Possible Equivalence Between the Optimal Solutions of Least Squares Regularized by L0 Norm and Penalized by L0 Norm

Mila Nikolova

CMLA, CNRS, ENS Cachan, France

http://mnikolova.perso.math.cnrs.fr

nikolova@cmla.ens-cachan.fr

Sparse Days Meeting 2014 at CERFACS, Toulouse

June 5th-6th, 2014

1. Two optimization problems involving the ℓ_0 pseudo norm

$\boldsymbol{A = (a_1, \cdots, a_n)} \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ (matrix) $\boldsymbol{a_i \neq 0 \;\; \forall\, i}$ N > M $\boldsymbol{d} \in \mathbb{R}^{\mathsf{M}} \setminus \{0\}$ (data)

- A vector $\widehat{u} \in \mathbb{R}^{\mathsf{N}}$ is k-sparse if $\|\boldsymbol{u}\|_0 := \sharp\left\{\boldsymbol{i} \;:\; \boldsymbol{u}[i] \neq 0\right\} \leqslant \mathbf{k}$.

One looks for a sparse vector $\widehat{u}$ such that "$A\widehat{u} \approx d$".

Two popular optimization problems to find a sparse $\widehat{u}$:

$(\mathcal{R}_\beta)$ $$\mathcal{F}_\beta(u) = \|Au - d\|_2^2 + \beta\|u\|_0 \qquad \beta > 0$$ (regularized)

$(\mathcal{C}_{\mathrm{k}})$ $$\min_{u\in\mathbb{R}^{\mathsf{N}}} \|Au - d\|_2^2 \quad \text{subject to} \quad \|u\|_0 \leqslant \mathrm{k}$$ (constrained)

- These are NP hard (combinatorial) nonconvex problems.
- $(\mathcal{R}_\beta)$ and $(\mathcal{C}_{\mathrm{k}})$ are often considered as "equivalent"...

Our goal:
Clarify the relationship between the optimal solution sets of $(\mathcal{R}_\beta)$ and $(\mathcal{C}_{\mathrm{k}})$.

Applications: signal and image processing, sparse coding, compression, dictionary building, compressive sensing, machine learning, model selection, classification...

$\|\cdot\|_0$ has served as a regularizer or as a penalty for a long time

- Markov random fields, MAP $\mathcal{F}_\beta(u) = \|Au - d\|_2^2 + \beta\|Du\|_0$
  Geman & Geman (1984), Besag (1986) - labeled images, stochastic algorithms
  Robini & Reissman (2012) - **global convergence / computation speed (!)**

- Subset selection via $(\mathcal{R}_\beta)$ - numerous algorithms - c.f. textbook Miller (2002)

- $(\mathcal{C}_\mathrm{k})$ – natural sparse coding constraint. Also the best K-term approximation [DeVore 1998)].

- Sparse-Land, M $<$ N - strong assumptions on A (RIP, spark, etc.) / various approximations.
  A huge amount of papers with approximating algorithms, e.g. Haupt & Nowak (06), Blumensath & Davies (08), Tropp (10), Zhang et al (12), Beck & Eldar (13)

  Typical assumption: every subset of K different columns of A is linearly independent

  Compressive Sensing Resources , references and software, , http://dsp.rice.edu/cs, 2013

Many algorithms, limited knowledge on the sets of optimal solutions

Notation and definitions

$\mathbb{I}_{\mathsf{N}} := (\{1, \cdots, \mathsf{N}\}, >)$ $\quad \mathbb{I}^0_{\mathsf{N}} := (\{0, 1, \cdots, \mathsf{N}\}, >)$ (totally strictly ordered)

- $\mathbf{supp}\,(\boldsymbol{u}) := \left\{ \boldsymbol{i} \in \mathbb{I}_{\mathsf{N}} \;:\; \boldsymbol{u}[\boldsymbol{i}] \neq \mathbf{0} \right\}$ $\quad \|\boldsymbol{u}\|_0 = \sharp\,\mathbf{supp}\,(\boldsymbol{u})$.
- $A_\omega := \left(a_{\omega[1]}, \cdots, a_{\omega[\sharp\,\omega]}\right) \in \mathbb{R}^{\mathsf{M} \times \sharp\,\omega}$, $\quad A^T_\omega$ is the transposed of A_ω

  $u_\omega := \left(u\big[\omega[1]\big], \cdots, u\big[\omega[\sharp\,\omega]\big]\right)^T \in \mathbb{R}^{\sharp\,\omega}$
- $\boxed{\|\cdot\| := \|\cdot\|_2\ .}$
- The <u>optimal solution sets</u> $\{\widehat{u}\}$ of problems $(\mathcal{R}_\beta)$ and $(\mathcal{C}_{\mathrm{k}})$:

$(\mathcal{R}_\beta) \qquad \mathcal{F}_\beta(\widehat{u}) \leqslant \mathcal{F}_\beta(u) \quad \forall\, u \in \mathbb{R}^{\mathsf{N}}$

$(\mathcal{C}_{\mathrm{k}}) \qquad \|A\widehat{u} - d\|^2 \leqslant \|Au - d\|^2 \quad \forall\, (\widehat{u}, u) \in \left\{ u \in \mathbb{R}^{\mathsf{N}} \;:\; \|u\|_0 \leqslant \mathrm{k} \right\}$

- Definition: Problems $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$ are *equivalent* for $\mathrm{k} \in \mathbb{I}_{\mathsf{M}}$ and $\beta > 0$ if

  $\widehat{u}$ – optimal solution of $(\mathcal{C}_{\mathrm{k}})$ $\quad \Leftrightarrow \quad$ $\widehat{u}$ – global minimizer of $\mathcal{F}_\beta$ for $\beta \in \overline{\beta}^{\mathrm{k}} \subset (0, +\infty)$

- Assumption $\mathbf{H}$: $\boxed{\mathbf{rank}(A) = \mathsf{M} < \mathsf{N}}$ systematically holds, no further remainder !

2. Preliminary results

Goal: Derive simple tests relating the optimal solution sets of $(\mathcal{R}_\beta)$ and of $(\mathcal{C}_{\mathrm{k}})$.

A constrained quadratic optimization problem: an useful tool

Given $\boldsymbol{d} \in \mathbb{R}^{\mathsf{M}}$ and $\boldsymbol{\omega} \subset \mathbb{I}_{\mathsf{N}}^{\mathbf{0}}$:

$$(\mathcal{P}_\omega) \qquad \min_{u \in \mathbb{R}^{\mathsf{N}}} \|Au - d\|^2 , \quad \text{subject to} \quad u[i] = 0 \quad \forall\, i \in \mathbb{I}_{\mathsf{N}} \setminus \omega \, .$$

$$\Big[\ \widehat{u} \in \mathbb{R}^{\mathsf{N}} \text{ solves } (\mathcal{P}_\omega)\ \Big] \quad \Leftrightarrow \quad \Big[\ \widehat{u} = Z_\omega\left(\widehat{u}_\omega\right) \quad \text{and} \quad \widehat{u}_\omega \quad \text{solves} \quad \min_{v \in \mathbb{R}^{\sharp\,\omega}} \|A_\omega v - d\|^2 \Big].$$

where $u := Z_\omega(u_\omega), \quad u[i] = \begin{cases} 0 & \text{if } \ i \notin \omega \ , \\ u_\omega[k] & \text{for the unique } k \text{ such that } \ \omega[k] = i \ . \end{cases}$

Problem $(\mathcal{P}_\omega)$ always has solutions, for any $\omega \subset \mathbb{I}_{\mathsf{N}}^0$.

Any (local/global) solution of $(\mathcal{R}_\beta)$ and of $(\mathcal{C}_{\mathrm{k}})$ is also a solution of $(\mathcal{P}_\omega)$ for an ω.

2.1. On the optimal solution sets of problem $(\mathcal{C}_{\mathrm{k}})$

$$\Theta_{\mathrm{k}} := \bigcup_{\mathrm{n}=0}^{\mathrm{k}} \left\{ \omega \subset \mathbb{I}_{\mathsf{N}} \ : \ \sharp\omega = \mathrm{n} \right\} \quad \forall\, \mathrm{k} \in \mathbb{I}_{\mathsf{N}}^{0} \ .$$

Θ_{k} lists the supports of all k-sparse solutions of $(\mathcal{C}_{\mathrm{k}})$ $\Rightarrow$ problem $(\mathcal{C}_{\mathrm{k}})$ reads as

$$(\mathcal{C}_{\mathrm{k}}) \qquad \min_{\omega\in\Theta_{\mathrm{k}}} \left\{ \|A\widetilde{u} - d\|^2 \ : \ \widetilde{u} \ \text{ solves } (\mathcal{P}_{\omega}) \text{ for } \ \omega \in \Theta_{\mathrm{k}} \right\}$$

Lemma 1. For any $\mathrm{k} \in \mathbb{I}_{\mathsf{N}}^{0}$, the optimal solution set of problem $(\mathcal{C}_{\mathrm{k}})$ is nonempty.

The optimal set of problem $(\mathcal{C}_{\mathrm{k}})$:

$$\boxed{\widehat{\mathrm{C}}_{\mathrm{k}} := \left\{ \widehat{u} \in \mathbb{R}^{\mathsf{N}} \ : \ \widehat{u} \ \text{ is an optimal solution of } \ (\mathcal{C}_{\mathrm{k}}) \right\} \ .}$$

$\forall\, \mathrm{k} \in \mathbb{I}_{\mathsf{N}}^{0}$ $\forall\, d \in \mathbb{R}^{\mathsf{M}}$ the optimal value θ_{k} of problem $(\mathcal{C}_{\mathrm{k}})$ is unique and finite:

$$\boxed{\theta_{\mathrm{k}} := \|A\widehat{u} - d\|^2 \quad \text{where} \quad \widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{k}} \ .}$$

Lemma 2. $\theta_d(0) = \|d\|^2$ and $\theta_{\mathrm{k}-1} \geqslant \theta_{\mathrm{k}} \quad \forall\, \mathrm{k} \geqslant 1$, and $\theta_{\mathrm{k}} = 0 \quad \forall\, \mathrm{k} \geqslant \mathsf{M}$.

Lemma 3. For $\mathrm{k} \in \mathbb{I}_{\mathsf{M}}$, let $(\mathcal{C}_{\mathrm{k}})$ have an optimal solution $\widehat{u}$ obeying

$$\|\widehat{\boldsymbol{u}}\|_{\mathbf{0}} = \mathbf{k} - \mathbf{n} \quad \text{for} \quad \mathbf{n} \geqslant \mathbf{1}\ .$$

Then $\quad \boldsymbol{A}\widehat{\boldsymbol{u}} = \boldsymbol{d}$.

Further, $\widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{m}}$ and $\theta_{\mathrm{m}} = 0$ for any $\mathrm{m} \geqslant \mathrm{k} - \mathrm{n}$.

In words, $\widehat{u}$ is an unconstrained minimizer of $(\mathcal{C}_{\mathrm{k}})$.

We do not know if Lemma 3 holds if $\mathbf{H}$ fails.

Corollary 1. For some $\mathrm{k} \in \mathbb{I}^0_{\mathsf{M}}$, let $\widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{k}}$. One has the implication

$$\theta_{\mathrm{k}} > 0 \quad \Rightarrow \quad \|\widehat{u}\|_0 = \mathrm{k}\ .$$

Example: $d = Au$ for $\|u\|_0 = \mathrm{k} < \mathsf{M} \quad \Rightarrow \quad \widehat{u} = u, \quad \theta_{\mathrm{k}} = 0 \quad \text{and} \quad d \in \mathbb{R}^{\mathrm{k}} \subsetneqq \mathbb{R}^{\mathsf{M}}$.

Given A satisfying $\mathbf{H}$ and $d \neq 0$, set

$$\boxed{\mathsf{L} := \min \left\{ \mathbf{k} \in \mathbb{I}_{\mathsf{M}} \ : \ \boldsymbol{\theta}_{\mathbf{k}} = \mathbf{0} \right\} .}$$

- $\diamond$ $\mathsf{L} \geqslant 1$ is uniquely defined;
- $\diamond$ L depends on d and on A;
- $\diamond$ Typically: $\mathsf{L} = \dim(d)$ and for noisy data $\dim(d) = \mathsf{M}$.

Lemma 4. Let $\mathbf{H}$ hold. If $\widehat{u}$ is an optimal solution of $(\mathcal{C}_{\mathsf{L}})$, then $\theta_{\mathsf{L}} = 0$ and $\|\widehat{u}\|_0 = \mathsf{L}$.

Good news:

Lemma 5. For $\mathrm{k} \in \mathbb{I}_{\mathsf{L}}^0$, let $\widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{k}}$. Then

$$\mathrm{rank}\left(A_{\,\mathrm{supp}\,(\widehat{u})}\right) = \sharp\,\mathrm{supp}\,(\widehat{u}) \ .$$

$$\Omega_{\mathrm{k}} := \Big\{ \, \omega \subset \mathbb{I}_{\mathsf{N}} \ : \ \sharp\,\omega = \mathrm{k} = \mathrm{rank}\,(A_\omega) \, \Big\} \quad \forall\, \mathrm{k} \in \mathbb{I}^0_{\mathsf{M}} \quad \text{and} \quad \Omega := \bigcup_{\mathrm{k}=0}^{\mathsf{M}} \Omega_{\mathrm{k}} \ .$$

Theorem 1.

$\widehat{\boldsymbol{u}}$ is an optimal solution of $(\boldsymbol{\mathcal{C}}_{\mathbf{k}})$ for $\mathbf{k} \in \mathbb{I}^{\mathbf{0}}_{\mathsf{L}} \quad \Rightarrow \quad \|\widehat{\boldsymbol{u}}\|_{\mathbf{0}} = \mathbf{k}$ and $\mathbf{supp}\,(\widehat{\boldsymbol{u}}) \in \boldsymbol{\Omega}_{\mathbf{k}}$.

$$(\mathcal{C}_{\mathrm{k}}) \qquad \|A\widehat{u} - d\|^2 = \min_{\omega \in \Omega_{\mathrm{k}}} \Big\{ \|A\widetilde{u} - d\|^2 \ : \ \widetilde{u} \ \text{ solves } (\mathcal{P}_\omega) \text{ for } \ \omega \in \Omega_{\mathrm{k}} \Big\} \quad \forall\, \mathrm{k} \in \mathbb{I}^0_{\mathsf{L}} \ .$$

$$\sharp\,\Omega_{\mathrm{k}} \ll \sharp\,\Theta_{\mathrm{k}}$$

$$\theta_{\mathrm{k}} = \min \Big\{ \|A_\omega u_\omega - d\|^2 \ : \ \omega \in \Omega_{\mathrm{k}} \Big\} \quad \forall\, \mathrm{k} \in \mathbb{I}^0_{\mathsf{L}} \ .$$

- $\widehat{\mathrm{C}}_{\mathrm{k}} \cap \widehat{\mathrm{C}}_{\mathrm{n}} = \varnothing \quad \mathrm{k} \neq \mathrm{n} \quad \forall\, (\mathrm{k},\mathrm{n}) \in \mathbb{I}^0_{\mathsf{L}}$ and $\widehat{\mathrm{C}}_{\mathsf{L}} \subset \widehat{\mathrm{C}}_{\mathrm{k}} \quad \forall\, \mathrm{k} > \mathsf{L}$;
- For $\mathrm{k} \geqslant \mathsf{L}$ one has unconstrained optimal solutions – no interest.
- If $\mathsf{L} = \mathsf{M}$, then $(\mathcal{C}_{\mathsf{M}})$ is unconstrained and has $\sharp\,\Omega_{\mathsf{M}}$ optimal solutions – no interest.

2.2. Connections between problems $(\mathcal{R}_\beta)$ and $(\mathcal{C}_{\mathbf{k}})$

A minimizer $\widehat{u}$ of $\mathcal{F}_\beta$ is strict if $\exists$ neighborhood $\mathcal{O} \subset \mathbb{R}^{\mathsf{N}}$, $\widehat{u} \in \mathcal{O}$, so that $\mathcal{F}_\beta(\widehat{u}) < \mathcal{F}_\beta(u)\ \forall\, u \in \mathcal{O}$.

$$\widehat{\mathrm{F}}_\beta := \left\{\widehat{u} \in \mathbb{R}^{\mathsf{N}} \ :\ \widehat{u} \ \text{ is a global minimizer of } \ \mathcal{F}_\beta \ \text{ for } \ \beta > 0\right\} \ .$$

Theorem 2. (Nikolova 2013) Let $d \in \mathbb{R}^{\mathsf{M}}$ and $\beta > 0$. Then (Here $\mathbf{H}$ is no needed)

(i) $\widehat{\mathrm{F}}_\beta \neq \varnothing$ (i.e. $\mathcal{F}_\beta$ has global minimizers) ;

(ii) $\widehat{u} \in \widehat{\mathrm{F}}_\beta \ \Rightarrow \ \widehat{u}$ is a <u>strict</u> (local) minimizer of $\mathcal{F}_\beta$ (i.e. $\mathrm{supp}\,(\widehat{u}) \in \Omega$).

Theorem 3. Let $\beta > 0$.

$\widehat{u}$ is a <u>strict</u> (local) minimizer of $\mathcal{F}_\beta \ \Leftrightarrow \ \widehat{u} \in \mathrm{U} := \bigcup_{\omega \in \Omega} \{\widetilde{u} \in \mathbb{R}^{\mathsf{N}} \ :\ \widetilde{u}$ solves $(\mathcal{P}_\omega)$ for $\omega \in \Omega\}$.

$$\mathrm{U} = \bigcup_{\mathrm{k}=0}^{\mathsf{M}} \mathrm{U}_{\mathrm{k}} \quad \text{where} \quad \mathrm{U}_{\mathrm{k}} := \bigcup_{\omega \in \Omega} \left\{ \widetilde{u} \ : \ \widetilde{u} \text{ solves } (\mathcal{P}_\omega) \text{ for } \omega \in \Omega \text{ and } \|\widetilde{u}\|_0 = \mathrm{k} \right\} .$$

$$\widehat{\mathrm{C}}_{\mathrm{k}} \subset \mathrm{U}_{\mathrm{k}} \quad \forall\, \mathrm{k} \in \mathbb{I}_{\mathsf{L}}^{0} .$$

Lemma 6. Let $\beta > 0$ and let $\widehat{u}$ be an optimal solution of $(\mathcal{C}_{\mathrm{k}})$ for $\mathrm{k} \in \mathbb{I}_{\mathsf{L}}^{0}$.

$$\begin{aligned} \mathcal{F}_\beta(\widehat{u}) &= \theta_{\mathrm{k}} + \beta \mathrm{k} \quad \forall\, \widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{k}} \ ; \\ \mathcal{F}_\beta(\widetilde{u}) &> \mathcal{F}_\beta(\widehat{u}) \qquad \forall\, \widetilde{u} \in \mathrm{U}_{\mathrm{k}} \setminus \widehat{\mathrm{C}}_{\mathrm{k}} \ . \end{aligned}$$

$$\mathrm{n} \in \{\mathsf{L}+1, \cdots, \mathsf{M}\} \ \text{ and } \ \widetilde{u} \in \mathrm{U}_{\mathrm{n}} \quad \Rightarrow \quad \mathcal{F}_\beta(\widetilde{u}) > \theta_{\mathsf{L}} + \beta \mathsf{L} = \mathcal{F}(\overline{u}) \quad \text{where } \theta_{\mathsf{L}} = 0.$$

2.4. Joint optimality conditions for $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$

Proposition 1. Let $\beta > 0$. Then

$$\widehat{u} \text{ is a global minimizer of } \mathcal{F}_\beta \quad (\text{i.e. } \widehat{u} \in \widehat{\mathrm{F}}_\beta) \quad \Rightarrow \quad \widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{k}} \text{ for } \mathrm{k} := \|\widehat{u}\|_0 \in \mathbb{I}^0_{\mathsf{L}} .$$

$\widehat{\mathrm{C}}$ is the collection of all optimal solutions of problems $(\mathcal{C}_{\mathrm{k}})$ for $\mathrm{k} \in \mathbb{I}^0_{\mathsf{L}}$:

$$\widehat{\mathrm{C}} := \bigcup_{\mathrm{k}=0}^{\mathsf{L}} \widehat{\mathrm{C}}_{\mathrm{k}}$$

Corollary 2. $\beta > 0$ and $\widehat{u} \in \widehat{\mathrm{F}}_\beta \quad \Rightarrow \quad \widehat{u} \in \widehat{\mathrm{C}}$.

$\widehat{\mathrm{F}}$ is the collection of all global minimizers of $\mathcal{F}_\beta$ for all $\beta > 0$:

$$\widehat{\mathrm{F}} := \bigcup_{\beta>0} \widehat{\mathrm{F}}_\beta .$$

$$\Rightarrow \quad \widehat{\mathrm{F}} \subset \widehat{\mathrm{C}}$$

Crucial result:

Theorem 4. Let $\mathrm{k} \in \mathbb{I}_{\mathsf{L}}^{0}$ and $\beta > 0$. Suppose that $\widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{k}}$ and that

$$\mathcal{F}_{\beta}(\overline{u}) - \mathcal{F}_{\beta}(\widehat{u}) > 0 \qquad \forall\, \overline{u} \in \widehat{\mathrm{C}} \setminus \widehat{\mathrm{C}}_{\mathrm{k}} \ .$$

Then

$$u \in \mathbb{R}^{\mathsf{N}} \setminus \widehat{\mathrm{C}}_{\mathrm{k}} \quad \Rightarrow \quad \mathcal{F}(u) > \mathcal{F}(\widehat{u}) = \theta_{\mathrm{k}} + \beta \mathrm{k} \ \ \forall\, \widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{k}} \ ;$$

- Theorem 4 shows how to compare the optimal solution sets of $(\mathcal{R}_{\beta})$ and of $(\mathcal{C}_{\mathrm{k}})$.

3. Conditions for equivalence between $(\mathcal{C}_{\mathsf{K}})$ and $(\mathcal{R}_\beta)$

3.1. Critical parameter values

Hint: Let $\widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{k}}$ and let $\overline{u} \in \widehat{\mathrm{C}}_{\mathrm{k}+1}$.

$$0 = \mathcal{F}_\beta(\widehat{u}) - \mathcal{F}_\beta(\overline{u}) = \big(\theta_{\mathrm{k}} + \beta \mathrm{k}\big) - \big(\theta_{\mathrm{k}+1} + \beta(\mathrm{k}+1)\big) = \theta_{\mathrm{k}} - \theta_{\mathrm{k}+1} - \beta \quad \Rightarrow \quad \beta = \theta_{\mathrm{k}} - \theta_{\mathrm{k}+1} \,.$$

Set $\beta_0^- = \beta_0 := \|d\|^2 - \theta_1 > 0 \;\; \beta_0^+ = \beta_{-1} = +\infty$ and $\beta_{\mathrm{k}}^+ = \beta_{\mathrm{k}} = 0 \;\; \forall\, \mathrm{k} \geqslant \mathsf{L}$.

For any $\mathrm{k} \in \mathbb{I}_{\mathsf{L}}$ define

$$\beta_{\mathrm{k}}^- := \max_{\mathrm{n}=1}^{\mathsf{L}-\mathrm{k}} \frac{\theta_{\mathrm{k}} - \theta_{\mathrm{k}+\mathrm{n}}}{\mathrm{n}} \quad \text{and} \quad \beta_{\mathrm{k}}^+ := \min_{\mathrm{n}=1}^{\mathrm{k}} \frac{\theta_{\mathrm{k}-\mathrm{n}} - \theta_{\mathrm{k}}}{\mathrm{n}}$$

Theorem 5. Assume that $\beta_{\mathrm{k}}^- \leqslant \beta_{\mathrm{k}}^+$. Then

$\mathcal{F}_\beta$ has a global minimizer at $\widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{k}}$ if and only if $\beta \in [\beta_{\mathrm{k}}^-, \beta_{\mathrm{k}}^+]$

Corollary 3. If $\beta_{\mathrm{k}}^- > \beta_{\mathrm{k}}^+$ then $\widehat{\mathrm{C}}_{\mathrm{k}} \cap \widehat{\mathrm{F}} = \varnothing$ (i.e. $\forall\, \beta > 0$, no global minimizer of $\mathcal{F}_\beta$ is in $\widehat{\mathrm{C}}_{\mathrm{k}}$).

3.2. General equivalence results

$$E := \{\mathrm{k} \in \mathbb{I}_{\mathsf{L}}^{0} \ : \ \beta_{\mathrm{k}}^{-} < \beta_{\mathrm{k}}^{+}\} = \{\mathrm{k}_1, \ldots, \mathrm{k}_{\mathrm{m}}\} \quad \mathrm{m} \leqslant \mathsf{L}$$

Note that $E \neq \varnothing$ for any $d \neq 0$.

Proposition 2. For any $\mathrm{k}_{\mathrm{n}} \in E$ it holds

$$\ldots \beta_{\mathrm{k}_{\mathrm{n}+2}}^{+} = \beta_{\mathrm{k}_{\mathrm{n}+1}}^{-} < \beta_{\mathrm{k}_{\mathrm{n}+1}}^{+} = \beta_{\mathrm{k}_{\mathrm{n}}}^{-} < \beta_{\mathrm{k}_{\mathrm{n}}}^{+} = \beta_{\mathrm{k}-1}^{-} < \beta_{\mathrm{k}-1}^{+} = \beta_{\mathrm{k}-2}^{-} \ldots$$

$$\mathrm{Closure}\left(\bigcup_{\mathrm{k} \in E} (\beta_{\mathrm{k}}^{-}, \ \beta_{\mathrm{k}}^{+}) \right) = \mathbb{R}_{+}$$

Theorem 6. For any $\mathrm{k} \in E$ one has

$\widehat{\boldsymbol{u}}$ is an optimal solution of $(\boldsymbol{\mathcal{C}}_{\mathbf{k}})$ $\iff$ $\widehat{\boldsymbol{u}}$ is a global minimizer of $\boldsymbol{\mathcal{F}}_{\boldsymbol{\beta}}$ for $\boldsymbol{\beta} \in (\boldsymbol{\beta}_{\mathbf{k}}^{-}, \boldsymbol{\beta}_{\mathbf{k}}^{+})$.

3.3. Necessary and sufficient conditions for full equivalence

$$\boxed{\boldsymbol{\beta_k} := \boldsymbol{\theta}_{\mathbf{k}} - \boldsymbol{\theta}_{\mathbf{k+1}}\ , \quad \forall\, \mathbf{k} \in \mathbb{I}_{\mathsf{L}}^{\mathbf{0}}\ .}$$

Lemma 7. $E = \mathbb{I}_{\mathsf{L}}^{0} \iff \{\beta_{\mathrm{k}}\}_{\mathrm{k}=0}^{\mathsf{L}}$ is strictly decreasing

$\{\beta_{\mathrm{k}}\}_{\mathrm{k}=0}^{\mathsf{L}}$ is strictly decreasing $\Rightarrow$ $\boldsymbol{\beta}_{\mathbf{k}}^{-} = \boldsymbol{\beta}_{\mathbf{k}}$ and $\boldsymbol{\beta}_{\mathbf{k}}^{+} = \boldsymbol{\beta}_{\mathbf{k-1}}$ $\quad \forall\, \mathrm{k} \in \mathbb{I}_{\mathsf{L}}^{0}$

Theorem 7. For any $\mathrm{k} \in \mathbb{I}_{\mathsf{L}}^{0}$ one has

$\widehat{\boldsymbol{u}}$ is an optimal solution of $(\boldsymbol{\mathcal{C}}_{\mathbf{k}})$ $\overset{(\star)}{\iff}$ $\widehat{\boldsymbol{u}}$ is a global minimizer of $\boldsymbol{\mathcal{F}}_{\boldsymbol{\beta}}$ for $\boldsymbol{\beta} \in (\boldsymbol{\beta}_{\mathbf{k}}, \boldsymbol{\beta}_{\mathbf{k-1}})$.

if and only if

$$\boldsymbol{\beta}_{\mathbf{0}} > \boldsymbol{\beta}_{\mathbf{1}} > \cdots > \boldsymbol{\beta}_{\mathsf{L}-\mathbf{1}} > \boldsymbol{\beta}_{\mathsf{L}}\ . \qquad (\maltese)$$

$\beta = \beta_{\mathrm{k}} \quad \Rightarrow \quad \widehat{\mathrm{C}}_{\mathrm{k}} \cup \widehat{\mathrm{C}}_{\mathrm{k+1}} \subset \widehat{\mathrm{F}}_{\beta}$ whereas $\widehat{\mathrm{C}}_{\mathrm{k}} \cap \widehat{\mathrm{C}}_{\mathrm{k+1}} = \varnothing \quad \Rightarrow \quad (\star)$ fails.

$(\mathcal{R}_{\beta})$ and $(\mathcal{C}_{\mathrm{k}})$ are fully equivalent if and only if $\mathrm{k} \mapsto \beta_{\mathrm{k}}$ is strictly decreasing on $\mathbb{I}_{\mathsf{M}}^{0}$.

3.4. More facts

The condition (✠) in Theorem 7 reads as

$$\beta_{k-1} > \beta_k \quad \forall\, k \in \mathbb{I}_L^0 \quad \Leftrightarrow \quad \theta_k < \frac{1}{2}\left(\theta_{k-1} + \theta_{k+1}\right) \quad \forall\, k \in \mathbb{I}_L^0 .$$

For d living in open subsets in $\mathbb{R}^M$ it can fail for some $k \in \mathbb{I}_L^0 \setminus E$.

Proposition 3. Let $\bigcup_{n=k+1}^{m-1} \widehat{C}_n \not\subset \widehat{F}$ and $\widehat{C}_k \cup \widehat{C}_m \in \widehat{F}$. Then

$$\overline{\beta}_{k,m} := \frac{\theta_k - \theta_m}{m - k} \quad \Rightarrow \quad \overline{\beta}_{k,m} = \beta_k^- \quad \text{and} \quad \overline{\beta}_{k,m} = \beta_m^+$$

Example. Let $\beta_{k-1} < \beta_k$. Then $\forall\, \beta > 0$, $\mathcal{F}_\beta$ does not have global minimizers in $\widehat{C}_k$.

$$\min_{n=2}^{k} \beta_{k-n} \geqslant \beta_{k-2} > \overline{\beta}_{k-1,k+1} := \frac{\theta_{k-1} - \theta_{k+1}}{2} > \beta_{k+1} \geqslant \max_{n=1}^{L-k} \beta_{k+n} \quad \Rightarrow$$

$\widehat{u}$ is global minimizer of $\mathcal{F}_\beta$ for $\beta \in (\overline{\beta}, \beta_{k-2})$ $\Leftrightarrow$ $\widehat{u}$ is optimal solution of $(\mathcal{C}_{k-1})$ and $\|\widehat{u}\|_0 = k - 1$.

$\widehat{u}$ is global minimizer of $\mathcal{F}_\beta$ for $\beta \in (\beta_{k+1}, \overline{\beta})$ $\Leftrightarrow$ $\widehat{u}$ is optimal solution of $(\mathcal{C}_{k+1})$ and $\|\widehat{u}\|_0 = k + 1$.

- For $\beta = \overline{\beta}_{k-1,k+1}$ any $\widehat{u} \in \widehat{C}_{k-1} \cup \widehat{C}_{k+1}$ is a global minimizer of $\mathcal{F}_\beta$.

4. On the critical parameter values

Any optimal solution $\widehat{u}$ of $(\mathcal{R}_\beta)$ and $(\mathcal{C}_{\rm k})$ is a solutions of $(\mathcal{P}_\omega)$ for $\omega := \operatorname{supp}(\widehat{u})$ such that $\operatorname{rank}(A_\omega) = \sharp\,\omega$, i.e. for $\omega \in \Omega$. Then $\widehat{u}$ is of the form

$$\widehat{u} = Z_\omega(\widehat{u}_\omega) \quad \text{where} \quad \widehat{u}_\omega = \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T d \ .$$

These solutions should in general be sensitive to noise in the data d since they solve a non-regularized least-squares problem $(\mathcal{P}_\omega)$.

$\Pi_\omega := A_\omega \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T$. Then

$$\widehat{u} \in \widehat{\mathrm{C}}_{\,\mathrm{k}} \quad \text{and} \quad \omega := \operatorname{supp}(\widehat{u}) \quad \Rightarrow \quad \theta_{\,\mathrm{k}} = \|(I - \Pi_\omega)d\|^2 \quad \forall\, \mathrm{k} \in \mathbb{I}^0_{\mathsf{L}} \ .$$

Set

$$\mathrm{E_k} := \bigcup_{\omega \in \Omega_{\mathrm{k}}} \ker\left(A_\omega^T\right) \quad \text{and} \quad \mathrm{G_k} := \bigcup_{\omega \in \Omega_{\mathrm{k}}} \operatorname{range}(A_\omega) \ .$$

Clearly, $\mathrm{G}_{\mathsf{M}} = \mathbb{R}^{\mathsf{M}}$ and $\mathrm{E}_{\mathsf{M}} = \varnothing$ by $\mathbf{H}$.

Lemma 8. Let $\mathbf{H}$ hold and $d \neq 0$.

- One has $\quad \theta_d(0) > \theta_{\mathrm{k}} \quad \forall\, \mathrm{k} \geqslant 1$.
- Let $\mathrm{k} \in \{1, \cdots, \mathsf{M}-1\}$. Then $\quad \left[d \in \mathbb{R}^{\mathsf{M}} \setminus \mathrm{G}_{\mathrm{k}} \quad \Leftrightarrow \quad \theta_{\mathrm{k}} > 0\ . \right]$

  (d is not in the range of any k full column submatrix $\Leftrightarrow \quad \cdots$)
- Let $\mathrm{k} \in \{2, \cdots, \mathsf{L}-1\}$. Then $\quad \left[d \in \mathbb{R}^{\mathsf{M}} \setminus \mathrm{E}_{\mathrm{k}} \quad \Rightarrow \quad \theta_{\mathrm{k}-1} > \theta_{\mathrm{k}}\ . \right]$

Corollary 4. $\quad d \in \mathbb{R}^{\mathsf{M}} \setminus (\mathrm{E}_2 \cup \mathrm{G}_{\mathsf{L}-1}) \quad \Rightarrow \quad \beta_k > 0 \quad \forall\, \mathrm{k} \in \mathbb{I}^0_{\mathsf{L}-1} \;$ and $\; \beta_{\mathrm{k}} = 0 \;\; \forall\, \mathrm{k} \geqslant \mathsf{L}$.

Remark. E_2 is a finite union of vector subspaces of dimension $\mathsf{M}-2$.
$\mathrm{G}_{\mathsf{L}-1}$ is a finite union of vector subspaces of dimension $\mathsf{L}-1 \leqslant \mathsf{M}-1$.
Therefore, $\mathrm{E}_2 \cup \mathrm{G}_{\mathsf{L}-1}$ is closed in $\mathbb{R}^{\mathsf{M}}$ and its Lebesgue measure in $\mathbb{R}^{\mathsf{M}}$ is null.
Conversely, $\mathbb{R}^{\mathsf{M}} \setminus (\mathrm{E}_2 \cup \mathrm{G}_{\mathsf{L}-1})$ form an open dense subset of $\mathbb{R}^{\mathsf{M}}$: the data we usually have.

No general guarantees that $\mathbf{k} \mapsto \boldsymbol{\beta}_{\mathbf{k}}$ is strictly decreasing...

5. Numerical tests

$$A = \begin{pmatrix} 13.94 & 16.36 & 4.88 & -3.09 & -15.42 & 1.31 & -3.18 & -12.13 & -4.26 & -10.09 \\ 7.06 & -6.48 & -9.07 & -8.37 & -2.72 & -17.42 & -5.83 & -3.81 & 3.87 & -1.80 \\ 11.63 & 6.73 & -4.75 & -6.28 & 3.42 & 6.68 & -1.64 & 13.23 & 9.03 & -20.27 \\ -7.54 & 12.74 & -6.66 & 5.01 & 4.84 & 8.98 & -9.35 & 3.85 & 7.18 & 4.09 \\ 3.22 & -10.40 & -5.02 & 16.70 & 9.53 & -5.49 & 11.88 & -3.62 & 17.36 & 7.34 \end{pmatrix}$$

$$u_o = \begin{pmatrix} 0 & \mathbf{4} & 0 & 0 & 0 & \mathbf{9} & 0 & 0 & \mathbf{3} & 0 \end{pmatrix}^T .$$

The components of A follow a nearly normal distribution with variance 10.

H holds because $\mathrm{rank}(A) = \mathsf{M} = 5$. Problem $(\mathcal{C}_\mathsf{M})$ has $\sharp\,\Omega_\mathsf{M} = 252$ optimal solutions.

Next we consider

(a) Noise-free data, $d = Au_o$

(b) Noisy data 1, $d = Au_o$+noise1

(c) Noisy data 2, $d = Au_o$+noise2

All results – calculated using an exhaustive combinatorial search.

(a) Noise-free data: $d = Au_o = \begin{pmatrix} 64.45 & -171.09 & 114.13 & 153.32 & -38.93 \end{pmatrix}^T$.

$\widehat{u} := u_0$ is an optimal solution to problems $(\mathcal{C}_3)$, $(\mathcal{C}_4)$ and $(\mathcal{C}_5)$ with $\theta_3 = \theta_4 = \theta_5 = 0$.

$$\beta_0 = 6.373 \times 10^4, \quad \beta_1 = 3777, \quad \beta_2 = 3968, \quad \beta_3 = 0 \ .$$

$\beta_2 > \beta_1$ and L $= 3 \quad \Rightarrow \quad$ no global minimizer of $\mathcal{F}_\beta$ with $\ell_0 \in \{2, 4, 5\}$, $\forall\, \beta > 0$

$$\beta_0 > \overline{\beta}_{1,3} := \frac{1}{2}(\ \theta_d(1) - \ \theta_d(3)) = 3872.46 > \beta_3 \ .$$

$\widehat{\mathrm{C}}_3 = \{\ \widehat{\mathrm{F}}_\beta \ : \ \beta \in (\beta_3, \overline{\beta}_{1,3})\}$ and $\widehat{\mathrm{C}}_1 = \{\ \widehat{\mathrm{F}}_\beta \ : \ \beta \in (\overline{\beta}_{1,3}, \beta_0)\}$.

k	Optimal solution of $(\mathcal{C}_\mathrm{k})$										Global min. of $\mathcal{F}_\beta$	θ_k
3	0	**4**	0	0	0	**9**	0	0	**3**	0	$\beta \in (0, \overline{\beta}_{1,3})$	0
2	0	**3.25**	0	0	0	**9.29**	0	0	0	0	**no**	3968
1	0	0	0	0	0	**11.76**	0	0	0	0	$\beta \in (\overline{\beta}_{1,3}, \beta_0)$	7745
0	0	0	0	0	0	0	0	0	0	0	$\beta > \beta_0$	7.147×10^4

(b) Noisy data 1, noise $\approx$ normal, centered, i.i.d. and SNR$= 32.32$ dB:

$$d = \left(\begin{array}{ccccc} 69.13 & -171.95 & 113.74 & 150.27 & -36.09 \end{array} \right)^T .$$

$$\beta_0 = 6.315 \times 10^4 \quad \beta_1 = 3973 \quad \beta_2 = 4003 \quad \beta_3 = 36.25 \quad \beta_4 = 0.0681 \quad \beta_5 = 0 \ .$$

$\beta_2 > \beta_1 \quad \Rightarrow \quad$ no global minimizer of $\mathcal{F}_\beta$ with $\ell_0 = 2$, $\forall\ \beta > 0$

$$\beta_0 > \overline{\beta}_{1,3} := \frac{1}{2}(\theta_1 - \theta_3) = 3987.6848 > \beta_3 > \beta_4 > \beta_5$$

$\widehat{\mathrm{C}}_3 = \left\{\ \widehat{\mathrm{F}}_\beta \ :\ \beta \in (\beta_3, \overline{\beta}_{1,3}) \right\}$ and $\widehat{\mathrm{C}}_1 = \left\{\ \widehat{\mathrm{F}}_\beta \ :\ \beta \in (\overline{\beta}_{1,3}, \beta_0) \right\}$.

k	θ_{k}	Optimal solution of $(\mathcal{C}_{\mathrm{k}})$										Global min. of $\mathcal{F}_\beta$
4	0.0681	0	**4.40**	0	0	0	**8.71**	**0.54**	0	**2.95**	0	$\beta \in (\beta_4, \beta_3)$
3	36.31	0	**4.09**	0	0	0	**8.88**	0	0	**3.01**	0	$\beta \in (\beta_3, \overline{\beta}_{1,3})$
2	4039	0	**3.33**	0	0	0	**9.17**	0	0	0	0	**no**
1	8012	0	0	0	0	0	**11.71**	0	0	0	0	$\beta \in (\overline{\beta}_{1,3}, \beta_0)$
0	7.1×10^4	0	0	0	0	0	0	0	0	0	0	$\beta > \beta_0$

(c) Noisy data 2, noise ≈ normal, centered, i.i.d. and SNR= 25.74 dB:

$$d = \begin{pmatrix} 66.67 & -169.08 & 101.56 & 149.38 & -39.50 \end{pmatrix}^T .$$

$$\beta_0 = 6.029 \times 10^4 \quad \beta_1 = 3825 \quad \beta_2 = 3037 \quad \beta_3 = 72.73 \quad \beta_4 = 0.0259 \quad \beta_5 = 0 .$$

$\{\beta_{\mathrm{k}}\}$ is strictly decreasing, so $(\mathcal{R}_\beta)$ and $(\mathcal{C}_{\mathrm{k}})$ are "fully" equivalent.

k	θ_{k}	Optimal solution of $(\mathcal{C}_{\mathrm{k}})$										Global min. of $\mathcal{F}_\beta$
4	0.0259	0	**8.54**	0	0	**4.49**	**4.90**	**2.73**	0	0	0	$\beta \in (\beta_4, \beta_3)$
3	72.76	0	**3.93**	0	0	0	**8.70**	0	0	**2.63**	0	$\beta \in (\beta_3, \beta_2)$
2	3110	0	**3.27**	0	0	0	**8.95**	0	0	0	0	$\beta \in (\beta_2, \beta_1)$
1	6935	0	0	0	0	0	**11.44**	0	0	0	0	$\beta \in (\beta_1, \beta_0)$
0	6.7×10^4	0	0	0	0	0	0	0	0	0	0	$\beta > \beta_0$

Overall, we have found numerically more cases with $\{\beta_{\mathrm{k}}\}$ strictly decreasing.

6. Summary

It is assumed that $\mathrm{rank}(A) = \mathsf{M}$ (i.e., A has full row rank) and that data $d \in \mathbb{R}^{\mathsf{M}} \setminus \{0\}$.

We defined $\mathsf{L} \leqslant \mathsf{M}$ as the least integer k so that the optimal value θ_{k} of $(\mathcal{C}_{\mathrm{k}})$ is null.

New facts on the optimal solution sets of problem $(\mathcal{C}_{\mathrm{k}})$

- For any $\mathrm{k} \leqslant \mathsf{L}$, any optimal solution $\widehat{u}$ of $(\mathcal{C}_{\mathrm{k}})$ obeys $\|\widehat{u}\|_0 = \mathrm{k}$ and the columns of A indexed by the support of $\widehat{u}$ are linearly independent;

- For any $\mathrm{k} > \mathsf{L}$, any optimal solution of $(\mathcal{C}_{\mathsf{L}})$ is also an optimal solution of $(\mathcal{C}_{\mathrm{k}})$.

New facts on the global minimizers of $(\mathcal{R}_\beta)$

- Any global minimizer $\widehat{u}$ of $\mathcal{F}_\beta$ belongs to the optimal solution set of $(\mathcal{C}_{\mathrm{k}})$ for $\mathrm{k} = \|\widehat{u}\|_0 \leqslant \mathsf{L}$;

- We exhibit necessary and sufficient conditions for β so that the set of the global minimizers of $\mathcal{F}_\beta$ coincides with the optimal solutions of $(\mathcal{C}_{\mathrm{k}})$.

Equivalence results

- There is a subset $\varnothing \neq E = \{\mathrm{k}_1, \ldots, \mathrm{k}_{\mathrm{m}}\} \subset \{0, \ldots, \mathsf{L}\}$ and critical values $\{\beta_{\mathrm{k}_{\mathrm{n}}}\}_{\mathrm{n}=1}^{\mathrm{m}}$ so that $\forall\, \mathrm{k}_{\mathrm{n}} \in E$, the optimal set of $(C_{\mathrm{k}_{\mathrm{n}}})$ is the set of the global minimizers of $\mathcal{F}_\beta$, $\forall\, \beta \in (\beta_{\mathrm{k}_{\mathrm{n}}}, \beta_{\mathrm{k}_{\mathrm{n}-1}})$.

  For $\beta = \beta_{\mathrm{k}_{\mathrm{n}}}$ the global minimizers of $\mathcal{F}_\beta$ contain the optimal sets of $(C_{\mathrm{k}_{\mathrm{n}}})$ and $(C_{\mathrm{k}_{\mathrm{n}+1}})$.

- $E = \{0, \ldots, \mathsf{L}\}$ if and only if $\{\beta_{\mathrm{k}} := \theta_{\mathrm{k}} - \theta_{\mathrm{k}+1}\}_{\mathrm{k}=0}^{\mathsf{L}}$ is strictly decreasing. I.e., $\forall\, \mathrm{k} \in \{0, \ldots, \mathsf{L}\}$ the optimal set of $(\mathcal{C}_{\mathrm{k}})$ equals the global minimizers of $\mathcal{F}_\beta$, $\forall\, \beta \in (\beta_{\mathrm{k}}, \beta_{\mathrm{k}-1})$.
  In this case problems $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$ are "fully" equivalent.

L and all β_{k} depend on A and on d .

7. Conclusions and open question

– $(\mathcal{R}_\beta)$ and $(\mathcal{C}_\mathrm{k})$ are NP-hard but some contemporary algorithms can find optimal solutions.

– We have fully clarified the relationship between their optimal solution sets.

– Partial equivalence between the optimal sets of $(\mathcal{C}_\mathrm{k})$ and $(\mathcal{R}_\beta)$ always holds.

– Many numerical tests (not shown) suggest that "full" equivalence is rather frequent.

– Under mild assumptions on A (not given) the optimal set of any $(\mathcal{C}_\mathrm{k})$, $\mathrm{k} \leqslant \mathsf{M} - 1$ is unique.

– Our comparative results can clarify the choice between models $(\mathcal{R}_\beta)$ and $(\mathcal{C}_\mathrm{k})$ in applications.

– Also, they can help the conception and the convergence analysis of algorithms.

– Given A, can we have good statistical models for the optimal value of problems $(\mathcal{C}_\mathrm{k})$?

– Are there matrices A so that "full" equivalence holds for almost every d ?

– Extensions to synthesis penalties of the form $\|Gu\|_0$ seems important.

Thank you for your attention!

Some references

- A. Beck and Y. C. Eldar, Sparsity constrained nonlinear optimization: Optimality conditions and algorithms, arXiv:1203.4580v1, (2012).
- T. Blumensath and M. Davies, Normalized iterative hard thresholding: Guaranteed stability and performance, IEEE Journal of Selected Topics in Signal Processing, 4 (2010).
- A. M. Bruckstein, D. L. Donoho, and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Review, 51 (2009).
- R. A. DeVore, Nonlinear approximation, Acta Numerica, 7 (1998).
- Y. C. Eldar and G. Kutyniok, Compressed Sensing: Theory and Applications, Cambridge Univ. Press, 2012.
- R. Gribonval and M. Nielsen, Sparse representations in unions of bases, IEEE Transactions on Information Theory, (2003).
- M. 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).
- M. C. Robini and P.-J. Reissman, From simulated annealing to stochastic continuation: a new trend in combinatorial optimization, J. Global Optim., 56 (2013)
- J. Tropp, Just relax: convex programming methods for identifying sparse signals in noise, IEEE Transactions on Information Theory, 52 (2006).