CEL0: A CONTINUOUS ALTERNATIVE TO ℓ_0 PENALTY.

Emmanuel Soubies† Laure Blanc-Féraud† Gilles Aubert‡

† Université de Nice Sophia Antipolis, CNRS, Laboratoire I3S, UMR 7271, France.
‡ Université de Nice Sophia Antipolis, CNRS, Laboratoire J.A.D, UMR 7351, France.

Context & Notations

We consider the following ℓ_0 penalized least squares problem,

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

where $A \in \mathbb{R}^{M\times N}$, $M \ll N$, $d \in \mathbb{R}^M$. This *NP-hard* combinatorial problem is encountered in signal processing areas as coding, compressed sensing, variable selection, source separation, learning...

In this work, we propose a new *continuous nonsmooth nonconvex penalty* allowing an *exact reformulation* of problem (1).

Notations:
- $\mathbb{I}_N = \{1,\cdots,N\}$, $a_i \in \mathbb{R}^M$ denotes the ith column of $A \in \mathbb{R}^{M\times N}$,
- $\sigma(x) := \{i \in \mathbb{I}_N;\ x_i \neq 0\}$ defines the support of $x \in \mathbb{R}^N$,
- $\sigma^-(x) := \left\{i \in \sigma(x) : |x_i| < \sqrt{2\lambda}/\|a_i\|\right\}$ a part of the support.

The Continuous Exact ℓ_0 (CEL0) penalty [6]

Replacing the ℓ_0 pseudo norm in (1) by the CEL0 penalty,

$$\Phi_{\text{CEL0}}(x) := N\lambda - \sum_{i\in\mathbb{I}_N} \frac{\|a_i\|^2}{2}\left(|x_i| - \frac{\sqrt{2\lambda}}{\|a_i\|}\right)^2 \mathbb{1}_{\{|x_i|\leq\frac{\sqrt{2\lambda}}{\|a_i\|}\}},$$

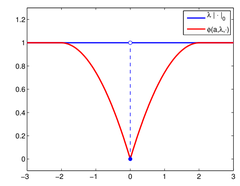


Fig. 1: ℓ_0 and CEL0 penalties.

leads to a *tight continuous relaxation* of the objective function G_{ℓ_0} denoted G_{CEL0},

$$G_{\text{CEL0}}(x) := \frac{1}{2}\|Ax-d\|_2^2 + \Phi_{\text{CEL0}}(x)\,. \tag{2}$$

When A *is orthogonal*, G_{CEL0} is the *convex hull* of G_{ℓ_0}—*i.e.* $G_{\text{CEL0}}(x) = G_{\ell_0}^{\star\star}(x)$—and the sparsest solution $x^\star$ is given by the following hard thresholding rule,

$$\forall i \in \mathbb{I}_N,\ x_i^\star = \frac{z_i}{\|a_i\|^2}\mathbb{1}_{\{|z_i|>\sqrt{2\lambda}\|a_i\|\}} \quad \text{where} \quad z = A^T d\,. \tag{3}$$

Links between minimizers of G_{CEL0} and G_{ℓ_0} (for any $A \in \mathbb{R}^{M\times N}$)

Theorem 1 (link between global minimizers)

(i) the set of global minimizers of G_{ℓ_0} is included in the set of global minimizers of G_{CEL0},

$$\arg\min_{x\in\mathbb{R}^N} G_{\ell_0}(x) \subseteq \arg\min_{x\in\mathbb{R}^N} G_{\text{CEL0}}(x) \tag{4}$$

(ii) conversely if $\hat{x} \in \mathbb{R}^N$ is a global minimizer of G_{CEL0}, $\hat{x}^0$ defined by

$$\forall i \in \mathbb{I}_N, \quad \hat{x}_i^0 = \hat{x}_i \mathbb{1}_{\{|\hat{x}_i|\geq\frac{\sqrt{2\lambda}}{\|a_i\|}\}}\,, \tag{5}$$

is a global minimizer of G_{ℓ_0} and $G_{\text{CEL0}}(\hat{x}) = G_{\text{CEL0}}(\hat{x}^0) = G_{\ell_0}(\hat{x}^0)$.

Each global minimizer of G_{ℓ_0} is a global minimizer of G_{CEL0}.
The set of global minimizers of G_{CEL0} is nonempty (since the one of G_{ℓ_0} is nonempty from [4, Th 4.4 (i)]).

From each global minimizer of G_{CEL0} one can easily identify a global minimizer of G_{ℓ_0} by simple thresholding.

Theorem 2 (link between local minimizers)

Let G_{CEL0} have a local minimum at $\hat{x} \in \mathbb{R}^N$. Then $\hat{x}^0$ (defined by (5)) is a local minimizer of G_{ℓ_0} and

$$G_{\text{CEL0}}(\hat{x}) = G_{\text{CEL0}}(\hat{x}^0) = G_{\ell_0}(\hat{x}^0)$$

- Th 2 extends of the second point of Th 1 to all local minimizers.
- Some local minimizers of G_{ℓ_0} are not critical point for G_{CEL0}. In particular, G_{CEL0} *eliminates* the local minimizers $\hat{x}$ of G_{ℓ_0} such that $\sigma^-(\hat{x}) \neq \emptyset$.

Conclusion One can address problem (1) by minimizing the *continuous* functional G_{CEL0} instead of G_{ℓ_0} since the global minimizers of G_{CEL0} *contain* those of G_{ℓ_0} and that G_{CEL0} has *"less" local minimizers* than G_{ℓ_0}.

Numerical Illustrations

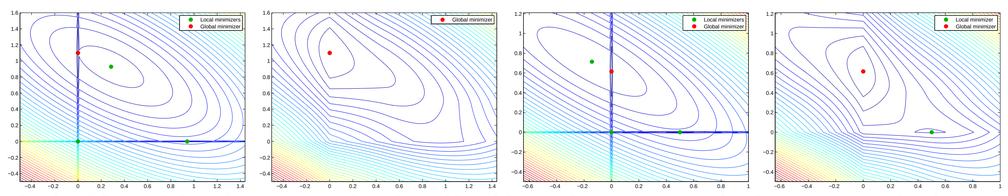
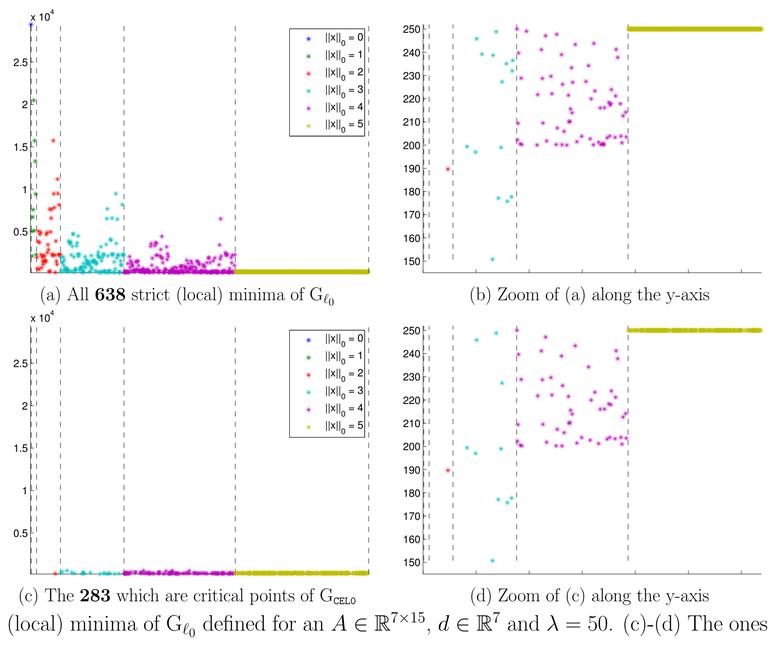


Fig. 2: Isolevels of G_{ℓ_0} and G_{CEL0} for two examples where $N = M = 2$. Green points are local minimizers and red ones are global.

From [4], we know that a strict (local) minimizer of G_{ℓ_0} is easy to compute by choosing a support $\omega \in \Omega_{max}$ and solving the normal equations restricted to this support where,

$$\Omega_{max} = \bigcup_{r=0}^{M} \Omega_r \quad \text{where} \quad \Omega_r = \left\{\omega \subset \mathbb{I}_N \ :\ \#\omega = r = \operatorname{rank}(A_\omega)\right\}. \tag{6}$$

Hence, for a given matrix A and vector d, one can compute all the strict (local) minimizers of G_{ℓ_0} since Ω_{max} is a finite set and check which ones are also *critical points* for G_{CEL0} (i.e. verify the conditions of [6, Lemma 4.1]).



(a) All **638** strict (local) minima of G_{ℓ_0} (b) Zoom of (a) along the y-axis

(c) The **283** which are critical points of G_{CEL0} (d) Zoom of (c) along the y-axis

Fig. 3: (a)-(b) Strict (local) minima of G_{ℓ_0} defined for an $A \in \mathbb{R}^{7\times 15}$, $d \in \mathbb{R}^7$ and $\lambda = 50$. (c)-(d) The ones which are also critical points of G_{CEL0}.

Minimizing G_{CEL0}

The *continuity* of G_{CEL0} allows to use nonsmooth nonconvex algorithms for its minimization,

(Alg) - *DC programming* [3]
- *MM algorithms* (*e.g.* IRL1 [5])
- *Forward-Backward splitting* [1]

$\Rightarrow$ converge to a **critical point** of G_{CEL0}

Lemma (critical points of G_{CEL0} and minimizers of G_{ℓ_0})

Let $\hat{x} \in \mathbb{R}^N$ be a critical point of G_{CEL0} verifying $\sigma^-(\hat{x}) = \emptyset$. Then it is a (local) minimizer of G_{ℓ_0} and $G_{\text{CEL0}}(\hat{x}) = G_{\ell_0}(\hat{x})$.

Macro algorithm [6, Algo 1]

- adds an outer loop to any nonsmooth nonconvex algorithm, *e.g.* see (Alg).
- moves iteratively from a critical point $\hat{x}$ of G_{CEL0} to another one, while decreasing the cost function, until $\sigma^-(\hat{x}) = \emptyset$.
- converges to a minimizer of G_{ℓ_0} [6, Th 5.1]

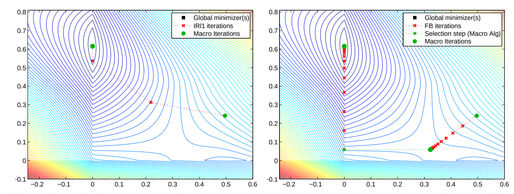


Fig. 4: 2D illustrations of the Macro algorithm using IRL1 (left) and FBS (right).

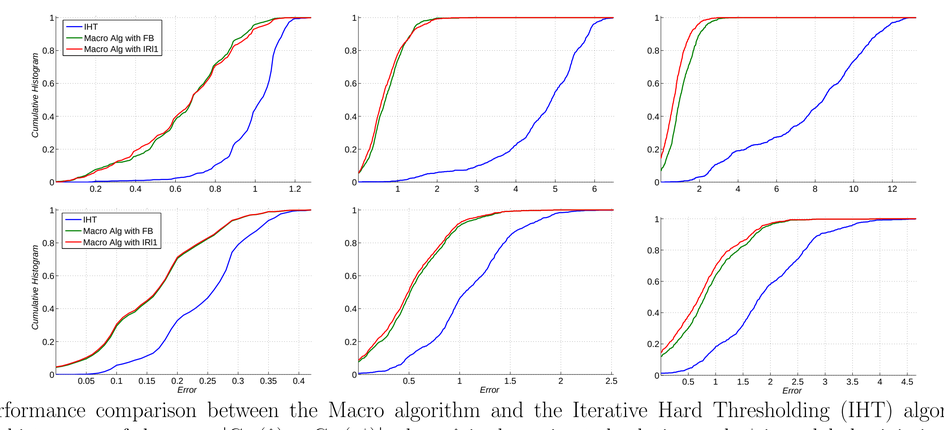


Fig. 5: Performance comparison between the Macro algorithm and the Iterative Hard Thresholding (IHT) algorithm [2, 1]. Cumulative histograms of the error $|G_{\ell_0}(\hat{x}) - G_{\ell_0}(x^\star)|$ where $\hat{x}$ is the estimated solution and $x^\star$ is a global minimizer of G_{ℓ_0}.The histograms are computed from 1000 random matrices $A \in \mathbb{R}^{7\times 15}$ and $d \in R^7$ generated from an uniform distribution. From left to right: $\lambda = 0.1$, $\lambda = 0.5$ and $\lambda = 1$. Top line: initialization $x_0 = A^T d$. Bottom line: initialization $x_0 = x_{\ell_1}^\star$.

References

[1] 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(1-2):91–129, 2013.

[2] Thomas Blumensath and Mike E Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5-6):629–654, 2008.

[3] Gilles Gasso, Alain Rakotomamonjy, and Stéphane Canu. Recovering sparse signals with a certain family of nonconvex penalties and DC programming. *IEEE Transactions on Signal Processing*, 57(12):4686–4698, 2009.

[4] 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(2):904–937, 2013.

[5] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock. On Iteratively Reweighted Algorithms for Nonsmooth Nonconvex Optimization in Computer Vision. *SIAM Journal on Imaging Sciences*, 8(1):331–372, 2015.

[6] E. Soubies, L. Blanc-Féraud, and G. Aubert. A Continuous Exact ℓ_0 penalty (CEL0) for least squares regularized problems. *Preprint hal-01102492. Accepted for publication in SIAM Journal on Imaging Sciences*, 2015.