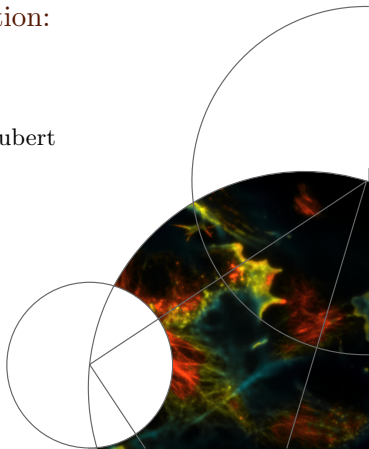


Some results on sparse L2-L0 reconstruction:
Continuous Exact ℓ_0 penalties

Laure Blanc-Féraud,
joint work with Emmanuel Soubies and Gilles Aubert





Introduction

Many signal processing areas are concerned with **sparse solution recovery** :
compressed sensing, variable selection, source separation, learning...

- Linear observation : $d = Ax$
 - d : observed data, vector in \mathbb{R}^M
 - x unknown data to be estimated in \mathbb{R}^N
 - A observation matrix, $M \times N$ matrix.

usually $M < N$, the system is undertermined, A is ill-conditioned,
observation are noisy

- Least square solution $\hat{x} = \arg \min_{x \in \mathbb{R}^N} \|Ax - d\|_2^2$
- Regularization : sparse signal hypothesis modeled by considering " ℓ_0 -norm" constraints :

$$\|x\|_0 \leq k \text{ where } \|x\|_0 = \# \{x_i, i = 1, \dots, N : x_i \neq 0\}$$





Example : compressed sensing

One can recover a signal from far fewer measurements than Nyquist rates if

- the signal is sparse on a given dictionary Ψ : sparsifying matrix
- each measurement picks up little information about each component.

Φ : sensing matrix.

$A = \Phi\Psi$. Sensing and sparsifying matrix must be incoherent.

Reconstruction [Donoho et al 06, Candès et al 06] :

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} \|x\|_1 \text{ subject to } Ax = d$$

or subject to $\|Ax - d\|_2^2 \leq \epsilon$

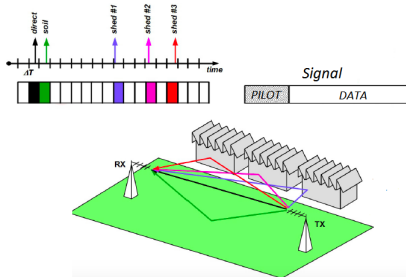
Examples : Φ Discrete Fourier and Ψ identity

Φ Random Gaussian and Ψ any basis.



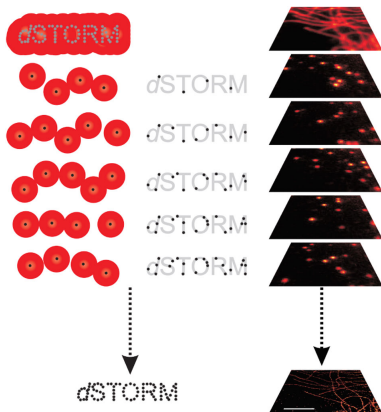
Experiment description

In matrix vector formulation $d = Ah$ with $A_{i,j} = s_{i-j}$ and h is sparse.





Example in image superresolution : single molecule reconstruction



In fluorescence microscopy, superresolution can be obtained by single molecule illumination (PALM, STORM)

- A models the 2D diffraction convolution due to the optics,
- d is one acquired image,
- x the deconvolved image only containing sparse impulse spots.



Introduction

A prior known on the **sparsity** of the solution

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} \|Ax - d\|_2^2 \quad \text{subject to} \quad \|x\|_0 \leq k$$

A prior known on the **noise** variance

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} \|x\|_0 \quad \text{subject to} \quad \|Ax - d\|_2^2 \leq \epsilon$$

No prior

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} G_{\ell_0}(x) := \frac{1}{2} \|Ax - d\|_2^2 + \lambda \|x\|_0$$

Non continuous, nonconvex optimization problems (NP hard).





A quick tour on existing methods for sparse approximation

Convex relaxation

Replacing the ℓ_0 by the ℓ_1 convex norm :

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} \frac{1}{2} \|Ax - d\|_2^2 + \lambda \|x\|_1$$

→ take advantages of **convex optimization** algorithms (*Basic Pursuit, LASSO, simplex or interior point methods, Forward-Backward splitting algorithms (ISTA)* and further variants)...

$$x^{k+1} = \text{prox}_{\gamma\lambda\|\cdot\|_1} \left(x^k - \gamma A^t (Ax^k - d) \right)$$

Exact sparse recovery

- Under sufficient (but restrictive) conditions on A , sparse signals can be **exactly recovered** by ℓ_1 minimization [Donoho et al 06, Candès et al 06].
- **Restricted isometry property (RIP)** : all K -column sub-matrices of A are *nearly* orthogonal.





A quick tour on existing methods for sparse approximation

Greedy algorithms

Idea : starting from a null solution, the most relevant components are added iteratively one by one

- *Matching Pursuit* (MP) [Mallat et al 93], *Orthogonal Matching Pursuit* (OMP) [Pati et al 93] and further variants
- *Bayesian OMP* [Herzet et al 10], *Single Best Replacement* (SBR) [Soussen et al 11] and further variants.

Exact sparse recovery

Similar recovery results as for the ℓ_1 relaxation have been proved for OMP (under conditions on A) [Tropp 04].





A quick tour on existing methods for sparse approximation

Continuous nonsmooth nonconvex penalties

$$\|x\|_0 = \sum_{i=1}^N |x_i|_0 \approx \sum_{i=1}^N \phi(x_i)$$

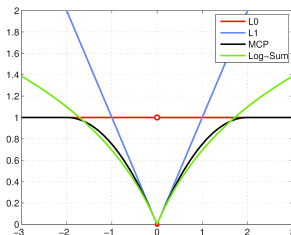


FIGURE: Plot of ℓ_0 (red), MCP (black), LogSum (green), ℓ_1 (blue).

Continuous penalties which better approximate ℓ_0 than the convex ℓ_1 -norm \rightarrow better behaviour on many applications.





A quick tour on existing methods for sparse approximation

Continuous nonsmooth nonconvex penalties

- *Adaptive Lasso* [Zou 06]
- *Nonnegative Garrote* [Breiman 95]
- *Log-Sum Penalty* [Candès et al 08]
- *Smoothly Clipped Absolute Deviation (SCAD)* [Fan and Li 01]
- *Minimax Concave Penalty (MCP)* [Zhang 10]
- ℓ_p -norms $0 < p < 1$ [Chartrand 07, Foucart and Lai 09]
- *Smoothed ℓ_0 -norm Penalty (SL0)* [Mohimani et al 09]



II.2) A quick tour on existing methods for sparse approximation

Iterative Hard Thresholding algorithm [Blumensath and Davies 08]

- Forward-Backward Splitting (FBS) algorithm

$$\mathbf{x}^{k+1} = \text{prox}_{\gamma\lambda\|\cdot\|_0} \left(\mathbf{x}^k - \gamma A^t (A\mathbf{x}^k - d) \right)$$

- Convergence to a local minimizer of G_{ℓ_0} for any A [Attouch et al 13]

Exact reformulations

- *Class of smooth nonconvex penalties* \rightarrow asymptotic connections with the ℓ_2 - ℓ_0 criteria [Chouzenoux et al 13]
- *Reformulation as DC programs* [Le Thi et al 14, Le Thi et al 15]
- *Equivalence of ℓ_0 - and ℓ_p -norm ($0 < p \leq 1$) minimization under linear equalities or inequalities* [Fung and Mangasarian 11]
- *Reformulation PMNE* \rightarrow global optimum for problems of reasonable size [Bourguignon et al 15]





Goal of the work

Question concerning smooth approximations

→ What is the link between the $\ell_2 - \ell_0$ problem and the approximate ones?

→ Is it possible to find **continuous** functions which **approximate exactly** the $\ell_2 - \ell_0$ problem in the sense that :

- they preserve the global minimizers
- local minimizers of the approximate problem are also local minimizers of the $\ell_2 - \ell_0$ problem

for **any** A matrix.



Outline of the talk

-
- ➊ 1D analysis : the CEL0 penalization
 - ➋ CEL0 nD extension and properties w.r.t the $\ell_2 - \ell_0$ problem
 - ➌ How to find other exact continuous $\ell_2 - \ell_0$ approximation ?
 - ➍ Using FBS algorithm : the CEL0 threshold
 - ➎ Numerical Illustrations and comparisons
-



I/ 1D case analysis

For the **one dimensional** case,

$$\hat{u} = \arg \min_{u \in \mathbb{R}} g(u) := \frac{1}{2}(au - d)^2 + \lambda |u|_0,$$

where $a > 0$ and $d \in \mathbb{R}$.

$$|u|_0 = \begin{cases} 0 & \text{if } u = 0, \\ 1 & \text{if } u \neq 0. \end{cases} \quad (1)$$



I/ 1D case analysis

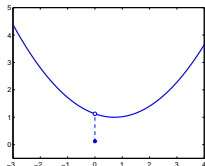
$$g(u) = \frac{1}{2}(au - d)^2 + \lambda|u|_0$$

- if $u = 0$ then $g(0) = \frac{1}{2}(d)^2$
- The minimum could be reached at $\hat{u} = 0$, the value is $g(\hat{u}) = \frac{1}{2}(d)^2$

- if $u \neq 0$ then $g(u) = \frac{1}{2}(au - d)^2 + \lambda$

- The minimum is reached at $\hat{u} = \frac{d}{a}$ and the value is $g(\hat{u}) = \lambda$

if $|d| \leq \sqrt{2\lambda}$ then $\hat{u} = 0$



if $|d| \geq \sqrt{2\lambda}$ then $\hat{u} = \frac{d}{a}$

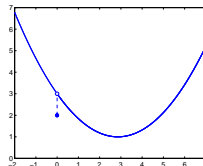
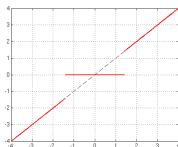


FIGURE: Plot of g (blue) for $a = 0.7$, $\lambda = 1$ and $d = 0.5$ (left) or $d = 2$ (right).



I/ Methods/Principles to find an approximate solution

The solution is given by the **Hard Threshold** function $\hat{u} = \left(\frac{d}{a}\right) \mathbb{1}_{\{|d| > \sqrt{2\lambda}\}}$



Combinatorial problem in the n-dimensional case.



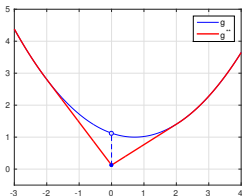
II/ 1D and orthogonal cases analysis

For the **one dimensional** case,

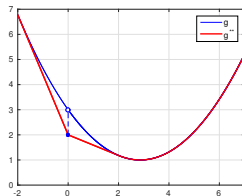
$$\hat{u} = \arg \min_{u \in \mathbb{R}} g(u) := \frac{1}{2}(au - d)^2 + \lambda|u|_0,$$

where $a > 0$ and $d \in \mathbb{R}$.

- The solution is given by $\hat{u} = \left(\frac{d}{a}\right) \mathbb{1}_{\{|d| > \sqrt{2\lambda}\}}$
- We will compute the **convex hull** of this 1D nonconvex function



(a) $|d| < \sqrt{2\lambda}$ ($d = 0.5$)



(b) $|d| > \sqrt{2\lambda}$ ($d = 2$)

FIGURE: Plot of g (blue) and its convex hull (red).





II/ 1D and orthogonal cases analysis

The Legendre-Fenchel Transform :

For $f : E \rightarrow \bar{\mathbb{R}}$,

$$f^*(u^*) = \sup_{u \in E} \{u^*u - f(u)\}$$

The biconjugate f^{**} is the *convex hull* or *convex envelope* of f (i.e. the largest lower-semi-continuous (l.s.c) convex function less than f).



II/ 1D and orthogonal cases analysis

The Legendre-Fenchel Transform :

For $f : E \rightarrow \bar{\mathbb{R}}$,

$$f^*(u^*) = \sup_{u \in E} \{u^*u - f(u)\}$$

The biconjugate f^{**} is the *convex hull* or *convex envelope* of f (i.e. the largest lower-semi-continuous (l.s.c) convex function less than f).

For the *one dimensional* case,

$$\hat{u} = \arg \min_{u \in \mathbb{R}} g(u) := \frac{1}{2}(au - d)^2 + \lambda|u|_0,$$

where $a > 0$ and $d \in \mathbb{R}$ we can show that,

$$g^{**}(u) = \frac{1}{2}(au - d)^2 + \phi(a, \lambda; u),$$

$$\phi(a, \lambda; u) = \lambda - \frac{a^2}{2} \left(|u| - \frac{\sqrt{2\lambda}}{a} \right)^2 \mathbb{1}_{\{|u| \leq \frac{\sqrt{2\lambda}}{a}\}}$$





II/ 1D and orthogonal cases analysis

The Legendre-Fenchel Transform :

For $f : E \rightarrow \bar{\mathbb{R}}$,

$$f^*(u^*) = \sup_{u \in E} \{u^*u - f(u)\}$$

The biconjugate f^{**} is the *convex hull* or *convex envelope* of f (i.e. the largest lower-semi-continuous (l.s.c) convex function less than f).

For the *one dimensional* case,

$$\hat{u} = \arg \min_{u \in \mathbb{R}} g(u) := \frac{1}{2}(au - d)^2 + \lambda|u|_0,$$

where $a > 0$ and $d \in \mathbb{R}$ we can show that,

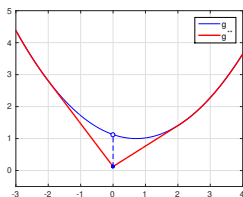
$$g^{**}(u) = \frac{1}{2}(au - d)^2 + \phi(a, \lambda; u),$$

$$\phi(a, \lambda; u) = \lambda - \frac{a^2}{2} \left(|u| - \frac{\sqrt{2\lambda}}{a} \right)^2 \mathbb{1}_{\{|u| \leq \frac{\sqrt{2\lambda}}{a}\}}$$

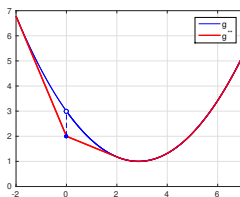




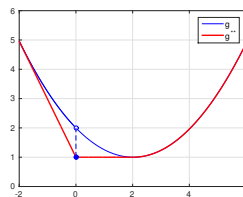
II/ 1D and orthogonal cases analysis



(a) $|d| < \sqrt{2\lambda}$ ($d = 0.5$)



(b) $|d| > \sqrt{2\lambda}$ ($d = 2$)



(c) $|d| = \sqrt{2\lambda}$ ($d = \sqrt{2}$)

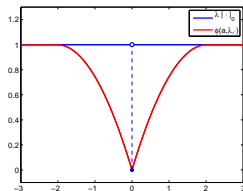


FIGURE: Top : Plot of g (blue) and g^{**} (red) for $a = 0.7$, $\lambda = 1$ and different d .
Bottom : plot of $\lambda | \cdot |_0$ (blue) and $\phi(a, \lambda; \cdot)$ for $a = 0.7$ and $\lambda = 1$.





II/ 1D and orthogonal cases analysis

Convex Hull in ND for A orthogonal (*i.e.* $A^T A$ diagonal)

Let $a_i \in \mathbb{R}^M$ be the i th column of $A \in \mathbb{R}^{M \times N}$, for $i \in \mathbb{I}_N = \{1, \dots, N\}$. The quadratic term of G_{ℓ_0} can be decomposed as,

$$\frac{1}{2} \|Ax - d\|^2 = \frac{1}{2} \|d - \hat{d}\|^2 + \frac{1}{2} \|Dx - \tilde{z}\|^2,$$

where $D \in \mathbb{R}^{N \times N}$ is a diagonal matrix with $D_{ii} = \|a_i\|$, $\hat{d} = AD^{-2}A^T d$ and $\tilde{z} = D^{-1}A^T d$.

- $\frac{1}{2} \|d - \hat{d}\|^2 \rightarrow$ independent of x
- $\frac{1}{2} \|Dx - \tilde{z}\|^2 \rightarrow$ additively separable $\left(\frac{1}{2} \sum_{i \in \mathbb{I}_N} (\|a_i\| x_i - \tilde{z}_i)^2 \right)$

In the **orthogonal case**, the computation of the Biconjugate $G_{\ell_0}^{**}$ (*convex hull* of G_{ℓ_0}) is reduced to the resolution of N **one dimensional problems**





II/ 1D and orthogonal cases analysis

Convex Hull in ND for A orthogonal (i.e. $A^T A$ diagonal)

Let $a_i \in \mathbb{R}^M$ be the i th column of $A \in \mathbb{R}^{M \times N}$, for $i \in \mathbb{I}_N = \{1, \dots, N\}$. The quadratic term of G_{ℓ_0} can be decomposed as,

$$\frac{1}{2} \|Ax - d\|^2 = \frac{1}{2} \|d - \hat{d}\|^2 + \frac{1}{2} \|Dx - \tilde{z}\|^2,$$

where $D \in \mathbb{R}^{N \times N}$ is a diagonal matrix with $D_{ii} = \|a_i\|$, $\hat{d} = AD^{-2}A^T d$ and $\tilde{z} = D^{-1}A^T d$.

- $\frac{1}{2} \|d - \hat{d}\|^2 \rightarrow$ independent of x
- $\frac{1}{2} \|Dx - \tilde{z}\|^2 \rightarrow$ additively separable $\left(\frac{1}{2} \sum_{i \in \mathbb{I}_N} (\|a_i\| x_i - \tilde{z}_i)^2 \right)$

In the **orthogonal case**, the computation of the Biconjugate $G_{\ell_0}^{**}$ (convex hull of G_{ℓ_0}) is reduced to the resolution of N **one dimensional problems**





II/ 1D and orthogonal cases analysis

Convex Hull in ND for A orthogonal (i.e. $A^T A$ diagonal)

Let $a_i \in \mathbb{R}^M$ be the i th column of $A \in \mathbb{R}^{M \times N}$, for $i \in \mathbb{I}_N = \{1, \dots, N\}$. The quadratic term of G_{ℓ_0} can be decomposed as,

$$\frac{1}{2} \|Ax - d\|^2 = \frac{1}{2} \|d - \hat{d}\|^2 + \frac{1}{2} \|Dx - \tilde{z}\|^2,$$

where $D \in \mathbb{R}^{N \times N}$ is a diagonal matrix with $D_{ii} = \|a_i\|$, $\hat{d} = AD^{-2}A^T d$ and $\tilde{z} = D^{-1}A^T d$.

- $\frac{1}{2} \|d - \hat{d}\|^2 \rightarrow$ independent of x
- $\frac{1}{2} \|Dx - \tilde{z}\|^2 \rightarrow$ additively separable $\left(\frac{1}{2} \sum_{i \in \mathbb{I}_N} (\|a_i\| x_i - \tilde{z}_i)^2 \right)$

In the orthogonal case, the computation of the Biconjugate $G_{\ell_0}^{**}$ (convex hull of G_{ℓ_0}) is reduced to the resolution of N one dimensional problems





II/ 1D and orthogonal cases analysis

Convex Hull in ND for A orthogonal (*i.e.* $A^T A$ diagonal)

Let $a_i \in \mathbb{R}^M$ be the i th column of $A \in \mathbb{R}^{M \times N}$, for $i \in \mathbb{I}_N = \{1, \dots, N\}$. The quadratic term of G_{ℓ_0} can be decomposed as,

$$\frac{1}{2} \|Ax - d\|^2 = \frac{1}{2} \|d - \hat{d}\|^2 + \frac{1}{2} \|Dx - \tilde{z}\|^2,$$

where $D \in \mathbb{R}^{N \times N}$ is a diagonal matrix with $D_{ii} = \|a_i\|$, $\hat{d} = AD^{-2}A^T d$ and $\tilde{z} = D^{-1}A^T d$.

- $\frac{1}{2} \|d - \hat{d}\|^2 \rightarrow$ independent of x
- $\frac{1}{2} \|Dx - \tilde{z}\|^2 \rightarrow$ additively separable $\left(\frac{1}{2} \sum_{i \in \mathbb{I}_N} (\|a_i\| x_i - \tilde{z}_i)^2 \right)$

In the **orthogonal case**, the computation of the **Biconjugate** $G_{\ell_0}^{**}$ (*convex hull* of G_{ℓ_0}) is reduced to the resolution of N **one dimensional problems**





II/ 1D and orthogonal cases analysis

The *convex hull* of G_{ℓ_0} in the orthogonal case

By identification with the previous one dimensional case we show that,

$$G_{\ell_0}^{**}(x) = \frac{1}{2} \|Ax - d\|^2 + \sum_{i \in \mathbb{I}_N} \phi(\|a_i\|, \lambda; x_i)$$

The CEL0 penalty

$$\Phi_{\text{CEL0}}(x) := \sum_{i \in \mathbb{I}_N} \phi(\|a_i\|, \lambda, x_i)$$

And in a general case ? (A arbitrary matrix of $\mathbb{R}^{M \times N}$)





II/ 1D and orthogonal cases analysis

The *convex hull* of G_{ℓ_0} in the orthogonal case

By identification with the previous one dimensional case we show that,

$$G_{\ell_0}^{**}(x) = \frac{1}{2} \|Ax - d\|^2 + \sum_{i \in \mathbb{I}_N} \phi(\|a_i\|, \lambda; x_i)$$

The CEL0 penalty

$$\Phi_{\text{CEL0}}(x) := \sum_{i \in \mathbb{I}_N} \phi(\|a_i\|, \lambda, x_i)$$

And in a general case ? (A arbitrary matrix of $\mathbb{R}^{M \times N}$)





III/ Links between $\ell_2 - \ell_0$ and ℓ_2 -CEL0 problems

Recall the $\ell_2 - \ell_0$ and ℓ_2 -CEL0 functionals (for an arbitrary matrix A) :

$$G_{\ell_0}(x) := \frac{1}{2} \|Ax - d\|^2 + \lambda \|x\|_0$$

$$G_{\text{CEL0}}(x) = \frac{1}{2} \|Ax - d\|^2 + \Phi_{\text{CEL0}}(x)$$

Properties

- $G_{\text{CEL0}}(x)$ is continuous
- $G_{\text{CEL0}}(x)$ is convex w.r.t. to each component

Notations

- $\sigma(x) = \{i \in \mathbb{I}_N; x_i \neq 0\} \subseteq \mathbb{I}_N$, the support of $x \in \mathbb{R}^N$
- $\forall x \in \mathbb{R}^N$, $\sigma^-(x) := \left\{i \in \sigma(x) : |x_i| \leq \sqrt{2\lambda}/\|a_i\|\right\}$, a part of the support





VI/ Links $G_{\ell_0} \leftrightarrow G_{\text{CELO}}$ - *The global minimizers of G_{CELO}*

Theorem (Link between global minimizers of G_{ℓ_0} and G_{CELO})

Let $d \in \mathbb{R}^M$ and $\lambda > 0$,

- (i) the set of *global minimizers of G_{ℓ_0}* is *included in the set of global minimizers of G_{CELO}* ,

$$\arg \min_{x \in \mathbb{R}^N} G_{\ell_0}(x) \subseteq \arg \min_{x \in \mathbb{R}^N} G_{\text{CELO}}(x) \quad (2)$$

- (ii) conversely if $\hat{x} \in \mathbb{R}^N$ is a *global minimizer of G_{CELO}* , \hat{x}^0 defined by

$$\forall i \in \mathbb{I}_N, \quad \hat{x}_i^0 = \hat{x}_i \mathbb{1}_{\left\{|\hat{x}_i| \geq \frac{\sqrt{2\lambda}}{\|a_i\|}\right\}}, \quad (3)$$

is a *global minimizer of G_{ℓ_0}* and $G_{\text{CELO}}(\hat{x}) = G_{\text{CELO}}(\hat{x}^0) = G_{\ell_0}(\hat{x}^0)$.

Proof is based on :

- G_{CELO} underestimates G_{ℓ_0} ,
- G_{CELO} is convex w.r.t each component
- Description of the minimizers of G_{ℓ_0} [Nikolova 13]





VI/ Links $G_{\ell_0} \leftrightarrow G_{\text{CELO}}$ - *The global minimizers of G_{CELO}*

Theorem (Link between global minimizers of G_{ℓ_0} and G_{CELO})

Let $d \in \mathbb{R}^M$ and $\lambda > 0$,

- (i) *the set of global minimizers of G_{ℓ_0} is included in the set of global minimizers of G_{CELO} ,*

$$\arg \min_{x \in \mathbb{R}^N} G_{\ell_0}(x) \subseteq \arg \min_{x \in \mathbb{R}^N} G_{\text{CELO}}(x) \quad (2)$$

- (ii) *conversely if $\hat{x} \in \mathbb{R}^N$ is a global minimizer of G_{CELO} , \hat{x}^0 defined by*

$$\forall i \in \mathbb{I}_N, \quad \hat{x}_i^0 = \hat{x}_i \mathbb{1}_{\left\{|\hat{x}_i| \geq \frac{\sqrt{2\lambda}}{\|a_i\|}\right\}}, \quad (3)$$

is a global minimizer of G_{ℓ_0} and $G_{\text{CELO}}(\hat{x}) = G_{\text{CELO}}(\hat{x}^0) = G_{\ell_0}(\hat{x}^0)$.

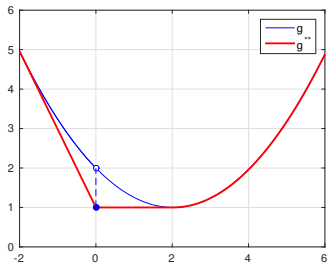
Proof is based on :

- G_{CELO} underestimates G_{ℓ_0} ,
- G_{CELO} is convex w.r.t each component
- Description of the minimizers of G_{ℓ_0} [Nikolova 13]

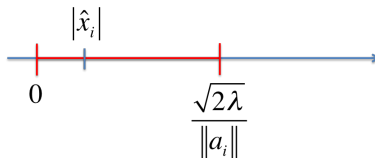




III/ Links $G_{\ell_0} \leftrightarrow G_{\text{CELO}}$ - *The global minimizers of G_{CELO}*



(a) Plot of g (blue) and g^{**} (red) when $\frac{1}{2}d^2 = \lambda$



(b) Illustration \hat{x}^0 thresholding



II.4) Links $G_{\ell_0} \leftrightarrow G_{\text{CELO}}$ - *The local minimizers of G_{CELO}*

Theorem (Link between **local minimizers** of G_{ℓ_0} and G_{CELO})

$\hat{x} \in \mathbb{R}^N$ local minimizer of $G_{\text{CELO}} \implies \hat{x}^0$ local minimizer of G_{ℓ_0}

and

$$G_{\text{CELO}}(\hat{x}) = G_{\text{CELO}}(\hat{x}^0) = G_{\ell_0}(\hat{x}^0)$$



II.4) Links $G_{\ell_0} \leftrightarrow G_{\text{CELO}}$ - *The local minimizers of G_{CELO}*

Theorem (Link between **local minimizers** of G_{ℓ_0} and G_{CELO})

$$\hat{x} \in \mathbb{R}^N \text{ local minimizer of } G_{\text{CELO}} \implies \hat{x}^0 \text{ local minimizer of } G_{\ell_0}$$

and

$$G_{\text{CELO}}(\hat{x}) = G_{\text{CELO}}(\hat{x}^0) = G_{\ell_0}(\hat{x}^0)$$

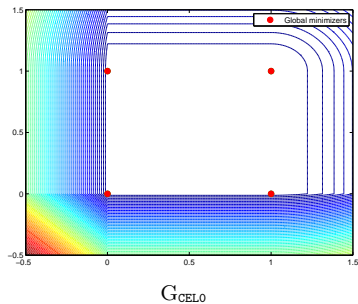
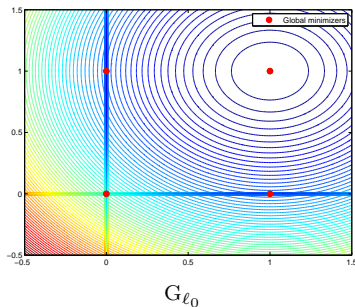
~~NO~~ NO converse property

While G_{CELO} and G_{ℓ_0} share the **same global minimizers**, G_{CELO} has potentially **fewer local minimizers** than G_{ℓ_0}





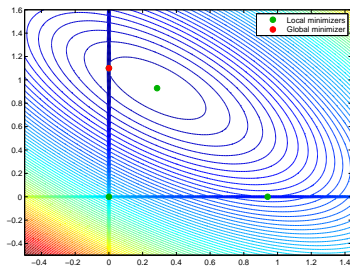
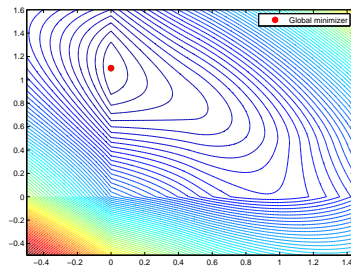
II.4) Links $G_{\ell_0} \leftrightarrow G_{\text{CELO}}$ - *Illustration on 2D examples*



$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad d = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \lambda = 0.5$$



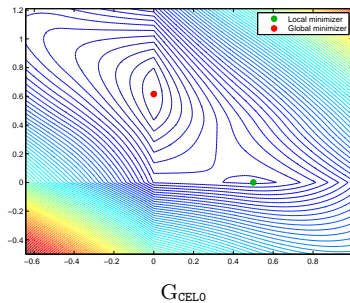
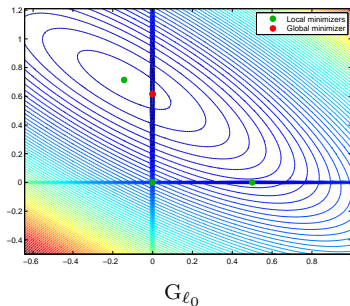
II.4) Links $G_{\ell_0} \leftrightarrow G_{\text{CELO}}$ - *Illustration on 2D examples*


 G_{ℓ_0}

 G_{CELO}

$$A = \begin{pmatrix} 0.5 & 2 \\ 2 & 1 \end{pmatrix}, \quad d = \begin{pmatrix} 2 \\ 1.5 \end{pmatrix}, \quad \lambda = 0.5$$



II.4) Links $G_{\ell_0} \leftrightarrow G_{\text{CELO}}$ - *Illustration on 2D examples*



$$A = \begin{pmatrix} 3 & 2 \\ 1 & 3 \end{pmatrix}, \quad d = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \lambda = 1$$



II.5) On a unified view of exact ℓ_0 penalties

Question

Is there any other *continuous* penalties leading to a relaxations of G_{ℓ_0} with the same properties as G_{CELO} ?

Objective

We consider continuous approximation of the ℓ_0 -norm of the form :

$$\Phi(x) = \sum_{i=1}^N \phi_i(x_i) \text{ and the associated } \tilde{G}(x) = \frac{1}{2} \|Ax - d\|_2^2 + \Phi(x).$$

Determination of **necessary and sufficient** conditions on ϕ_i such that $\forall d \in \mathbb{R}^M$,

$$\arg \min_{x \in \mathbb{R}^N} \tilde{G}(x) = \arg \min_{x \in \mathbb{R}^N} G_{\ell_0}(x), \quad (\text{P1})$$

$$\hat{x} \text{ (local) minimizer of } \tilde{G} \implies \hat{x} \text{ (local) minimizer of } G_{\ell_0}. \quad (\text{P2})$$





II.5) On a unified view of exact ℓ_0 penalties

One dimensional study - Notations

- $g_0(u) := \frac{1}{2}(au - d)^2 + \lambda|u|_0,$
- $\tilde{g}(u) := \frac{1}{2}(au - d)^2 + \phi(u),$
- B : a finite subset of points of \mathbb{R} on which ϕ is not differentiable.

Additional assumptions

- $\min_{u \in \mathbb{R}} g_0(u) = \min_{u \in \mathbb{R}} \tilde{g}(u),$
- ϕ is twice differentiable on $\mathbb{R} \setminus B.$

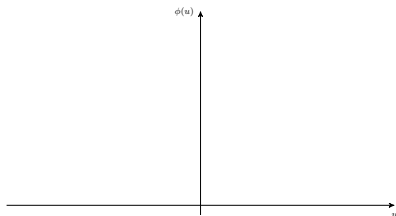


II.5) On a unified view of exact ℓ_0 penalties

Theorem (NS conds for P1)

\tilde{g} has property P1 $\forall d \in \mathbb{R}$ iff ϕ verifies :

- $\phi(0) = 0$,
- $\forall u \in \mathbb{R} \setminus \left(-\frac{\sqrt{2\lambda}}{a}, \frac{\sqrt{2\lambda}}{a}\right)$,
 $\phi(u) = \lambda|u|_0 = \lambda$,
- $\forall u \in \left(-\frac{\sqrt{2\lambda}}{a}, \frac{\sqrt{2\lambda}}{a}\right) \setminus \{0\}$,
 $\phi(u) > \phi_{\text{CELO}}(a, \lambda; u)$



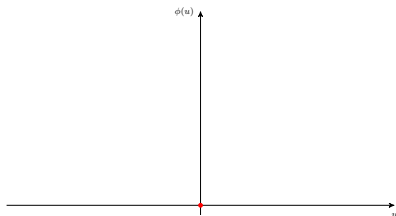


II.5) On a unified view of exact ℓ_0 penalties

Theorem (NS conds for P1)

\tilde{g} has property P1 $\forall d \in \mathbb{R}$ iff ϕ verifies :

- $\phi(0) = 0$,
- $\forall u \in \mathbb{R} \setminus \left(-\frac{\sqrt{2\lambda}}{a}, \frac{\sqrt{2\lambda}}{a}\right)$,
 $\phi(u) = \lambda|u|_0 = \lambda$,
- $\forall u \in \left(-\frac{\sqrt{2\lambda}}{a}, \frac{\sqrt{2\lambda}}{a}\right) \setminus \{0\}$,
 $\phi(u) > \phi_{\text{CELO}}(a, \lambda; u)$



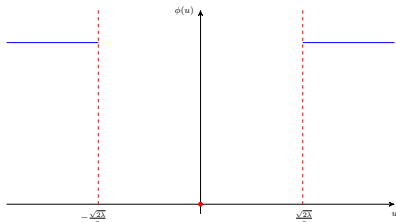


II.5) On a unified view of exact ℓ_0 penalties

Theorem (NS conds for P1)

\tilde{g} has property P1 $\forall d \in \mathbb{R}$ iff ϕ verifies :

- $\phi(0) = 0$,
- $\forall u \in \mathbb{R} \setminus \left(-\frac{\sqrt{2\lambda}}{a}, \frac{\sqrt{2\lambda}}{a}\right)$,
 $\phi(u) = \lambda|u|_0 = \lambda$,
- $\forall u \in \left(-\frac{\sqrt{2\lambda}}{a}, \frac{\sqrt{2\lambda}}{a}\right) \setminus \{0\}$,
 $\phi(u) > \phi_{\text{CELO}}(a, \lambda; u)$



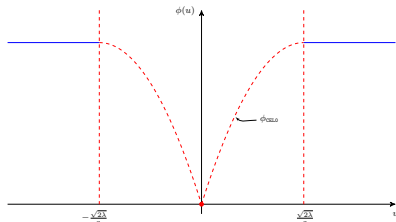


II.5) On a unified view of exact ℓ_0 penalties

Theorem (NS conds for P1)

\tilde{g} has property P1 $\forall d \in \mathbb{R}$ iff ϕ verifies :

- $\phi(0) = 0$,
- $\forall u \in \mathbb{R} \setminus \left(-\frac{\sqrt{2\lambda}}{a}, \frac{\sqrt{2\lambda}}{a}\right)$,
 $\phi(u) = \lambda|u|_0 = \lambda$,
- $\forall u \in \left(-\frac{\sqrt{2\lambda}}{a}, \frac{\sqrt{2\lambda}}{a}\right) \setminus \{0\}$,
 $\phi(u) > \phi_{\text{CELO}}(a, \lambda; u)$



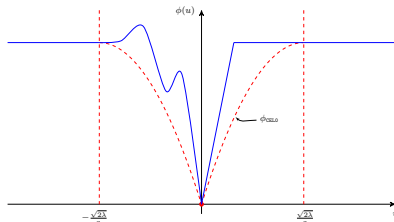


II.5) On a unified view of exact ℓ_0 penalties

Theorem (NS conds for P1)

\tilde{g} has property P1 $\forall d \in \mathbb{R}$ iff ϕ verifies :

- $\phi(0) = 0$,
- $\forall u \in \mathbb{R} \setminus \left(-\frac{\sqrt{2\lambda}}{a}, \frac{\sqrt{2\lambda}}{a}\right)$,
 $\phi(u) = \lambda|u|_0 = \lambda$,
- $\forall u \in \left(-\frac{\sqrt{2\lambda}}{a}, \frac{\sqrt{2\lambda}}{a}\right) \setminus \{0\}$,
 $\phi(u) > \phi_{\text{CELO}}(a, \lambda; u)$





II.5) On a unified view of exact ℓ_0 penalties

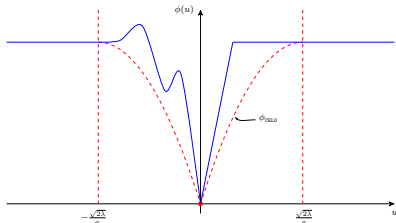
Theorem (NS conds for P1 and P2)

\tilde{g} has property P1 and P2 $\forall d \in \mathbb{R}$ iff in addition to the previous conditions, ϕ verifies :

- $\forall u \in B \setminus \{0\}, \lim_{\substack{v \rightarrow u \\ v < u}} \phi'(v) > \lim_{\substack{v \rightarrow u \\ v > u}} \phi'(v)$

- $\forall u \in (\beta^-, \beta^+) \setminus B, \phi''(u) < -a^2$

for $\beta^- \in \left[-\frac{\sqrt{2\lambda}}{a}, 0\right)$ and $\beta^+ \in \left(0, \frac{\sqrt{2\lambda}}{a}\right]$.





II.5) On a unified view of exact ℓ_0 penalties

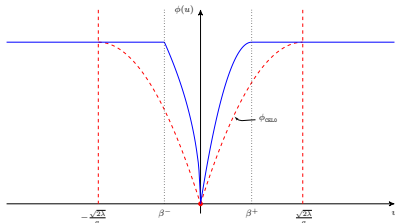
Theorem (NS conds for P1 and P2)

\tilde{g} has property P1 and P2 $\forall d \in \mathbb{R}$ iff in addition to the previous conditions, ϕ verifies :

- $\forall u \in B \setminus \{0\}, \lim_{\substack{v \rightarrow u \\ v < u}} \phi'(v) > \lim_{\substack{v \rightarrow u \\ v > u}} \phi'(v)$

- $\forall u \in (\beta^-, \beta^+) \setminus B, \phi''(u) < -a^2$

for $\beta^- \in \left[-\frac{\sqrt{2\lambda}}{a}, 0\right)$ and $\beta^+ \in \left(0, \frac{\sqrt{2\lambda}}{a}\right]$.





II.5) On a unified view of exact ℓ_0 penalties

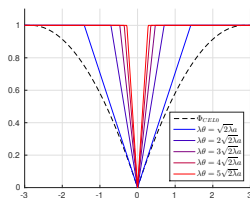
Examples using state of the art penalties

Penalty	Def ϕ_i	P1	P2	Conditions
Cap- ℓ_1 [Zhang 09]	$\lambda \min \{ \theta_i x_i , 1 \}$	✓	✗	$\lambda \theta_i \geq \sqrt{2\lambda} \ a_i\ $
SCAD [Fan and Li 01]	$\begin{cases} \bar{\lambda}_i u & \text{if } u \leq \bar{\lambda}_i, \\ \frac{2\gamma_i \bar{\lambda}_i u - \bar{\lambda}_i^2 - u^2}{2(\gamma_i - 1)} & \text{if } \bar{\lambda}_i < u \leq \gamma_i \bar{\lambda}_i, \\ \frac{(\gamma_i + 1) \bar{\lambda}_i^2}{2} & \text{if } u > \gamma_i \bar{\lambda}_i \end{cases}$	✓	✗	$\frac{(\gamma_i + 1) \bar{\lambda}_i^2}{2} = \lambda$ $2 < \gamma_i \leq \frac{1}{\ a_i\ } - 1$
MCP [Zhang 10]	$\begin{cases} \frac{\gamma_i \bar{\lambda}_i^2}{2} & \text{if } u > \gamma_i \bar{\lambda}_i \\ \left(\bar{\lambda}_i u - \frac{u^2}{2\gamma_i} \right) & \text{if } u \leq \gamma_i \bar{\lambda}_i \end{cases}$	✓	✓	$\frac{\gamma_i \bar{\lambda}_i^2}{2} = \lambda$ $\gamma_i < \frac{1}{\ a_i\ ^2}$
Trunc- ℓ_p	$\lambda \min \{ \theta_i x_i ^{p_i}, 1 \}$	✓	✓	$\theta_i \geq \left(\frac{\ a_i\ ^2}{p_i(1-p_i)\lambda} \right)^{p_i/2}$

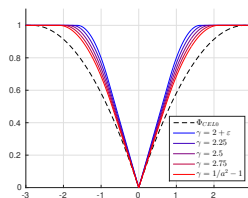




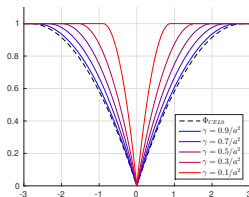
II.5) On a unified view of exact ℓ_0 penalties



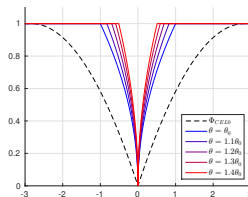
(c) Capped- ℓ_1



(d) SCAD



(e) MCP



(f) Truncated- ℓ_p

FIGURE: Examples of penalties for which P1 (Top) or P1 and P2 (Bottom) hold for $a = 0.5$, $\lambda = 1$ and $d = 1.8$.





II.5) On a unified view of exact ℓ_0 penalties

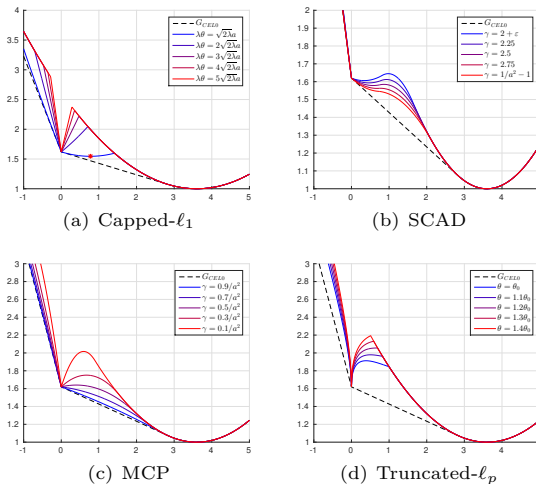


FIGURE: Examples of penalties for which P1 (Top) or P1 and P2 (Bottom) hold for $a = 0.5$, $\lambda = 1$ and $d = 1.8$.





IV/ Minimizing the ℓ_2 -CEL0 functional

Nonsmooth nonconvex algorithms

The **continuity of G_{CEL0}** allows to use recent *nonsmooth nonconvex* algorithms to minimize (indirectly) G_{ℓ_0} ,

- *Difference of Convex* (DC) functions programming [Gasso et al 09]
- *Majorization-Minimization* (MM) algorithms (*e.g.* Iteratively Reweighted ℓ_1 (IRL1) [Ochs et al 14])
- *Forward-Backward splitting* (GIST [Gong et al 13], [Attouch et al 13])



V/ Minimizing the ℓ_2 -CEL0 functional - *FB Algorithm*

$$G_{\text{CEL0}}(x) := \frac{1}{2} \|Ax - d\|^2 + \Phi_{\text{CEL0}}(x)$$

Forward-Backward Splitting Algorithm

$$x^{k+1} \in \text{prox}_{\gamma^k \Phi_{\text{CEL0}}(\cdot)} \left(x^k - \gamma^k A^T (Ax^k - d) \right),$$

where $0 < \underline{\gamma} < \gamma^k < \bar{\gamma} < \frac{1}{L}$ with L the Lipschitz constant of the gradient of the quadratic term.

- If the functional has the Kurdyka-Lojaseiwicz (KL) property, the generated sequence converges to a critical point of the minimized functional [Attouch et al 13]
- With $\text{prox}_{\gamma|\cdot|_0} \Rightarrow$ *Iterative Hard Thresholding* [Attouch et al 13] previously introduced by [Blumensath and Davies 08]





II.6) Minimizing the CEL0 functional - *Proximal operator*

Proximal operator for $\|\cdot\|_0$ and Φ_{CEL0}

$$\text{prox}_{\gamma f(\cdot)}(y) := \arg \min_{x \in \mathbb{R}^N} \frac{1}{2\gamma} \|x - y\|^2 + f(x).$$

Since $\|\cdot\|_0$ and Φ_{CEL0} are additively separable, this problem is reduced to N independent 1D problems,

$$\begin{aligned} \text{prox}_{\gamma|\cdot|_0}(u) &= u \mathbb{1}_{\{|u| > \sqrt{2\gamma\lambda}\}} + \{0, u\} \mathbb{1}_{\{|u| = \sqrt{2\gamma\lambda}\}} \\ \text{prox}_{\gamma\phi(a, \lambda; \cdot)}(u) &= \begin{cases} \text{sign}(u) \min \left(|u|, (|u| - \sqrt{2\lambda\gamma}a)_+ / (1 - a^2\gamma) \right) & \text{if } a^2\gamma < 1, \\ u \mathbb{1}_{\{|u| > \sqrt{2\gamma\lambda}\}} + \{0, u\} \mathbb{1}_{\{|u| = \sqrt{2\gamma\lambda}\}} & \text{if } a^2\gamma \geq 1. \end{cases} \end{aligned} \quad (4)$$



II.6) Minimizing the CEL0 functional - *Proximal operator*

Proximal operator for $\|\cdot\|_0$ and Φ_{CEL0}

$$\text{prox}_{\gamma f(\cdot)}(y) := \arg \min_{x \in \mathbb{R}^N} \frac{1}{2\gamma} \|x - y\|^2 + f(x).$$

Since $\|\cdot\|_0$ and Φ_{CEL0} are additively separable, this problem is reduced to N independent 1D problems,

$$\begin{aligned} \text{prox}_{\gamma \|\cdot\|_0}(u) &= u \mathbb{1}_{\{|u| > \sqrt{2\gamma\lambda}\}} + \{0, u\} \mathbb{1}_{\{|u| = \sqrt{2\gamma\lambda}\}} \\ \text{prox}_{\gamma \phi(a, \lambda; \cdot)}(u) &= \begin{cases} \text{sign}(u) \min(|u|, (|u| - \sqrt{2\gamma\lambda}a)_+ / (1 - a^2\gamma)) & \text{if } a^2\gamma < 1, \\ u \mathbb{1}_{\{|u| > \sqrt{2\gamma\lambda}\}} + \{0, u\} \mathbb{1}_{\{|u| = \sqrt{2\gamma\lambda}\}} & \text{if } a^2\gamma \geq 1. \end{cases} \end{aligned} \quad (4)$$

$$\left. \begin{aligned} \gamma &\in \left(0, \frac{1}{\|A\|^2}\right) \\ \|A\| &= \sup_{\|x\|=1} \|Ax\| \geq \max_{i \in \mathbb{I}_N} \|a_i\| \end{aligned} \right\} \Rightarrow \forall i \in \mathbb{I}_N, \|a_i\|^2 \gamma < 1$$





IV/ Minimizing the ℓ_2 -CEL0 functional - *Proximal operator*

Proximal operator for $\|\cdot\|_0$, Φ_{CEL0} and $\|\cdot\|_1$.

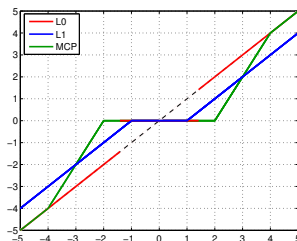


FIGURE: Proximal operator of ℓ_0 (red), Φ_{CEL0} (green), ℓ_1 (blue).



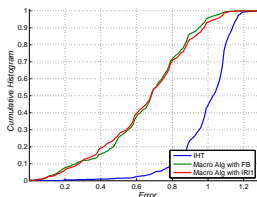
V/ Numerical Illustrations and comparisons – *Minimizing G_{ℓ_0}*

Experiment description

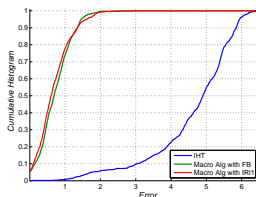
- generation of 1000 random $A \in \mathbb{R}^{7 \times 15}$ and $d \in \mathbb{R}^7$ from an uniform distribution
- resolution using the IHT algorithm for G_{ℓ_0} and the IRL1 or FBS algorithm G_{CELO} with the same initialization
- for each generation and each algorithm used, computation of the error $|G_{\ell_0}(\hat{x}) - G_{\ell_0}(x^*)|$ where x^* is a global minimizer of G_{ℓ_0} and G_{CELO} computed by combinatorial search on this small example



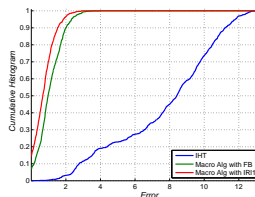
V/ Numerical Illustrations and comparisons – *Minimizing G_{ℓ_0}*



(a) $\lambda = 0.1$



(b) $\lambda = 0.5$



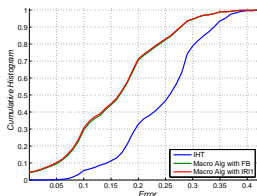
(c) $\lambda = 1$

FIGURE: Cumulative histograms of the error $|G_{\ell_0}(\hat{x}) - G_{\ell_0}(x^*)|$. Algorithms initialization : $x^0 = A^T d$. The experiment is repeated for three values of λ (0.1, 0.5 and 1). IHT algorithm (blue), macro algorithm combined with IR11 (red) or FB (green).

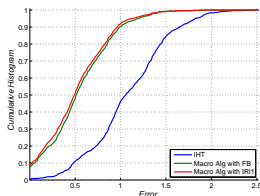




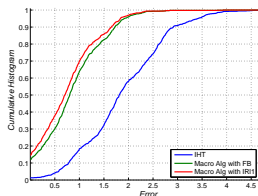
V/ Numerical Illustrations and comparisons – *Minimizing G_{ℓ_0}*



(a) $\lambda = 0.1$



(b) $\lambda = 0.5$



(c) $\lambda = 1$

FIGURE: Cumulative histograms of the error $|G_{\ell_0}(\hat{x}) - G_{\ell_0}(x^*)|$. Algorithms initialization : $x^0 = x_{\ell_1}$ the solution of the ℓ_1 relaxed problem. The experiment is repeated for three values of λ (0.1, 0.5 and 1). IHT algorithm (blue), macro algorithm combined with IR11 (red) or FB (green).





V/ Numerical Illustrations and comparisons – *Channel estimation*

Experiment description

- signal at the receiver = composition of scaled and delayed versions of the transmitted signal,
- channel modelled as a finite set of K delays,
- the signal is pilot-aided and symbols of the pilot are known at the receiver,

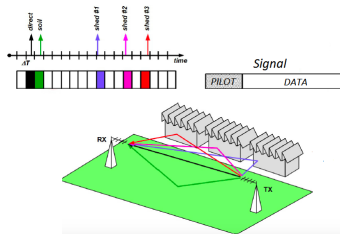


FIGURE: Channel estimation problem

Problem formulation :

$$y = Sh + n$$

where

- S is a matrix defined by the symbols of the pilot,
- h is a sparse vector containing K nonzero entries corresponding to the unknown delays,
- y is the received signal,
- n is an additive noise.





V/ Numerical Illustrations and comparisons – *Channel estimation*

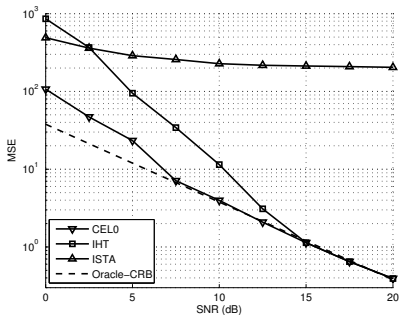


FIGURE: MSE vs. SNR for channel estimation ($K = 9$ and $N=600$) using ℓ_1 minimization (FBS with soft threshold) and ℓ_0 minimization (IHT or CEL0 with FBS). The dashed line represents the Oracle-CRB.

Joint work with

- Adilson Chinatto, Cynthia Junqueira, João M. T. Romano from University of Campinas, Brazil
- Pascal Larzabal and Jean-Pierre Barbot from ENS Cachan, France.





Take home message

Take home message

- To minimize G_{ℓ_0} replace it by G_{CEL0} ,
- It does not change the global minimizers,
- It avoids some locals minimizers (does not change the others),
- In the framework of IHT (FBS) \implies replace the Hard threshold by the continuous CEL0 threshold,
- G_{CEL0} is continuous, allowing to use recent nonsmooth nonconvex algorithms
- G_{CEL0} is an “optimal” continuous relaxation of G_{ℓ_0} in a certain sense

Ongoing work

- Algorithms to minimize efficiently the CEL0 functional,
- Application to Signal/Image problems (Sparse spikes deconvolution, Direction Of Arrivals estimation, Photo Activation Localization Microscopy...)
- Generalization to non-quadratics data fidelity terms.





MERCI



Ecole PEYRESQ en Traitement du Signal, Image et Vision.





References I



AMIR BECK AND MARC TEOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.



HEDY ATTOUCH, JÉRÔME BOLTE, AND BENAR FUX SVAITER, *Convergence of descent methods for semi-algebraic and tame problems : proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods*, Mathematical Programming, 137 (2013), pp. 91–129.



THOMAS BLUMENSATH AND MIKE E DAVIES, *Iterative thresholding for sparse approximations*, Journal of Fourier Analysis and Applications, 14 (2008), pp. 629–654.



SÉBASTIEN BOURGUIGNON, JORDAN NININ, HERVÉ CARFANTAN AND MARCEL MONGEAU, *Optimisation exacte de critères parcimonieux en norme ℓ_0 par programmation mixte en nombres entiers*, Colloque GRETSI 2015.



LEO BREIMAN, *Better subset regression using the nonnegative garrote*, Technometrics, 37 (1995), pp. 373–384.



EMMANUEL J CANDÈS, JUSTIN ROMBERG, AND TERENCE TAO, *Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information*, Information Theory, IEEE Transactions on, 52 (2006), pp. 489–509.



EMMANUEL J CANDÈS, MICHAEL B WAKIN, AND STEPHEN P BOYD, *Enhancing sparsity by reweighted ℓ_1 minimization*, Journal of Fourier analysis and applications, 14 (2008), pp. 877–905.



RICK CHARTRAND, *Exact reconstruction of sparse signals via nonconvex minimization*, Signal Processing Letters, IEEE, 14 (2007), pp. 707–710.





References II



SCOTT SHAOBING CHEN, DAVID L DONOHO, AND MICHAEL A SAUNDERS, *Atomic decomposition by basis pursuit*, SIAM journal on scientific computing, 20 (1998), pp. 33–61.



EMILIE CHOUZENOUX, ANNA JEZIERSKA, JEAN-CHRISTOPHE PESQUET AND HUGUES TALBOT, *A majorize-minimize subspace approach for ℓ_2 - ℓ_0 image regularization*, SIAM Journal on Imaging Sciences, 6 (2013), pp.563–591.



FRANK H CLARKE, *Optimization and nonsmooth analysis*, vol. 5, Siam, 1990.



INGRID DAUBECHIES, MICHEL DEFRISE, AND CHRISTINE DE MOL, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications on Pure and Applied Mathematics, 57 (2004), pp. 1413–1457.



DAVID L DONOHO, *For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution*, Communications on Pure and Applied Mathematics, 59 (2006), pp. 797–829.



JIANQING FAN AND RUNZE LI, *Variable selection via nonconcave penalized likelihood and its oracle properties*, Journal of the American Statistical Association, 96 (2001), pp. 1348–1360.



SIMON FOUCART AND MING-JUN LAI, *Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$* , Applied and Computational Harmonic Analysis, 26 (2009), pp. 395–407.



G.M. FUNG AND O.L. MANGASARIAN, *Equivalence of minimal ℓ_0 - and ℓ_p -norm solutions of linear equalities, inequalities and linear programs for sufficiently small p* , Journal of optimization theory and applications, 151 (2011), pp. 1–10.





References III



GILLES GASSO, ALAIN RAKOTOMAMONJY, AND STÉPHANE CANU, *Recovering sparse signals with a certain family of nonconvex penalties and DC programming*, Signal Processing, IEEE Transactions on, 57 (2009), pp. 4686–4698.



PINGHUA GONG, CHANGSHUI ZHANG, ZHAOSONG LU, JIANHUA HUANG, AND JIEPING YE, *A General Iterative Shrinkage and Thresholding Algorithm for Non-convex Regularized Optimization Problems*, in Proceedings of The 30th International Conference on Machine Learning, 2013, pp. 37–45.



CÉDRIC HERZET AND ANGÉLIQUE DRÉMEAU, *Bayesian Pursuit Algorithms*, in Proceedings of European Signal Processing conference (EUSIPCO), Aalborg, Denmark, August 2010.



HOAI AN LE THI, HOAI MINH LE AND TAO PHAM DINH, *Feature selection in machine learning : an exact penalty approach using a Difference of Convex function Algorithm*, Machine Learning, 2014, pp. 1–24.



HOAI AN LE THI, TAO PHAM DINH, HOAI MINH LE AND XUAN THANH VO, *DC approximation approaches for sparse optimization*, European Journal of Operational Research, 244 (2015), pp. 26–46.



STÉPHANE G MALLAT AND ZHIFENG ZHANG, *Matching pursuits with time-frequency dictionaries*, Signal Processing, IEEE Transactions on, 41 (1993), pp. 3397–3415.



HOSEIN MOHIMANI, MASSOUD BABAIE-ZADEH AND CHRISTIAN JUTTEN, *A fast approach for overcomplete sparse decomposition based on smoothed ℓ_0 norm*, Signal Processing, IEEE Transactions on, 57 (2008), pp. 289–301.



MILA NIKOLOVA, *Description of the minimizers of least squares regularized with ℓ_0 -norm. Uniqueness of the global minimizer*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 904–937.





References IV



P. OCHS, A. DOSOVITSKIY, T. BROX, AND T. POCK, *An iteratively reweighted Algorithm for Non-smooth Non-convex Optimization in Computer Vision*, Technical Report, (2014).



YAGYENSH CHANDRA PATI, RAMIN REZAIHAFAR, AND PS KRISHNAPRASAD, *Orthogonal matching pursuit : Recursive function approximation with applications to wavelet decomposition*, in Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on, IEEE, 1993, pp. 40–44.



CHARLES SOUSSEN, JÉRÔME IDIER, DAVID BRIE, AND JUNBO DUAN, *From Bernoulli–Gaussian deconvolution to sparse signal restoration*, Signal Processing, IEEE Transactions on, 59 (2011), pp. 4572–4584.



R. TIBSHIRANI, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society, 46 (1996), pp. 431–439.



JOEL A TROPP, *Greed is good : Algorithmic results for sparse approximation*, Information Theory, IEEE Transactions on, 50 (2004), pp. 2231–2242.



CUN-HUI ZHANG, *Multi-stage convex relaxation for learning with sparse regularization*, Advances in Neural Information Processing Systems, (2009), pp. 1929–1936.



CUN-HUI ZHANG, *Nearly unbiased variable selection under minimax concave penalty*, The Annals of Statistics, (2010), pp. 894–942.



HUI ZOU, *The adaptive lasso and its oracle properties*, Journal of the American statistical association, 101 (2006), pp. 1418–1429.

