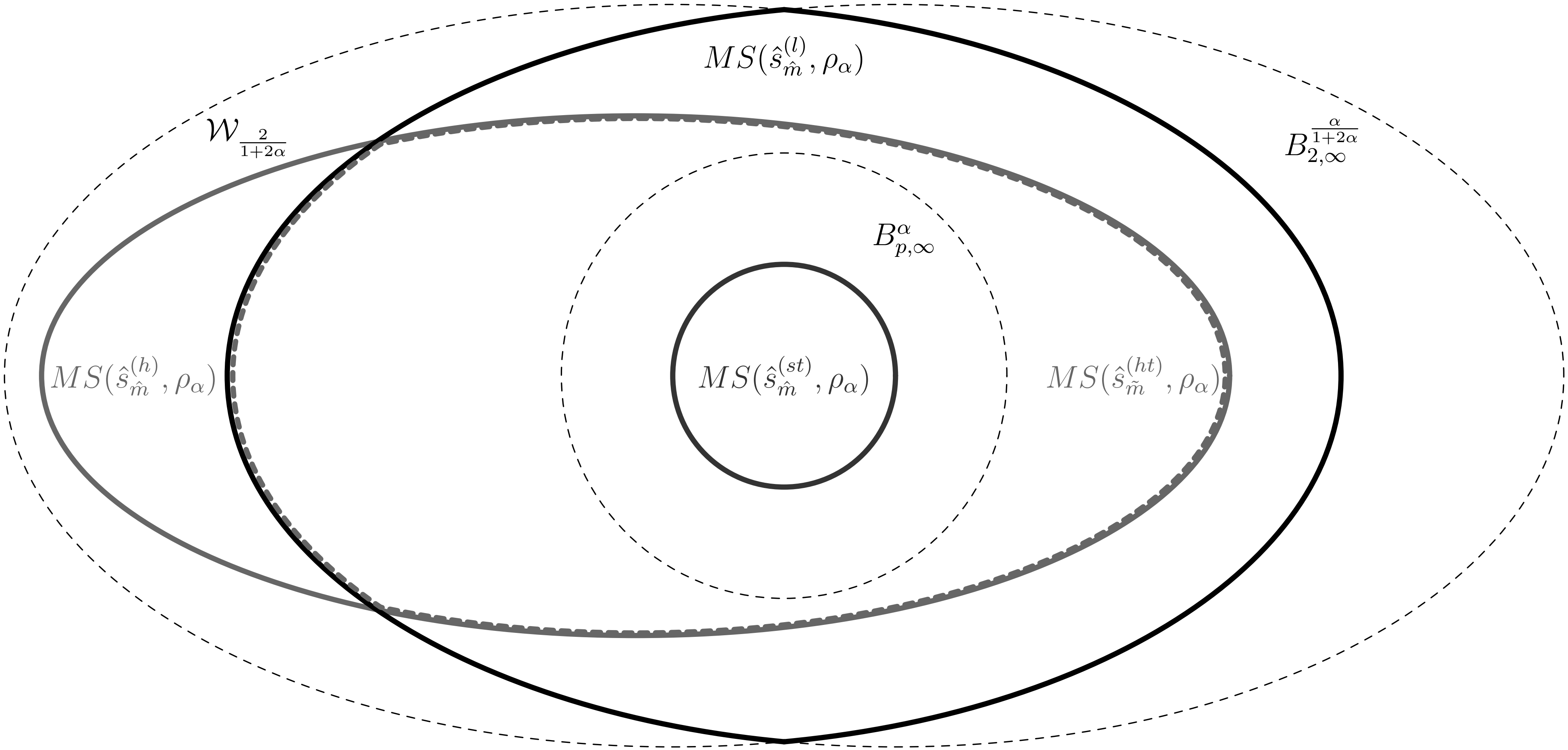
Collection intermédiaire

- Collection à taille contrôlée proposée par P. Massart:

$$I_m = \{(j,k) \in I, \quad I \cap \{(j',k)\}_k = \{(j',k)\}_k, \quad \forall j' \leq J,$$
$$\text{et } |\{(j',k)\}_k| = \frac{2^J}{(j-J)^\theta}, \quad \forall j' > J\}$$

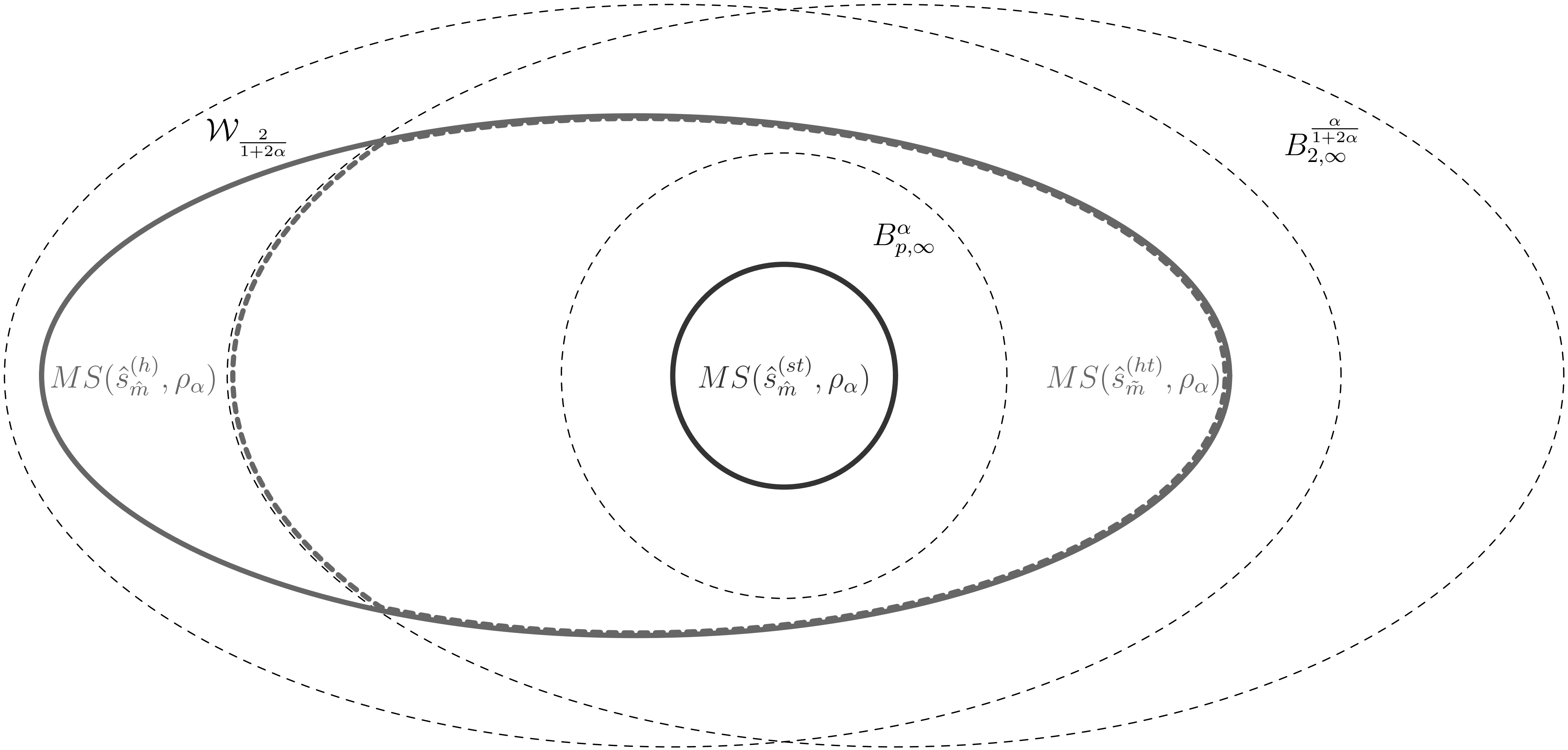
  pour tout J dans $\mathbb{N}$.
- Intermédiaire entre seuillage et linéaire.
- Pénalisation avec λ_N constant suffit.
- Maxiset $\mathcal{A}^\alpha$ complétement caractérisé en terme d'approximation mais pas de caractérisation fonctionnelle.
- Propriété intéressante: $\mathcal{B}^\alpha_{p,\infty} \subsetneq \mathcal{A}^\alpha$ pour $p \geq \max(1, 2/(2\alpha+1))$.
- Estimateur est minimax pour ces espaces de Besov (et adaptatif).
- Non calculable $\rightarrow$ calculable en se restreignant au coefficients à des échelles plus petite que $J_0(N)$.
- Prix à payer: maxiset $=$ intersection du maxiset précédent avec $\mathcal{B}^{\alpha/(1+2\alpha)}_{2,\infty}$...

Comparaison – 1



$\lambda_N = \lambda_0 \log(N)$ et $\max(1, 2\left(\frac{1}{1+2\alpha} + 2\alpha\right)^{-1}) \leq p \leq 2$

Comparaison – 2



$$\lambda_N = \lambda_0 \text{ et } \max(1, 2\left(\frac{1}{1+2\alpha} + 2\alpha\right)^{-1}) \leq p \leq 2$$

Conclusion

- Étude des maxisets des estimateurs par sélection de modèles.
- Exemples avec des estimateurs en ondelettes.
- Lien très fort entre estimation et théorie de l'approximation.
- Bilan: On estime bien les fonctions qu'on approche bien et uniquement celles-ci!
- Autres résultats:
  - Espaces correspondants à des vitesses non (log) polynomiales: pas de caractérisation classique mais pas de problèmes.
  - Fonctionne pour des dictionnaire de bases orthogonales (bandelettes).
- Maxiset: prisme intéressant pour comparer des estimateurs...