Estimation de densité par pénalisation ℓ_1 minimale

Erwan Le Pennec

LPMA, Univ. Paris Diderot / SELECT, INRIA Saclay - Univ. Paris Sud

en collaboration avec Karine Bertin (Univ. de Valparaiso, Chili)

et Vincent Rivoirard (Univ. Paris Sud - ENS Paris)

Novembre 2009

L'évènement Ω_γ

- Contraintes :

$$\eta_k^\gamma = \sqrt{2\gamma \log p}\frac{\hat{\sigma}_k}{\sqrt{n}} + \frac{14}{3}\gamma \log p \frac{\|\phi_k\|_\infty}{n}$$

- Évènement Ω_γ de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma-1}$:

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

- De plus sous Ω_γ :

$$\hat{\sigma}_k \leq \sigma_k + 2\sqrt{2\gamma \log p}\frac{\|\phi_k\|_\infty}{\sqrt{n}}$$

$$\|\eta^{\gamma'}\|_\infty \leq \sqrt{2\gamma' \log p}\frac{\sigma_k}{\sqrt{n}} + 4\left(\sqrt{\gamma\gamma'} + \frac{7}{6}\gamma'\right) \log p \frac{\|\phi_k\|_\infty}{n}$$

Estimateurs

- Contraintes :

$$\eta_k^{\gamma_0} = \sqrt{2\gamma_0 \log p}\frac{\hat{\sigma}_k}{\sqrt{n}} + \frac{14}{3}\gamma_0 \log p \frac{\|\phi_k\|_\infty}{n}.$$

- Estimateur de Dantzig adaptatif :

$$\hat{\lambda}^{D,\gamma_0} = \operatorname{argmin}\|\lambda\|_{\ell_1} \quad \text{sous} \quad \forall k, |(G\lambda)_k - \hat{\beta}_k| \leq \eta_k^{\gamma_0}.$$

- Estimateur de Lasso adaptatif :

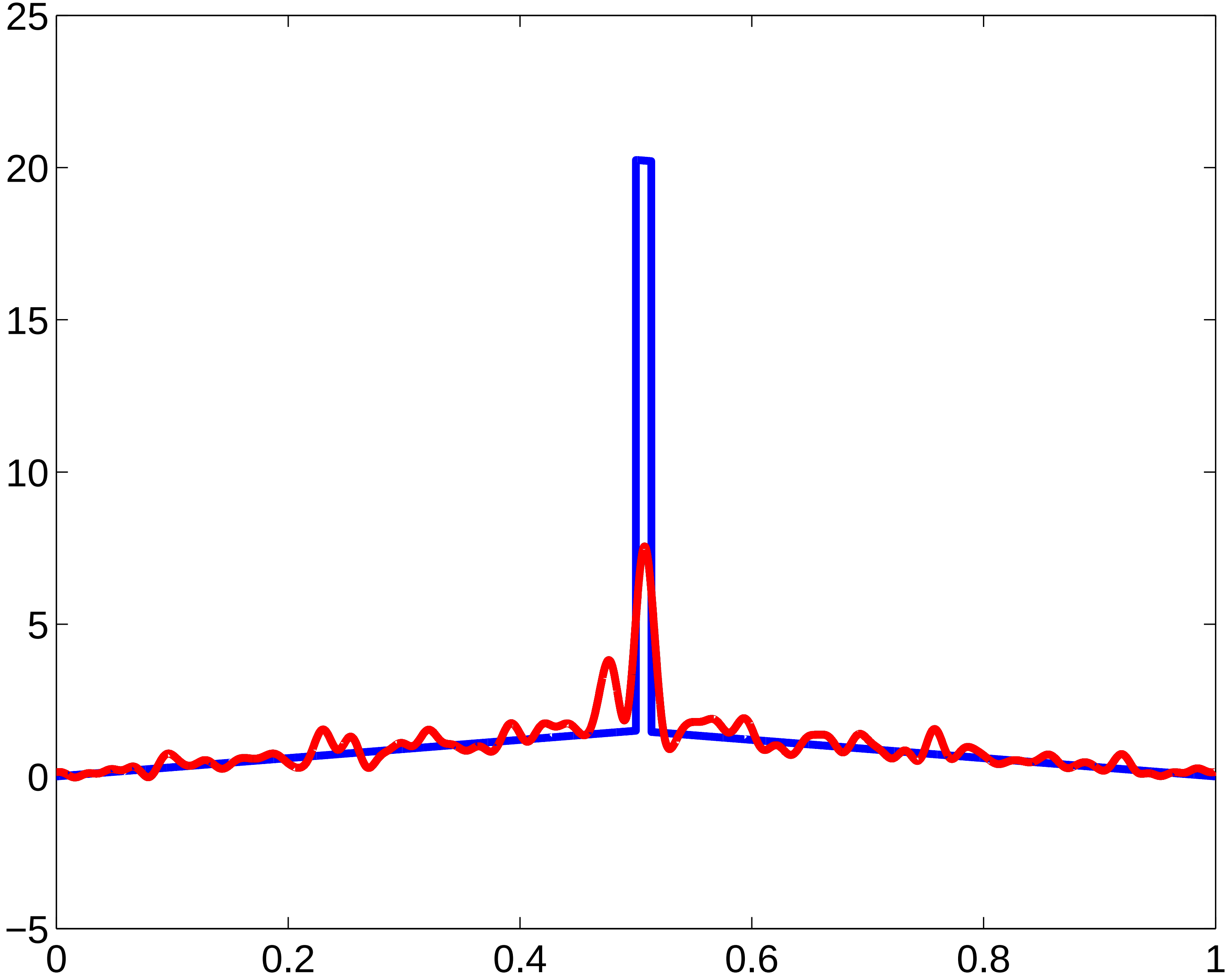
$$\hat{\lambda}^{L,\gamma_0} = \operatorname{argmin}\frac{1}{2}\left(-2\lambda^*\hat{\beta} + \lambda^* G\lambda + \sum_{k=1}^{p} \eta_k^{\gamma_0}|\lambda_k|\right).$$

- Estimateur non adaptatif en remplaçant $\eta_k^{\gamma_0}$ par $\tilde{\eta}_k^{\gamma_0}$

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

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

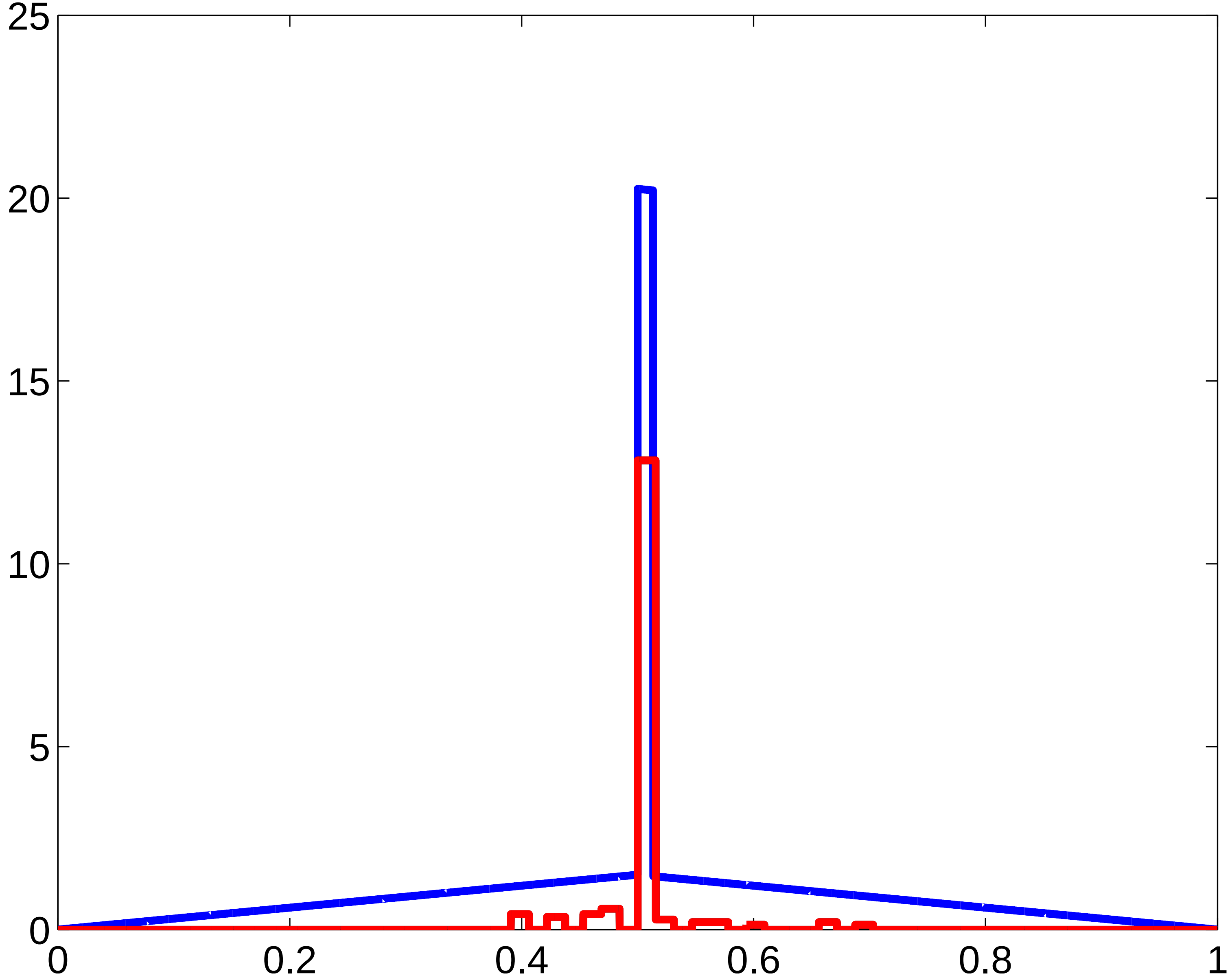
Fourier



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

$$\|\hat{\lambda}\|_0 = 67$$

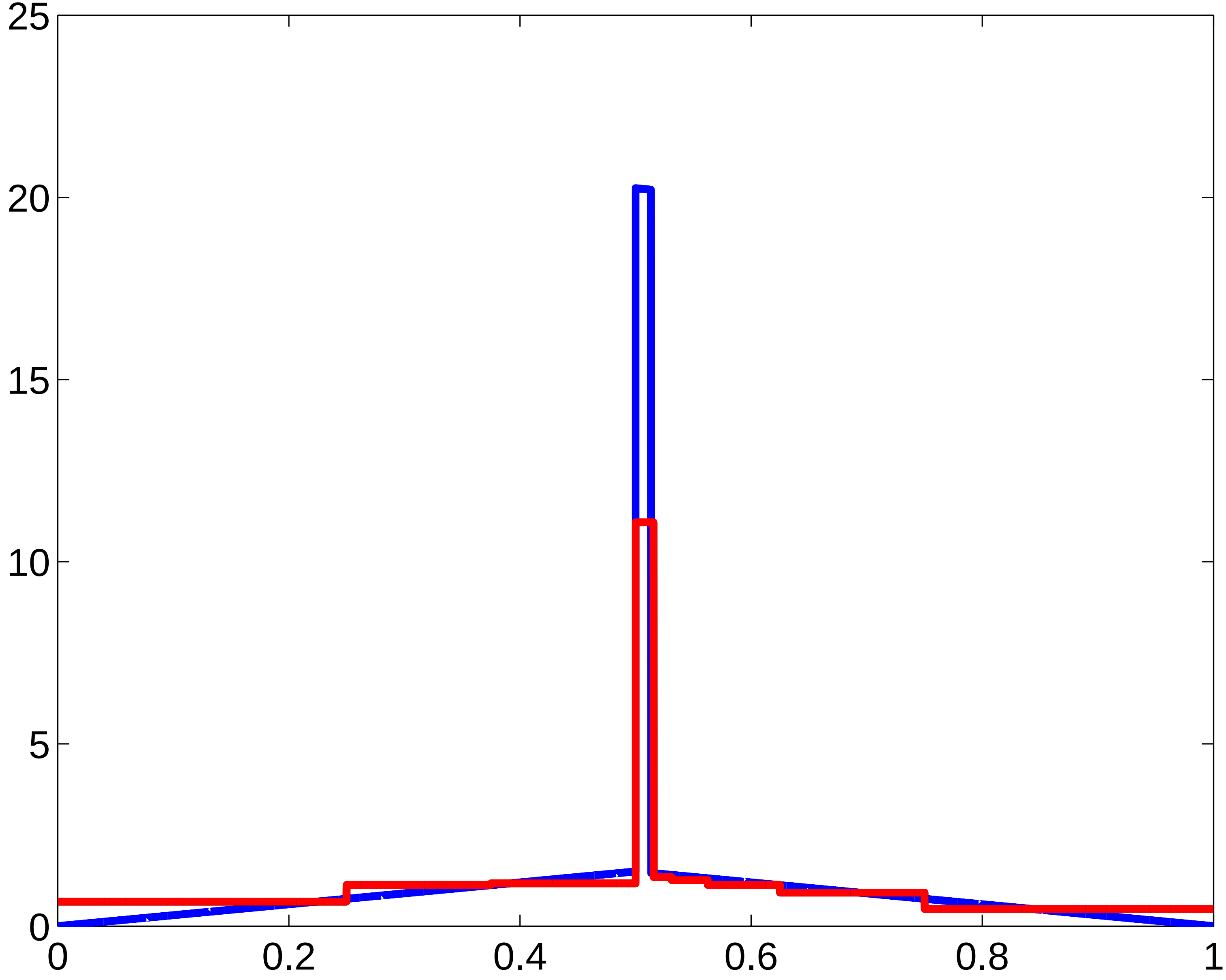
Boxes



$$\|f - \hat{f}\|^2 = 1.51$$

$$\|\hat{\lambda}\|_0 = 7$$

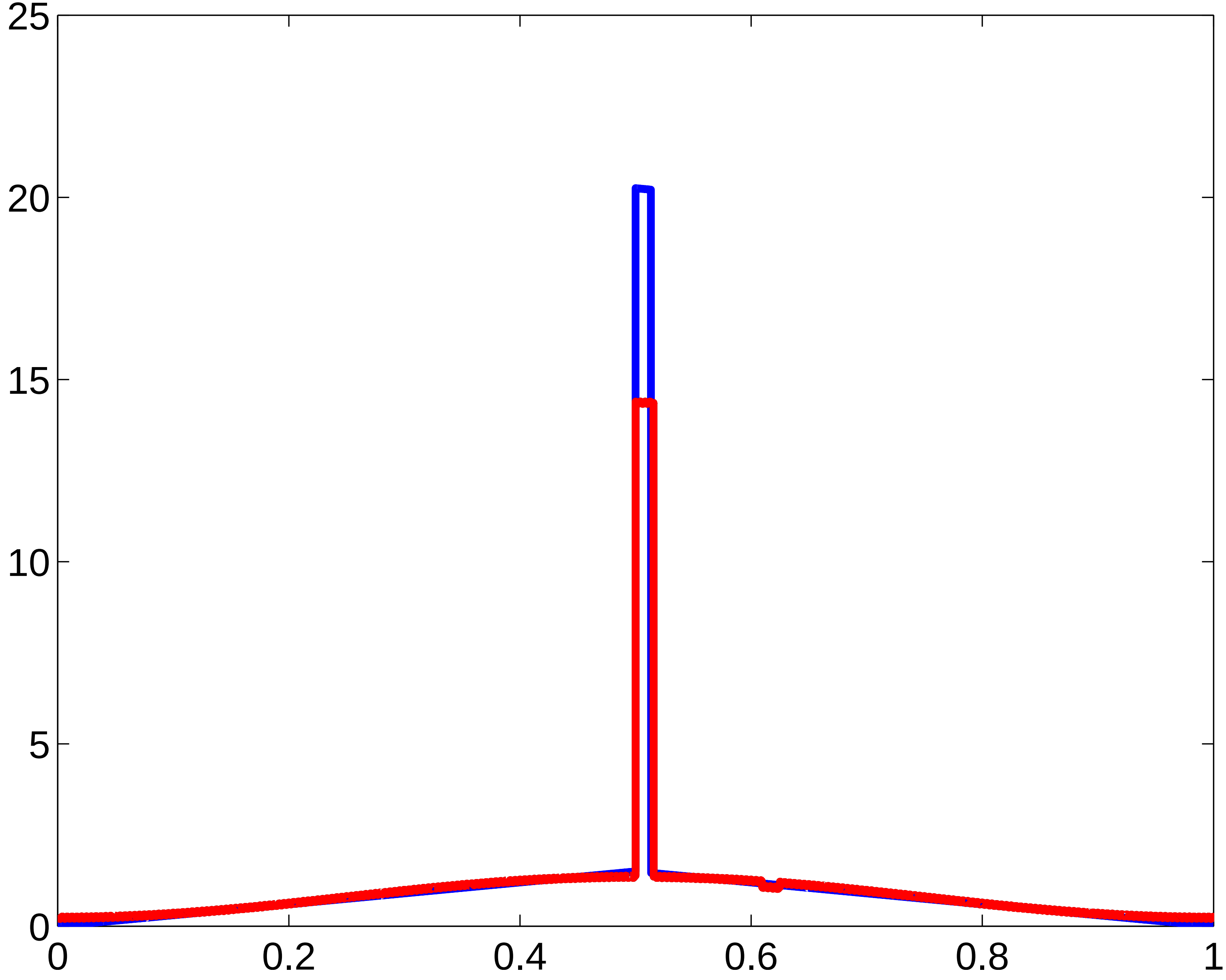
Haar



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

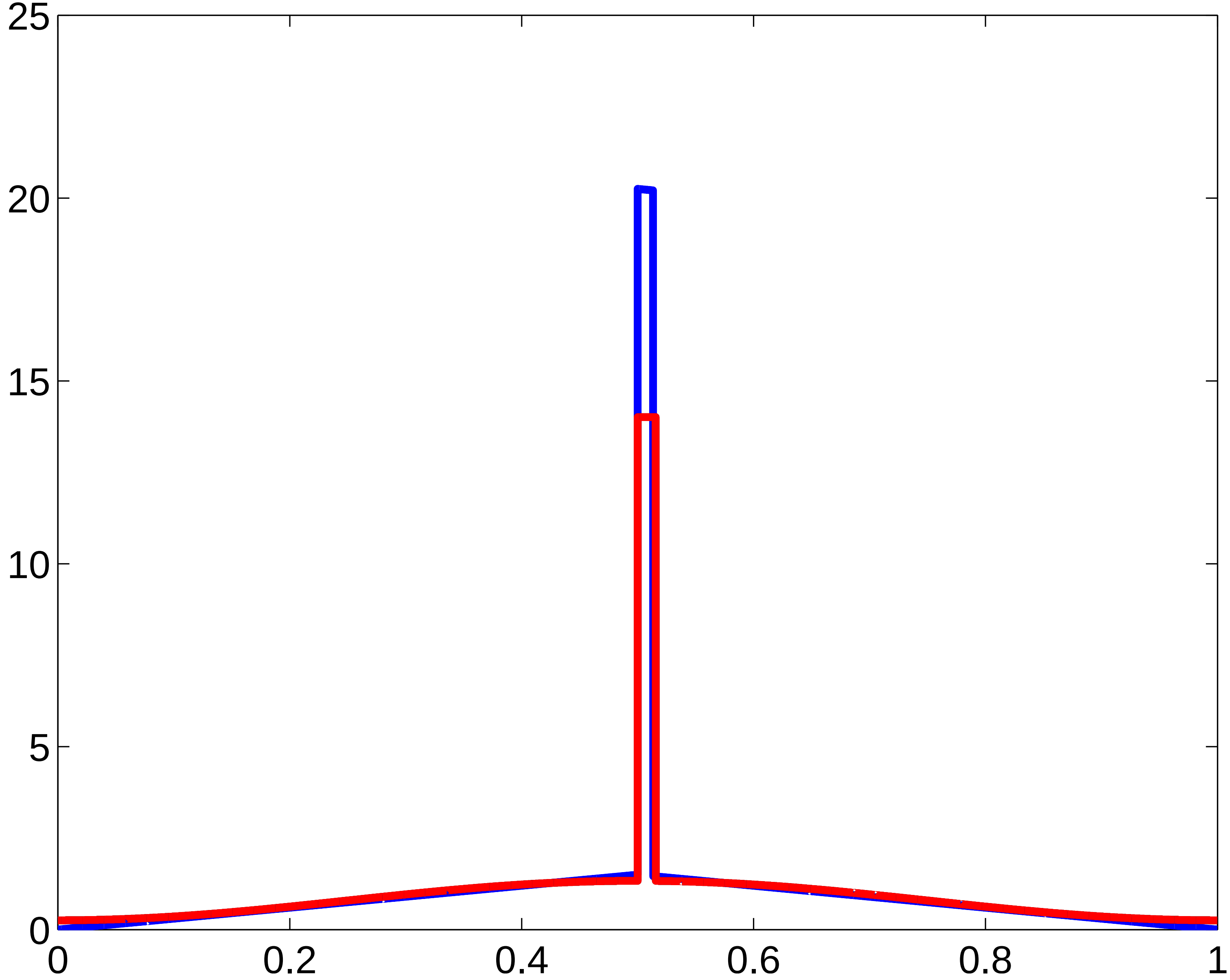
$\|\hat{\lambda}\|_0 = 8$

Fourier + Boxes



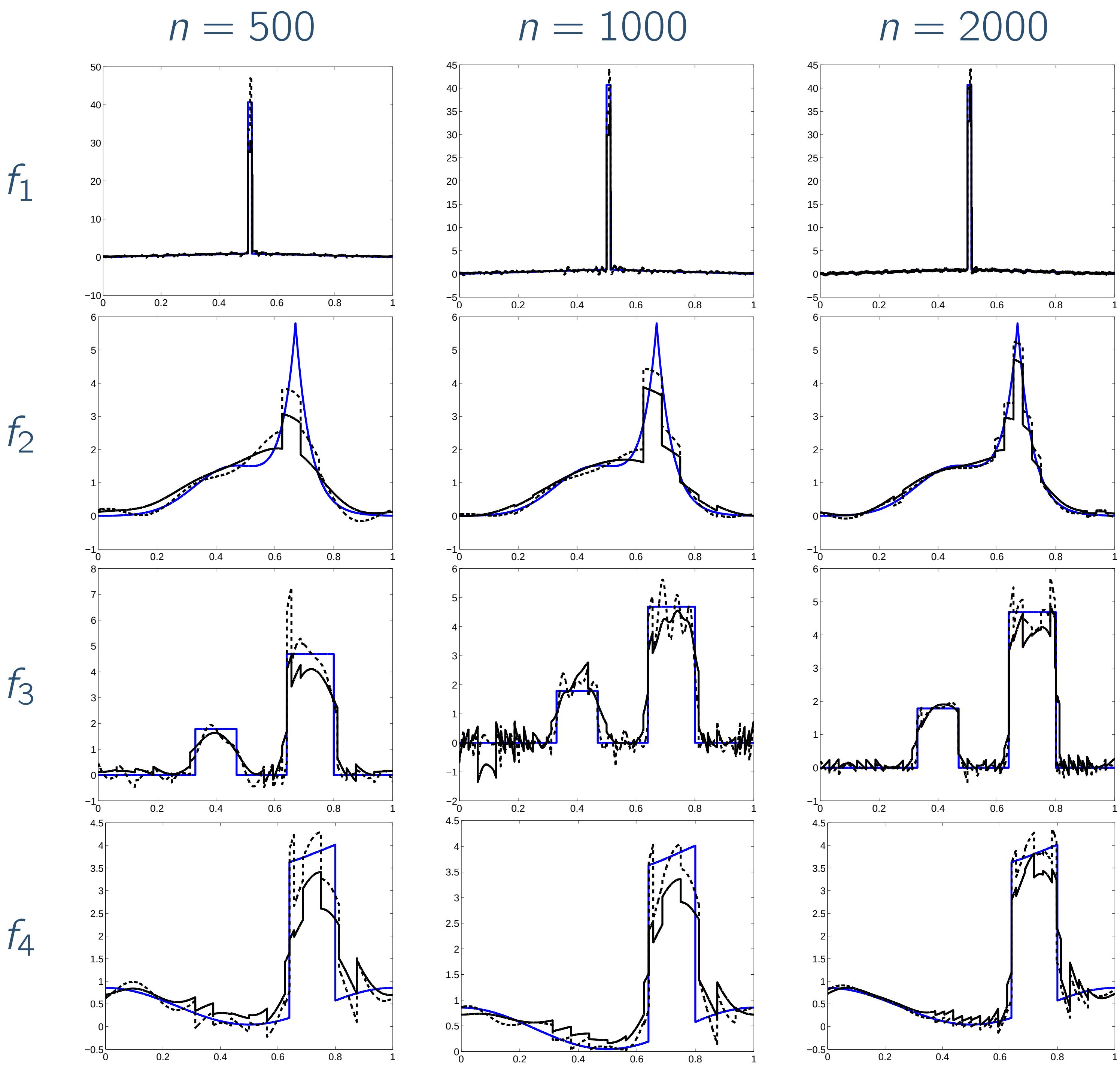
$$\|f - \hat{f}\|^2 = 0.74$$
$$\|\hat{\lambda}\|_0 = 5$$

Fourier + Boxes + Haar

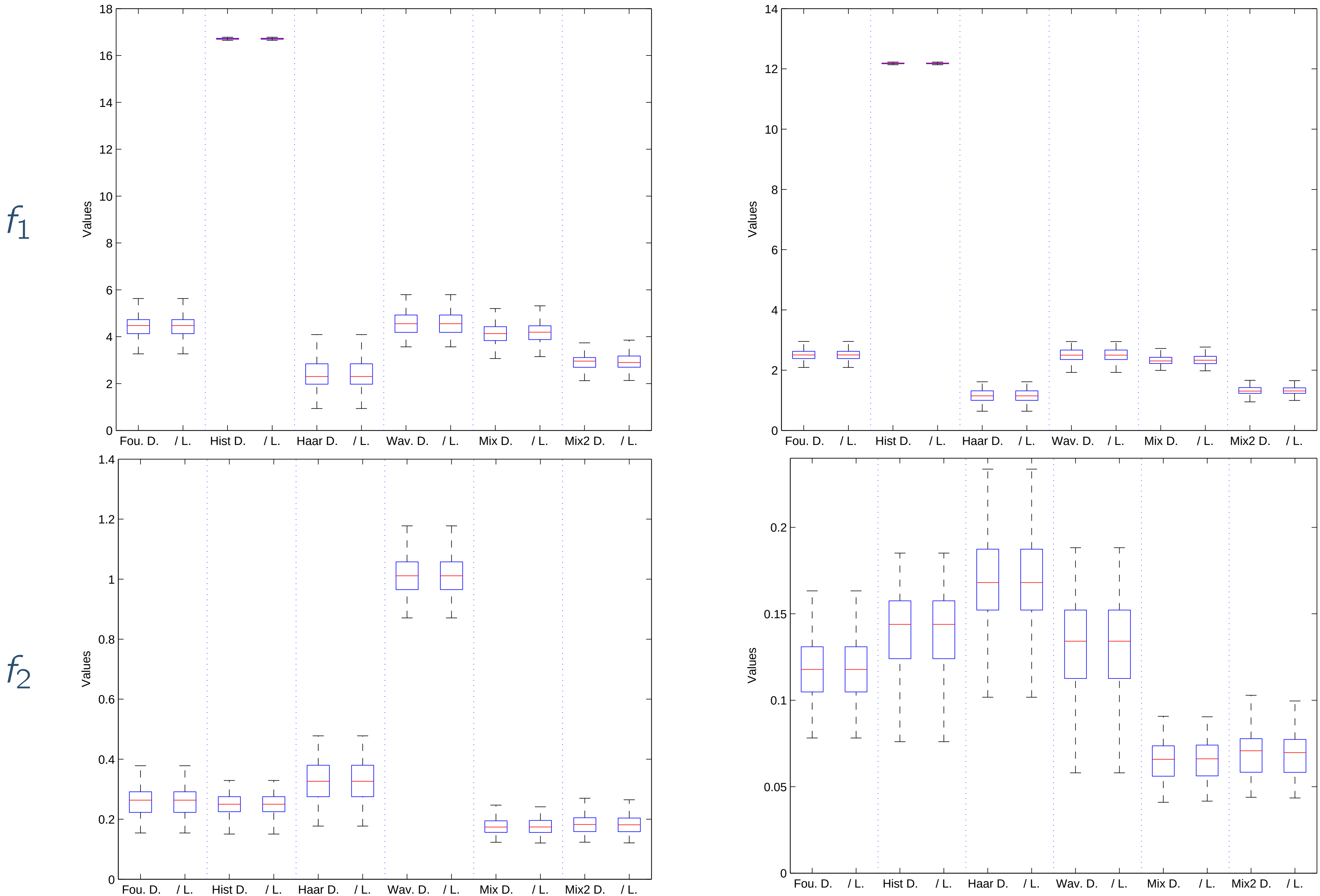


$$\|f - \hat{f}\|^2 = 0.75$$
$$\|\hat{\lambda}\|_0 = 4$$

Dantzig et Dantzig+LS



Dantzig / Lasso f_1/f_2

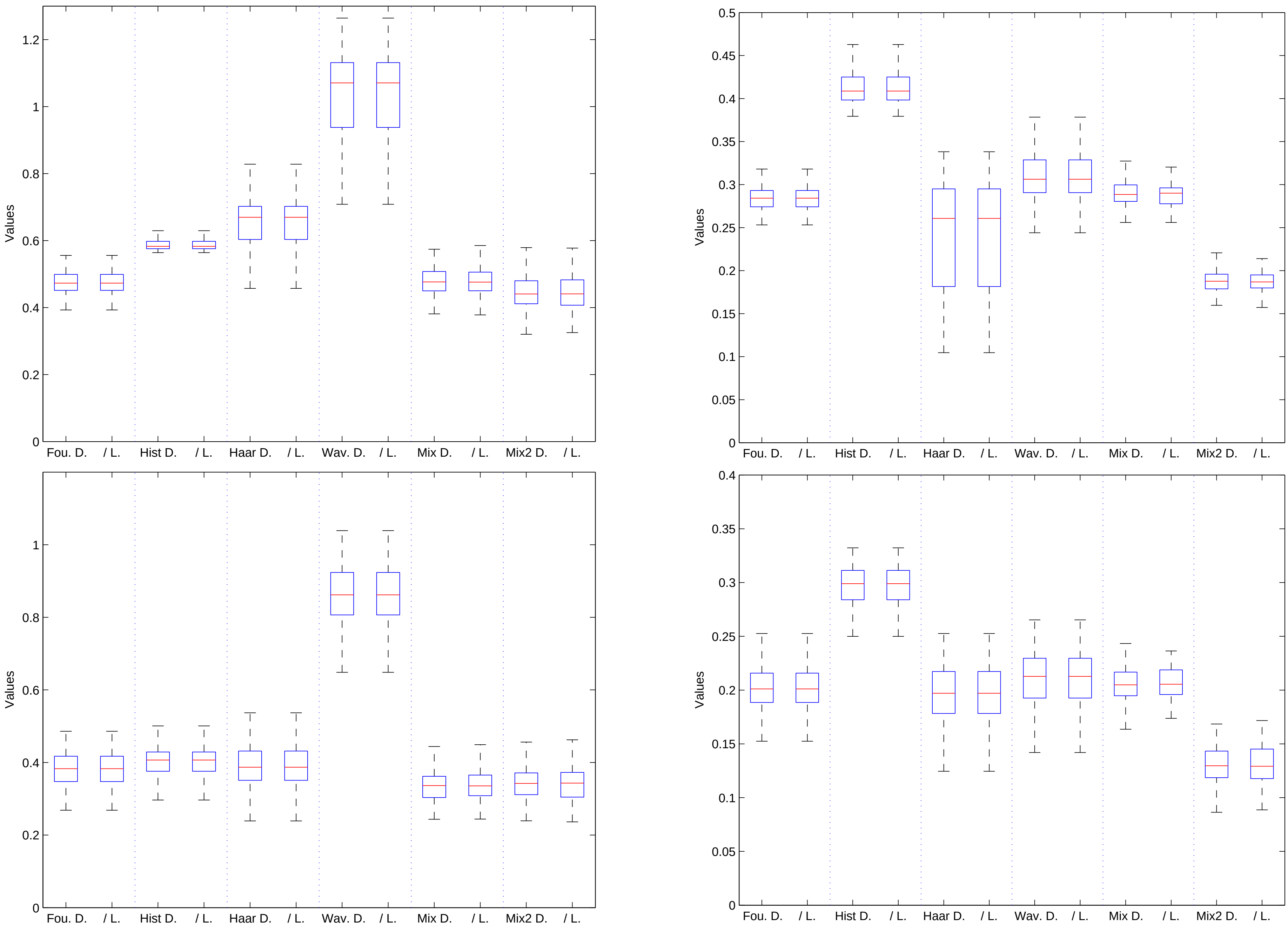


$n = 500$ $n = 2000$

Dantzig / Lasso f_3/f_4

f_3

f_4

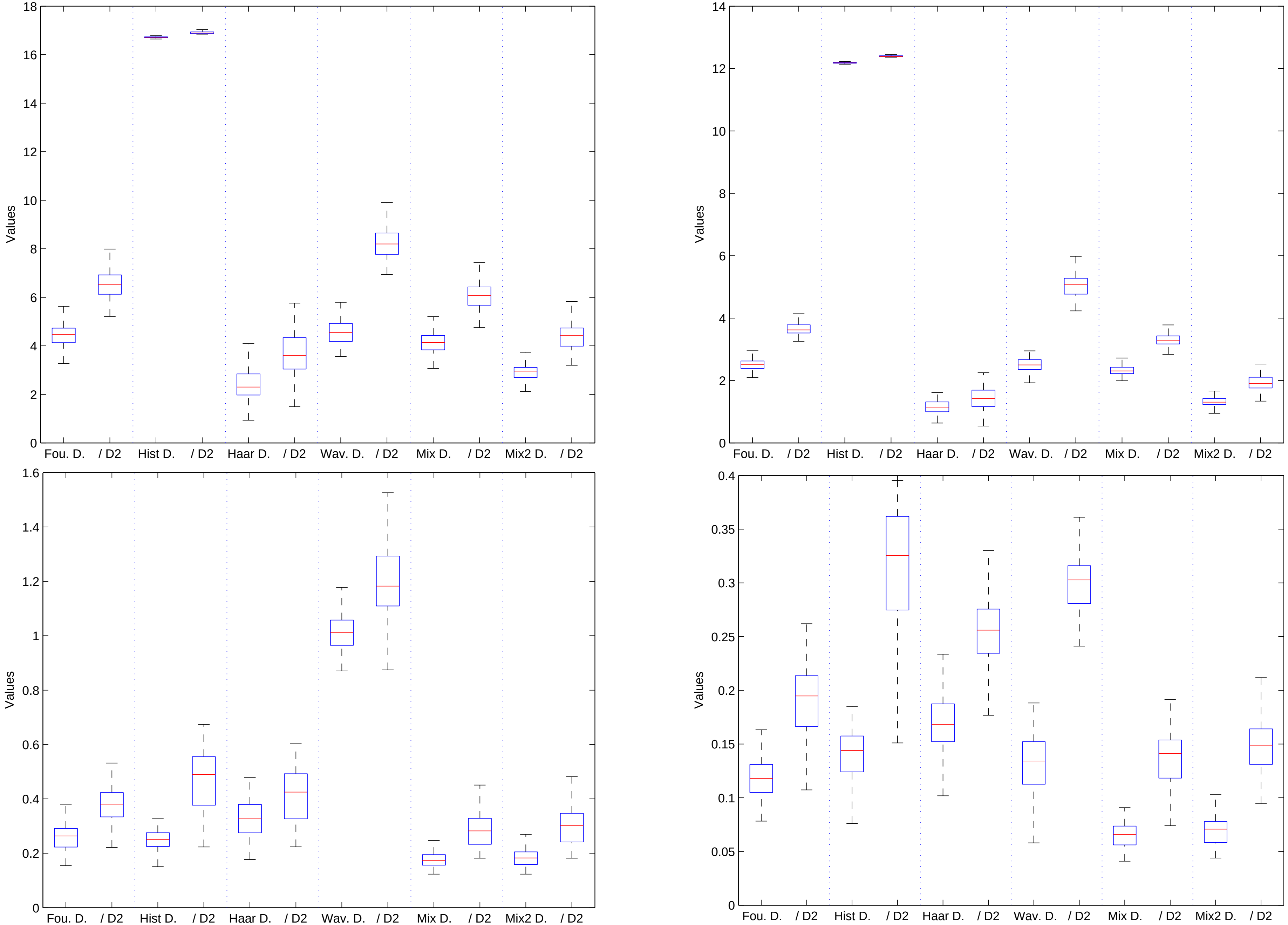


$n = 500$ $n = 2000$

Dantzig / Non adaptive Dantzig f_1/f_2

f_1

f_2

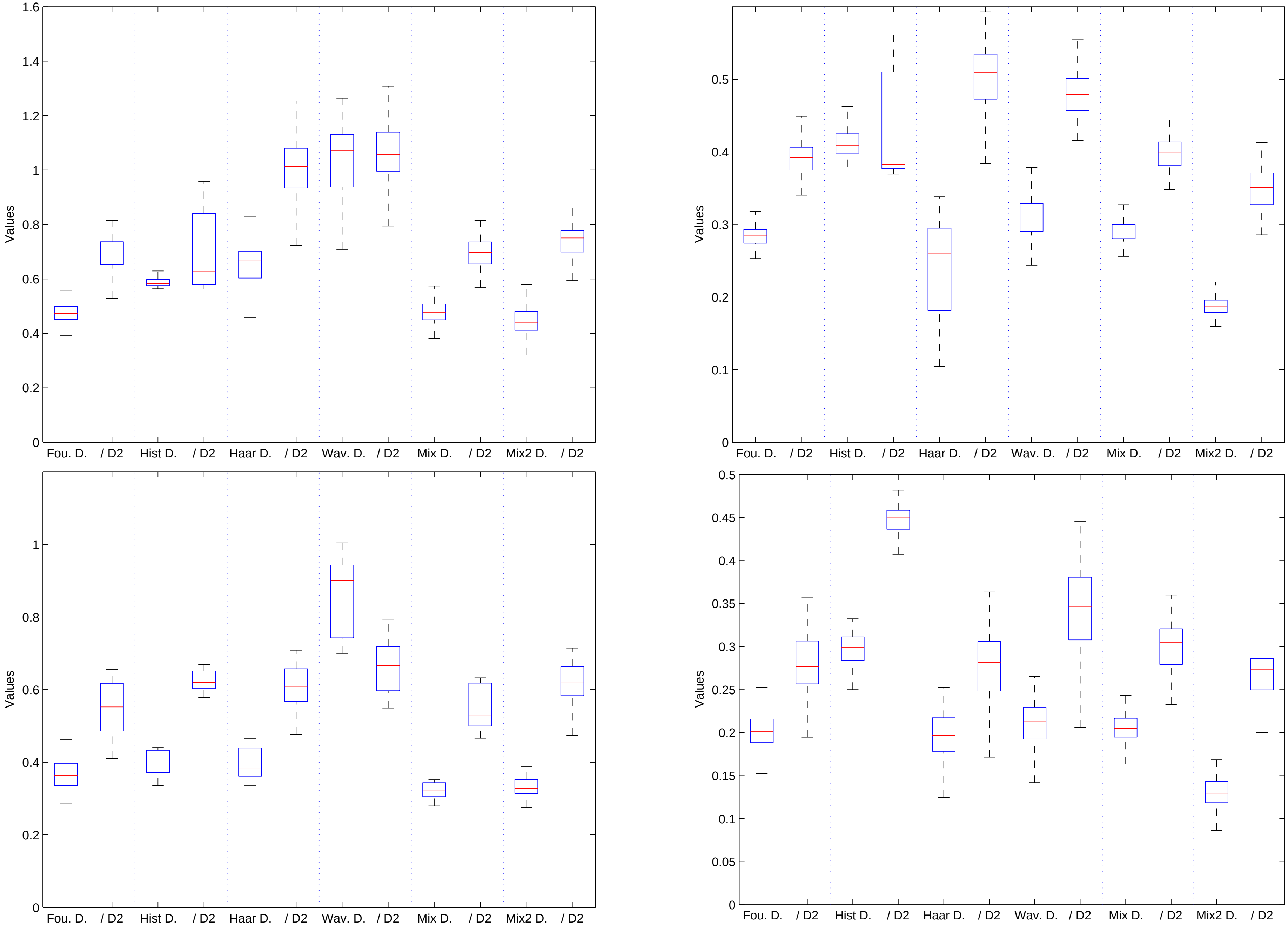


$n = 500$ $n = 2000$

Dantzig / Non adaptive Dantzig f_3/f_4

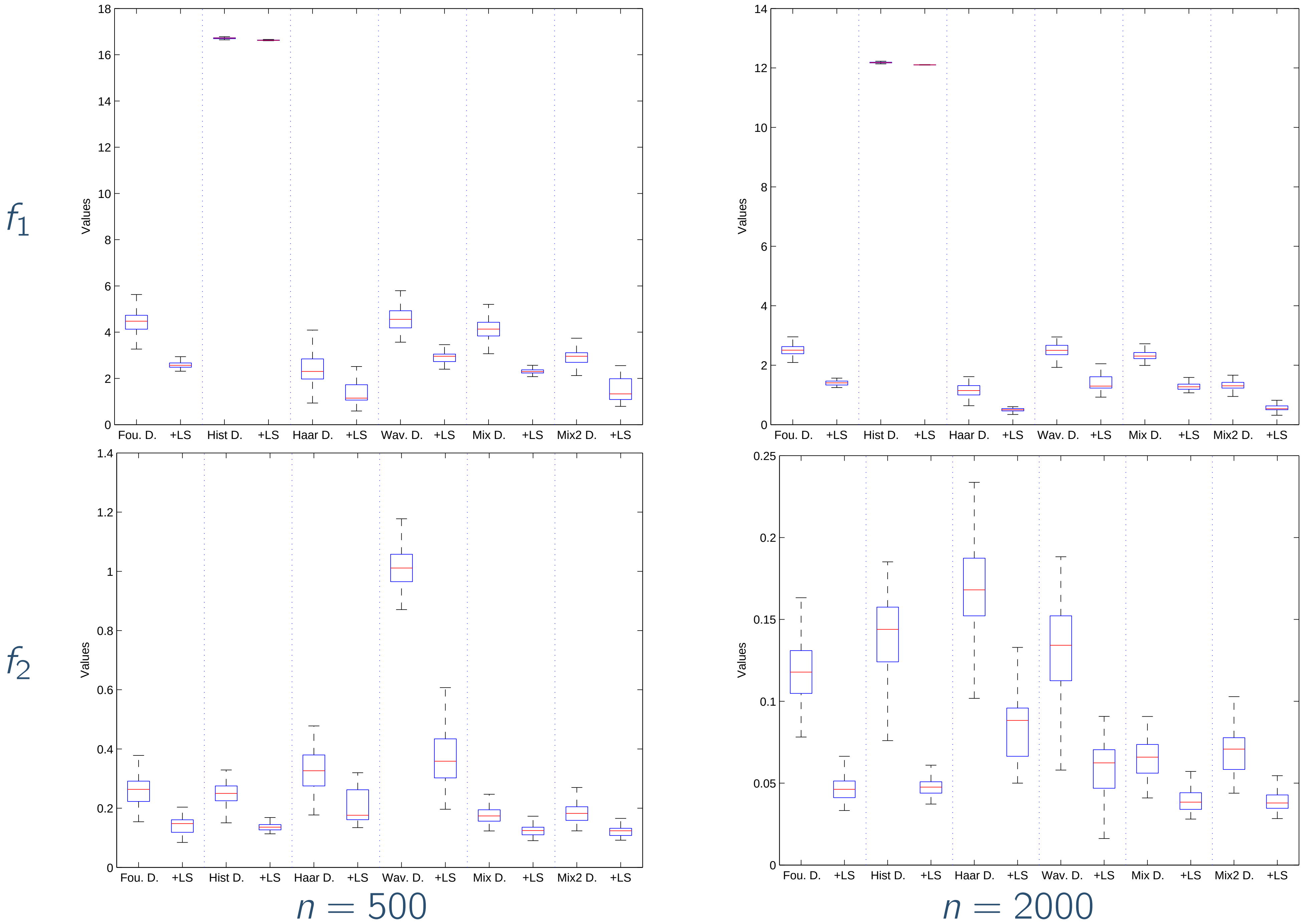
f_3

f_4



$n = 500$ $\quad$ $n = 2000$

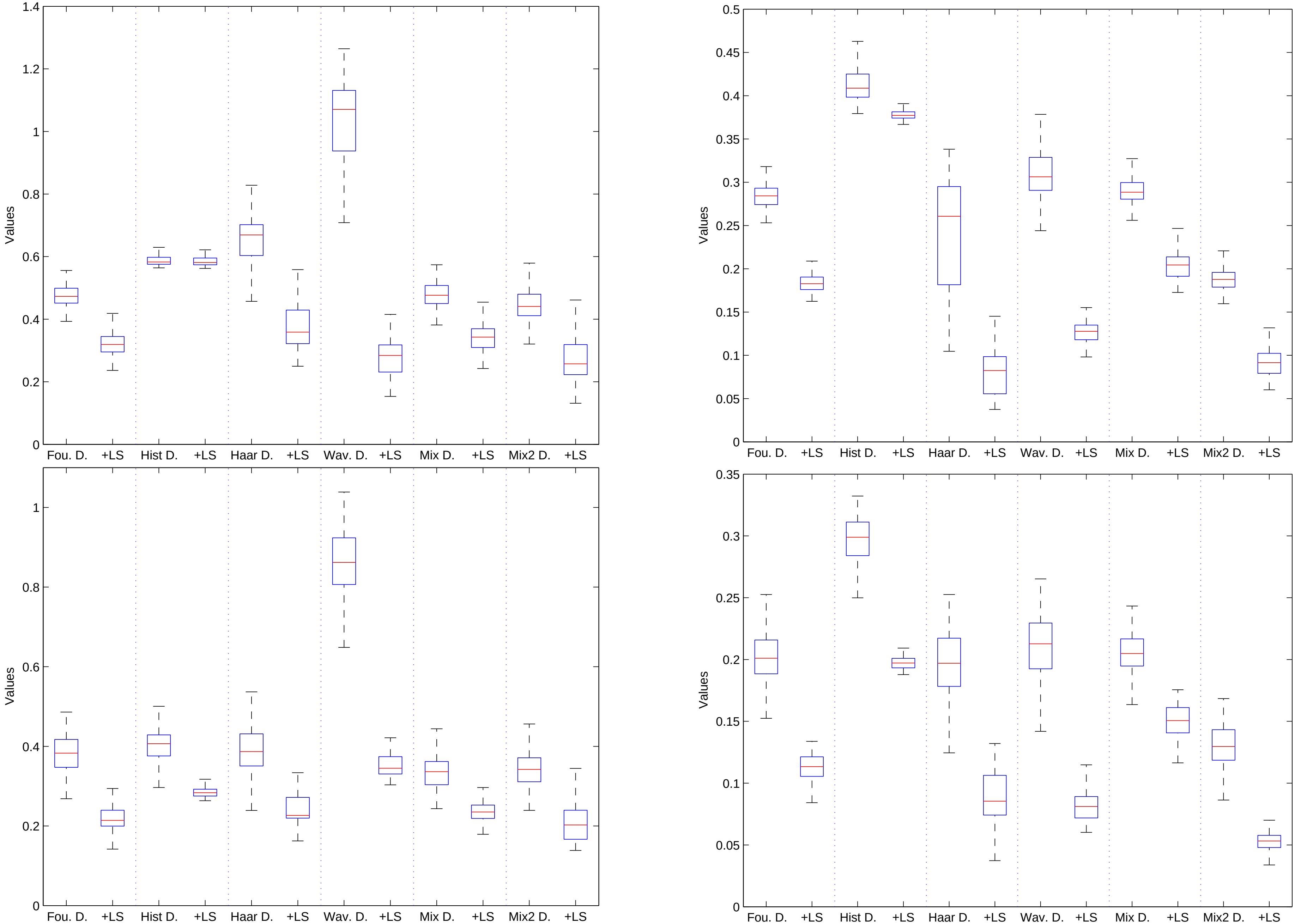
Dantzig / Dantzig+LS f_1/f_2



Dantzig / Dantzig+LS f_3 / f_4

f_3

f_4



$n = 500$ $n = 2000$

Théorème Dantzig

Théorème 1. *Sous Ω_γ, événement de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma-1}$, si la condition*

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

est satisfaite avec $\kappa_{J_0} > 0$ et $\mu_{J_0} \geq 0$ alors pour tout $\lambda \in \mathbb{R}^p$ et tout $\alpha > 0$

$$\begin{aligned}\|\hat{f}^{D,\gamma_0} - f_0\|_2^2 \leq \inf_{\lambda\in\mathbb{R}^p}\Bigg\{&\|f_\lambda - f_0\|_2^2 + 2\alpha\left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{\|\lambda_{J_0^C}\|_{\ell_1}^2}{|J_0|}\\ &+ \alpha\left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{(\|\hat{\lambda}^{D,\gamma_0}\|_{\ell_1} - \|\lambda\|_{\ell_1})_+^2}{|J_0|}\\ &+ 4\left(\frac{1}{\alpha} + \frac{1}{\kappa_{J_0}^2}\right)|J_0|(\|\eta^{\gamma_0}\|_{\ell_\infty} + \|\eta^{\gamma}\|_{\ell_\infty})^2\Bigg\}.\end{aligned}$$

De plus, toujours sous Ω_γ,

$$\begin{aligned}(\|\eta^{\gamma_0}\|_{\ell_\infty} + \|\eta^{\gamma}\|_{\ell_\infty})^2 \leq\ & 8(\gamma + \gamma_0)\log p \frac{\sup_k \sigma_k^2}{n}\\ &+ 64\log^2 p(\gamma\gamma_0 + \frac{49}{36}\gamma_0^2)\frac{\sup_k \|\phi_k\|_\infty^2}{n^2}\end{aligned}$$

Corollaire

Corollaire 1. *Sous Ω_γ, événement de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma-1}$, si la condition*

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

est satisfaite avec $\kappa_{J_0} > 0$ et $\mu_{J_0} \geq 0$ alors si $f_\lambda = P_{\mathcal{D}} f_0$ et si $\gamma \leq \gamma_0$, pour tout $\alpha > 0$

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

De plus, toujours sous Ω_γ,

$$(\|\eta^{\gamma_0}\|_{\ell_\infty} + \|\eta^\gamma\|_{\ell_\infty})^2 \leq 8(\gamma + \gamma_0)\log p \frac{\sup_k \sigma_k^2}{n} + 64\log^2 p(\gamma\gamma_0 + \frac{49}{36}\gamma_0^2)\frac{\sup_k \|\phi_k\|_\infty^2}{n^2}$$

Théorème Dantzig/Lasso

Théorème 2. *Sous Ω_γ, événement de proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma-1}$, si la condition*

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

est satisfaite avec $\kappa_{J_0} > 0$ et $\mu_{J_0} \geq 0$ alors pour tout $\alpha > 0$

$$\begin{aligned}\|\|\hat{f}^{D,\gamma_0} - f_0\|_2^2 - \|\hat{f}^{D,\gamma_0} - f_0\|_2^2 \leq \Bigg\{ &\alpha\left(1 + \frac{2\mu_{J_0}}{\kappa_{J_0}}\right)^2 \frac{\|\lambda^{L,\gamma_0}_{J_0^C}\|_{\ell_1}^2}{|J_0|} \\ &+ 4\left(\frac{1}{\alpha} + \frac{1}{\kappa_{J_0}^2}\right)|J_0|(\|\eta^{\gamma_0}\|_{\ell_\infty} + \|\eta^{\gamma}\|_{\ell_\infty})^2\Bigg\}.\end{aligned}$$

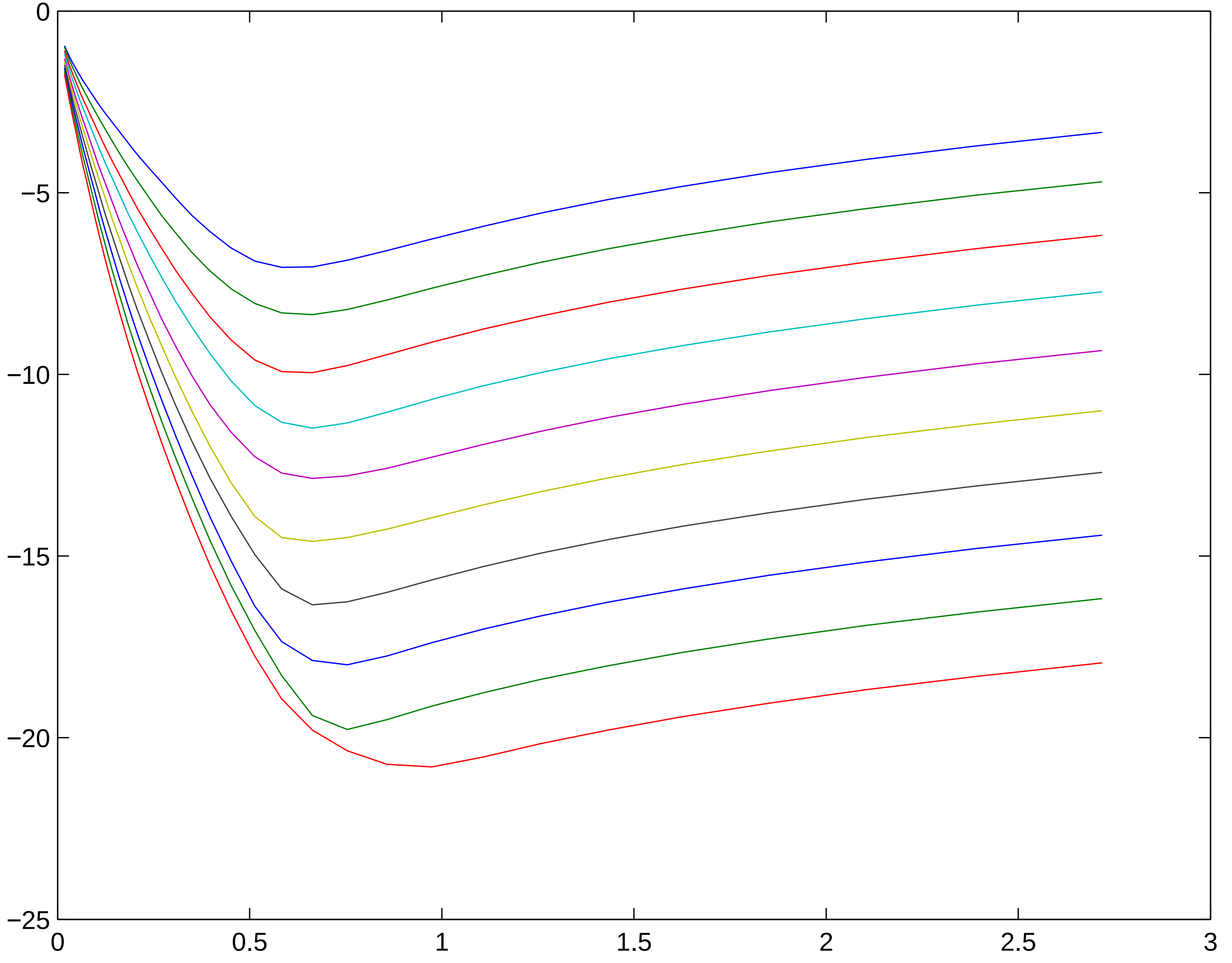
De plus, toujours sous Ω_γ,

$$\begin{aligned}(\|\eta^{\gamma_0}\|_{\ell_\infty} + \|\eta^{\gamma}\|_{\ell_\infty})^2 \leq\ & 8(\gamma + \gamma_0)\log p \frac{\sup_k \sigma_k^2}{n} \\ &+ 64 \log^2 p(\gamma\gamma_0 + \frac{49}{36}\gamma_0^2)\frac{\sup_k \|\phi_k\|_\infty^2}{n^2}\end{aligned}$$

Pénalité minimale

- Comment choisir γ_0 ?
- Inégalité oracle pour tout γ_0 !
- γ_0 trop grand : surlissage (biais) / γ_0 trop petit : instabilité (variance).
- Pb pour γ_0 grand visible dans l'inégalité oracle.
- Pb pour γ_0 petit plus subtil : présence du terme Dantzig $((\|\lambda^{S,\gamma_0}\|_{\ell_1} - \|\lambda\|_{\ell_1})_+)$ difficile à contrôler.
- Bon choix : $\gamma_0 = 1$!
- Si $\gamma_0 > 1$: Avec proba $\geq 1 - 2\left(\frac{1}{p}\right)^{\gamma_0}$, inégalité oracle valide sans le terme Dantzig pour tout λ tel que $f_\lambda = P_{\mathcal{D}} f_0$.
- Si $\gamma_0 < 1$: Théorème montrant dans un cas simple que l'estimateur est mauvais. (Estimation de la densité uniforme sur $[0, 1]$ dans la base de Haar).

Pénalité minimale



Hypothèse locale

- Pour J_0 fixé,

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

  est satisfaite avec $\kappa_{J_0} > 0$ et $\mu_{J_0} \geq 0$.
- Est-ce une hypothèse raisonnable ?
- Cas simple : $\mathcal{D}$ famille orthonormée $\Longrightarrow$ $\kappa_{J_0} = 1$ and $\mu_{J_0} = 0$ convient.

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

Hypothèses globales

- $\mathcal{D} \simeq$ orthonormée du point de vue de la norme ($\|f_\lambda\|_2 \simeq \|\lambda\|_{\ell_2}$)

$$\phi_{\min} = \min_{\lambda \neq 0} \frac{\|f_\lambda\|_2^2}{\|\lambda\|_{\ell_2}^2} \qquad \text{et} \qquad \phi_{\max} = \max_{\lambda \neq 0} \frac{\|f_\lambda\|_2^2}{\|\lambda\|_{\ell_2}^2}$$

- $|J_0| = s < p\frac{\phi_{\min}}{\phi_{\max}+\phi_{\min}} \implies$

$$\kappa_{J_0} = \sqrt{\phi_{\min}}\left(1 - \sqrt{\frac{\phi_{\max}}{\phi_{\min}}}\sqrt{\frac{s}{p-s}}\right) \quad \text{et} \quad \mu_{J_0} = \sqrt{\phi_{\max}}\sqrt{\frac{s}{p-s}}$$

- $\mathcal{D} \simeq$ orthonormée du point de vue du p.s. ($\langle f_{\lambda_J}, f_{\lambda_{J'}} \rangle \simeq 0$ if $J \cap J' \neq \emptyset$)

$$\theta = \max_{J \cap J' \neq \emptyset} \max_{\substack{\lambda \\ \lambda_J \neq 0 \\ \lambda_{J'} \neq 0}} \frac{\langle f_{\lambda_J}, f_{\lambda_{J'}} \rangle}{\|\lambda_J\|_{\ell_2}\|\lambda_{J'}\|_{\ell_2}}$$

- $|J_0| = s < \min(\frac{p}{2}, p\frac{\phi_{\min}^2}{\theta^2+\phi_{\min}^2}) \implies$

$$\kappa_{J_0} = \sqrt{\phi_{\min}}\left(1 - \frac{\theta}{\phi_{\min}}\right) \quad \text{et} \quad \mu_{J_0} = \frac{\theta}{\sqrt{\phi_{\min}}}$$

Hypothèses sur le support

- Restriction sur la taille du support :

$$\phi_{\min}(l) = \min_{|J|\leq l} \min_{\lambda_J \neq 0} \frac{\|f_{\lambda_J}\|_2^2}{\|\lambda_J\|_{\ell_2}^2} \qquad \text{et} \qquad \phi_{\max}(l) = \max_{|J|\leq l} \max_{\lambda_J \neq 0} \frac{\|f_{\lambda_J}\|_2^2}{\|\lambda_J\|_{\ell_2}^2}$$

$$\theta_{l,l'} = \max_{\substack{|J|\leq l \\ |J'|\leq l' \\ J\cap J' \neq \emptyset}} \max_{\substack{\lambda \\ \lambda_J \neq 0 \\ \lambda_{J'} \neq 0}} \frac{\langle f_{\lambda_J}, f_{\lambda_{J'}} \rangle}{\|\lambda_J\|_{\ell_2} \|\lambda_{J'}\|_{\ell_2}}$$

- $|J_0| = s \leq p/2$, $s \leq l \leq p-s$ et $l\phi_{\min}(s+l) > s\phi_{\max}(l) \implies$

$$\kappa_{J_0} = \sqrt{\phi_{\min}(s+l)} \left(1 - \sqrt{\frac{\phi_{\max}(l)}{\phi_{\min}(s+l)}} \sqrt{\frac{s}{l}}\right) \quad \text{et} \quad \mu_{J_0} = \sqrt{\phi_{\max}(l)} \sqrt{\frac{s}{l}}$$

- $|J_0| = s \leq p/2$, $s \leq l \leq p-s$ et $s\theta^2_{s,s+l} < l\phi_{\min}(s+l)^2 \implies$

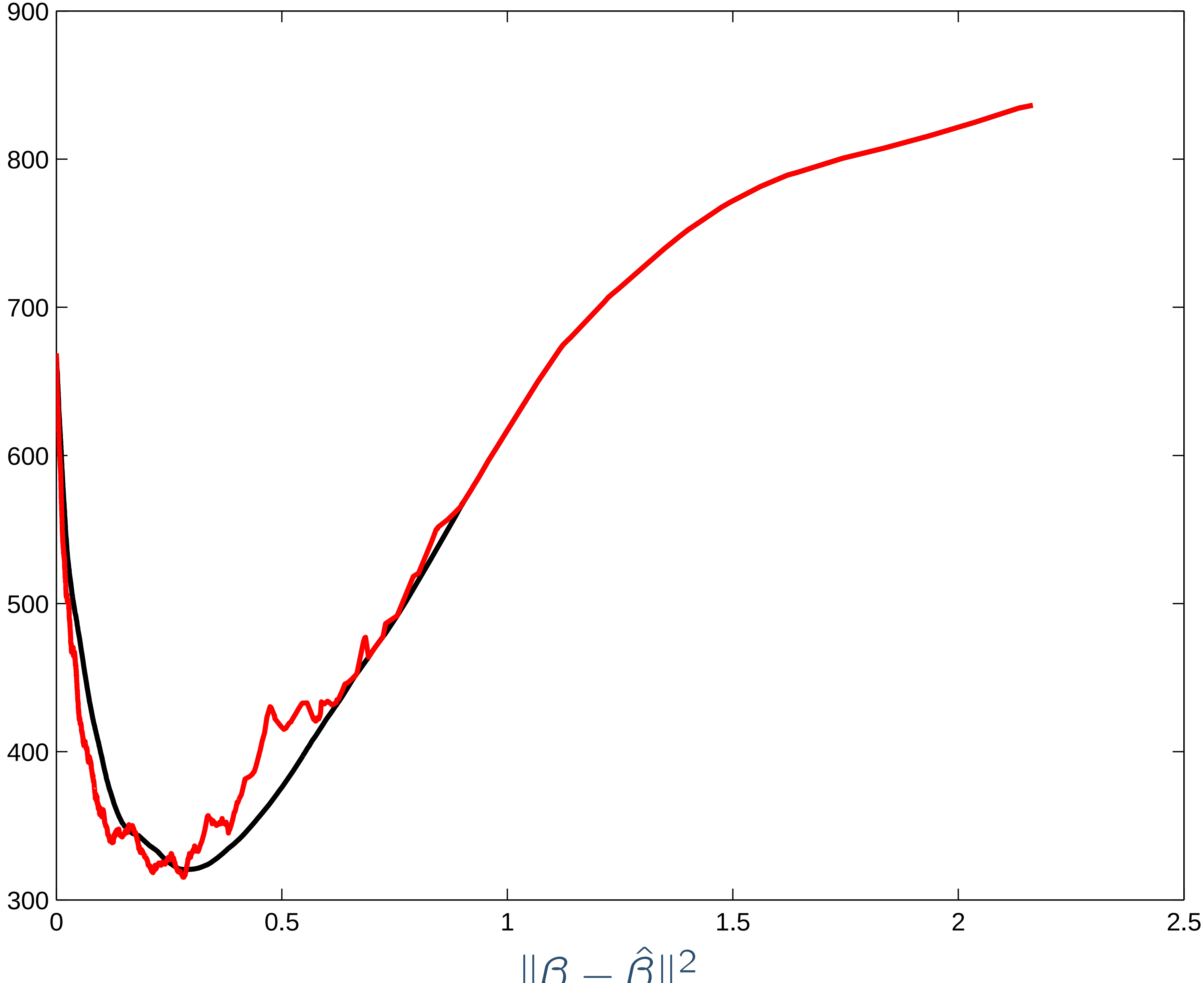
$$\kappa_{J_0} = \sqrt{\phi_{\min}(s+l)} \left(1 - \frac{\theta_{s,s+l}}{\phi_{\min}(s+l)} \sqrt{\frac{s}{l}}\right) \quad \text{et} \quad \mu_{J_0} = \frac{\theta_{s,s+l}}{\sqrt{\phi_{\min}(s+l)}}$$

- RIP $\implies$ condition précédente pour $l = s$.

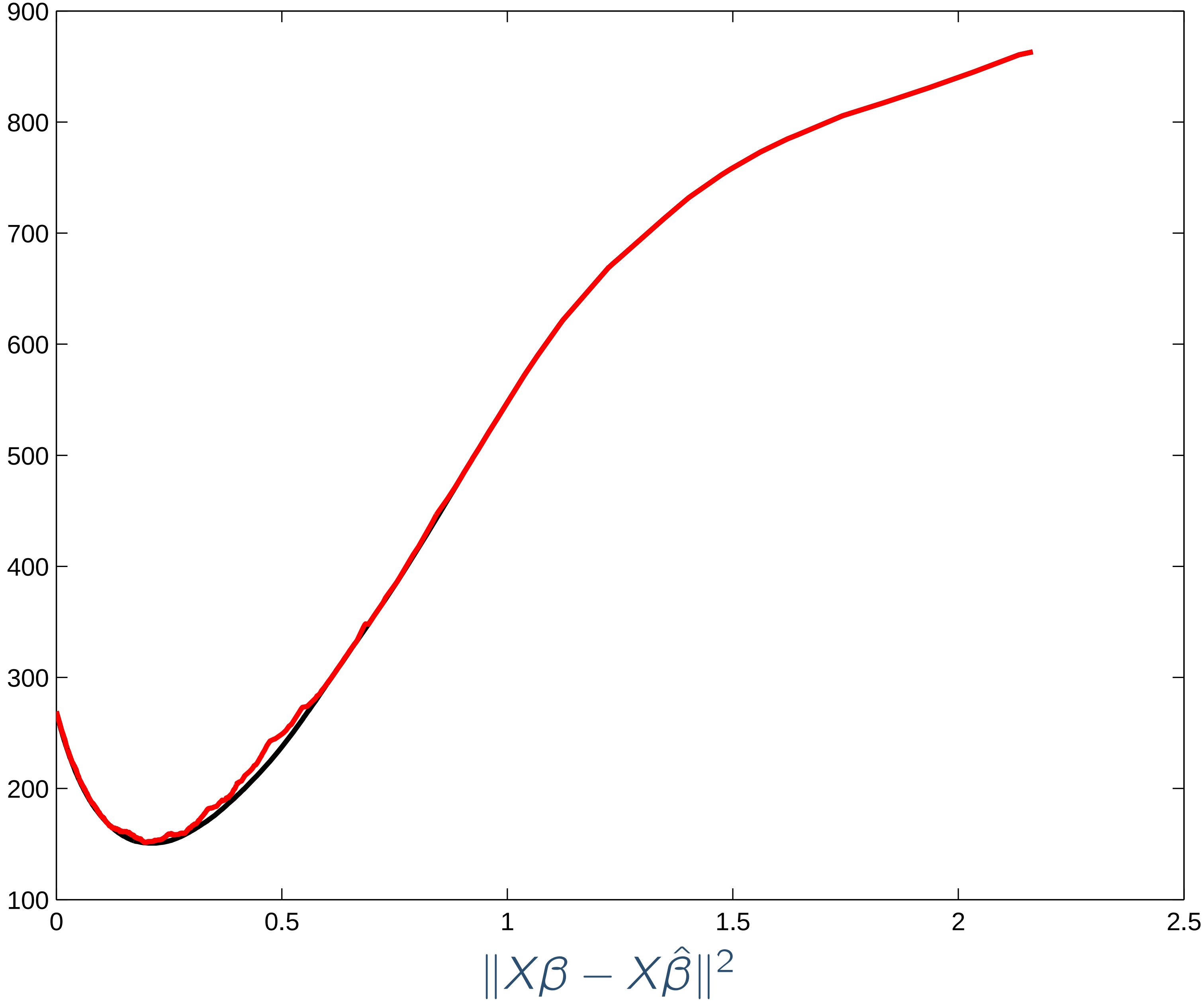
Conclusion

- Nouvelle méthode d'estimation de densité par pénalisation ℓ_1.
- Utilisation d'inégalités de concentration fine pour obtenir une pénalisation adaptative.
- Inégalité oracle en probabilité sur l'estimateur.
- Calibration de la pénalité : pénalité minimale.
- Hypothèses assez faibles sur la structure du dictionnaire.
- 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.

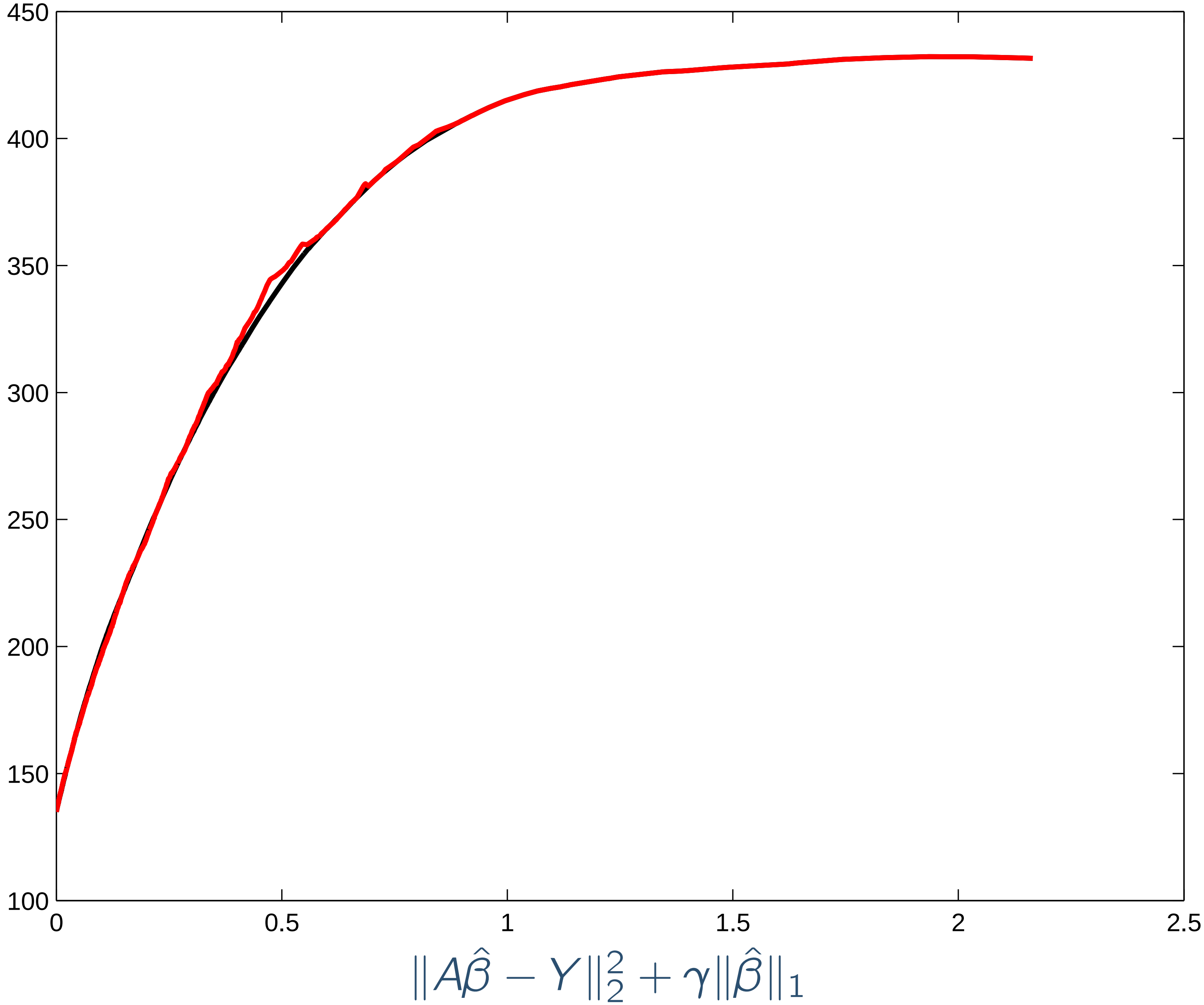
Lasso vs Dantzig



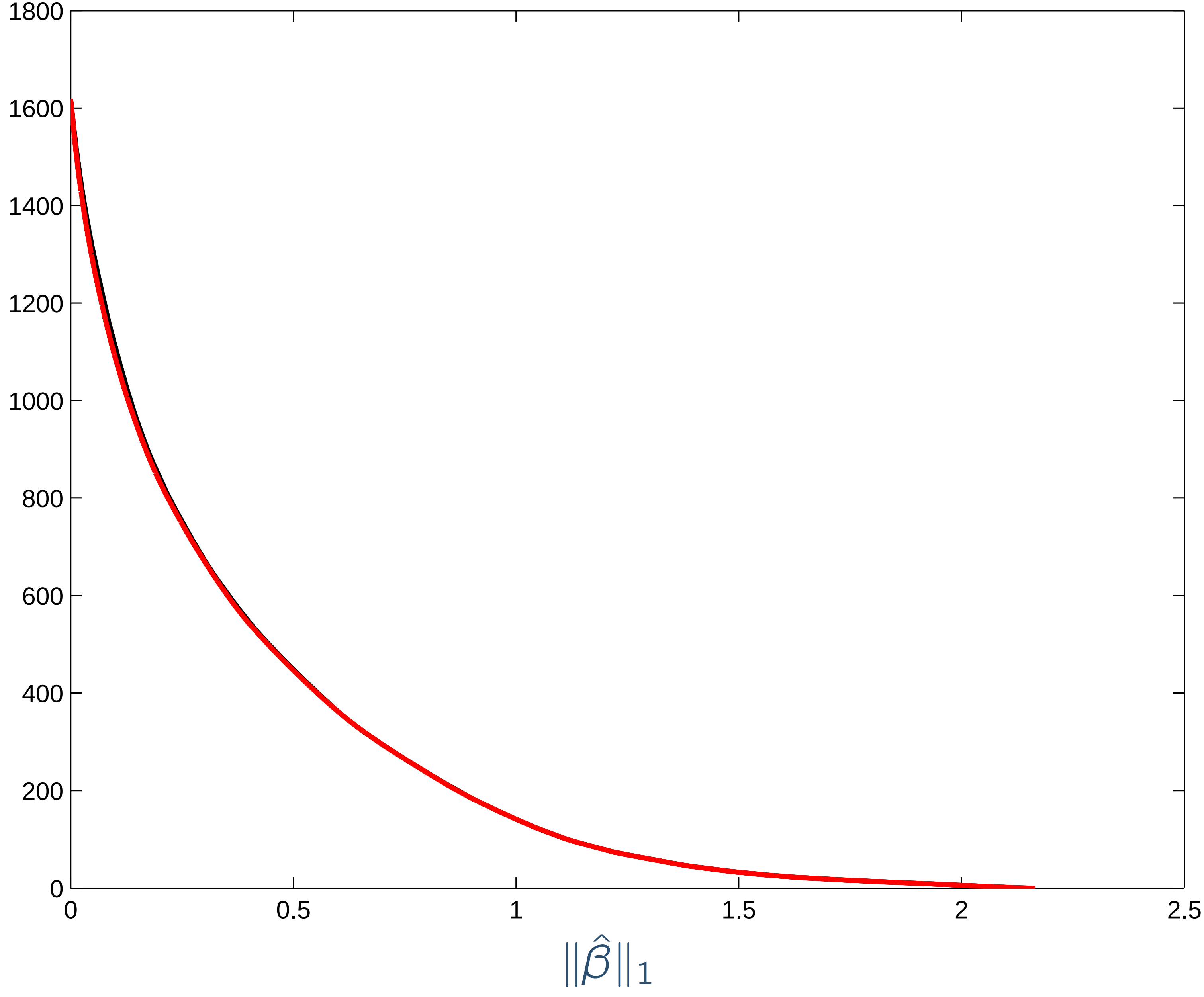
Lasso vs Dantzig



Lasso vs Dantzig



Lasso vs Dantzig



Lasso vs Dantzig