# Estimation de densité par pénalisation $\ell_1$ minimale

Erwan LE PENNEC
SELECT, INRIA Saclay - Univ. Paris Sud
en collaboration avec
Karine BERTIN (Univ. de Valparaiso, Chili)
et
Vincent RIVOIRARD (Univ. Paris Dauphine)

50 ans du LPMA - 07 décembre 2010

INRIA

# Estimation de densité par pénalisation $\ell_1$ minimale

Erwan Le Pennec
ex LPMA – Université Paris Diderot
en collaboration avec
Karine Bertin (ex LPMA – Université Pierre et Marie Curie)
et
Vincent Rivoirard (ex LPMA – Université Paris Diderot)

50 ans du LPMA - 07 décembre 2010

# LPMA

# LPMA




# LPMA

# LPMA







# Cadre

## Estimation de densité

- $X_1, \ldots, X_n$ i.i.d de loi $f_0(x)dx$.
- Hyp : $f_0 \in L^2$.
- Pas d'hypothèse $f_0 \in L^\infty$ et surtout pas $\|f_0\|_\infty$ connue.

## Approche dictionnaire

- Dictionnaire arbitraire $\mathcal{D} = \{\phi_k\}_{1 \leq k \leq p}$.
- Hyp : $\phi_k \in L^2 \cap L^\infty$ et $\|\phi_k\|_\infty$ connue.
- Notation : $f_\lambda = \sum_{k=1}^{p} \lambda_k \phi_k$.

## Objectifs

- Estimation de $f_0$ par une combinaison linéaire $f_\lambda$ du dictionnaire $\mathcal{D}$.
- Mesure de l'erreur en norme $L^2$.
- Oracle $\ell_0$ : $\|\hat{f} - f_0\|^2 \leq (1 + c)\left(\inf_{\lambda \in \mathbb{R}^p} \|f_\lambda - f_0\|^2 + C|\lambda|_0\right)$ ?
- Cas $p \gg n$ possible...

# Estimateur de Dantzig

## Produits scalaires

- Produits scalaires et espérances : $\beta_k = \langle f_0, \phi_k \rangle = \mathbb{E}\left(\phi_k(X)\right)$.
- Produits scalaires empiriques : $\hat{\beta}_k = \frac{1}{n}\sum_{i=1}^{n} \phi_k(X_i) \simeq \beta_k$.
- Pour $f_\lambda = \sum_{k=1}^{p} \lambda_k \phi_k$, $\langle f_\lambda, \phi_k \rangle = (G\lambda)_k$ où $G$ : matrice de Gram du dictionnaire $\mathcal{D}$ ($G_{k,k'} = \langle \phi_k, \phi_{k'} \rangle$).

## Contraintes et régularisation

- Recherche de $\lambda$ tel que $G\lambda \simeq \hat{\beta} \simeq \beta$.
- Moindres carrés : $\hat{\lambda}^{MC} = \mathrm{argmin} \| G\lambda - \hat{\beta} \|^2$. (Pb si $p \gg n$ !).
- Contraintes de Dantzig : $\forall k \in \{1, \ldots, p\}, |(G\lambda)_k - \hat{\beta}_k| \leq \eta_k$.
- $+$ Régularisation par la norme $\ell_1$ de $\lambda$.
- Estimateur de Dantzig :

$$\hat{\lambda}^D = \mathrm{argmin} \|\lambda\|_{\ell_1} \quad \text{sous} \quad \forall k \in \{1, \ldots, p\}, |(G\lambda)_k - \hat{\beta}_k| \leq \eta_k.$$

# Estimateur Lasso / SPADES

## Norme quadratique

- Erreur $L^2$ : $\|f - f_0\|_2^2 = \|f\|_2^2 - 2\mathbb{E}(f(X)) + \|f_0\|_2^2$.
- Contraste $L^2$ : $\gamma(f) = -2\frac{1}{n}\sum_{i=1}^{n} f(X_i) + \|f\|_2^2 \simeq \|f - f_0\|_2^2 - \|f_0\|_2^2$.
- Pour $f_\lambda = \sum_{k=1}^{p} \lambda_k \phi_k$, $\gamma(f_\lambda) = -2\lambda^* \hat{\beta} + \lambda^* G \lambda$.

## Contraste et pénalisation

- Minimisation du contraste empirique pénalisé par $2\sum_k \eta_k |\lambda_k|$.
- SPADES / Estimateur Lasso :

$$\hat{\lambda}^L = \operatorname{argmin} \frac{1}{2}\left(-2\lambda^* \hat{\beta} + \lambda^* G \lambda\right) + \sum_{k=1}^{p} \eta_k |\lambda_k|.$$

## Lien Dantzig / Lasso

- Condition d'optimalité de premier ordre du Lasso
  $=$ Contraintes de Dantzig, $\forall k \in \{1, \ldots, p\}, |(G\lambda)_k - \hat{\beta}_k| \leq \eta_k$
- Lien fort entre les deux estimateurs...

# Bibliographie parcimonieuse

### Pénalisation $\ell_1$ en statistique

- Frank et Friedman (91–93) Bridge.
- Tibshirani et al. (94–96) Lasso.
- Donoho et al. (95–01) Basis Pursuit.

### Parcimonie et système sous déterminé

- Donoho (2004) Parcimonie et matrice gaussienne.
- Candes, Tao et Romberg (2004) Parcimonie et matrice de Fourier.

### Parcimonie et $\ell_1$ en statistique

- Candes et Tao (2007) Dantzig.
- Lasso/Dantzig, SPADES : Tsybakov, Bickel, Ritov, Bunea, Wegkamp
- Énorme littérature : Plan, Temlyakov, Lounici, van de Geer,...

# Inégalités de concentration

## Choix de $\eta_k$

- Comment choisir $\eta_k$ pour que $\mathbb{P}\left(|\beta_k - \hat{\beta}_k| > \eta_k\right) \leq 2e^{-u^2/2}$ ?
- Concentration d'un processus empirique simple autour de sa moyenne.

## Hoeffding

- $\eta_k = u\dfrac{\|\phi_k\|_\infty}{\sqrt{n}} \implies \mathbb{P}\left(|\beta_k - \hat{\beta}_k| > \eta_k\right) \leq 2e^{-u^2/2}$.
- Souvent utilisé mais majoration grossière n'utilisant pas $f_0$ et limitation sur la norme des $\phi_k$.

## Bernstein

- Utilisation de la variance : $\sigma_k^2 = \mathrm{Var}\,(\phi_k(X))$.
- $\eta_k = u\dfrac{\sigma_k}{\sqrt{n}} + \dfrac{u^2}{3}\dfrac{\|\phi_k\|_\infty}{n} \implies \mathbb{P}\left(|\beta_k - \hat{\beta}_k| > \eta_k\right) \leq 2e^{-u^2/2}$.
- Prise en compte fine de $f_0$ et limitation beaucoup moins forte sur la norme de $\phi_k$.

# Inégalités de concentration

## Bernstein (suite)

- Pb : Comme $f_0$, la variance est inconnue...
- Majoration possible mais perte du caractère fin :
  - $\sigma_k \leq \|\phi_k\|_\infty$ ($\simeq$ Hoeffding)
  - $\sigma_k \leq \|f\|_\infty^{1/2} \|\phi_k\|_2$ (Pb hypothèse sur $\|f\|_\infty$)

## Adaptation/Auto renormalisation

- Variance empirique : $\hat{\sigma}_k^2 = \dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n} \left(\phi_k(X_i) - \hat{\beta}_k\right)^2$
- $\eta_k = \tilde{u} \dfrac{\hat{\sigma}_k}{\sqrt{n}} + \dfrac{7}{2} \dfrac{\tilde{u}^2}{3} \dfrac{\|\phi_k\|_{\bar{\infty}}}{n-1} \implies \mathbb{P}\left(|\beta_k - \hat{\beta}_k| > \eta_k\right) \leq 2e^{-u^2/2}$ avec $\tilde{u} = u\sqrt{1 + \frac{2\log 2}{u^2}}$ et $\|\phi_k\|_{\bar{\infty}} = \sup \phi_k - \inf \phi_k \leq 2\|\phi_k\|_\infty$.
- Prise en compte adaptative de la variabilité de $f_0$.
- De plus, $\mathbb{P}\left(|\hat{\sigma}_k - \sigma_k| > u \frac{\|\phi_k\|_{\bar{\infty}}}{\sqrt{n-1}}\right) \leq 2e^{-u^2/2}$.

# Contrôle sur les coefficients avec grande probabilité

## Contrôle sur les coefficients

- Borne d'union avec $u = \sqrt{2\gamma \log p}$.
- $\eta_k^\gamma = \sqrt{2\gamma c_{\gamma,p} \log p}\frac{\hat{\sigma}_k}{\sqrt{n-1}} + \frac{7}{3}\gamma c_{\gamma,p} \log p \frac{\|\phi_k\|_{\bar{\infty}}}{n-1}$ avec $c_{\gamma,p} = 1 + \frac{\log 2}{\gamma \log p}$.

## L'évènement $\Omega_{\gamma_1}$

- Pour tout $\gamma_1 > 0$, évènement $\Omega_{\gamma_1}$ de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma_1 - 1}$ :

$$\Omega_{\gamma_1} = \{\forall k \in \{1, \dots, p\}, |\hat{\beta}_k - \beta_k| \leq \eta_k^{\gamma_1}\}$$

- De plus sous $\Omega_{\gamma_1}$ : $\forall k \in \{1, \dots, p\}, |\hat{\sigma}_k - \sigma_k| \leq \sqrt{2\gamma_1 \log p}\frac{\|\phi_k\|_{\bar{\infty}}}{\sqrt{n-1}}$

$$\sup_k \eta_k^\gamma = \|\eta^\gamma\|_\infty \leq \sqrt{2\gamma c_{\gamma,p} \log p}\frac{\sigma_k}{\sqrt{n}} + (2\sqrt{\gamma c_{\gamma,p}\gamma_1} + \frac{7}{3}\gamma c_{\gamma,p}) \log p \frac{\|\phi_k\|_{\bar{\infty}}}{n-1}$$

# Estimateurs

## Contrainte de Dantzig

- Pour $\gamma_0 > 0$, $\forall k \in \{1, \ldots, p\}$

$$|(G\lambda)_k - \hat{\beta}_k| \leq \eta_k^{\gamma_0} = \sqrt{2\gamma_0 c_{\gamma_0,p} \log p}\,\frac{\hat{\sigma}_k}{\sqrt{n-1}} + \frac{7}{3}\gamma_0 c_{\gamma_0,p} \log p\,\frac{\|\phi_k\|_{\bar{\infty}}}{n-1}.$$

## Estimateurs pénalisés adaptatifs

- Dantzig :
$\hat{\lambda}^{D,\gamma_0} = \operatorname{argmin}\|\lambda\|_{\ell_1}$ sous $\forall k \in \{1, \ldots, p\}, |(G\lambda)_k - \hat{\beta}_k| \leq \eta_k^{\gamma_0}$.
- Lasso : $\hat{\lambda}^{L,\gamma_0} = \operatorname{argmin} \frac{1}{2}\left(-2\lambda^*\hat{\beta} + \lambda^* G\lambda\right) + \sum_{k=1}^{p} \eta_k^{\gamma_0}|\lambda_k|$.

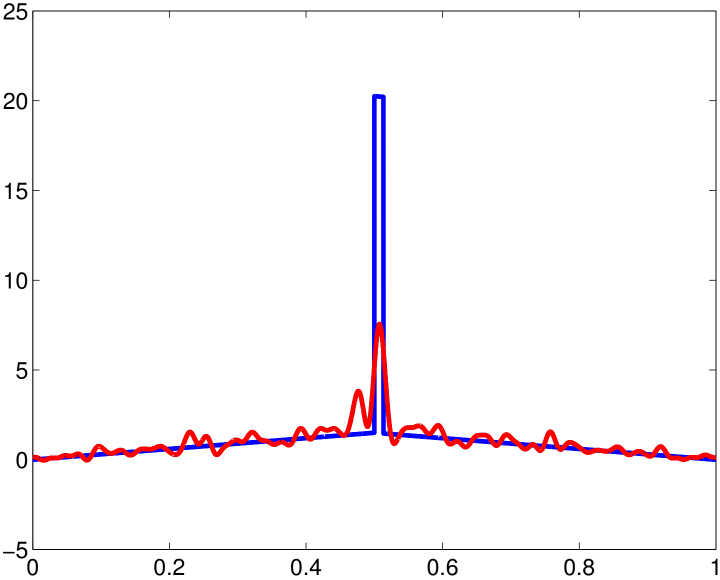
## Variantes

- Estimateurs non adaptatifs en remplaçant $\eta_k^{\gamma_0}$ par $\tilde{\eta}_k^{\gamma_0}$ :

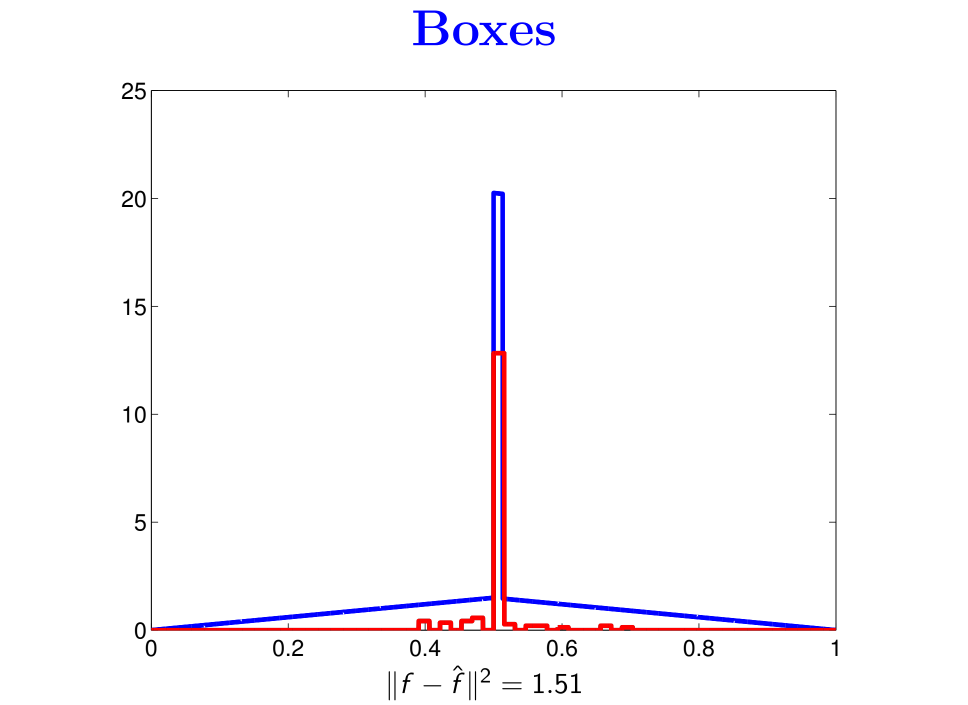
$$\tilde{\eta}_k^{\gamma_0} = \min\left(\sqrt{2\gamma_0 \log p}\,\frac{\|\phi_k\|_\infty}{\sqrt{n}}, \sqrt{2\gamma_0 \log p}\,\frac{\|f_0\|_\infty^{1/2}\|\phi_k\|_2}{\sqrt{n}} + \frac{2}{3}\gamma_0 \log p\,\frac{\|\phi_k\|_\infty}{n}\right)$$

- Amélioration possible avec une seconde étape de type moindre carré sur le support de $\hat{\lambda}$ (à la Gauss-Dantzig).

# Fourier

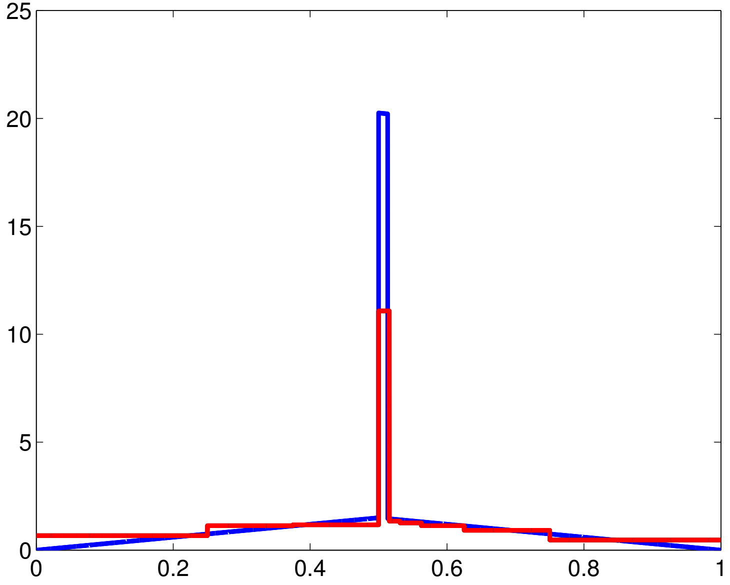



$\|f - \hat{f}\|^2 = 2.16$



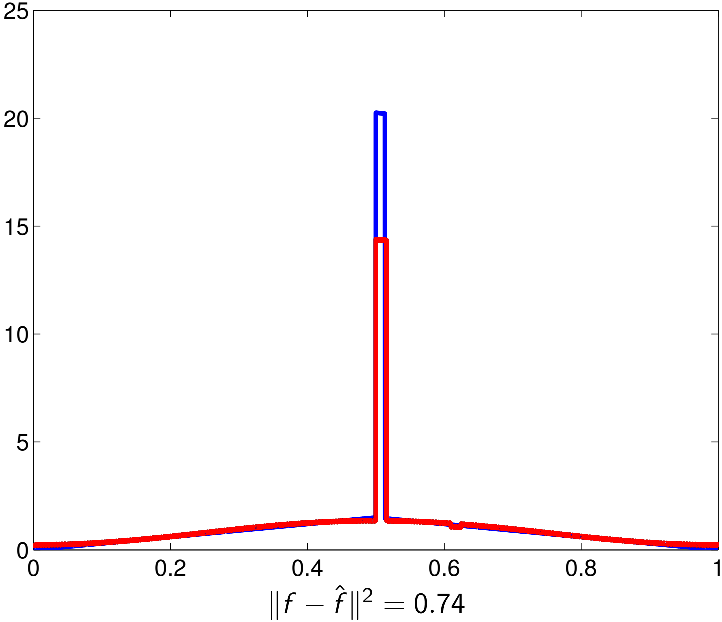
Boxes
25
20
15
10
5
0
0
0.2
0.4
0.6
0.8
1
$\|f - \hat{f}\|^2 = 1.51$

# Haar

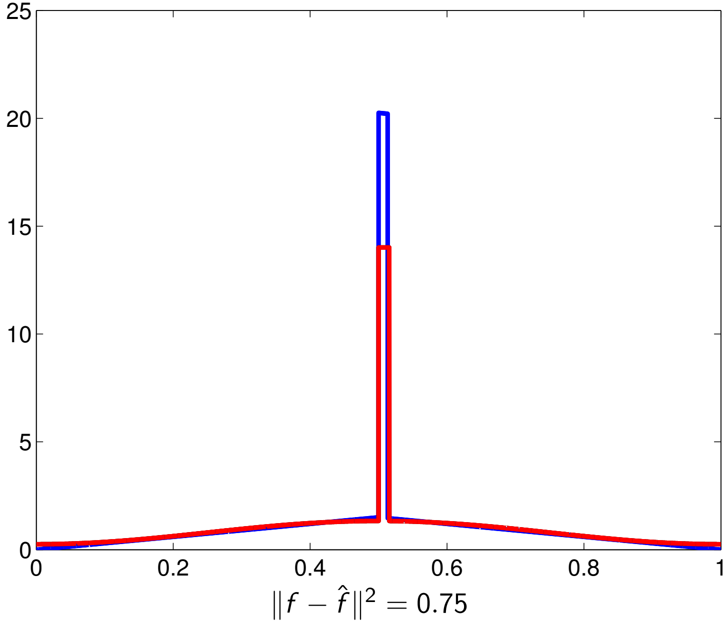



$\|f - \hat{f}\|^2 = 1.09$

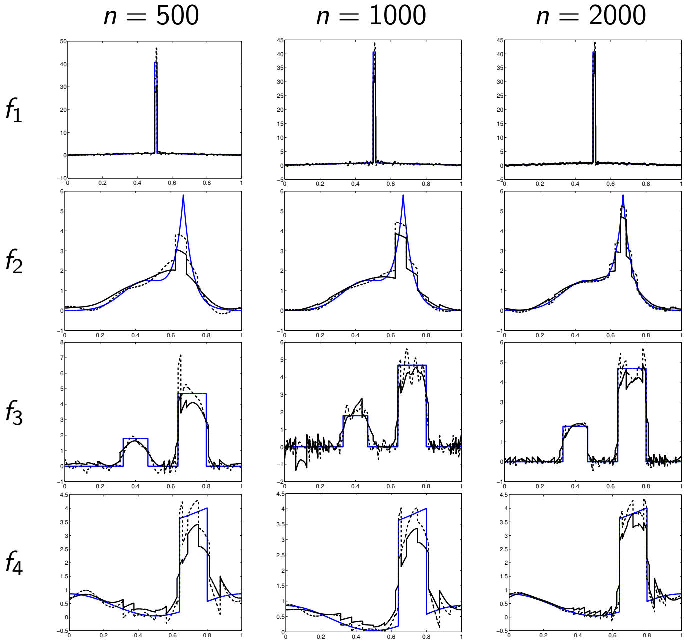
# Mix = Fourier + Boxes



# Mix2 = Fourier + Boxes + Haar



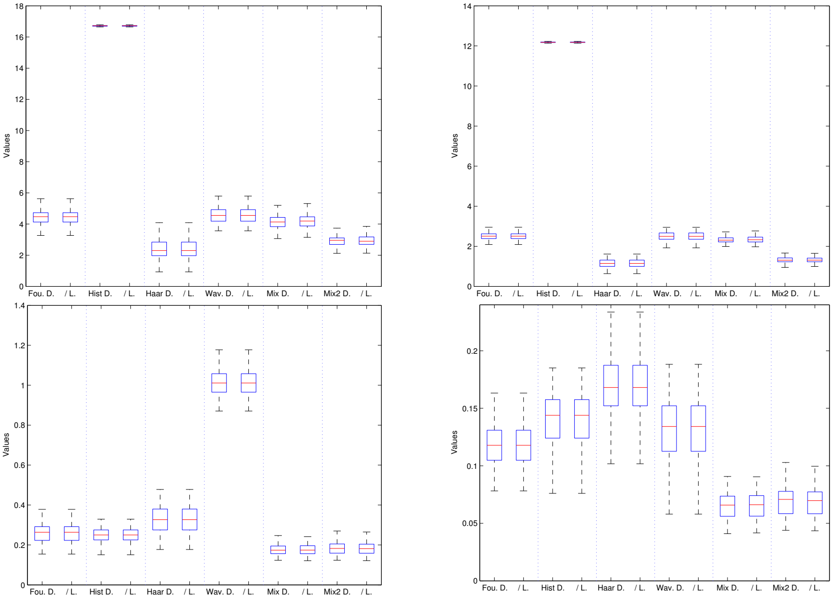
# Dantzig et Dantzig+LS



# Dantzig / Lasso $f_1/f_2$

$f_1$

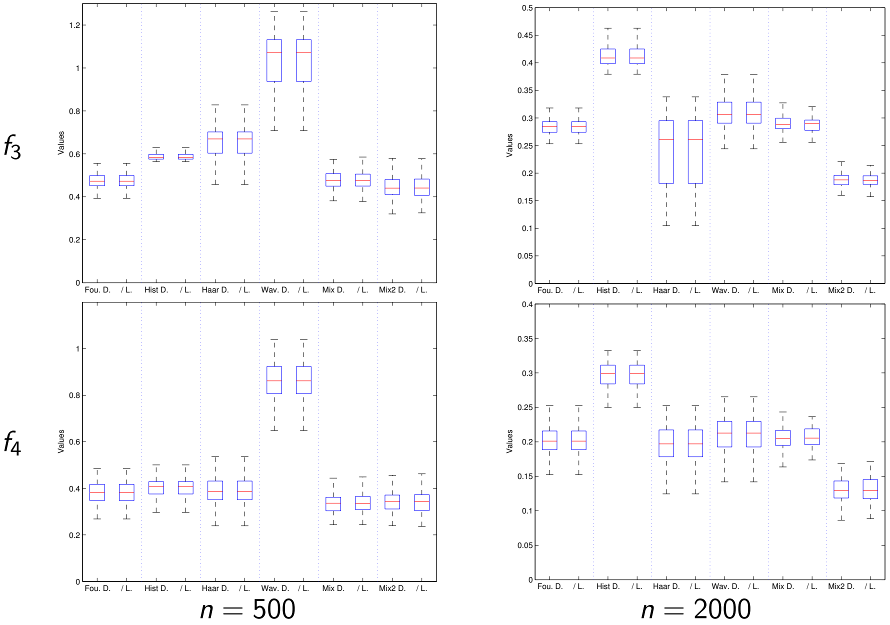
$f_2$



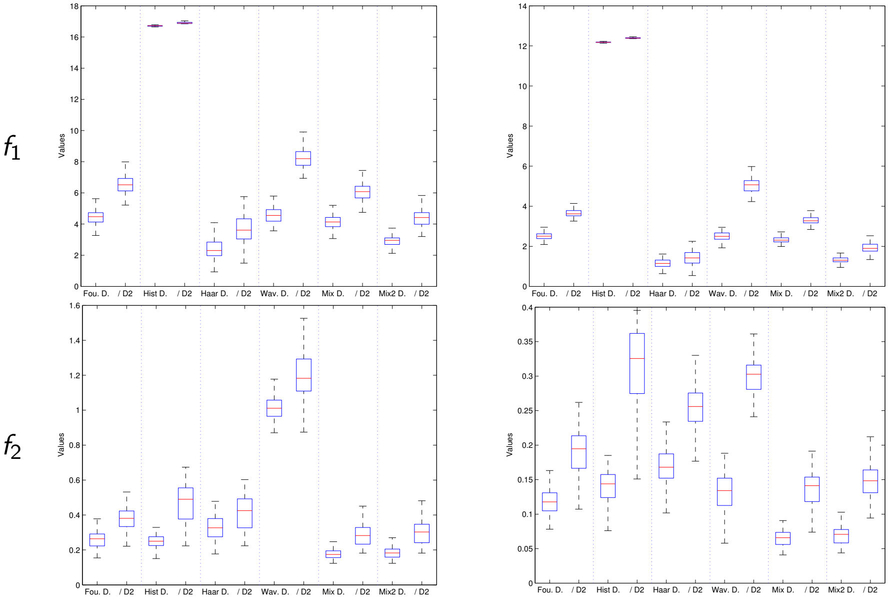

$n = 500$

$n = 2000$

# Dantzig / Lasso $f_3/f_4$

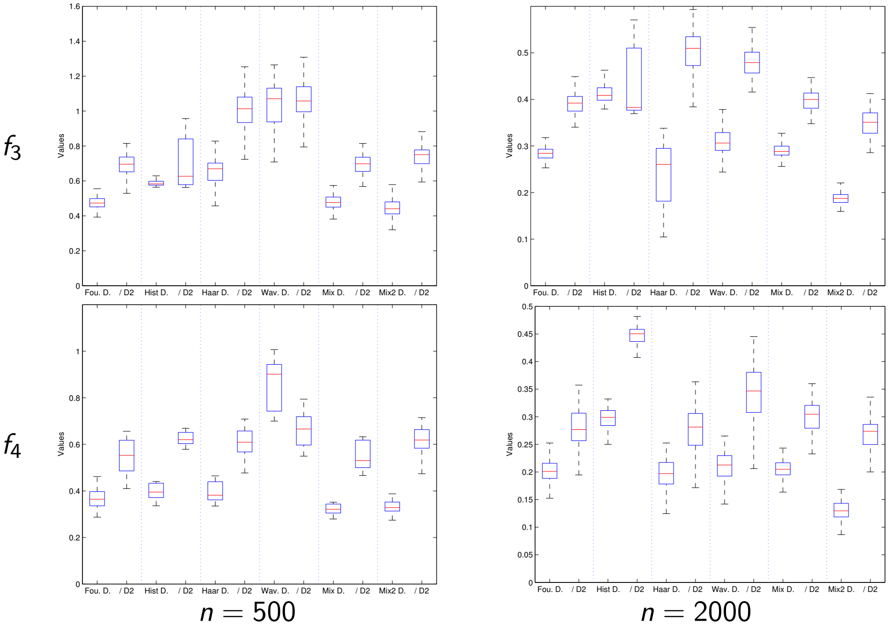


# Dantzig / Non adaptive D. $f_1/f_2$

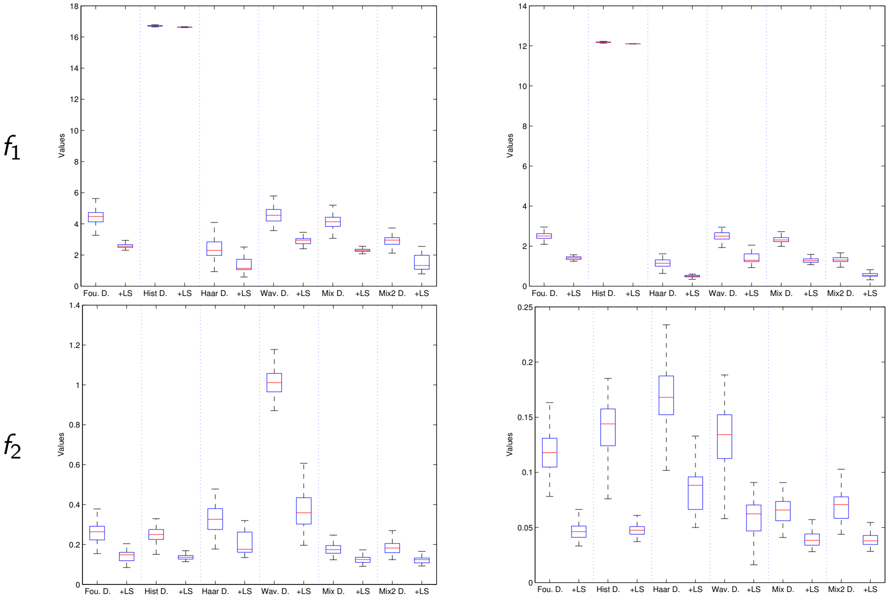



$n = 500$ $n = 2000$

# Dantzig / Non adaptive D. $f_3/f_4$

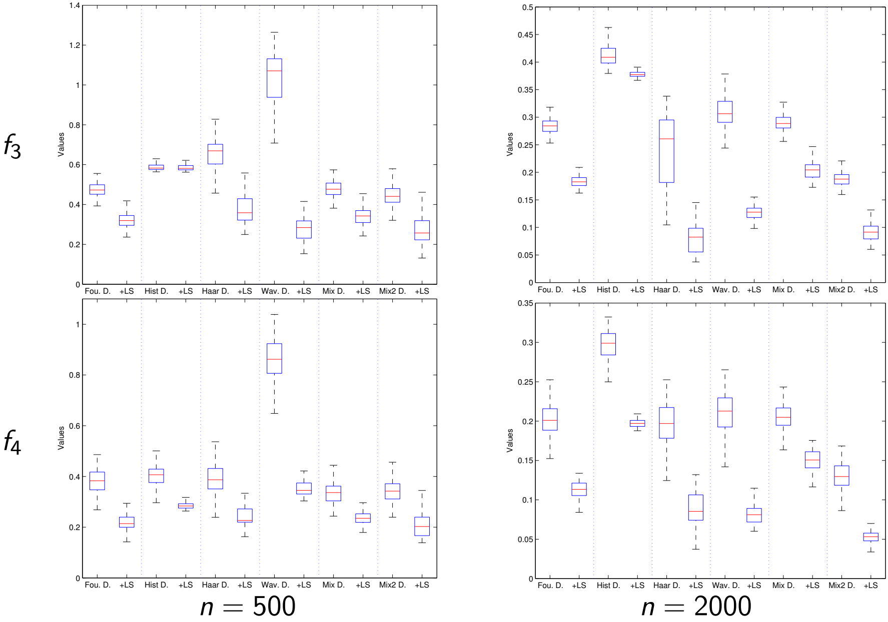


# Dantzig / Dantzig+LS $f_1/f_2$




$n = 500$ $n = 2000$

# Dantzig / Dantzig+LS $f_3/f_4$



# Hypothèses sur le dictionnaire

## Hypothèses de structure sur le dictionnaire

- Indispensable pour obtenir des résultats de type $\ell_0$.
- Variante autour des bases orthonormées...

## Hypothèse locale

- Hypothèse « minimale » pour obtenir un résultat.
- Soit $J_0$ un support, l'hypothèse locale $HL(J_0, \kappa_{J_0}, \mu_{J_0})$ est :

$$\|f_\lambda\|_2 \geq \frac{\kappa_{J_0}\|\lambda_{J_0}\|_{\ell_1} - \mu_{J_0}\left(\|\lambda_{J_0^C}\|_{\ell_1} - \|\lambda_{J_0}\|_{\ell_1}\right)_+}{\sqrt{|J_0|}}$$

  est vérifiée pour tout $\lambda$ avec $\kappa_{J_0} > 0$ et $\mu_{J_0} \geq 0$.
- Pour un dictionnaire $\mathcal{D}$ orthonormé, $\kappa_{J_0} = 1$ and $\mu_{J_0} = 0$ convient :

$$\|f_\lambda\|_2 = \|\lambda\|_{\ell_2} \geq \|\lambda_{J_0}\|_{\ell_2} \geq \frac{\|\lambda_{J_0}\|_{\ell_1}}{\sqrt{|J_0|}}$$

- RIP, RE et leurs variantes $\Longrightarrow$ $HL(J_0, \kappa_{J_0}, \mu_{J_0})$ dès que $|J_0| < c\frac{n}{\log n}$ avec $\kappa_{J_0}$ et $\mu_{J_0}$ ne dépendant que de la taille du support.

# Théorème Dantzig

## Théorème

- Sous $\Omega_{\gamma_1}$, événement de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma_1 - 1}$, si l'hypothèse locale $HL(J_0, \kappa_{J_0}, \mu_{J_0})$ est satisfaite alors pour tout $\lambda \in \mathbb{R}^p$

$$\|\hat{f}^{D,\gamma_0} - f_0\|_2^2 \leq \|f_\lambda - f_0\|_2^2 + 2\left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{\|\lambda_{J_0^C}\|_{\ell_1}^2}{|J_0|}$$
$$+ \left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{(\|\hat{\lambda}^{D,\gamma_0}\|_{\ell_1} - \|\lambda\|_{\ell_1})_+^2}{|J_0|}$$
$$+ 4\left(1 + \frac{1}{\kappa_{J_0}^2}\right) |J_0| (\|\eta^{\gamma_0}\|_{\ell_\infty} + \|\eta^{\gamma_1}\|_{\ell_\infty})^2 .$$

## Interprétation

- Biais ( Approx. par $f_\lambda$ $+$ Approx. en dehors de $J_0$ ) / Variance
- "Faisabilité" de la contrainte de Dantzig

# Théorème Dantzig

## Théorème

- Sous $\Omega_{\gamma_1}$, événement de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma_1 - 1}$, si l'hypothèse locale $HL(J_0, \kappa_{J_0}, \mu_{J_0})$ est satisfaite alors pour tout $\lambda \in \mathbb{R}^p$

$$\|\hat{f}^{D,\gamma_0} - f_0\|_2^2 \leq \boxed{\|f_\lambda - f_0\|_2^2} + 2\left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{\|\lambda_{J_0^C}\|_{\ell_1}^2}{|J_0|}$$

$$+ \left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{(\|\hat{\lambda}^{D,\gamma_0}\|_{\ell_1} - \|\lambda\|_{\ell_1})_+^2}{|J_0|}$$

$$+ 4\left(1 + \frac{1}{\kappa_{J_0}^2}\right) |J_0| (\|\eta^{\gamma_0}\|_{\ell_\infty} + \|\eta^{\gamma_1}\|_{\ell_\infty})^2 .$$

## Interprétation

- Biais ( Approx. par $f_\lambda$ + Approx. en dehors de $J_0$ ) / Variance
- "Faisabilité" de la contrainte de Dantzig

# Théorème Dantzig

## Théorème

- Sous $\Omega_{\gamma_1}$, événement de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma_1 - 1}$, si l'hypothèse locale $HL(J_0, \kappa_{J_0}, \mu_{J_0})$ est satisfaite alors pour tout $\lambda \in \mathbb{R}^p$

$$\|\hat{f}^{D,\gamma_0} - f_0\|_2^2 \leq \|f_\lambda - f_0\|_2^2 + 2\left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{\|\lambda_{J_0^C}\|_{\ell_1}^2}{|J_0|}$$
$$+ \left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{(\|\hat{\lambda}^{D,\gamma_0}\|_{\ell_1} - \|\lambda\|_{\ell_1})_+^2}{|J_0|}$$
$$+ 4\left(1 + \frac{1}{\kappa_{J_0}^2}\right) |J_0| (\|\eta^{\gamma_0}\|_{\ell_\infty} + \|\eta^{\gamma_1}\|_{\ell_\infty})^2 .$$

## Interprétation

- Biais ( Approx. par $f_\lambda$ + Approx. en dehors de $J_0$ ) / Variance
- "Faisabilité" de la contrainte de Dantzig

# Théorème Dantzig

## Théorème

- Sous $\Omega_{\gamma_1}$, événement de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma_1 - 1}$, si l'hypothèse locale $HL(J_0, \kappa_{J_0}, \mu_{J_0})$ est satisfaite alors pour tout $\lambda \in \mathbb{R}^p$

$$\|\hat{f}^{D,\gamma_0} - f_0\|_2^2 \leq \|f_\lambda - f_0\|_2^2 + 2\left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{\|\lambda_{J_0^C}\|_{\ell_1}^2}{|J_0|}$$
$$+ \left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{(\|\hat{\lambda}^{D,\gamma_0}\|_{\ell_1} - \|\lambda\|_{\ell_1})_+^2}{|J_0|}$$
$$+ 4\left(1 + \frac{1}{\kappa_{J_0}^2}\right) |J_0| (\|\eta^{\gamma_0}\|_{\ell_\infty} + \|\eta^{\gamma_1}\|_{\ell_\infty})^2 .$$

## Interprétation

- Biais ( Approx. par $f_\lambda$ + Approx. en dehors de $J_0$ ) / Variance
- "Faisabilité" de la contrainte de Dantzig

# Théorème Dantzig

## Théorème

- Sous $\Omega_{\gamma_1}$, événement de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma_1 - 1}$, si l'hypothèse locale $HL(J_0, \kappa_{J_0}, \mu_{J_0})$ est satisfaite alors pour tout $\lambda \in \mathbb{R}^p$

$$\|\hat{f}^{D,\gamma_0} - f_0\|_2^2 \leq \|f_\lambda - f_0\|_2^2 + 2\left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{\|\lambda_{J_0^C}\|_{\ell_1}^2}{|J_0|}$$
$$+ \left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{(\|\hat{\lambda}^{D,\gamma_0}\|_{\ell_1} - \|\lambda\|_{\ell_1})_+^2}{|J_0|}$$
$$+ 4\left(1 + \frac{1}{\kappa_{J_0}^2}\right)|J_0|(\|\eta^{\gamma_0}\|_{\ell_\infty} + \|\eta^{\gamma_1}\|_{\ell_\infty})^2 .$$

## Interprétation

- Biais ( Approx. par $f_\lambda$ + Approx. en dehors de $J_0$ ) / Variance
- "Faisabilité" de la contrainte de Dantzig

# Variance

## Terme de variance

- Dans le théorème : $4\left(1+\frac{1}{\kappa_{J_0}^2}\right)|J_0|\left(\|\eta^{\gamma_0}\|_{\ell_\infty}+\|\eta^{\gamma_1}\|_{\ell_\infty}\right)^2$
- Sous $\Omega_{\gamma_1}$,

$$\left(\|\eta^{\gamma_0}\|_{\ell_\infty}+\|\eta^{\gamma_1}\|_{\ell_\infty}\right)^2 \leq 8(\gamma_0 c_{\gamma_0,p}+\gamma_1 c_{\gamma_1,p})\log p\frac{\sup_k \sigma_k^2}{n-1}$$
$$+2\Bigg(2\left(\sqrt{\gamma_0 c_{\gamma_0,p}\gamma_1}+\sqrt{c_{\gamma_1,p}\gamma_1}\right) + \frac{7}{3}\left(\gamma_0 c_{\gamma_0,p}+\gamma_1 c_{\gamma_1,p}\right)\Bigg)^2$$
$$\times \log^2 p\frac{\sup_k \|\phi_k\|_\infty^2}{(n-1)^2}.$$

- Simplification dans le théorème !

# Théorème Dantzig simplifié

## Théorème

- Sous $\Omega_{\gamma_1}$, événement de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma_1 - 1}$, si l'hypothèse locale $HL(J_0, \kappa_{J_0}, \mu_{J_0})$ est satisfaite alors pour tout $\lambda \in \mathbb{R}^p$

$$\|\hat{f}^{D,\gamma_0} - f_0\|_2^2 \lesssim \|f_\lambda - f_0\|_2^2 + 2\left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{\|\lambda_{J_0^C}\|_{\ell_1}^2}{|J_0|}$$
$$+ \left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{(\|\hat{\lambda}^{D,\gamma_0}\|_{\ell_1} - \|\lambda\|_{\ell_1})_+^2}{|J_0|}$$
$$+ 32(\gamma_0 c_{\gamma_0,p} + \gamma_1 c_{\gamma_1,p})\left(1 + \frac{1}{\kappa_{J_0}^2}\right)|J_0| \log p \frac{\sup_k \sigma_k^2}{n}.$$

## Interprétation

- Biais ( Approx. par $f_\lambda$ + Approx. en dehors de $J_0$ ) / Variance
- "Faisabilité" de la contrainte de Dantzig

# Contrainte de Dantzig

## "Faisabilité" de la contrainte de Dantzig

- Dans le théorème : $\left(1+\frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{(\|\hat{\lambda}^{D,\gamma_0}\|_{\ell_1}-\|\lambda\|_{\ell_1})_+^2}{|J_0|}$
- Contrôle difficile : pas de lien a priori entre $\lambda$ et $\hat{\lambda}^{D,\gamma_0}$...
- Disparition si $\|\lambda\|_{\ell_1} \geq \|\hat{\lambda}^{D,\gamma_0}\|_{\ell_1}$.
- Disparition si $\lambda$ satisfait la contrainte de Dantzig.

## Corollaire

- Si $\gamma_1 \leq \gamma_0$, sous $\Omega_{\gamma_1}$, événement de proba $\geq 1-2\left(\frac{1}{p}\right)^{\gamma_1-1}$, si l'hypothèse locale $HL(J_0,\kappa_{J_0},\mu_{J_0})$ est satisfaite et si $f_{\lambda_0}=P_{\mathcal{D}}f_0$ alors

$$\|\hat{f}^{D,\gamma_0}-f_0\|_2^2 \lesssim \|P_{\mathcal{D}}f_0-f_0\|_2^2 + 2\left(1+\frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2\frac{\|\lambda_{0J_0^C}\|_{\ell_1}^2}{|J_0|}$$

$$+\,32(\gamma_0 c_{\gamma_0,p}+\gamma_1 c_{\gamma_1,p})\left(1+\frac{1}{\kappa_{J_0}^2}\right)|J_0|\log p\frac{\sup_k \sigma_k^2}{n}.$$

# Théorème Dantzig/Lasso

## Dantzig/Lasso

- Proximité des estimateurs $\Longrightarrow$ proximité des solutions ?
- Proximité déjà prouvée dans un sens car $\hat{f}^{L,\gamma_0}$ est faisable...
- Contrôle plus fin possible.

## Théorème (sans le terme de Dantzig)

- Sous $\Omega_{\gamma_1}$, événement de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma_1 - 1}$, si l'hypothèse locale $HL(J_0, \kappa_{J_0}, \mu_{J_0})$ est satisfaite alors

$$\left| \|\hat{f}^{D,\gamma_0} - f_0\|_2^2 - \|\hat{f}^{L,\gamma_0} - f_0\|_2^2 \right|$$

$$\leq 2\left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{\|\widehat{\lambda}^{L,\gamma_0}_{J_0^C}\|_{\ell_1}^2}{|J_0|}$$

$$+ 32(\gamma_0 c_{\gamma_0,p} + \gamma_1 c_{\gamma_1,p}) \left(1 + \frac{1}{\kappa_{J_0}^2}\right) |J_0| \log p \frac{\sup_k \sigma_k^2}{n}$$

# Pénalité minimale

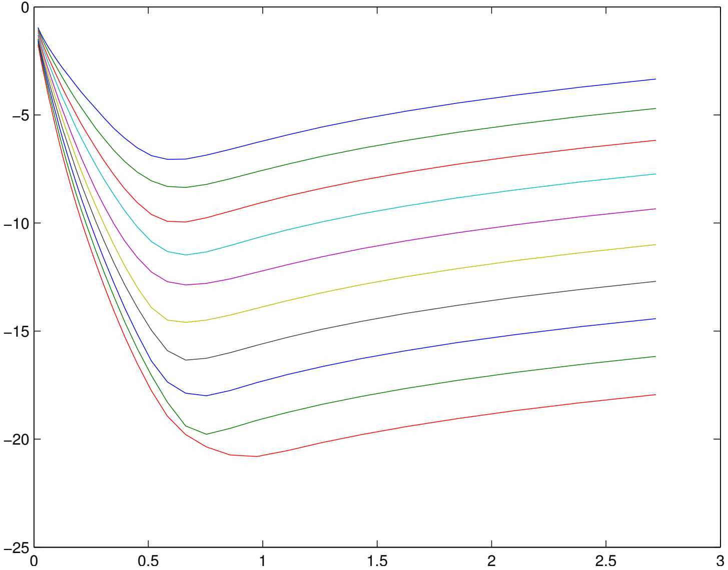
## Choix de $\gamma_0$

- Comment choisir $\gamma_0$ ?
- Paramètre de pénalisation :
  - $\gamma_0$ grand : surlissage (grand biais),
  - $\gamma_0$ petit : instabilité (grande variance).
- Inégalité oracle pour tout $\gamma_0$ :
  - $\gamma_0$ grand : surpénalisation de la variance dans l'inégalité oracle,
  - $\gamma_0$ petit : présence du terme Dantzig en $(\|\lambda^{S,\gamma_0}\|_{\ell_1} - \|\lambda\|_{\ell_1})_+)$ difficile à contrôler.

## Pénalité minimale

- Bon calibrage : $\gamma_0 = 1$ !
- Résultat théorique :
  - Si $\gamma_0 > 1$ : avec proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma_0 - 1}$, inégalité oracle sans le terme Dantzig pour tout $\lambda$ tel que $f_\lambda = P_{\mathcal{D}} f_0$.
  - Si $\gamma_0 < 1$ : théorème montrant que l'estimateur est mauvais (sur un cas simple d'estimation de la densité uniforme sur $[0, 1]$ dans la base de Haar).

# Pénalité minimale



# Conclusion

## Estimation de densité par Dantzig adaptatif

- Nouvelle méthode d'estimation de densité par pénalisation $\ell_1$ adaptative.
- Utilisation d'inégalités de concentration fine pour obtenir cette pénalisation adaptative.
- Inégalité oracle en probabilité sur l'estimateur.
- Calibration de la pénalité : pénalité minimale.
- Hypothèses assez faibles sur la structure du dictionnaire.
- Lien théorique avec le Lasso adaptatif.

## Perspectives

- Performance de la méthode en deux étapes.
- Passage à des résultats en espérance.
- Relaxation des hypothèses sur la structure par passage à des résultats en proba sur les signaux (modélisation probabiliste des signaux).