On the Local Minimizers of the CEL0 Relaxation

Emmanuel Soubies, Laure Blanc-Féraud, Gilles Aubert

On the Local Minimizers of the CEL0 Relaxation

Emmanuel Soubies
IRIT, Université de Toulouse, CNRS.
Email: emmanuel.soubies@irit.fr

Laure Blanc-Féraud
Université Côte d'Azur, CNRS, INRIA, I3S.
Email: laure.blancferaud@cnrs.fr

Gilles Aubert
Université Côte d'Azur, UNS, LJAD.
Email: gilles.aubert@unice.fr

***Abstract*—We study the strict local minimizers of the CEL0 functional, an exact continuous relaxation of the ℓ_0-regularized least-squares criterion. More precisely, we derive a necessary and sufficient condition for strict local optimality, recalling that global minimizers are strict. Moreover, we quantify the number of strict local (not global) minimizers of the initial functional that are eliminated by the relaxation.**

I. Introduction

Let $\mathbf{A} \in \mathbb{R}^{M\times N}$ with $M \ll N$ be an arbitrary linear operator and $\mathbf{y} \in \mathbb{R}^N$ be a measurement vector obtained through $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ for a sparse signal $\mathbf{x} \in \mathbb{R}^N$ (*i.e.,* $\|\mathbf{x}\|_0 \ll N$) and a vector of noise $\mathbf{n} \in \mathbb{R}^M$. The problem of recovering $\mathbf{x}$ from the data $\mathbf{y}$ has received a considerable attention in the context of compressed sensing [2]. To that end, one would like to solve

$$\hat{\mathbf{x}} \in \left\{ \arg\min_{\mathbf{x}\in\mathbb{R}^N} F_0(\mathbf{x}) := \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2 + \lambda\|\mathbf{x}\|_0 \right\}, \tag{1}$$

for a suitable choice of $\lambda > 0$. However, Problem (1) is non-convex and, furthermore, NP-hard, which makes its resolution a challenging task. Hence, it is customary to relax Problem (1) as

$$\hat{\mathbf{x}} \in \left\{ \arg\min_{\mathbf{x}\in\mathbb{R}^N} \tilde{F}(\mathbf{x}) := \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2 + \sum_{i=1}^{N}\phi_i(x_i) \right\}, \tag{2}$$

where $\phi_i : \mathbb{R} \to \mathbb{R}$, $i \in \{1,\dots,N\}$, are one-dimensional penalty functions. A variety of penalties have been proposed and studied over the past, going from the ℓ_1 convex relaxation [1] to continuous non-convex approximations of the ℓ_0 pseudo-norm [3, 6–8, 13–15]. Interestingly, some of them lead to exact continuous relaxations of the initial criterion (1) in the sense of Theorem 1 below [4, 11, 12]. In this work, we consider the continuous exact ℓ_0 (CEL0) relaxation that corresponds to the inferior limit of the class of exact relaxations derived in [12], and is defined by [11]

$$\phi_i(x) = \lambda - \frac{\|\mathbf{a}_i\|^2}{2}\left(|x| - \frac{\sqrt{2\lambda}}{\|\mathbf{a}_i\|}\right)^2 \mathbb{1}_{\left\{|x| \leq \frac{\sqrt{2\lambda}}{\|\mathbf{a}_i\|}\right\}}, \tag{3}$$

where $\mathbf{a}_i$ denotes the ith column of $\mathbf{A}$.

Theorem 1 (Links between (1) and (2)-(3) [4, 11])**.** *Let $\mathcal{L}_0$ (resp. $\tilde{\mathcal{L}}$) be the set of local minimizers of F_0 (resp. $\tilde{F}$). Let $\mathcal{G}_0 \subseteq \mathcal{L}_0$ (resp. $\tilde{\mathcal{G}} \subseteq \tilde{\mathcal{L}}$) be the corresponding subset of global minimizers. Then,*

1) *there exists a simple thresholding rule*[1] *$\mathcal{T} : \mathbb{R}^N \to \mathbb{R}^N$ such that for any $\mathbf{x} \in \tilde{\mathcal{L}}$, $\mathcal{T}(\mathbf{x}) \in \mathcal{L}_0$*
2) *$\mathcal{G}_0 \subseteq \tilde{\mathcal{G}}$.*

An important consequence of Theorem 1 is that, while each local minimizer of $\tilde{F}$ can be easily mapped to a local minimizer of F_0, the converse does not hold and some local (not global) minimizers of F_0 are removed by the relaxation $\tilde{F}$.

In the present communication, we make the following assumption.

Assumption 1. *Every pair $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)$ of global minimizers ($\hat{\mathbf{x}}_1 \neq \hat{\mathbf{x}}_2$) verify $\|\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2\|_0 > 1$.*

[1] Which is completely characterized in [11].

It is worth noting to mention that this assumption is not fulfilled only for a finitely number of λ values [10]. Then, we have the following corollary of Theorem 1.

Corollary 2. *Under Assumption 1, global minimizers of F_0 and $\tilde{F}$ coincide (i.e., $\mathcal{G}_0 = \tilde{\mathcal{G}}$). Moreover, they are stricts for both F_0 and $\tilde{F}$.*

Hence, under Assumption 1, being a strict local minimizer of $\tilde{F}$ is a necessary condition of global optimality for both $\tilde{F}$ and F_0. This motivates the analysis of the strict local minimizers of $\tilde{F}$.

II. The Strict Local Minimizers of $\tilde{F}$

Previous works [4, 11] were limited to the characterization of the critical points of $\tilde{F}$ and the study of the links between the minimizers of $\tilde{F}$ and F_0 (Theorem 1). Here, we go one-step further by providing, in Theorem 3, a necessary and sufficient condition to recognize critical points that are strict local minimizers of $\tilde{F}$.

Theorem 3 (Strict local optimality for $\tilde{F}$)**.** *A critical point $\mathbf{x} \in \mathbb{R}^N$ of $\tilde{F}$ is a strict (local) minimizer of $\tilde{F}$ if and only if*

- $\forall i \in \sigma_{\mathbf{x}}$, $|x_i| > \sqrt{2\lambda}/\|\mathbf{a}_i\|$,
- $\forall i \in \mathbb{I}_N\backslash\sigma_{\mathbf{x}}$, $|\langle \mathbf{a}_i, \mathbf{A}\mathbf{x}-\mathbf{y}\rangle| < \sqrt{2\lambda}\|\mathbf{a}_i\|$,
- $\mathrm{rank}(\mathbf{A}_{\sigma_{\mathbf{x}}}) = \#\sigma_{\mathbf{x}}$,

where $\mathbb{I}_N = \{1,\dots,N\}$, $\sigma_{\mathbf{x}} = \{i \in \mathbb{I}_N : |x_i| \neq 0\}$ and $\mathbf{A}_{\sigma_{\mathbf{x}}}$ is the restriction of $\mathbf{A}$ to the columns indexed by the elements of $\sigma_{\mathbf{x}}$.

III. Quantification of the Strict Local Minimizers

Given a small-size matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$, one can compute all strict local minimizers of F_0 by solving $(\mathbf{A}_\omega)^T\mathbf{A}_\omega\mathbf{x}_\omega = (\mathbf{A}_\omega)^T\mathbf{y}$ for any support $\omega \in \mathbb{I}_N$ such that $\mathrm{rank}(\mathbf{A}_\omega) = \#\omega$ [9, Theorem 3.2]. Moreover, as $\tilde{\mathcal{S}} \subseteq \mathcal{S}_0$ [11, Corollary 4.9] – where $\mathcal{S}_0$ (resp. $\tilde{\mathcal{S}}$) denotes the set of strict local minimizers of F_0 (resp. $\tilde{F}$) – Theorem 3 allows to extract all strict local minimizers of $\tilde{F}$ from those of F_0. We performed this experiment for $M = 5$ and $N = 10$. We report the results in Figure 1. One can make two observations,

- the situation is more favorable (more minimizers are removed by the relaxation) when $\mathbf{A}$ is generated from a Normal distribution (with good RIP properties). This observation is in line with the recent work [5] and deserves a deeper analysis,
- the behaviour for extremal values of λ is independent of the matrix $\mathbf{A}$. This is theoretically explained by Theorem 4.

Theorem 4. *Let $\mathcal{X}_{\mathrm{LS}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2$. Then, there exists $\lambda_0 > 0$ and $\lambda_\infty > 0$ such that*

1) *$\forall\lambda \in (\lambda_\infty, +\infty)$, $\tilde{\mathcal{S}} = \{\mathbf{0}_{\mathbb{R}^N}\}$,*
2) *$\forall\lambda \in (0, \lambda_0)$, $\tilde{\mathcal{S}} = (\mathcal{S}_0 \cap \mathcal{X}_{\mathrm{LS}})$.*

In other words, for a sufficiently small value of λ, only the strict minimizers of F_0 that solve the un-regularized least squares problem are minimizers of $\tilde{F}$. On the other hand, for a large λ, all strict local minimizers of F_0 are removed by $\tilde{F}$, except $\mathbf{0}_{\mathbb{R}^N}$ which (for such λ) is the global minimizer of both F_0 and $\tilde{F}$.

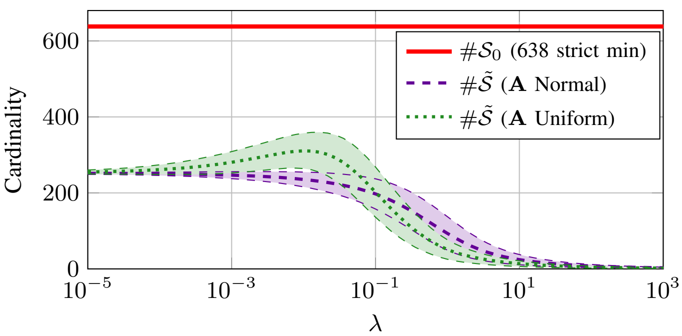

Fig. 1: Cardinality of $\tilde{\mathcal{S}}$ with respect to λ. The curves correspond to an average value (with standard deviation) over 2000 generations of $\mathbf{A} \in \mathbb{R}^{5\times 10}$ and $\mathbf{y} \in \mathbb{R}^5$ from a Normal (purple) or Uniform (green) distribution. As a reference, we plot the value of $\#\mathcal{S}_0$ that does not depend on λ [9, Remark 5]. Note that $\mathbf{A}$ and $\mathbf{y}$ are generated with the constraint that for any two supports $\omega \neq \omega'$ such that $\mathrm{rank}(\mathbf{A}_\omega) = \#\omega$ and $\mathrm{rank}(\mathbf{A}_{\omega'}) = \#\omega'$, the corresponding strict minimizers are different. This ensures that for each generation of $\mathbf{A}$, there is exactly $\binom{10}{5} = 252$ 5-sparse strict minimizers which corresponds to the limiting value for small λ (see Theorem 4).

References

[1] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.

[2] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling [a sensing/sampling paradigm that goes against the common knowledge in data acquisition]. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[3] Emmanuel J. Cands, Michael B. Wakin, and Stephen P. Boyd. Enhancing Sparsity by Reweighted l1 Minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, December 2008.

[4] Marcus Carlsson. On Convex Envelopes and Regularization of Non-Convex Functionals without moving Global Minima. *arXiv:1811.03439 [math]*, November 2018. arXiv: 1811.03439.

[5] Marcus Carlsson, Daniele Gerosa, and Carl Olsson. An unbiased approach to compressed sensing. *arXiv:1806.05283 [math]*, June 2018. arXiv: 1806.05283.

[6] E. Chouzenoux, A. Jezierska, J. Pesquet, and H. Talbot. A Majorize-Minimize Subspace Approach for l2-l0 Image Regularization. *SIAM Journal on Imaging Sciences*, 6(1):563–591, January 2013.

[7] Jianqing Fan and Runze Li. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

[8] Simon Foucart and Ming-Jun Lai. Sparsest solutions of underdetermined linear systems via lq-minimization for $0 < q \leq 1$. *Applied and Computational Harmonic Analysis*, 26(3):395–407, May 2009.

[9] M. Nikolova. Description of the Minimizers of Least Squares Regularized with l0-norm. Uniqueness of the Global Minimizer. *SIAM Journal on Imaging Sciences*, 6(2):904–937, January 2013.

[10] Mila Nikolova. Relationship between the optimal solutions of least squares regularized with l0 -norm and constrained by k-sparsity. *Applied and Computational Harmonic Analysis*, 41(1):237–265, July 2016.

[11] E. Soubies, L. Blanc-Fraud, and G. Aubert. A Continuous Exact l0 Penalty (CEL0) for Least Squares Regularized Problem. *SIAM Journal on Imaging Sciences*, 8(3):1607–1639, January 2015.

[12] E. Soubies, L. Blanc-Fraud, and G. Aubert. A Unified View of Exact Continuous Penalties for l2-l0 Minimization. *SIAM Journal on Optimization*, 27(3):2034–2060, January 2017.

[13] Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, April 2010.

[14] Tong Zhang. Multi-stage Convex Relaxation for Learning with Sparse Regularization. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 1929–1936. Curran Associates, Inc., 2009.

[15] Hui Zou. The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476):1418–1429, December 2006.