Adaptive Dantzig density estimation

E. Le Pennec (LPMA, Paris Diderot / SELECT, INRIA Saclay)

with K. Bertin (Universidad de Valparaíso)
and V. Rivoirard (Paris Sud / ENS Ulm)

→ ℓ_1-penalization

LASSO Tibshirani (94 → 96) Regression shrinkage and selection

Basis Pursuit Donoho (95 → 2001) Decomposition in an overcomplete dictionary

LASSO	BP
design matrix, Regressor $X = [X_1 \dots X_p]$	Dictionary $\Phi = [\phi_1, \dots, \phi_p]$
observation $Y = X\beta + \varepsilon$	observation $Y = \Phi\beta + \varepsilon$
$\hat{\beta} = \operatorname{argmin} \|Y - \phi\beta\|_2^2$ sous $\|\beta\|_1 \le \eta$	$\hat{\beta} = \operatorname{argmin} \frac{1}{2}\|Y - \Phi\beta\|_2^2 + \lambda \|\beta\|_1$
Connexion with Stage Wise	Matching pursuit

Justification through numerical experiments + efficient minimization algorithm

→ Sparsity:

2004 Donoho $\min \|a\|_1$ under $\|y - \Phi\beta\|_2 \le \eta$ (Gaussian matrix)

Candes / Tao / Romberg $\min \|a\|_1$ under $y = \phi\beta$ (Fourier)

with $p \gg n$: → ~~retrieve~~ find the solution if the solution is sparse

→ Dantzig. Candes / Tao (2007)

$$\min \|\beta\|_1 \text{ sous } \|\Phi^*(Y - \Phi\beta)\| \le \eta$$

$Y = \phi\beta + \varepsilon$ with ε Gaussian White Noise

⇒ Statistical results even if $p \gg n$

→ on β (~~coefficient identification~~ regression)

→ on $\Phi\beta$ (prediction)

⇒ Huge litterature since ....

→ Dantzig, Lasso, ...
→ Regression: support, sign, ...
→ prediction
→ regression on a dictionary
→ weakening of the dictionary assumptions

II Density estimation with dictionaries and concentration inequalities

Setting $X_1 \dots X_n$ iid $\sim f_0(x)\,dx$

Assumption $f_0 \in L^2$ (but not $f \in L^\infty$ and above all $\|f\|_\infty$ unknown)

$\Rightarrow$ Construction of an estimate $\hat{f}$ of $f_0 \dots$

as a linear combination of atoms $\{\varphi_\ell\}$ of a dictionary $\mathcal{D}$

$$f_\lambda = \sum_{\ell=1}^{p} \lambda_\ell \varphi_\ell$$

(mild) Assumptions on the dictionary
- $\varphi_\ell \in L^2$ (often $\|\varphi_\ell\|_2^2 = 1$ but not required)
- $\varphi_\ell \in L^\infty$ and $\|\varphi_\ell\|_\infty$ known
- $p \gg n$ possible.

Basic tools
- $E(\varphi_\ell(X)) = \langle f_0, \varphi_\ell \rangle = \beta_\ell$
- $$\|f_0 - f\|_2^2 = \|f_0\|_2^2 - 2\langle f_0, f\rangle + \|f\|_2^2$$
$$= \|f_0\|_2^2 - 2E(f(X)) + \|f\|_2^2$$

Empirical counterparts (Empirical processes)
- $$\hat{\beta}_\ell = \frac{1}{n}\sum_{i=1}^{n} \varphi_\ell(X_i) \qquad \left(E(\hat{\beta}_\ell) = \langle f_0, \varphi_\ell\rangle\right)$$
- $\forall \lambda \in \mathbb{R}^p$ $$\langle f_\lambda, \varphi_\ell \rangle = \sum_{\ell'=1}^{p} \lambda_{\ell'} \langle \varphi_{\ell'}, \varphi_\ell \rangle$$
$$= (G\lambda)_\ell$$

where G is the Gram matrix associated to $\mathcal{D}$

$$G_{\ell',\ell} = \langle \varphi_{\ell'}, \varphi_\ell \rangle$$

• $\gamma_\lambda = -2 \sum_{i=1}^{n} b_\lambda(x_i) + \|b_\lambda\|_2^2$ $\quad E(\gamma_\lambda) = \|b_0 - b\|_2^2 - \|b_0\|^2$

$$\gamma_\lambda = -2\lambda^* \beta + \lambda^* G \lambda$$

<u>L^∞ constraints on the coefficients:</u>

search for λ such that $(G\lambda)_\ell \simeq \hat{\beta}_\ell$

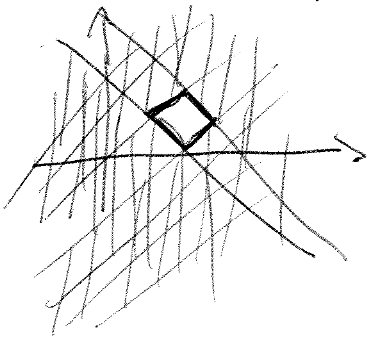
(i.e. $\langle b_\lambda, \varphi_\ell \rangle \simeq \langle b_0, \varphi_\ell \rangle$)

simple solution: fix $\eta > 0$ and search for λ such that

$$\|G\lambda - \hat{\beta}\|_\infty \leq \eta$$

→ Geometric constraint on λ given by hyperplanes

$$\hat{\beta}_\ell - \eta \leq (G\lambda)_\ell = \sum_{\ell'=1}^{p} \langle \varphi_{\ell'}, \varphi_\ell \rangle \lambda_{\ell'} \leq \hat{\beta}_\ell + \eta$$



→ "Often" "many" solutions to this problem

→ need of regularization

minimization of a "norm" under this constraint

• minimization of $\|\lambda\|_0$: • sparse solution

. calculability issue / only exploration algorithm

• minimization of $\|\lambda\|_1$: • sparse solution (under good assumptions)

. No calculability issue : convex analysis

• minimization of $\|\lambda\|_2$. non sparse solution ...

Dantzig:

$$\hat\lambda^D = \operatorname{argmin} \|\lambda\|_1 \quad \text{under} \quad \|G\lambda - \hat\beta\|_\infty \leq \eta$$

$$\hat\beta^D = \beta_{\hat\lambda^D}$$

Associated Lasso estimate

"Lasso" = Penalized least-square : $\frac{1}{2}\gamma_\lambda + \eta\|\lambda\|_1$

$$\frac{1}{2}\left[-2\lambda^*\hat\beta + \lambda^* G\lambda\right] + \eta\|\lambda\|_1$$

Minimization $\rightarrow$ Karush-Kuhn-Tucker condition

first order condition: $\|G\lambda - \hat\beta\|_\infty \leq \eta$

$\Rightarrow \hat\lambda^L = \operatorname{argmin} \frac{1}{2}\left[-2\lambda^*\hat\beta + \lambda^* G\lambda\right] + \eta\|\lambda\|_1$

is such that $\|G\hat\lambda^L - \hat\beta\|_\infty \leq \eta$

$\rightarrow$ Close relationship between Dantzig and Lasso!

Rq: Dantzig $\neq$ $\operatorname{argmin} \frac{1}{2}\|G\lambda - \hat\beta\|_2^2 + \eta\|\lambda\|_1$

$\rightsquigarrow \|G^2\lambda - G\hat\beta\|_\infty \leq \eta$

hard to interpret...

choice of η

Control of $|\tilde\beta_\ell - \hat\beta_\ell|$ is required

. link between the true coefficient $\beta_\ell = \langle f_0, \varphi_\ell\rangle$ and the empirical coefficients $\hat\beta_\ell = \frac{1}{n}\sum_{i=1}^n \varphi_\ell(X_i)$ defining the constraint.

. The "true" parameter should satisfy the constraint with high probability!

Uncertainty inequality:

$$P(|\hat{\beta}_\ell - \beta_\ell| > \eta_\ell) \leq 2e^{-v^2/2}$$

→ Hoeffding $\quad \eta_\ell = v \dfrac{\|\varphi_\ell\|_\infty}{\sqrt{n}}$

→ Bernstein $\quad \eta_\ell = v \dfrac{\sigma_\ell}{\sqrt{n}} + \dfrac{v^2}{3} \dfrac{\|\varphi_\ell\|_\infty}{n}$

σ_ℓ? : $\sigma_\ell \leq \|\varphi_\ell\|_\infty$ → Hoeffding is better

$\sigma_\ell \leq \|\beta\|_\infty^{1/2} \|\varphi_\ell\|_2$ → but $\|\beta\|_\infty$ is unknown

→ Adaptivity

$$\hat{\sigma}_\ell^2 = \frac{1}{2n(n-1)} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \left(\varphi_\ell(X_i) - \varphi_\ell(X_j)\right)^2$$

Key: $E\left(\left(\varphi_\ell(X_i) - \varphi_\ell(X_j)\right)^2\right) = 2\sigma_\ell^2$ (U-stat)

$$\tilde{\sigma}_\ell^2 = \left(\hat{\sigma}_\ell + v \frac{\|\varphi_\ell\|_\infty}{\sqrt{n}}\right)^2 + 3v^2 \frac{\|\varphi_\ell\|_\infty^2}{n}$$

$$\eta_\ell = v \frac{\tilde{\sigma}_\ell}{\sqrt{n}} + \frac{v^2}{3} \frac{\|\varphi_\ell\|_\infty}{n}$$

Th: $P(|\hat{\beta}_\ell - \beta_\ell| > \eta_\ell) \leq 2e^{-v^2/2}$

Simultaneous control : union bound

$$v = \sqrt{2\gamma \log p}$$

$$\Rightarrow P(\forall \ell, |\hat{\beta}_\ell - \beta_\ell| \leq \eta_\ell) \geq 1 - 2\left(\frac{1}{p}\right)^{\gamma - 1}$$

$\Rightarrow \eta_\ell = \sqrt{2\gamma}\, \|\varphi_\ell\|_\infty \sqrt{\dfrac{\log p}{n}}$ (fixed)

$\eta_\ell = \sqrt{2\gamma}\, \tilde{\sigma}_\ell \sqrt{\dfrac{\log p}{n}} + \dfrac{2\gamma}{3} \|\varphi_\ell\|_\infty \dfrac{\log p}{n}$ (adaptive)

Control of η_ℓ:

With high probability,

$$\sqrt{\gamma}\,\sigma_\ell \sqrt{\frac{\log p}{n}} + \frac{2\gamma}{3}\|\varphi_\ell\|_\infty \frac{\log p}{n} \leq \eta_\ell \leq 4\sqrt{\gamma}\,\sigma_\ell \sqrt{\frac{\log p}{n}} + 10\gamma\,\|\varphi_\ell\|_\infty \frac{\log p}{n}$$

$$\rightsquigarrow \quad \eta_\ell^2 \simeq 2\gamma\,\sigma_\ell^2 \,\frac{\log p}{n} \quad \text{(variance term!)}$$

Adaptive Dantzig and adaptive Lasso:

$$\hat{\lambda}^D = \operatorname{argmin} \|\lambda\|_1 \quad \text{s.t.} \quad |(G\lambda - \hat{\beta})_\ell| \leq \eta_\ell$$

$$\hat{\lambda}^L = \operatorname{argmin} \quad \frac{1}{2}\left[-2\lambda^* \hat{\beta} + \lambda^* G\lambda\right] + \sum \eta_\ell |\lambda_\ell|$$

$\Rightarrow$ The larger η_ℓ, the less stringent the constraint
the more one regularizes

Numerical algorithm:
. Convex programming $\rightarrow$ no issues
$\rightarrow$ off the shelves algorithms.

. Homotopy path type algorithm (LARS)
for the two cases

III Assumptions on the dictionaries and performances

<u>Estimators</u> $\hat{f}^D = f_{\hat{\lambda}^D}$ and $\hat{f}^L_{\hat{\lambda}^L}$

$\rightsquigarrow$ How to analyze their performances?

<u>Assumptions on the dictionaries are required!</u>
Essentially $\{\varphi_\ell\}$ should behave "as" a orthonormal basis!

<u>$\{\varphi_\ell\}$ orthonormal basis.</u> Dantzig = Lasso = Soft thresholding
$\rightsquigarrow$ No issue.

<u>$\{\varphi_\ell\} \simeq$ orthonormal basis with respect to the norm</u> $\|f_\lambda\|_2^2 \simeq \|\lambda\|_2^2$

$$\rightsquigarrow \quad \phi_{min} \|\lambda\|^2 \leq \|f_\lambda\|_2^2 \leq \phi_{max} \|\lambda\|_2^2$$
$$= \lambda^* G \lambda$$

$\rightarrow$ ~~exact~~ restriction on the eigenvalues of the Gram matrix

$\rightsquigarrow \triangle \neq$ Frame $\rightsquigarrow$ ~~Riesz~~ Riesz basis (synthesis coefficients $\neq$ analysis coeffs)

<u>$\{\varphi_\ell\} \simeq$ orthonormal basis with respect to the correlation</u> $\langle \varphi_{\ell'}, \varphi_\ell \rangle \simeq \delta_{\ell',\ell}$

$$\rightsquigarrow \max_{\ell \neq \ell'} |\langle \varphi_{\ell'}, \varphi_\ell \rangle| \leq M(\mathcal{D}) \quad \text{mutual coherence}$$

$\rightsquigarrow$ restriction on the off-diagonal values of G

$\rightsquigarrow$ reconstruction if $\|\lambda\|_0 \leq \frac{1}{2}\left(1 + \frac{1}{M(\mathcal{D})}\right)$

Similar hypotheses with support size constraints

- $\phi_{min}(s)\,\|\lambda\|_2^2 \le \|f_\lambda\|_2^2 \le \phi_{max}(s)\,\|\lambda\|_2^2$

  if $\|\lambda\|_0 \le s$

- $|\langle f_\lambda, f_{\lambda'} \rangle| \le \theta_{s,\ell}\, \|f_\lambda\|_2 \|f_{\lambda'}\|_2$

  $= \lambda^* G \lambda'$

  if $\|\lambda\|_0 \le s$, $\|\lambda'\|_0 \le \ell$

  and $J_\lambda \cap J_{\lambda'} = \emptyset$ $\Leftarrow$ off diagonal restriction

  support of λ

Global hypothesis: $\exists\ s, \ell$ such that
- $1 \le s \le \frac{p}{2}$
- $\ell \ge s$
- $\ell + s \le p$

and H1) $\ell\, \phi_{min}(s+\ell) > s\, \phi_{max}(\ell)$

H2) $\phi_{min}(s+\ell) > \theta_{\ell, s+\ell}$

Th: If H_1 or H_2 holds for s, ℓ then with probability $\ge 1 - \left(\frac{1}{p}\right)^{\gamma' - 1}$

$\forall \lambda \in \mathbb{R}^p$, $\forall \alpha > 0$, $\forall J_0$ such that $|J_0| \le s$

$$\|\hat{f}^D - f_0\|^2 \le \|f_\lambda - f_0\|^2 + \alpha \frac{\Lambda(\lambda, J_0^c)^2}{s} \left(1 + \frac{2\nu_{s,\ell}\sqrt{s}}{\kappa_{s,\ell}}\right)^2$$

$$+ 4 s \left(\frac{1}{\alpha} + \frac{1}{\kappa_{s,\ell}^2}\right) \left[\|\eta_\gamma\|_\infty + \|\eta_\gamma\|_\infty\right]^2$$

with $\Lambda(\lambda, J_0^c) = \|\lambda_{J_0^c}\|_1 + \frac{(\|\hat{\lambda}^D\|_1 - \|\lambda\|_1)_+}{2}$

and ν and κ as in the slides....

H1) $N_{\Delta,\ell} = \sqrt{\frac{q_{max}(\ell)}{\ell}}$ and $K_{\Delta,\ell} = \sqrt{q_{min}(s+\ell)}\left[1 - \sqrt{\frac{\Delta\, q_{max}(\ell)}{\ell\, q_{min}(s+\ell)}}\right]$

H2) $N_{\Delta,\ell} = \frac{\theta_{\ell,s+\ell}}{\sqrt{\ell\, q_{min}(s+\ell)}}$ and $K_{\Delta,\ell} = \sqrt{q_{min}(s+\ell)}\left[1 - \frac{\theta_{\ell,s+\ell}}{q_{min}(s+\ell)}\sqrt{\frac{\Delta}{\ell}}\right]$

Rq $K_{\Delta,\ell} = \sqrt{q_{min}(s+\ell)} - \sqrt{\Delta}\, N_{\Delta,\ell}$

Comments: • No restriction on λ or on J_0 (besides the size...)

• $\|\beta_\lambda - \beta_0\|_2^2 \rightsquigarrow$ classical approximation term
• $\Delta(\|\eta_{\gamma'}\|_\infty + \|\eta_\gamma\|_\infty)^2 \rightsquigarrow$ variance term on $J \simeq (\gamma+\gamma')\, \|\sigma_a^2\|_\infty \frac{\log p}{n}\, \Delta$
• $\frac{\|\lambda_{J_0^c}\|_1^2}{\Delta} \rightsquigarrow$ approximation term outside J_0

$\rightsquigarrow \lambda_{(a)} \sim \left(\frac{1}{\ell}\right)^\gamma \Rightarrow \frac{\|\lambda_{J_0^c}\|_1}{\Delta} \simeq \|\lambda_{J_0^c}\|_2^2$

and J_0 best subset

• $\frac{(\|\hat\lambda^D\|_1 - \|\lambda\|_1)_+}{2\Delta} \rightsquigarrow$ Dantzig $\Rightarrow$ no control if $\|\lambda\|_1 \le \|\hat\lambda^D\|_1$

$\rightsquigarrow$ link with the constraint!

• Key issue: γ' appears instead of γ

$\Rightarrow$ almost possible to take the expectation

$\rightsquigarrow$ only issue is the $(\|\hat\lambda^D\|_1 - \|\lambda\|_1)_+$ term...

Special case If $\beta_0 = \beta_{\lambda_0}$ or $\beta_{\lambda_0} = P\beta$ then with $P \ge 1 - 2\left(\frac{1}{p}\right)^\gamma$

$G\lambda_0 = \beta_0$ satisfies the Dantzig constraint.

$\rightarrow \|\hat\lambda^D\|_1 \le \|\lambda_0\|_1$

$\rightarrow$ the $(\ )_+$ term disappears ~~[illegible]~~

$\rightarrow$ choice of $\gamma > 1$ will be shown interesting...

$$\|\hat\beta^D - \beta_0\|^2 \le \|\beta_{\lambda_0} - \beta_0\|^2 + \alpha \frac{\|\lambda_{J_0^c}\|_1^2}{\Delta}\left(1 + \frac{2N_{\Delta,\ell}\sqrt{\Delta}}{K_\Delta}\right)^2$$
$$+ 4\Delta\left(\frac{1}{\alpha} + \frac{1}{K_{\Delta,\ell}^2}\right)\left[\|\eta_\gamma\|_\infty + \|\eta_{\gamma'}\|_\infty\right]^2$$

"Equivalence" Dantzig / Lasso

$$\left| \|\hat{f}^D - f\|_2^2 - \|\hat{f}^L - f\|_2^2 \right| \leq 4\Delta \left(\|\eta_{\gamma'}\|_\infty + \|\eta_\gamma\|_\infty \right)^4 \left(\frac{1}{\beta} + \frac{1}{\kappa^2} \right)$$

$$+ \frac{\|\hat{\lambda}^L_{J_0^c}\|}{\Delta} \left(1 + 2 \frac{\nu\sqrt{\Delta}}{\kappa} \right)$$

Rk: $\hat{\lambda}^L$ always satisfies the Dantzig constraint so no (l_1 lem...

Sketch of proof $\Rightarrow$ "stronger" theorem with minimal assumptions (van de Geer + Bühlmann)

Local assumptions:

$$\|f_\lambda\|_2 \geq \kappa_{J_0} \|\lambda_{J_0}\|_2 \quad \text{if} \quad \|\lambda_{J_0^c}\|_1 \leq c_0 \|\lambda_{J_0}\|_1$$

$$\rightsquigarrow \quad \|f_\lambda\|_2 \geq \kappa_{J_0} \|\lambda_{J_0}\|_2 - \mu_0 \left(\|\lambda_{J_0^c}\|_1 - \|\lambda_{J_0}\|_1 \right)_+ \quad \boxed{LA(J_0, \kappa_{J_0}, \mu_{J_0})}$$

Rq $\|\lambda_{J_0}\|_2$ could be replaced by $\|\lambda_{J_0}\|_1 / \sqrt{J_0}$

Rq H1 or H2 $\Rightarrow$ $LA(J_0, \kappa_{J_0}\mu)$ $\forall J_0$, $|J_0| = \Delta$

Th With Proba $\geq 1 - \left(\frac{1}{p}\right)^{\gamma' - 1}$

$\forall J_0$ such that $LA(J_0, \kappa_{J_0}, \mu_{J_0})$

$\forall \lambda \in \mathbb{R}^p$, $\forall \beta > 0$

$$\|\hat{f}^D - f\|_2^2 \leq \|f_\lambda - f\|_2^2 + \cancel{\alpha} \frac{\Lambda(\lambda, J_0^c)^2}{|J_0|} \left(1 + \frac{2\mu_{J_0}\sqrt{|J_0|}}{\kappa_{J_0}} \right)^2$$

$$+ 4|J_0| \left(\frac{1}{\alpha} + \frac{1}{\kappa_{J_0}^2} \right) \left(\|\eta_{\gamma'}\|_\infty + \|\eta_\gamma\|_\infty \right)^2$$

$\Rightarrow$ Much finer "granularity", the constants may depends on the set J_0

$\rightsquigarrow$ automatic adaptation.

$\Rightarrow$ Numerical illustration ...

Proof:

- $\|b_\lambda - b_0\|_2^2 = \|\hat{b}^D - b_0\|_2^2 + \|b_\lambda + \hat{b}^D\|_2^2 + 2\langle \hat{b}^D - b_0, b_\lambda - \hat{b}^D \rangle$

- $\|b_\lambda - \hat{b}^D\|_2^2 = \|b_\Delta\|_2^2$ where $\Delta = \lambda - \hat{\lambda}^D$

- $\langle \hat{b}^D - b_0, b_\lambda - \hat{b}^D \rangle = \sum_{\ell=1}^{p} (\lambda_\ell - \hat{\lambda}^D_\ell)\left((G\hat{\lambda}^D)_\ell - \beta_{0,\ell}\right)$

$$\leq \|\Delta\|_1 \ \|(G\hat{\lambda}^D)_\ell - \beta_{0,\ell}\|_\infty$$

- $\|(G\hat{\lambda}^D)_\ell - \beta_{0,\ell}\|_\infty \leq \|(G\hat{\lambda}^D)_\ell - \hat{\beta}_\ell\|_\infty + \|\hat{\beta}_\ell - \beta_{0,\ell}\|_\infty$

$$\leq \|\eta_\ell\|_\infty + \|\eta_{\ell'}\|_\infty \quad \text{with proba } 1 - C_1\left(\frac{1}{p}\right)^{1-\frac{\delta}{1+\varepsilon}}$$

- $\|\hat{b}^D - b_0\|_2^2 \leq \|b_\lambda - b_0\|^2 + 2(\|\eta_\gamma\|_\infty + \|\eta_{\gamma'}\|_\infty)\|\Delta\|_1 - \|b_\Delta\|^2$

lemma: $\|\Delta_{J^c}\|_1 \leq \|\Delta_J\|_1 + 2\|\lambda_{J^c}\|_1 + (\|\hat{\lambda}^D\|_1 - \|\lambda\|_1)_+$

Proof: $\|\hat{\lambda}^D\|_1 \leq \|\lambda\|_1 + (\|\hat{\lambda}^D\|_1 - \|\lambda\|_1)_+$

$\|\Delta_J - \lambda_J\|_1 + \|\Delta_{J^c} - \lambda_{J^c}\|_1 \leq \|\lambda_J\|_1 + \|\lambda_{J^c}\|_1 + (\|\hat{\lambda}^D\|_1 - \|\lambda\|_1)_+$

$\|\lambda_J\|_1 - \|\Delta_J\|_1 + \|\Delta_{J^c}\| - \|\lambda_{J^c}\| \leq \|\lambda_J\|_1 + \|\lambda_{J^c}\|_1 + (\|\hat{\lambda}^D\|_1 - \|\lambda\|_1)_+$

QED!

$$\|\Delta_{J_0^c}\|_1 - \|\Delta_{J_0}\|_1 \leq 2\|\lambda_{J_0^c}\|_1 + (\|\hat{\lambda}^D\|_1 - \|\lambda\|_1)_+ = \Lambda(\lambda, J_0^c)$$

$$\|b_\Delta\|_2 \geq \kappa_{J_0} \frac{\|\Delta_{J_0}\|_1}{\sqrt{J_0}} - 2\mu_{J_0}\Lambda(\lambda, J_0^c) \quad \text{LA!}$$

$$\Rightarrow \|\Delta_{J_0}\|_1 \leq \frac{\sqrt{J_0}}{\kappa_{J_0}}\|b_\Delta\|_2 + 2\frac{\mu_{J_0}}{\kappa_{J_0}}\sqrt{J_0}\,\Lambda(\lambda, J_0^c)$$

$$\|\Delta\|_1 \le \|\Delta_{J_0}\|_1 + \|\Delta_{J_0^c}\|_1$$

$$\le 2\|\Delta_{J_0}\|_1 + 2\Lambda(\lambda, J_0^c)$$

$$\le \frac{2\sqrt{J_0}}{\kappa_{J_0}} \|f_\Delta\|_2 + 2\Lambda(\lambda, J_0^c)\left[1 + \frac{2N_{J_0}\sqrt{|J_0|}}{\kappa_{J_0}}\right]$$

We use then
$$2(\|\eta_\gamma\|_\infty + \|\eta_{\gamma'}\|_\infty)\frac{2\sqrt{J_0}}{\kappa_{J_0}}\|f_\Delta\|^2$$

$$\le \frac{4|J_0|(\|\eta_\gamma\|_\infty + \|\eta_{\gamma'}\|_\infty)^2}{\kappa_{J_0}^2} + \|f_\Delta\|^2$$

So that

$$2(\|\eta_\gamma\|_\infty + \|\eta_{\gamma'}\|_\infty)\|\Delta\|_1 - \|f_\Delta\|^2$$

$$\le \frac{4|J_0|(\|\eta_\gamma\|_\infty + \|\eta_{\gamma'}\|_\infty)^2}{\kappa_{J_0}^2} + 4(\|\eta_\gamma\|_\infty + \|\eta_{\gamma'}\|_\infty)\Lambda(\lambda, J_0^c)\left(1 + \frac{2N_0\sqrt{|J_0|}}{\kappa_{J_0}}\right)$$

$$\le 4|J_0|\left(\frac{1}{\alpha} + \frac{1}{\kappa_{J_0}^2}\right)(\|\eta_\gamma\|_\infty + \|\eta_{\gamma'}\|_\infty)^2$$

$$+ \alpha\frac{\Lambda(\lambda, J_0^c)^2}{|J_0|}\left(1 + \frac{2N_{J_0}\sqrt{|J_0|}}{\kappa_{J_0}}\right)^2$$

QED!

IV Minimal penalization and least-square Dantzig

- When dealing with penalties an important issue lies in the calibration.

Th $\Rightarrow$ γ should be taken as close of 1 as possible

Q. Is $\gamma < 1$ a "bad" idea?

A: Yes, we can prove that $\gamma < 1$ leads to bad performances ... in some cases...

Th: $f = \mathbb{1}_{[0,1]}$ φ_ℓ Haar basis $p = n$

$$\gamma > 1 \Rightarrow E(\|\hat{f}^D - f\|_2^2) \leq \gamma_\gamma \left(\frac{\log n}{n}\right)$$

$$\gamma < 1 \Rightarrow \exists \delta < 1 \quad E(\|\hat{f}^D - f\|_2^2) \geq \frac{1}{n^\delta}(1 + o_n(1))$$

$\Rightarrow$ Minimal penalty means that amongst our family the choice of γ is optimal: a smaller γ leads to a "bad" estimator

$\Rightarrow$ Numerical illustration.

- $\Rightarrow$ $\ell_1 \sim$ biased on the coefficients

Two-step procedure Gauss-Dantzig / LS-Dantzig

Reestimation of the coefficients on the support of $\hat{\lambda}^D$

$$\left(\hat{\lambda}^{D+LS}\right)_{\hat{J}^D} = G^{-1}_{\hat{J}^D}\left(\hat{\beta}_\ell\right)_{\ell \in \hat{J}^D}$$

$\rightarrow$ Numerical improvement.