Relationships between necessary optimality conditions for the ℓ_2-ℓ_0 minimization problem.

Emmanuel Soubies

Imaging in Paris Seminar - October 3 2019

Signal and Communication Group, IRIT, Université de Toulouse, CNRS

L. Blanc-Féraud

G. Aubert


MULTISCALE MODEL. SIMUL.
Vol. 4, No. 3, pp. 960–991
© 2005 Society for Industrial and Applied Mathematics
ANALYSIS OF THE RECOVERY OF EDGES IN IMAGES AND SIGNALS BY MINIMIZING NONCONVEX REGULARIZED LEAST-SQUARES*
MILA NIKOLOVA†
Bounds on the Minimizers of (nonconvex) Regularized Least-Squares
Mila Nikolova
CMLA, ENS Cachan, CNRS, PRES UniverSud
61 Av. President Wilson, F-94230 Cachan, France
nikolova@cmla.ens-cachan.fr
SIAM J. SCI. COMPUT.
Vol. 27, No. 3, pp. 937–966
ANALYSIS OF HALF-QUADRATIC MINIMIZATION METHODS FOR SIGNAL AND IMAGE RECOVERY*
MILA NIKOLOVA† AND MICHAEL K. NG‡
C. R. Acad. Sci. Paris, Ser. I
Mathematical Analysis
Solve exactly an under determined linear system by minimizing least squares regularized with an ℓ_0 penalty
Résoudre exactement un système sous-déterminé en minimisant des moindres carrés régularisés avec ℓ_0
Mila Nikolova
Applied and Computational Harmonic Analysis
Relationship between the optimal solutions of least squares regularized with ℓ_0-norm and constrained by k-sparsity
Mila Nikolova
CMLA, CNRS, ENS Cachan, Université Paris-Saclay, 94230 Cachan, France
SIAM J. IMAGING SCIENCES
Vol. 6, No. 2, pp. 904–937
© 2013 Society for Industrial and Applied Mathematics
Description of the Minimizers of Least Squares Regularized with ℓ_0-norm. Uniqueness of the Global Minimizer*
Mila Nikolova†
Abstract. We have an M × N real-valued arbitrary matrix A (e.g., a dictionary) with M < N and data d describing the sought-after object with the help of A. This work provides an in-depth analysis of

Outline of the talk

1. Introduction ℓ_2-ℓ_0 Minimization
2. Necessary optimality conditions
3. Relationship between optimality conditions
4. Quantifying “optimal” points
5. Algorithms and necessary optimality conditions
6. Concluding remarks

Introduction ℓ_2-ℓ_0 Minimization

Formulation

The ℓ_2-ℓ_0 minimization problem

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

- $\mathbf{A} \in \mathbb{R}^{M\times N}$ with $M \ll N$,
- Sparsity is modeled with the ℓ_0 **pseudo-norm**:

$$\|\mathbf{x}\|_0 = \sharp\left\{x_i,\ i \in [1,\dots,N] : x_i \neq 0\right\}$$

- **Non-convex** and **NP-Hard** problem [Natarajan, 1995, Nguyen et al., 2019]

Formulation

The ℓ_2-ℓ_0 minimization problem

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

- $\mathbf{A} \in \mathbb{R}^{M\times N}$ with $M \ll N$,
- Sparsity is modeled with the ℓ_0 **pseudo-norm**:

$$\|\mathbf{x}\|_0 = \sharp\{x_i,\ i \in [1,\ldots,N] : x_i \neq 0\}$$

- **Non-convex** and **NP-Hard** problem [Natarajan, 1995, Nguyen et al., 2019]

Applications

- Inverse problems,
- Statistical regression,
- Machine learning,
- Compressed sensing ...

A Brief Literature Review

Convex relaxations

Greedy algorithms

Iterative thresholding algorithms

Global optimization

A Brief Literature Review

Convex relaxations

- Basic Pursuit De-Noising [Chen et al., 2001], LASSO, [Tibshirani, 1996],

$$\hat{\mathbf{x}} \in \left\{ \arg\min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{A}\mathbf{x} - d\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}$$

  Under some **conditions** (RIP [Candes et al., 2006, Candès and Wakin, 2008], incoherence [Donoho, 2006, Gribonval and Nielsen, 2003] ...)

  $\rightarrow$ **exact recovery** by ℓ_1-minimization,
- The convex non-convex strategy [Selesnick and Farshchian, 2017, Selesnick, 2017]

Greedy algorithms

Iterative thresholding algorithms

Global optimization

A Brief Literature Review

Convex relaxations

Greedy algorithms

Idea: add **one by one** non-zero components to the solution:

- Matching Pursuit (MP) [Mallat and Zhang, 1993], Orthogonal Matching Pursuit (OMP) [Pati et al., 1993], Orthogonal Least Squares (OLS) [Chen et al., 1991] ...

  $\rightarrow$ under some **conditions, optimality** guarantees for OMP [Tropp, 2004] and OLS [Soussen et al., 2013],
- Forward-backward extensions: *Single Best Replacement* (SBR) [Soussen et al., 2011] ...

Iterative thresholding algorithms

Global optimization

A Brief Literature Review

Convex relaxations

Greedy algorithms

Iterative thresholding algorithms

- Iterative hard thresholding (IHT) [Blumensath and Davies, 2009],
- Subspace pursuit [Dai and Milenkovic, 2009],
- Hard thresholding pursuit [Foucart, 2011],
- Compressive Sampling Matching Pursuit (CoSaMP) [Needell and Tropp, 2009],

Global optimization

A Brief Literature Review

Convex relaxations

Greedy algorithms

Iterative thresholding algorithms

Global optimization

Mixed integer programming together with branch and bounds algorithms [Bourguignon et al., 2016] $\rightarrow$ limited to **moderate size** problems,

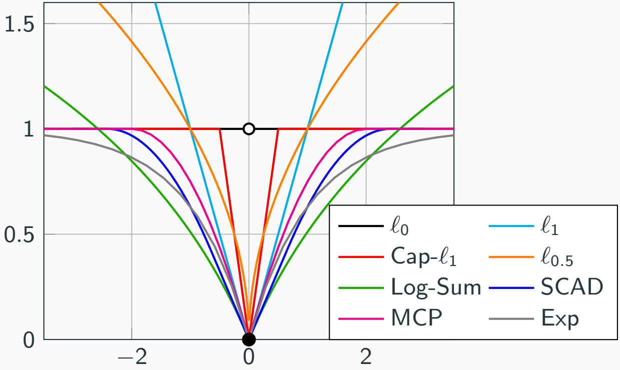
Continuous non-convex relaxations of the ℓ_0-norm

$$\hat{\mathbf{x}} \in \left\{ \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \Phi(\mathbf{x}) \right\}.$$

- Adaptive Lasso [Zou, 2006],
- Nonnegative Garrote [Breiman, 1995],
- Exponential approximation [Mangasarian, 1996],
- Log-Sum Penalty [Candès et al., 2008],
- Smoothly Clipped Absolute Deviation (SCAD) [Fan and Li, 2001],
- Minimax Concave Penalty (MCP) [Zhang, 2010],
- ℓ_p-norms $0 < p < 1$ [Chartrand, 2007, Foucart and Lai, 2009]
- Smoothed ℓ_0-norm Penalty (SL0) [Mohimani et al., 2009],
- Class of smooth non-convex penalties [Chouzenoux et al., 2013],
- Smoothed norm ratio [Repetti et al., 2015, Cherni et al., 2019].

Continuous non-convex relaxations of the ℓ_0-norm

$$\hat{\mathbf{x}} \in \left\{ \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2 + \Phi(\mathbf{x}) \right\}.$$



A Brief Literature Review

Continuous non-convex relaxations of the ℓ_0-norm

$$\hat{\mathbf{x}} \in \left\{ \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \Phi(\mathbf{x}) \right\}.$$

There exist a class of penalties Φ that lead to **exact continuous relaxation** of the ℓ_2-ℓ_0 functional in the sense that their **global minimizers coincide** [Soubies et al., 2017, Carlsson, 2019]

Motivation of this work

NP-hardness implies that

- one cannot expect, in general, to attain an optimal point
- verifying the optimality of a point $\hat{\mathbf{x}}$ is also, in general, intractable

Interest in studying the "**restrictiveness**" of **tractable** necessary **(but not sufficient)** optimality conditions

Motivation of this work

NP-hardness implies that

- one cannot expect, in general, to attain an optimal point
- verifying the optimality of a point $\hat{\mathbf{x}}$ is also, in general, intractable

Interest in studying the "**restrictiveness**" of **tractable** necessary **(but not sufficient)** optimality conditions

Some notations

- $\mathbb{I}_N = \{1, \ldots, N\}$,
- $\sigma_{\mathbf{x}} = \{i \in \mathbb{I}_N : x_i \neq 0\}$ denotes the support of $\mathbf{x} \in \mathbb{R}^N$,
- $\mathbf{x}_\omega \in \mathbb{R}^{\#\omega}$ is the restriction of $\mathbf{x} \in \mathbb{R}^N$ to the elements indexed by ω
- $\mathbf{A}_\omega \in \mathbb{R}^{M \times \#\omega}$ is the restriction of $\mathbf{A} \in \mathbb{R}^{M \times N}$ to the columns indexed by ω
- $\mathbf{a}_i = \mathbf{A}_{\{i\}} \in \mathbb{R}^M$

Necessary optimality conditions

Local optimality

Definition (Local optimality)

A point $\mathbf{x} \in \mathbb{R}^N$ is a **local minimizer** of F_0 if and only if

$$\mathbf{x} \in \left\{ \arg\min_{\mathbf{u} \in \mathbb{R}^N} \|\mathbf{A}\mathbf{u} - \mathbf{y}\|^2 \ \text{ s.t. } \ \sigma_{\mathbf{u}} \subseteq \sigma_{\mathbf{x}} \right\},$$

or, equivalently, if $\mathbf{x}$ is such that

$$\langle \mathbf{a}_i, \mathbf{A}\mathbf{x} - \mathbf{y} \rangle = 0 \quad \forall i \in \sigma_{\mathbf{x}}$$

Local optimality

Definition (Local optimality)

A point $\mathbf{x} \in \mathbb{R}^N$ is a **local minimizer** of F_0 if and only if

$$\mathbf{x} \in \left\{ \arg\min_{\mathbf{u} \in \mathbb{R}^N} \|\mathbf{A}\mathbf{u} - \mathbf{y}\|^2 \ \text{ s.t. } \ \sigma_{\mathbf{u}} \subseteq \sigma_{\mathbf{x}} \right\},$$

or, equivalently, if $\mathbf{x}$ is such that

$$\langle \mathbf{a}_i, \mathbf{A}\mathbf{x} - \mathbf{y} \rangle = 0 \quad \forall i \in \sigma_{\mathbf{x}}$$

- Local minimizers of F_0 are **independent** of λ,
- When $\mathrm{rank}(\mathbf{A}) < N$ (*e.g.*, $M < N$), local minimizers of F_0 are **uncountable**,
- An important subset contains **strict** local minimizers,

$$\exists \varepsilon > 0, \ \forall \mathbf{u} \in \mathcal{B}_2(\mathbf{x}, \varepsilon), \ F_0(\mathbf{x}) < F_0(\mathbf{u}).$$

- Indeed, **global** minimizers of F_0 are **strict** [Nikolova, 2013].

Local optimality

Theorem (Strict local optimality for F_0 [Nikolova, 2013])
A local minimizer $\mathbf{x} \in \mathbb{R}^N$ *of* F_0 *is strict if and only if* $\mathrm{rank}(\mathbf{A}_{\sigma_{\mathbf{x}}}) = \#\sigma_{\mathbf{x}}$.

Local optimality

Theorem (Strict local optimality for F_0 [Nikolova, 2013])
A local minimizer $\mathbf{x} \in \mathbb{R}^N$ of F_0 is strict if and only if $\operatorname{rank}(\mathbf{A}_{\sigma_\mathbf{x}}) = \#\sigma_\mathbf{x}$.

- A **strict (local) minimizer of F_0** can be easily computed:
  1. choose a support $\omega \in \Omega_{\max}$ where
  $$\Omega_{\max} = \bigcup_{r=0}^{M} \Omega_r \quad \text{and} \quad \Omega_r := \{\omega \in \mathbb{I}_N \ : \ \#\omega = r = \operatorname{rank}(\mathbf{A}_\omega)\} \ (\Omega_0 = \emptyset)$$
  2. solve $(\mathbf{A}_\omega)^T \mathbf{A}_\omega \mathbf{x}_\omega = (\mathbf{A}_\omega)^T \mathbf{y}$

$\Rightarrow$ Given $\mathbf{A}$ and $\mathbf{y}$ we can compute **all** the strict (local) minimizers of F_0 by solving the **restricted normal equations** $\forall \omega \in \Omega_{max}$

- $\#\Omega_{\max}$ is **finite** (but huge)

Definition (Partial support coordinate-wise points [Beck and Hallak, 2018])

A local minimizer $\mathbf{x} \in \mathbb{R}^N$ of F_0 is said to be **partial support coordinate-wise (CW)** optimal for F_0 if it verifies

$$F_0(\mathbf{x}) \leq \min\{F_0(\mathbf{u}) : \mathbf{u} \in \{\mathbf{u}_\mathbf{x}^-, \mathbf{u}_\mathbf{x}^{\mathrm{swap}}, \mathbf{u}_\mathbf{x}^+\}\}$$

where $\mathbf{u}_\mathbf{x}^-$, $\mathbf{u}_\mathbf{x}^{\mathrm{swap}}$, and $\mathbf{u}_\mathbf{x}^+$ are local minimizers of F_0 with supports

- $\sigma_{\mathbf{u}_\mathbf{x}^-} = \sigma_\mathbf{x} \backslash \{i_\mathbf{x}\}$
- $\sigma_{\mathbf{u}_\mathbf{x}^{\mathrm{swap}}} = \sigma_\mathbf{x} \backslash \{i_\mathbf{x}\} \cup \{j_\mathbf{x}\}$
- $\sigma_{\mathbf{u}_\mathbf{x}^+} = \sigma_\mathbf{x} \cup \{j_\mathbf{x}\}$

for

$$i_\mathbf{x} \in \left\{ \arg\min_{k \in \sigma_\mathbf{x}} |x_k| \right\}, \quad j_\mathbf{x} \in \left\{ \arg\max_{k \in (\sigma_\mathbf{x})^c} |\langle \mathbf{a}_k, \mathbf{A}\mathbf{x} - \mathbf{y} \rangle| \right\}.$$

L-Stationarity

Definition (L-stationarity [Tropp, 2006, Beck and Hallak, 2018])

A point $\mathbf{x} \in \mathbb{R}^N$ is said to be L-stationary for F_0 ($L > 0$), if

$$\mathbf{x} \in \left\{ \arg\min_{\mathbf{u} \in \mathbb{R}^N} \frac{1}{2} \| T_L(\mathbf{x}) - \mathbf{u} \|^2 + \frac{\lambda}{L} \|\mathbf{u}\|_0 \right\},$$

where $T_L(\mathbf{x}) = \mathbf{x} - L^{-1} \mathbf{A}^T (\mathbf{A}\mathbf{x} - \mathbf{y})$.

L-Stationarity

Definition (L-stationarity [Tropp, 2006, Beck and Hallak, 2018])
A point $\mathbf{x} \in \mathbb{R}^N$ is said to be L-stationary for F_0 ($L > 0$), if

$$\mathbf{x} \in \left\{ \arg\min_{\mathbf{u} \in \mathbb{R}^N} \frac{1}{2} \| T_L(\mathbf{x}) - \mathbf{u} \|^2 + \frac{\lambda}{L} \|\mathbf{u}\|_0 \right\},$$

where $T_L(\mathbf{x}) = \mathbf{x} - L^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{y})$.

- For $L \geq \|\mathbf{A}\|^2$, L-stationarity points are fixed points of the IHT algorithm [Blumensath and Davies, 2009, Attouch et al., 2013].

Exact continuous relaxations: Motivation

Are there **continuous** relaxations of F_0 of the form

$$\tilde{F}(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \sum_{i=1}^{N} \phi_i(x_i),$$

such that, for all $\mathbf{y} \in \mathbb{R}^M$,

$$\arg\min_{\mathbf{x}\in\mathbb{R}^N} \tilde{F}(\mathbf{x}) = \arg\min_{\mathbf{x}\in\mathbb{R}^N} F_0(\mathbf{x}), \quad \text{(P1)}$$

$$\mathbf{x} \text{ local minimizer of } \tilde{F} \implies \mathbf{x} \text{ local minimizer of } F_0 \quad \text{(P2)}$$

Exact continuous relaxations: Motivation

Are there **continuous** relaxations of F_0 of the form

$$\tilde{F}(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \sum_{i=1}^{N} \phi_i(x_i),$$

such that, for all $\mathbf{y} \in \mathbb{R}^M$,

$$\arg\min_{\mathbf{x}\in\mathbb{R}^N} \tilde{F}(\mathbf{x}) = \arg\min_{\mathbf{x}\in\mathbb{R}^N} F_0(\mathbf{x}), \qquad \text{(P1)}$$

$$\mathbf{x} \text{ local minimizer of } \tilde{F} \implies \mathbf{x} \text{ local minimizer of } F_0 \qquad \text{(P2)}$$

- Properties (P1) and (P2) imply that **local optimality for $\tilde{F}$** is a **necessary optimality condition for F_0**.
- Moreover, there is no converse property for (P2)
  $\longrightarrow$ $\tilde{F}$ can potentially **remove** local (not global) minimizers of F_0

Conditions based on exact relaxations

Theorem ([Soubies et al., 2017])

Properties (P1) and (P2) are satisfied $\forall \mathbf{y} \in \mathbb{R}^M$ *if and only if* Φ *verifies the following conditions:* $\forall i \in \{1, \ldots, N\}$,

$\phi_i(0) = 0$,

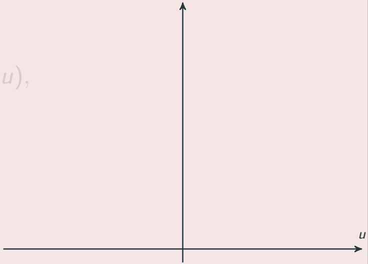
$\forall u \in \mathbb{R} \setminus (\beta^{i-}, \beta^{i+}),\ \phi_i(u) = \lambda$,

$\forall u \in (\beta^{i-}, \beta^{i+}) \setminus \{0\},\ \phi_i(u) > \phi_{celo}(\|\mathbf{a}_i\|, \lambda; u)$,

$\forall u \in B^i \setminus \{0\},\ \lim_{\substack{v \to u \\ v < u}} \phi_i'(v) > \lim_{\substack{v \to u \\ v > u}} \phi_i'(v)$,

$\forall u \in]\beta^{i-}, \beta^{i+}[\setminus B^i$,

$$\begin{cases} \phi_i''(u) \leq -\|\mathbf{a}_i\|^2 \text{ and} \\ \forall \varepsilon > 0,\ \exists v_\varepsilon \in (u - \varepsilon, u + \varepsilon) \\ \qquad \text{s.t.}\ \ \phi_i''(v_\varepsilon) < -\|\mathbf{a}_i\|^2 \end{cases}$$



Moreover, **global minimizers** *of* $\tilde{F}$ *are* **strict**.

Theorem ([Soubies et al., 2017])

Properties (P1) and (P2) are satisfied $\forall \mathbf{y} \in \mathbb{R}^M$ *if and only if* Φ *verifies the following conditions:* $\forall i \in \{1, \ldots, N\}$,

$\phi_i(0) = 0$,

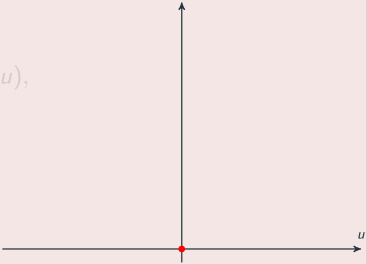
$\forall u \in \mathbb{R} \setminus (\beta^{i-}, \beta^{i+}),\ \phi_i(u) = \lambda$,

$\forall u \in (\beta^{i-}, \beta^{i+}) \setminus \{0\},\ \phi_i(u) > \phi_{celo}(\|\mathbf{a}_i\|, \lambda; u)$,

$\forall u \in B^i \setminus \{0\},\ \lim_{\substack{v \to u \\ v < u}} \phi_i'(v) > \lim_{\substack{v \to u \\ v > u}} \phi_i'(v)$,

$\forall u \in]\beta^{i-}, \beta^{i+}[\setminus B^i$,

$$\begin{cases} \phi_i''(u) \leq -\|\mathbf{a}_i\|^2 \text{ and} \\ \forall \varepsilon > 0,\ \exists v_\varepsilon \in (u - \varepsilon, u + \varepsilon) \\ \qquad \text{s.t.}\ \ \phi_i''(v_\varepsilon) < -\|\mathbf{a}_i\|^2 \end{cases}$$



Moreover, **global minimizers** *of* $\tilde{F}$ *are* **strict**.

Conditions based on exact relaxations

Theorem ([Soubies et al., 2017])

Properties (P1) and (P2) are satisfied $\forall \mathbf{y} \in \mathbb{R}^M$ *if and only if* Φ *verifies the following conditions:* $\forall i \in \{1, \dots, N\}$,

$\phi_i(0) = 0$,

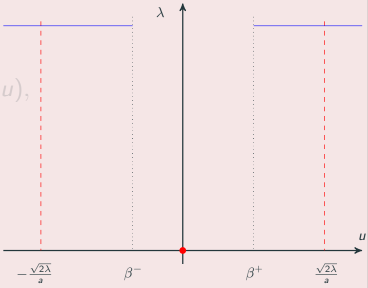
$\forall u \in \mathbb{R} \setminus (\beta^{i-}, \beta^{i+}),\ \phi_i(u) = \lambda$,

$\forall u \in (\beta^{i-}, \beta^{i+}) \setminus \{0\},\ \phi_i(u) > \phi_{celo}(\|\mathbf{a}_i\|, \lambda; u)$,

$\forall u \in B^i \setminus \{0\},\ \lim_{\substack{v \to u \\ v<u}} \phi_i'(v) > \lim_{\substack{v \to u \\ v>u}} \phi_i'(v)$,

$\forall u \in]\beta^{i-}, \beta^{i+}[\setminus B^i$,

$$\begin{cases} \phi_i''(u) \leq -\|\mathbf{a}_i\|^2 \text{ and} \\ \forall \varepsilon > 0,\ \exists v_\varepsilon \in (u - \varepsilon, u + \varepsilon) \\ \qquad \text{s.t.} \quad \phi_i''(v_\varepsilon) < -\|\mathbf{a}_i\|^2 \end{cases}$$



Moreover, **global minimizers** *of* $\tilde{F}$ *are* **strict**.

Conditions based on exact relaxations

Theorem ([Soubies et al., 2017])
Properties (P1) and (P2) are satisfied $\forall \mathbf{y} \in \mathbb{R}^M$ *if and only if* Φ *verifies the following conditions:* $\forall i \in \{1, \dots, N\}$,

$\phi_i(0) = 0,$

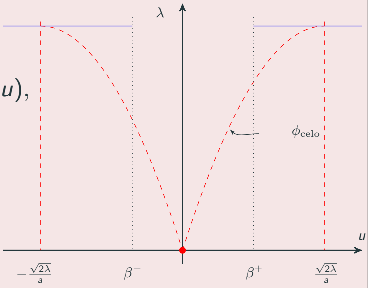
$\forall u \in \mathbb{R} \setminus (\beta^{i-}, \beta^{i+}),\ \phi_i(u) = \lambda,$

$\forall u \in (\beta^{i-}, \beta^{i+}) \setminus \{0\},\ \phi_i(u) > \phi_{celo}(\|\mathbf{a}_i\|, \lambda; u),$

$\forall u \in B^i \setminus \{0\},\ \lim_{\substack{v \to u \\ v < u}} \phi_i'(v) > \lim_{\substack{v \to u \\ v > u}} \phi_i'(v),$

$\forall u \in]\beta^{i-}, \beta^{i+}[\setminus B^i,$

$$\begin{cases} \phi_i''(u) \leq -\|\mathbf{a}_i\|^2 \text{ and} \\ \forall \varepsilon > 0,\ \exists v_\varepsilon \in (u - \varepsilon, u + \varepsilon) \\ \qquad \text{s.t. } \phi_i''(v_\varepsilon) < -\|\mathbf{a}_i\|^2 \end{cases}$$



Moreover, **global minimizers** *of* $\tilde{F}$ *are* **strict**.

Conditions based on exact relaxations

Theorem ([Soubies et al., 2017])

Properties (P1) and (P2) are satisfied $\forall \mathbf{y} \in \mathbb{R}^M$ *if and only if* Φ *verifies the following conditions:* $\forall i \in \{1, \dots, N\}$,

$\phi_i(0) = 0$,

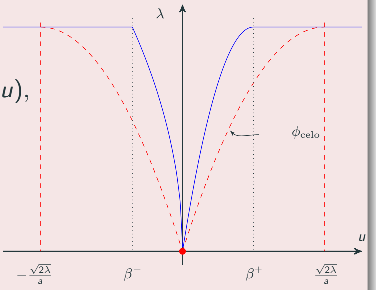
$\forall u \in \mathbb{R} \setminus (\beta^{i-}, \beta^{i+}),\ \phi_i(u) = \lambda$,

$\forall u \in (\beta^{i-}, \beta^{i+}) \setminus \{0\},\ \phi_i(u) > \phi_{celo}(\|\mathbf{a}_i\|, \lambda; u)$,

$\forall u \in B^i \setminus \{0\},\ \lim_{\substack{v \to u \\ v<u}} \phi_i'(v) > \lim_{\substack{v \to u \\ v>u}} \phi_i'(v)$,

$\forall u \in]\beta^{i-}, \beta^{i+}[\setminus B^i$,

$$\begin{cases} \phi_i''(u) \leq -\|\mathbf{a}_i\|^2 \text{ and} \\ \forall \varepsilon > 0,\ \exists v_\varepsilon \in (u-\varepsilon, u+\varepsilon) \\ \qquad \text{s.t. } \phi_i''(v_\varepsilon) < -\|\mathbf{a}_i\|^2 \end{cases}$$



Moreover, **global minimizers** *of* $\tilde{F}$ *are* **strict**.

Conditions based on exact relaxations

The continuous exact ℓ_0 penalty (CEL0) [Soubies et al., 2015]

$$\Phi_{\mathrm{celo}}(\mathbf{x}) = \sum_{i=1}^{N} \phi_{\mathrm{celo}}(\|a_i\|, \lambda; x_i),$$

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

Conditions based on exact relaxations

The continuous exact ℓ_0 penalty (CEL0) [Soubies et al., 2015]

$$\Phi_{\mathrm{celo}}(\mathbf{x}) = \sum_{i=1}^{N} \phi_{\mathrm{celo}}(\|a_i\|, \lambda; x_i),$$

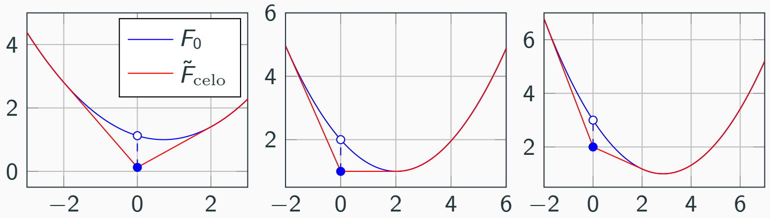
$$\phi_{\mathrm{celo}}(a, \lambda; u) = \lambda - \frac{a^2}{2}\left(|u| - \frac{\sqrt{2\lambda}}{a}\right)^2 1_{\{|u| \leq \frac{\sqrt{2\lambda}}{a}\}}$$

Properties of the CEL0 relaxation $\tilde{F}_{\mathrm{celo}}$

- **Inferior limit** of the derived class of penalties,
- **Convex hull** of F_0 when $\mathbf{A}$ has nonzero **orthogonal** columns,
- **Convex w.r.t. each variable** x_i for all $\mathbf{A} \in \mathbb{R}^{N \times M}$,
- Potentially **eliminates** the **larger amount** of local minimizers of F_0,

Properties of the CEL0 relaxation $\tilde{F}_{\text{celo}}$

- **Inferior limit** of the derived class of penalties,
- **Convex hull** of F_0 when $\mathbf{A}$ has nonzero **orthogonal** columns,
- **Convex w.r.t. each variable** x_i for all $\mathbf{A} \in \mathbb{R}^{N \times M}$,
- Potentially **eliminates** the **larger amount** of local minimizers of F_0,



Conditions based on exact relaxations

Theorem (Link between global minimizers of F_0 and $\tilde{F}_{\mathrm{celo}}$)

(i) *The set of* **global minimizers** *of* F_0 *is* **included** *in the one of* $\tilde{F}_{\mathrm{celo}}$*,*

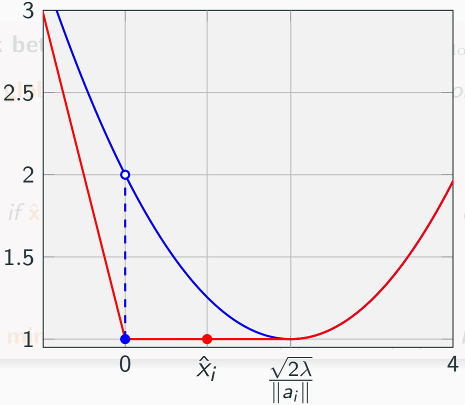
$$\arg\min_{\mathbf{x}\in\mathbb{R}^N} F_0(\mathbf{x}) \subseteq \arg\min_{\mathbf{x}\in\mathbb{R}^N} \tilde{F}_{\mathrm{celo}}(\mathbf{x})$$

(ii) *Conversely if* $\hat{\mathbf{x}} \in \mathbb{R}^N$ *is a* **global minimizer** *of* $\tilde{F}_{\mathrm{celo}}$*,* $\hat{\mathbf{x}}^0$ *defined by*

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

is a **global minimizer** *of* F_0 *and* $\tilde{F}_{\mathrm{celo}}(\hat{\mathbf{x}}) = \tilde{F}_{\mathrm{celo}}(\hat{\mathbf{x}}^0) = F_0(\hat{\mathbf{x}}^0)$*.*

Conditions based on exact relaxations



Conditions based on exact relaxations

Theorem (Link between global minimizers of F_0 and $\tilde{F}_{\text{celo}}$)

(i) *The set of* **global minimizers** *of F_0 is* **included** *in the one of $\tilde{F}_{\text{celo}}$,*

$$\arg\min_{\mathbf{x}\in\mathbb{R}^N} F_0(\mathbf{x}) \subseteq \arg\min_{\mathbf{x}\in\mathbb{R}^N} \tilde{F}_{\text{celo}}(\mathbf{x})$$

(ii) *Conversely if $\hat{\mathbf{x}} \in \mathbb{R}^N$ is a* **global minimizer** *of $\tilde{F}_{\text{celo}}$, $\hat{\mathbf{x}}^0$ defined by*

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

is a **global minimizer** *of F_0 and $\tilde{F}_{\text{celo}}(\hat{\mathbf{x}}) = \tilde{F}_{\text{celo}}(\hat{\mathbf{x}}^0) = F_0(\hat{\mathbf{x}}^0)$.*

Theorem (Link between local minimizers of F_0 and $\tilde{F}_{\text{celo}}$)

$$\hat{\mathbf{x}} \in \mathbb{R}^N \text{ local minimizer of } \tilde{F}_{\text{celo}} \implies \hat{\mathbf{x}}^0 \text{ local minimizer of } F_0$$

and

$$\tilde{F}_{\text{celo}}(\hat{\mathbf{x}}) = \tilde{F}_{\text{celo}}(\hat{\mathbf{x}}^0) = F_0(\hat{\mathbf{x}}^0)$$

Conditions based on exact relaxations

Assumption 1

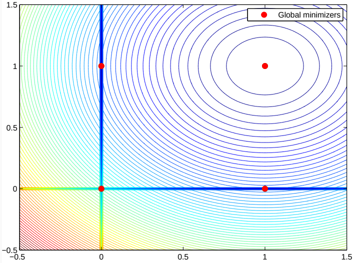
When F_0 does not admit a unique global minimizer, every pair $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)$ of global minimizers ($\hat{\mathbf{x}}_1 \neq \hat{\mathbf{x}}_2$) verify $\|\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2\|_0 > 1$.

Conditions based on exact relaxations

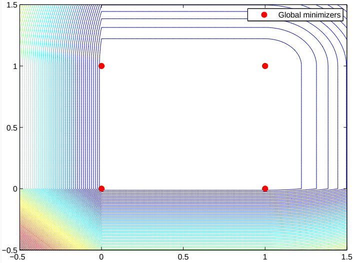
Assumption 1
When F_0 does not admit a unique global minimizer, every pair $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)$ of global minimizers $(\hat{\mathbf{x}}_1 \neq \hat{\mathbf{x}}_2)$ verify $\|\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2\|_0 > 1$.

Corollary ([Soubies et al., 2019])
Under Assumption 1, global minimizers of F_0 and $\tilde{F}_{\mathrm{celo}}$ coincide. Moreover, they are stricts for both F_0 and $\tilde{F}_{\mathrm{celo}}$.

Conditions based on exact relaxations



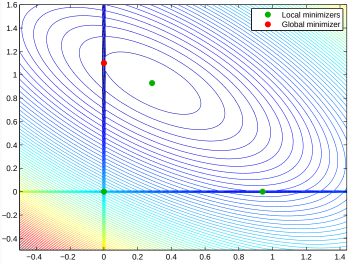
F_0



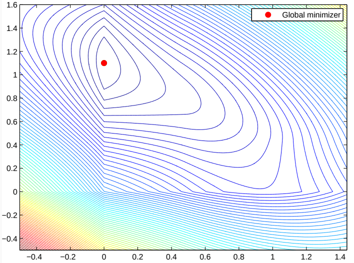
$\tilde{F}_{\mathrm{celo}}$

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

Conditions based on exact relaxations



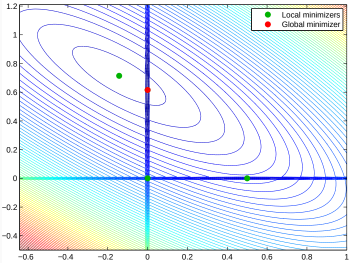
F_0

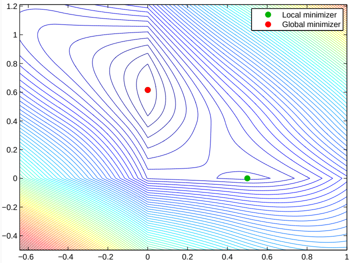


$\tilde{F}_{\mathrm{celo}}$

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

Conditions based on exact relaxations





F_0 $\tilde{F}_{\text{celo}}$

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

Relationship between optimality conditions

Relationship between optimality conditions

We introduced four necessary (not sufficient) optimality conditions for F_0

- Strict local optimality for F_0
- L-stationarity
- Partial support coordinate-wise optimality
- Strict local optimality for $\tilde{F}$

Are there any **inclusion properties** between the sets of points associated to these conditions ?

Relationship between optimality conditions

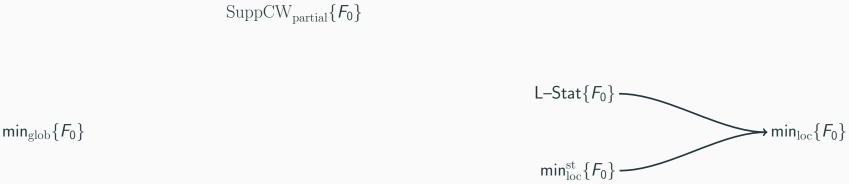
$\min_{\text{glob}}\{F_0\}$ $\min_{\text{loc}}\{F_0\}$

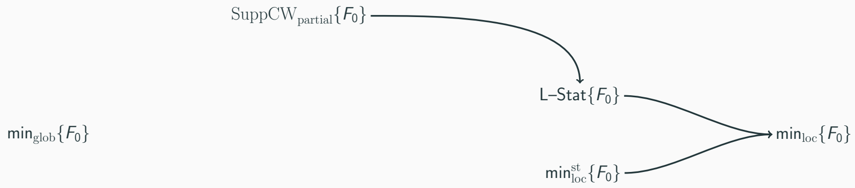
Relationship between optimality conditions

Relationship between optimality conditions

Theorem (L-stationary $\Rightarrow$ $\min_{\mathrm{loc}}\{F_0\}$ [Beck and Hallak, 2018])
Let $\mathbf{x} \in \mathbb{R}^N$ *be a L-stationary point of* F_0 *for some* $L > 0$*. Then* $\mathbf{x}$ *is a local minimizer of* F_0*.*

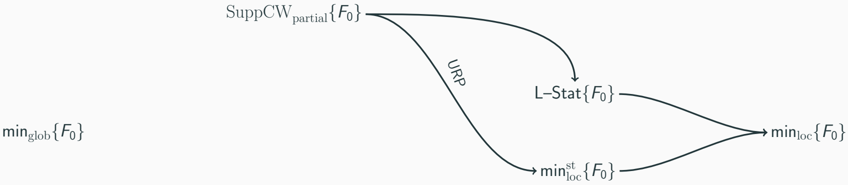
Relationship between optimality conditions





Theorem ($\mathrm{SuppCW}_{\mathrm{partial}}\{F_0\} \Rightarrow$ L-stationary [Beck and Hallak, 2018])
If $\mathbf{x} \in \mathbb{R}^N$ is a partial support CW point of F_0, then it is a L-stationary point of F_0 for any $L \geq \|A\|^2$.

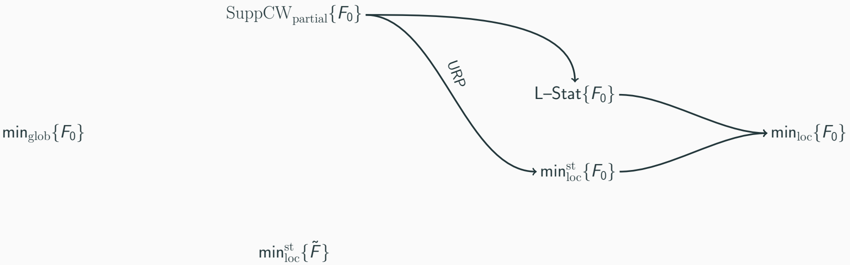
Relationship between optimality conditions



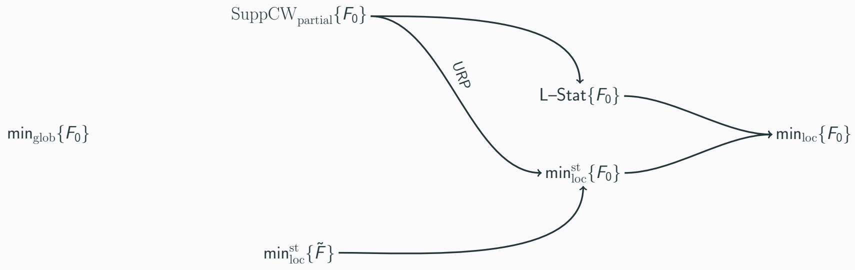
Theorem ($\mathrm{SuppCW}_{\mathrm{partial}}\{F_0\} \Rightarrow \min^{\mathrm{st}}_{\mathrm{loc}}\{F_0\}$ [Soubies et al., 2019])
Let $\mathbf{A}$ *satisfy the unique representation property (URP)*[a]*. Let* $\mathbf{x} \in \mathbb{R}^N$ *be a partial support CW point of* F_0*. Then it is a strict local minimizer of* F_0*.*

[a] A matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ satisfies the URP [Gorodnitsky and Rao, 1997] if any $\min\{M, N\}$ columns of $\mathbf{A}$ are linearly independent.

Relationship between optimality conditions

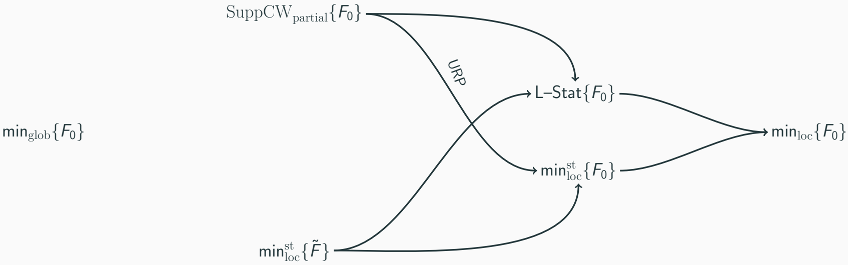


Relationship between optimality conditions



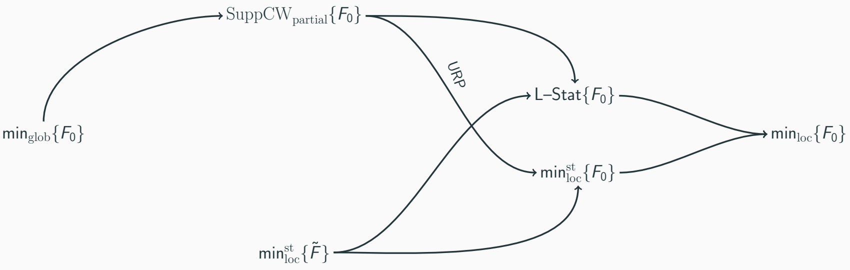
Theorem ($\min_{\mathrm{loc}}^{\mathrm{st}}\{\tilde{F}\} \Rightarrow \min_{\mathrm{loc}}^{\mathrm{st}}\{F_0\}$ [Soubies et al., 2015])
Let $\mathbf{x}$ be a strict local minimizer of $\tilde{F}$, then $\mathbf{x}$ is a strict local minimizer of F_0.

Relationship between optimality conditions



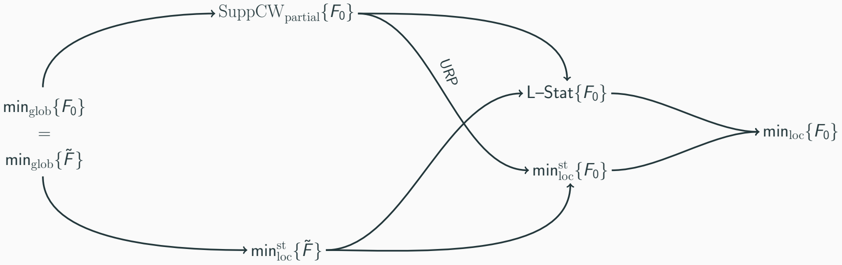
Theorem ($\min^{\mathrm{st}}_{\mathrm{loc}}\{\tilde{F}\} \Rightarrow$ L-stationary [Soubies et al., 2019])
Let $\mathbf{x} \in \mathbb{R}^N$ be a strict local minimizer of $\tilde{F}$. Then it is a L-stationary point of F_0 for any $L \geq \max_{i \in \mathbb{I}_N} \|\mathbf{a}_i\|^2$.

Relationship between optimality conditions



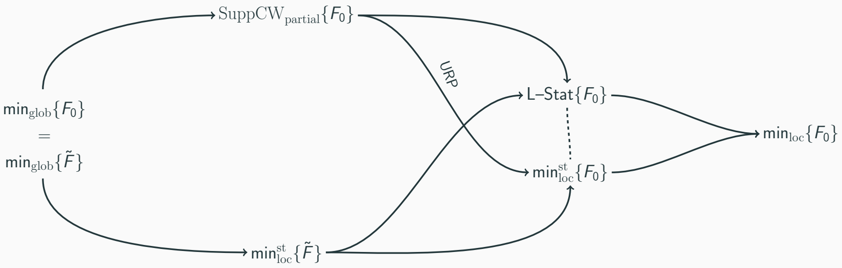
Theorem ($\min_{\mathrm{glob}}\{F_0\} \Rightarrow \mathrm{SuppCW}_{\mathrm{partial}}\{F_0\}$ [Beck and Hallak, 2018])
Let $\mathbf{x} \in \mathbb{R}^N$ be a global minimizer of F_0. Then it is partial support CW point of F_0.

Relationship between optimality conditions



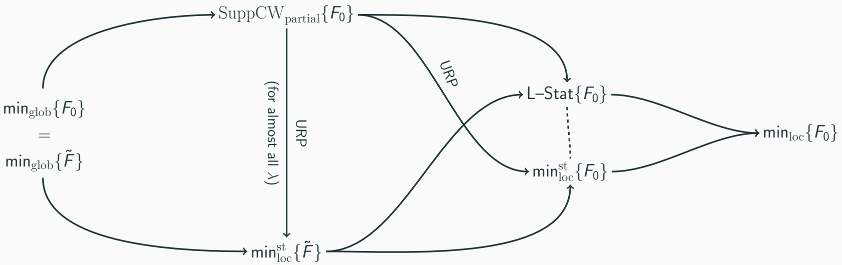
Theorem ($\min_{\mathrm{glob}}\{F_0\} = \min_{\mathrm{glob}}\{\tilde{F}\}$ [Soubies et al., 2019])
Global minimizers of F_0 and $\tilde{F}$ coincide. Moreover, they are stricts for both F_0 and $\tilde{F}$.

Relationship between optimality conditions



No inclusion property between L-Stat and $\min^{\mathrm{st}}_{\mathrm{loc}}\{F_0\}$ [Soubies et al., 2019]

Relationship between optimality conditions



Theorem ($\mathrm{SuppCW}_{\mathrm{partial}}\{F_0\} \Rightarrow \min^{\mathrm{st}}_{\mathrm{loc}}\{\tilde{F}\}$ [Soubies et al., 2019])
Let $\mathbf{A}$ satisfy the URP and have unit norm columns. Then, for all $\lambda \in \mathbb{R}_{>0} \setminus \Lambda$ (where Λ is a subset of $\mathbb{R}_{>0}$ whose Lebesgue measure is zero), each partial support CW point of F_0 is a strict local minimizer of $\tilde{F}$.

Quantifying “optimal” points

Quantifying "optimal" points

Objective

Let $\mathcal{S}_0$ be the set of strict local minimizers of F_0 and define the three following subsets

- $\mathcal{S}_{\mathrm{CW}} = \{\mathbf{x} \in \mathcal{S}_0 : \mathbf{x} \text{ partial support CW point}\}$
- $\tilde{\mathcal{S}} = \{\mathbf{x} \in \mathcal{S}_0 : \mathbf{x} \text{ strict local minimizer of } \tilde{F}\}$
- $\mathcal{S}_{\mathrm{L}} = \{\mathbf{x} \in \mathcal{S}_0 : \mathbf{x} \text{ L-stationary point}\}$

Then, our goal is to **quantify** the **cardinality** of these sets, *i.e.*, $\#\tilde{\mathcal{S}}$, $\#\mathcal{S}_{\mathrm{CW}}$, and $\#\mathcal{S}_{\mathrm{L}}$.

Quantifying "optimal" points

Experiment

Given $\mathbf{A} \in \mathbb{R}^{5\times 10}$ and $\mathbf{y} \in \mathbb{R}^5$, we proceed as follows:

1. Compute all strict local minimizers of $F_0 \to \mathcal{S}_0$ (independent of λ),
2. For each $\lambda \in \{\lambda_1, \dots, \lambda_P\}$, determine the subset of $\mathcal{S}_0$ that contains points verifying a given necessary optimality condition (*i.e.*, $\tilde{\mathcal{S}}$, $\mathcal{S}_{\mathrm{CW}}$, $\mathcal{S}_{\mathrm{L}}$),
3. Repeat 1-2 for different $\mathbf{A} \in \mathbb{R}^{5\times 10}$ and $\mathbf{y} \in \mathbb{R}^5$, and draw the average evolution of $\#\tilde{\mathcal{S}}$, $\#\mathcal{S}_{\mathrm{CW}}$, and $\#\mathcal{S}_{\mathrm{L}}$, with respect to λ.

Quantifying "optimal" points

Experiment

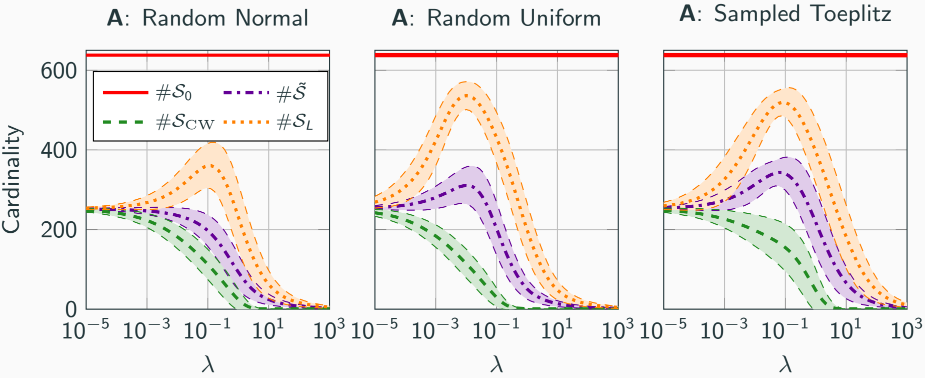
Given $\mathbf{A} \in \mathbb{R}^{5\times 10}$ and $\mathbf{y} \in \mathbb{R}^5$, we proceed as follows:

1. Compute all strict local minimizers of $F_0 \to \mathcal{S}_0$ (independent of λ),
2. For each $\lambda \in \{\lambda_1, \dots, \lambda_P\}$, determine the subset of $\mathcal{S}_0$ that contains points verifying a given necessary optimality condition (*i.e.*, $\tilde{\mathcal{S}}$, $\mathcal{S}_{\mathrm{CW}}$, $\mathcal{S}_{\mathrm{L}}$),
3. Repeat 1-2 for different $\mathbf{A} \in \mathbb{R}^{5\times 10}$ and $\mathbf{y} \in \mathbb{R}^5$, and draw the average evolution of $\#\tilde{\mathcal{S}}$, $\#\mathcal{S}_{\mathrm{CW}}$, and $\#\mathcal{S}_{\mathrm{L}}$, with respect to λ.

Considered scenario

1. The entries of $\mathbf{A}$ and $\mathbf{y}$ are drawn from a standard normal distribution,
2. The entries of $\mathbf{A}$ and $\mathbf{y}$ are drawn from a uniform distribution on $[0, 1]$,
3. $\mathbf{A}$ is a "sampled Toeplitz" matrix built from a Gaussian kernel with $\sigma^2 = 0.04$. The entries of $\mathbf{y}$ are drawn from a standard normal distribution.

Quantifying "optimal" points



Theorem ([Soubies et al., 2019])

Let $\mathcal{S}_0$ be the set of strict local minimizers of F_0. Let $\tilde{\mathcal{S}}$, $\mathcal{S}_{\mathrm{CW}}$, and $\mathcal{S}_{\mathrm{L}}$ be the subsets of $\mathcal{S}_0$ containing the strict local minimizers of $\tilde{F}$, the partial support CW points, and the L-stationary points, respectively. Finally, define

$$\mathcal{X}_{\mathrm{LS}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2, \tag{1}$$

the solution set of the un-penalized least-squares problem. Then, for all $\mathcal{S} \in \{\tilde{\mathcal{S}}, \mathcal{S}_{\mathrm{L}}, \mathcal{S}_{\mathrm{CW}}\}$, there exists (under the URP of $\mathbf{A}$ for $\mathcal{S}_{\mathrm{CW}}$) $\lambda_0 > 0$ and $\lambda_\infty > 0$ such that

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

Algorithms and necessary optimality conditions

Algorithms and necessary optimality conditions

One can expect that the **efficiency** of a given algorithm $\mathcal{A}$ to minimize F_0 depends on the "**restrictiveness**" of the necessary optimality condition it guarantees to converge to.

One can expect that the **efficiency** of a given algorithm $\mathcal{A}$ to minimize F_0 depends on the "**restrictiveness**" of the necessary optimality condition it guarantees to converge to.

Numerical Experiment

We consider four algorithms:

- **CowS**: the CW support optimatity (CowS) algorithm. Greedy method that converges to a **partial support CW** point [Beck and Hallak, 2018].
- **IHT**: the iterative hard thresholding (IHT) algorithm that ensures the convergence to an **L-stationary** point [Attouch et al., 2013] [Beck and Hallak, 2018] [Blumensath and Davies, 2009].
- **FBS-CEL0**: the forward-backward splitting (FBS) algorithm applied to the CEL0 relaxation $\tilde{F}$. FBS ensures the convergence to a **stationary point of** $\tilde{F}$ [Attouch et al., 2013].
- **IRL1-CEL0**: the iterative reweighted-ℓ_1 (IRL1) algorithm [Ochs et al., 2015] also used to obtain a **stationary point of** $\tilde{F}$.

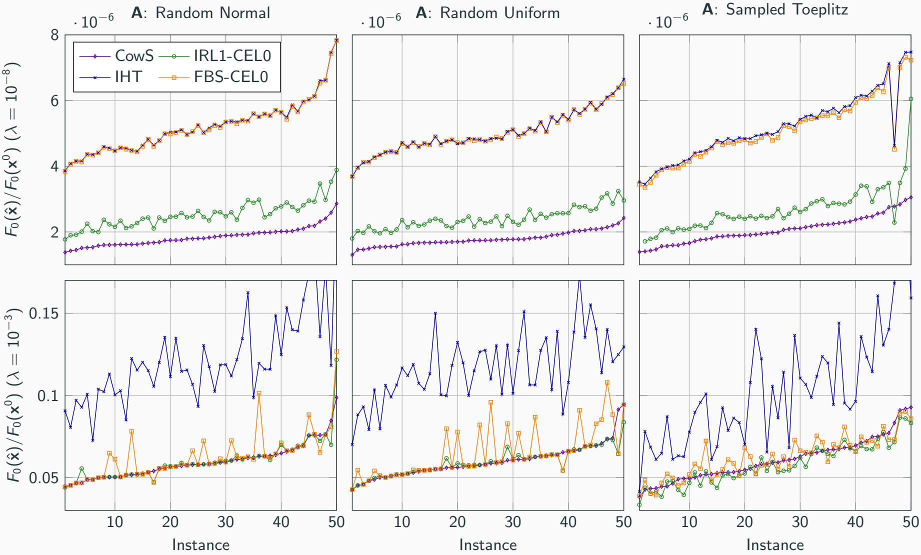
Numerical Experiment

- $K = 50$ instances of the problem (*i.e.*, instances of $\mathbf{A}$ and $\mathbf{y}$),
- $M = 100$, $N = 256$, $\lambda \in \{10^{-8}, 10^{-3}\}$,
- Initial point $\mathbf{x}^0 = \mathbf{0}_{\mathbb{R}^N}$,
- Generation of $\mathbf{A}$
  - i.i.d. entries drawn from a standard normal distribution,
  - i.i.d. entries drawn from a uniform distribution,
  - "sampled Toeplitz" matrix with a Gaussian kernel
- Measurements $\mathbf{y} \in \mathbb{R}^M$ are generated according to

$$\mathbf{y} = \mathbf{A}\mathbf{x}^\star + \mathbf{n},$$

where $\mathbf{x}^\star$ is a 30-sparse vector (*i.e.*, $\|\mathbf{x}^\star\|_0 = 30$) with non-zero entries drawn from a normal distribution. $\mathbf{n}$ is a vector of Gaussian noise with standard deviation 10^{-2}.

Algorithms and necessary optimality conditions



Concluding remarks

Concluding remarks

Support-based optimality conditions

Exact continuous relaxations

Concluding remarks

Support-based optimality conditions

- The **more restrictive** (stronger) among the conditions studied in this work
- **Trade-off** between restrictiveness and computational burden

Exact continuous relaxations

Concluding remarks

Support-based optimality conditions

- The **more restrictive** (stronger) among the conditions studied in this work
- **Trade-off** between restrictiveness and computational burden

Exact continuous relaxations

- **Open the door** to a variety of nonsmooth nonconvex optimization algorithms to minimize F_0,
- Although the derived inclusion properties plays in favor of greedy-based conditions, numerical experiments reveal that the associated algorithms are **comparable** in terms of their **ability to minimize F_0**.
  $\longrightarrow$ specific analysis of **fixed points** of algorithms that minimize $\tilde{F}$
- For **moderate-size problems**, exact continuous relaxations $\tilde{F}$ can be **globally minimized** using Lasserres hierarchies [Marmin et al., 2019]

New Insights on the Optimality Conditions of the ℓ_2-ℓ_0 Minimization Problem.
Submitted, 2019.
Emmanuel Soubies, Laure Blanc-Féraud and Gilles Aubert

Proximal Mapping for Symmetric Penalty and Sparsity.
SIAM Journal on Optimization 28-1, pp. 496-527, 2018.
Amir Beck and Nadav Hallak

Description of the Minimizers of Least Squares Regularized with ℓ_0-norm. Uniqueness of the Global Minimizer.
SIAM Journal on Imaging Science 6-2, pp. 904-937, 2013.
Mila Nikolova

A unified view of exact continuous penalties for ℓ_2-ℓ_0 minimization.
SIAM Journal on Optimization 27-3, pp. 2034-2060, 2017.
Emmanuel Soubies, Laure Blanc-Féraud and Gilles Aubert

A Continuous Exact ℓ_0 penalty (CEL0) for least squares regularized problem.
SIAM Journal on Imaging Science 8-3, pp. 1574-1606, 2015.
Emmanuel Soubies, Laure Blanc-Féraud and Gilles Aubert.

Thank you !

Attouch, H., Bolte, J., and Svaiter, B. F. (2013).
Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods.
Mathematical Programming, 137(1):91–129.

Beck, A. and Hallak, N. (2018).
Proximal Mapping for Symmetric Penalty and Sparsity.
SIAM Journal on Optimization, 28(1):496–527.

Blumensath, T. and Davies, M. E. (2009).
Iterative hard thresholding for compressed sensing.
Applied and Computational Harmonic Analysis, 27(3):265–274.

Bourguignon, S., Ninin, J., Carfantan, H., and Mongeau, M. (2016).
Exact Sparse Approximation Problems via Mixed-Integer Programming: Formulations and Computational Performance.
IEEE Transactions on Signal Processing, 64(6):1405–1419.

- Breiman, L. (1995).
Better Subset Regression Using the Nonnegative Garrote.
Technometrics, 37(4):373–384.

- Candes, E. J., Romberg, J., and Tao, T. (2006).
Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information.
IEEE Transactions on Information Theory, 52(2):489–509.

- Candès, E. J. and Wakin, M. B. (2008).
An Introduction To Compressive Sampling [A sensing/sampling paradigm that goes against the common knowledge in data acquisition].
IEEE Signal Processing Magazine, 25:21–30.

- Candès, E. J., Wakin, M. B., and Boyd, S. P. (2008).
Enhancing Sparsity by Reweighted ℓ_1 Minimization.
Journal of Fourier Analysis and Applications, 14(5):877–905.

Carlsson, M. (2019).
On Convex Envelopes and Regularization of Non-convex Functionals Without Moving Global Minima.
Journal of Optimization Theory and Applications, 183(1):66–84.

Chartrand, R. (2007).
Exact Reconstruction of Sparse Signals via Nonconvex Minimization.
IEEE Signal Processing Letters, 14(10):707–710.

Chen, S., Cowan, C. F. N., and Grant, P. M. (1991).
Orthogonal least squares learning algorithm for radial basis function networks.
IEEE Transactions on Neural Networks, 2(2):302–309.

Chen, S., Donoho, D., and Saunders, M. (2001).
Atomic Decomposition by Basis Pursuit.
SIAM Review, 43(1):129–159.

Cherni, A., Chouzenoux, E., Duval, L., and Pesquet, J.-C. (2019).
Forme lissée de rapports de normes lp/lq (SPOQ) pour la reconstruction des signaux avec pénalisation parcimonieuse.
In *GRETSI 2019*, Lille, France.

Chouzenoux, E., Jezierska, A., Pesquet, J., and Talbot, H. (2013).
A Majorize-Minimize Subspace Approach for $\ell_2 - \ell_0$ Image Regularization.
SIAM Journal on Imaging Sciences, 6(1):563–591.

Dai, W. and Milenkovic, O. (2009).
Subspace Pursuit for Compressive Sensing Signal Reconstruction.
IEEE Transactions on Information Theory, 55(5):2230–2249.

Donoho, D. L. (2006).
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(6):797–829.

Fan, J. and Li, R. (2001).
Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties.
Journal of the American Statistical Association, 96(456):1348–1360.

Foucart, S. (2011).
Hard Thresholding Pursuit: An Algorithm for Compressive Sensing.
SIAM Journal on Numerical Analysis, 49(6):2543–2563.

Foucart, S. and Lai, M.-J. (2009).
Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0<q \leq 1$.
Applied and Computational Harmonic Analysis, 26(3):395–407.

Gorodnitsky, I. F. and Rao, B. D. (1997).
Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm.
IEEE Transactions on Signal Processing, pages 600–616.

Gribonval, R. and Nielsen, M. (2003).
Sparse representations in unions of bases.
IEEE Transactions on Information Theory, 49(12):3320–3325.

Mallat, S. G. and Zhang, Z. (1993).
Matching pursuits with time-frequency dictionaries.
IEEE Transactions on Signal Processing, 41(12):3397–3415.

Mangasarian, O. L. (1996).
Machine Learning via Polyhedral Concave Minimization.
In Fischer, H., Riedmüller, B., and Schäffler, S., editors, *Applied Mathematics and Parallel Computing: Festschrift for Klaus Ritter*, pages 175–188. Physica-Verlag HD, Heidelberg.

Marmin, A., Castella, M., and Pesquet, J. (2019).
How to Globally Solve Non-convex Optimization Problems Involving an Approximate ℓ_0 Penalization.
In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5601–5605.

Mohimani, H., Babaie-Zadeh, M., and Jutten, C. (2009).
A Fast Approach for Overcomplete Sparse Decomposition Based on Smoothed ℓ_0-Norm.
IEEE Transactions on Signal Processing, 57(1):289–301.

Natarajan, B. (1995).
Sparse Approximate Solutions to Linear Systems.
SIAM Journal on Computing, 24(2):227–234.

Needell, D. and Tropp, J. A. (2009).
CoSaMP: Iterative signal recovery from incomplete and inaccurate samples.
Applied and Computational Harmonic Analysis, 26(3):301–321.

Nguyen, T. T., Soussen, C., Idier, J., and Djermoune, E.-H. (2019).
NP-hardness of ℓ_0 minimization problems: revision and extension to the non-negative setting.
In *International Conference on Sampling Theory and Applications (SampTa)*, Bordeaux.

Nikolova, M. (2013).
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.

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

Pati, Y. C., Rezaiifar, R., and Krishnaprasad, P. S. (1993).
Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition.
In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pages 40–44 vol.1.

Repetti, A., Pham, M. Q., Duval, L., Chouzenoux, É., and Pesquet, J. (2015).
Euclid in a Taxicab: Sparse Blind Deconvolution with Smoothed ℓ_1/ℓ_2 Regularization.
IEEE Signal Processing Letters, 22(5):539–543.

Selesnick, I. (2017).
Sparse Regularization via Convex Analysis.
IEEE Transactions on Signal Processing, 65(17):4481–4494.

Selesnick, I. and Farshchian, M. (2017).
Sparse Signal Approximation via Nonseparable Regularization.
IEEE Transactions on Signal Processing, 65(10):2561–2575.

Soubies, E., Blanc-Féraud, L., and Aubert, G. (2015).
A Continuous Exact ℓ_0 Penalty (CEL0) for Least Squares Regularized Problem.
SIAM Journal on Imaging Sciences, 8(3):1607–1639.

Soubies, E., Blanc-Féraud, L., and Aubert, G. (2017).
A Unified View of Exact Continuous Penalties for ℓ_2-ℓ_0 Minimization.
SIAM Journal on Optimization, 27(3):2034–2060.

Soubies, E., Blanc-Féraud, L., and Aubert, G. (2019).
New Insights on the Optimality Conditions of the l2-l0 Minimization Problem.

Soussen, C., Gribonval, R., Idier, J., and Herzet, C. (2013).
Joint K-Step Analysis of Orthogonal Matching Pursuit and Orthogonal Least Squares.
IEEE Transactions on Information Theory, 59(5):3158–3174.

Soussen, C., Idier, J., Brie, D., and Duan, J. (2011).
From Bernoulli–Gaussian Deconvolution to Sparse Signal Restoration.
IEEE Transactions on Signal Processing, 59(10):4572–4584.

Tibshirani, R. (1996).
Regression Shrinkage and Selection Via the Lasso.
Journal of the Royal Statistical Society: Series B (Methodological), 58(1):267–288.

Tropp, J. A. (2004).
Greed is good: algorithmic results for sparse approximation.
IEEE Transactions on Information Theory, 50(10):2231–2242.

Tropp, J. A. (2006).
Just relax: convex programming methods for identifying sparse signals in noise.
IEEE Transactions on Information Theory, 52(3):1030–1051.

Zhang, C.-H. (2010).
Nearly unbiased variable selection under minimax concave penalty.
The Annals of Statistics, 38(2):894–942.

Zou, H. (2006).
The Adaptive Lasso and Its Oracle Properties.
Journal of the American Statistical Association, 101(476):1418–1429.