Least squares regularized or constrained by L0: relationship between their global minimizers

Mila Nikolova

CMLA, CNRS, ENS Cachan, Université Paris-Saclay, France

nikolova@cmla.ens-cachan.fr

SIAM Minisymposium on Trends in the Mathematics of Signal Processing and Imaging

organized by Willi Freeden and Zuhair Nashed

Joint Mathematics Meetings, January 6 – 9, 2016, Seattle

Outline

1. Two optimization problems involving the ℓ_0 pseudo norm
2. Joint optimality conditions for $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_{\beta})$
3. Parameter values for equality between optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_{\beta})$
4. Equivalences between the global minimizers of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_{\beta})$
   - Partial equivalence
   - Quasi-complete equivalence
5. On the optimal values of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_{\beta})$
6. Cardinality of the optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_{\beta})$
   - Uniqueness of the optimal solutions of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_{\beta})$
7. Numerical tests
8. Conclusions and future directions

1. Two optimization problems involving the ℓ_0 pseudo norm

$A = (A_1, \cdots, A_N) \in \mathbb{R}^{M \times N}$ (matrix) N > M $d \in \mathbb{R}^M \setminus \{0\}$ (data)

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

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

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

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

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

◇ These are NP hard (combinatorial) nonconvex problems.

Our goal: [M. N., ACHA 2016].

Clarify the relationship between the global minimizers 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 (14)

  Typical assumptions: RIP or K spark(A) plus others (e.g. bounds on $\|A\|$ etc.)

Important progress in solving problems $(\mathcal{C}_\mathrm{k})$ and $(\mathcal{R}_\beta)$. The numerical schemes – common points.

$\Longrightarrow$ Explore the relationship between their optimal sets.

The optimal values / the optimal solution setsof problems $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$:

$$(\mathcal{C}_{\mathrm{k}}) \qquad \mathrm{c}_{\mathrm{k}} := \inf\left\{\|Au-d\|^2 \;\middle|\; u\in\mathbb{R}^{\mathsf{N}} \;\text{ and }\; \|u\|_0 \leqslant \mathrm{k}\right\}$$

$$\widehat{\mathrm{C}}_{\mathrm{k}} := \left\{\; u\in\mathbb{R}^{\mathsf{N}} \;\text{ and }\; \|u\|_0 \leqslant \mathrm{k} \;\middle|\; \|Au-d\|^2 = \mathrm{c}_{\mathrm{k}}\right\}$$

$$(\mathcal{R}_\beta) \qquad r_\beta := \inf\left\{\mathcal{F}_\beta(u) \mid u\in\mathbb{R}^{\mathsf{N}}\right\}$$

$$\widehat{\mathrm{R}}_\beta := \left\{u\in\mathbb{R}^{\mathsf{N}} \;\middle|\; \mathcal{F}_\beta(u) = r_\beta\right\}$$

Theorem 1 *For any* $d\in\mathbb{R}^{\mathsf{M}}$*:* $\quad \widehat{\mathrm{C}}_{\mathrm{k}} \neq \varnothing \;\; \forall\, \mathrm{k}$ $\quad$ *and* $\quad \widehat{\mathrm{R}}_\beta \neq \varnothing \;\; \forall\, \beta > 0$.

H 1 *Assumption :* $\boxed{\mathbf{rank}(A) = \mathsf{M} < \mathsf{N}}$ *no further reminder*

How to evaluate the extent of assumption dependent properties ?

Definition 1 *A property is generic on* $\mathbb{R}^{\mathsf{M}}$ *if it holds on a subset of* $\mathbb{R}^{\mathsf{M}}\setminus S$ *where* S *is closed in* $\mathbb{R}^{\mathsf{M}}$ *and its Lebesgue measure in* $\mathbb{R}^{\mathsf{M}}$ *is null.*

A generic property is stronger than a property that holds only with probability one.

- $\mathbb{I}_{\mathrm{n}} := \big(\{1,\ldots,\mathrm{n}\}, <\big)$ and $\mathbb{I}^0_{\mathrm{n}} := \big(\{0,1,\ldots,\mathrm{n}\}, <\big)$ (totally strictly ordered)
- $\mathsf{L} := \min\left\{\mathrm{k}\in\mathbb{I}_{\mathsf{N}} \;\middle|\; \mathrm{c}_{\mathrm{k}} = 0\right\}$ (uniquely defined) generically $\mathsf{L} = \mathsf{M}$

Main results

There is a strictly decreasing sequence $\{\beta_{\rm k}\}_{\rm k\in J} \equiv \{\beta_{\rm J_k}\}$ for $\rm J \subseteq \mathbb{I}_L$ such that

$\widehat{u}$ is global minimizer of $\mathcal{F}_\beta$ for $\beta \in \left(\beta_{\rm J_k}, \beta_{\rm J_{k-1}}\right)$ $\Longleftrightarrow$ $\widehat{u}$ is global minimizer of $(\mathcal{C}_{\rm J_k})$

Equivalently

$$\left\{\widehat{\rm R}_\beta \mid \beta \in \left(\beta_{\rm J_k}, \beta_{\rm J_{k-1}}\right)\right\} = \widehat{\rm C}_{\rm J_k} \quad \forall\, \rm k \in J$$

In a generic sense

$$\widehat{\rm R}_{\beta_{\rm J_k}} = \widehat{\rm C}_{\rm J_k} \cup \widehat{\rm C}_{\rm J_{k+1}}$$

- All $\beta_{\rm J_k}$'s are obtained from the optimal values $\rm c_k$'s of the problems $(\mathcal{C}_{\rm k})$, $\rm k \in \mathbb{I}_L^0$.
- The global minimizers of problems $(\mathcal{C}_{\rm k})$ and $(\mathcal{R}_\beta)$ are strict and generically uniques
- J is always nonempty
- For any $\rm n \in \mathbb{I}_L^0 \setminus J$ the global minimizers of $(\mathcal{C}_{\rm n})$ are not global minimizers of $(\mathcal{R}_\beta)$ $\forall\, \beta$
- When $\rm J = \mathbb{I}_L^0$, problems $(\mathcal{C}_{\rm k})$ and $(\mathcal{R}_\beta)$ are quasi-completely equivalent:
  $\left\{\widehat{\rm R}_\beta \mid \beta \in (\beta_{\rm k}, \beta_{\rm k-1})\right\} = \widehat{\rm C}_{\rm k} \quad \beta_k = \rm c_k - c_{k+1} \quad \forall\, k$

Notation

- $\| \,.\, \| := \| \,.\, \|_2$.
- $\mathrm{supp}\,(u) := \Big\{ i \in \mathbb{I}_{\mathsf{N}} \;:\; u[i] \neq 0 \Big\}$
- For any $\boldsymbol{\omega} \subset \mathbb{I}^{\mathbf{0}}_{\mathbf{N}}$

  $A_\omega := \big(A_{\omega_1}, \ldots, A_{\omega_{\sharp\,\omega}}\big) \in \mathbb{R}^{\mathsf{M}\times\,\sharp\,\omega}$, A^{T}_ω is the transposed of A_ω

  $u_\omega := \big(u_{\omega_1}, \ldots, u_{\omega_{\sharp\,\omega}}\big)^{\mathsf{T}} \in \mathbb{R}^{\sharp\,\omega}$

Definition 2 *Let* $f : \mathbb{R}^{\mathsf{N}} \to \mathbb{R}$ *and* $S \subseteq \mathbb{R}^{\mathsf{N}}$. *Consider the problem* $\min\{f(u) \mid u \in S\}$.

- $\widehat{u}$ *is a* strict *minimizer if there is a neighborhood* $\mathcal{O} \subset S$, $\widehat{u} \in \mathcal{O}$ *so that* $f(u) > f(\widehat{u})$ $\forall u \in \mathcal{O} \setminus \{\widehat{u}\}$.
- $\widehat{u}$ *is an* **isolated** *(local) minimizer if* $\widehat{u}$ *is the only minimizer in an open subset* $\mathcal{O}' \subset \mathcal{O}$

Outline

1. Two optimization problems involving the ℓ_0 pseudo norm

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

   – Preliminaries
   – On the global minimizers of problem $(\mathcal{C}_{\mathrm{k}})$
   – Necessary and sufficient conditions

3. Parameter values for equality between optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$

4. Equivalences between the global minimizers of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$

   – Partial equivalence

   – Quasi-complete equivalence

5. On the optimal values of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$

6. Cardinality of the optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$

   – Uniqueness of the optimal solutions of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$

7. Numerical tests

8. Conclusions and future directions

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

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

2.1 Preliminaries

A constrained quadratic optimization problem: given $d \in \mathbb{R}^\mathsf{M}$ and $\omega \subseteq \mathbb{I}_\mathsf{N}$

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

The convex problem $(\mathcal{P}_\omega)$ always has solutions, for any $\omega \in \mathbb{I}_\mathsf{N}^0$ and for any $d \in \mathbb{R}^\mathsf{M}$.

Some useful facts on the relation of $(\mathcal{P}_\omega)$ to $(\mathcal{C}_\mathrm{k})$ and $(\mathcal{R}_\beta)$ [M. N. SIIMS 2013]

$(\mathcal{R}_\beta)$ $\widehat{u}$ solves $(\mathcal{P}_\omega)$ for some $\omega \subset \mathbb{I}_\mathsf{N}^0$ $\quad\Leftrightarrow\quad$ $\widehat{u}$ is a (local) minimizer of $\mathcal{F}_\beta$, $\forall\, \beta > 0$

$\widehat{u}$ solves $(\mathcal{P}_\omega)$ for $\omega \subset \mathbb{I}_\mathsf{N}^0$ with $\mathrm{rank}(A_\omega) = \sharp\,\omega$ $\quad\Leftrightarrow\quad$ $\widehat{u} =$ strict (local) minimizer of $\mathcal{F}_\beta, \forall\, \beta$

$(\mathcal{C}_\mathrm{k})$ $\widehat{u}$ solves $(\mathcal{P}_\omega)$ for $\omega \subset \mathbb{I}_\mathsf{N}^0$ with $\sharp\,\omega = \mathrm{k}$ $\quad\Leftrightarrow\quad$ $\widehat{u}$ is a (local) minimizer of $(\mathcal{C}_\mathrm{k})$

Remark **1** *For any $\omega \subset \mathbb{I}_\mathsf{N}$ with $\mathrm{rank}(A_\omega) = \sharp\,\omega$, the minimizer $\widehat{u}$ of $(\mathcal{P}_\omega)$ is isolated.*

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

Lemma 1 $\mathrm{c}_0 = \|d\|^2$ *and* $\{\mathrm{c_k}\}_{\mathrm{k}\geqslant 0}$ *is decreasing with* $\mathrm{c_k} = 0 \;\; \forall\, \mathrm{k} \geqslant \mathsf{M}$.

Lemma 2 *For* $\mathrm{k} \in \mathbb{I}_\mathsf{M}$ *let* $(\mathcal{C}_\mathrm{k})$ *have a global minimizer* $\widehat{u}$ *obeying*

$$\|\widehat{u}\|_0 = \mathrm{k} - \mathrm{n} \quad \textit{for} \quad \mathrm{n} \geqslant 1 \ .$$

Then $\quad A\widehat{u} = d \ . \quad$ *Furthermore* $\quad \widehat{u} \in \widehat{\mathrm{C}}_\mathrm{m}$ *and* $\mathrm{c_m} = 0 \quad \forall\, \mathrm{m} \geqslant \mathrm{k} - \mathrm{n}$.

$$\boxed{\mathsf{L} := \min\left\{\mathrm{k} \in \mathbb{I}_\mathsf{M} \mid \mathrm{c_k} = 0\right\} \ .}$$

Example 1 *One has* $\mathsf{L} \leqslant \mathsf{M} - 1$ *if* $d = Au$ *for* $\|u\|_0 \leqslant \mathsf{M} - 1$.

Theorem 2 $\widehat{u} \in \widehat{\mathrm{C}}_\mathrm{k}$ *for* $\mathrm{k} \in \mathbb{I}^0_\mathsf{L} \implies \begin{cases} \|\widehat{u}\|_0 = \mathrm{k} = \mathrm{rank}\,(A_{\widehat{\sigma}}) \quad \text{for} \;\; \widehat{\sigma} := \mathrm{supp}\,(\widehat{u}) \\ \text{so} \;\; \widehat{u} \;\; \textit{is a } \underline{\textit{strict}} \; \textit{global minimizer of } (\mathcal{C}_\mathrm{k}). \end{cases}$

$\mathrm{k} \geqslant \mathsf{L} + 1 \quad \implies \quad \widehat{\mathrm{C}}_\mathsf{L} \subset \widehat{\mathrm{C}}_\mathrm{k}$.

Corollary 1 $\widehat{\mathrm{C}}_\mathrm{k} \cap \widehat{\mathrm{C}}_\mathrm{n} = \varnothing \quad \forall\, (\mathrm{k},\mathrm{n}) \in (\,\mathbb{I}^0_\mathsf{L}\,)^2$ *such that* $\mathrm{k} \neq \mathrm{n}$.

Example **2**

- $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ and $d = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

  $\widehat{u} = (0,0,1)^T = \widehat{\mathrm{C}}_1$ *(strict,* $\mathrm{rank}(A_3) = \|\widehat{u}\|_0$*)* $\implies$ $\mathrm{c}_1 = 0$ $\implies$ $\mathsf{L} = 1$.

  $\widehat{u} = (1,1,0)^T$ *is a strict global minimizer of* $(\mathcal{C}_2)$ *because* $\mathrm{rank}\left(A_{\mathrm{supp}\,(\widehat{u})}\right) = 2$ *and* $\mathrm{c}_2 = 0$.

- $A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$ and $d = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

  $\widehat{\mathrm{C}}_1 = \left\{(1,0,0,0)^T,\ (0,0,1,0)^T\right\}$ *(strict minimizers)* $\implies$ $\mathrm{c}_1 = 0$ $\implies$ $\mathsf{L} = 1$.

  For $\mathrm{k} \geqslant 2$ *all optimal solutions* $\not\in \widehat{\mathrm{C}}_1$ *are nonstrict and have the form* $\widehat{u} = (x, y, 1-x, -y)^T$, $x \in \mathbb{R} \setminus \{0,1\}$. *If* $y = 0$ *then* $\|\widehat{u}\|_0 = 2$ *and otherwise* $\|\widehat{u}\|_0 = 4$.

Remark **2** *By Theorem 2 the optimal value* c_k *of problem* $(\mathcal{C}_\mathrm{k})$ *for any* $\mathrm{k} \in \mathbb{I}^0_\mathsf{L}$ *obeys*

$$\mathrm{c}_\mathrm{k} = \min\left\{ \|A\widetilde{u} - d\|^2 \ \text{where } \widetilde{u} \in \mathbb{R}^\mathsf{N} \ \text{solves } (\mathcal{P}_\omega) \ \middle|\ \omega \in \Omega_\mathrm{k} \right\}$$

where $\quad \Omega_\mathrm{k} := \left\{ \omega \subset \mathbb{I}_\mathsf{N} \mid \sharp\omega = \mathrm{k} = \mathrm{rank}\,(A_\omega) \right\}.$

2.3. Necessary and sufficient conditions

Proposition 1 $\quad \widehat{u} \in \widehat{\mathrm{R}}_\beta \quad \Longrightarrow \quad \begin{cases} \widehat{u} \in \widehat{\mathrm{C}}_\mathrm{k} & \textit{where } \mathrm{k} := \|\widehat{u}\|_0 \in \mathbb{I}^0_\mathsf{L} \\ \widehat{\mathrm{C}}_\mathrm{k} \subseteq \widehat{\mathrm{R}}_\beta & \textit{for } \mathrm{k} := \|\widehat{u}\|_0 \in \mathbb{I}^0_\mathsf{L} \end{cases}$

The global minimizers of $\mathcal{F}_\beta$ are composed of some optimal sets $\widehat{\mathrm{C}}_\mathrm{k}$ for $\mathrm{k} \leqslant \mathsf{L}$.

$\widehat{\mathrm{C}}$ = the collection of all optimal solutions $\widehat{\mathrm{C}}_\mathrm{k}$ of problems $(\mathcal{C}_\mathrm{k})$ for all $\mathrm{k} \in \mathbb{I}^0_\mathsf{L}$;
$\widehat{\mathrm{R}}$ = the set of all global minimizers $\widehat{\mathrm{R}}_\beta$ of $\mathcal{F}_\beta$ for all $\beta > 0$

$$\widehat{\mathrm{C}} := \bigcup_{\mathrm{k}=0}^{\mathsf{L}} \widehat{\mathrm{C}}_\mathrm{k} \quad \text{and} \quad \widehat{\mathrm{R}} := \bigcup_{\beta>0} \widehat{\mathrm{R}}_\beta \ .$$

Theorem 3 $\quad \widehat{\mathrm{R}} \subset \widehat{\mathrm{C}}$.

When β ranges on $(0, +\infty)$, $\mathcal{F}_\beta$ can have *at most* $\mathsf{L}+1$ different sets of global minimizers which are optimal solutions of $(\mathcal{C}_\mathrm{k})$ for $\mathrm{k} \in \{0, \ldots, \mathsf{L}\}$.

Theorem 4 *For any* $\mathrm{k} \in \mathbb{I}^0_\mathsf{L}$ *one has:*

- $\widehat{\mathrm{C}}_\mathrm{k} \subseteq \widehat{\mathrm{R}}_\beta$ *if and only if* $\mathcal{F}_\beta(\overline{u}) - \mathcal{F}_\beta(\widehat{u}) \geqslant 0 \quad \forall\, \widehat{u} \in \widehat{\mathrm{C}}_\mathrm{k} \quad \forall\, \overline{u} \in \widehat{\mathrm{C}}$
- $\widehat{\mathrm{C}}_\mathrm{k} = \widehat{\mathrm{R}}_\beta$ *if and only if* $\mathcal{F}_\beta(\overline{u}) - \mathcal{F}_\beta(\widehat{u}) > 0 \quad \forall\, \widehat{u} \in \widehat{\mathrm{C}}_\mathrm{k} \quad \forall\, \overline{u} \in \widehat{\mathrm{C}} \setminus \widehat{\mathrm{C}}_\mathrm{k}$

Outline

1. Two optimization problems involving the ℓ_0 pseudo norm

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

3. Parameter values for equality between optimal sets of $(\mathcal{C}_\mathrm{k})$ and $(\mathcal{R}_\beta)$

   - The entire list of parameter values
   - Conditions for agreement between their optimal sets
   - The effective parameters values

4. Equivalences between the global minimizers of $(\mathcal{C}_\mathrm{k})$ and $(\mathcal{R}_\beta)$

   - Partial equivalence
   - Quasi-complete equivalence

5. On the optimal values of $(\mathcal{C}_\mathrm{k})$ and $(\mathcal{R}_\beta)$

6. Cardinality of the optimal sets of $(\mathcal{C}_\mathrm{k})$ and of $(\mathcal{R}_\beta)$

   - Uniqueness of the optimal solutions of $(\mathcal{C}_\mathrm{k})$ and of $(\mathcal{R}_\beta)$

7. Numerical tests

8. Conclusions and future directions

3. Parameter values for equality between optimal sets

3.1. The entire list of parameter values

Definition **3** (Critical parameter values)

$$\beta_k \;:=\; \max\left\{ \frac{c_k - c_{k+n}}{n} \;\middle|\; n \in \{1, \ldots, \mathsf{L} - k\} \right\} \quad \forall\, k \in \mathbb{I}^0_{\mathsf{L}-1} \quad \text{and} \quad \beta_{\mathsf{L}} = 0\,,$$

$$\beta^{\mathrm{U}}_k \;:=\; \min\left\{ \frac{c_{k-n} - c_k}{n} \;\middle|\; n \in \{1, \ldots, k\} \right\} \quad \forall\, k \in \mathbb{I}_{\mathsf{L}} \quad \text{and} \quad \beta^{\mathrm{U}}_0 \equiv \beta_{-1} := +\infty\,.$$

We have $\beta_{\mathsf{L}} = 0 < \beta^{\mathrm{U}}_{\mathsf{L}}$ and $\beta_0 < \beta^{\mathrm{U}}_0$.

The cases where $\beta_k < \beta^{\mathrm{U}}_k$ will be of particular interest.

Proposition **2** $\exists\, S$ – *finite union of vector subspaces of dimension* $\leqslant \mathsf{M} - 1$ *such that*

$$d \in \mathbb{R}^{\mathsf{M}} \setminus S \quad \Longrightarrow \quad \beta_k \neq \beta^{\mathrm{U}}_k \quad \forall\, k \in \mathbb{I}^0_{\mathsf{L}}\,.$$

$\beta_k \neq \beta^{\mathrm{U}}_k \quad \forall\, k \in \mathbb{I}^0_{\mathsf{L}}$ is a generic property.

3.2. Conditions for agreement between the optimal sets of $(\mathcal{C}_k)$ and $(\mathcal{R}_\beta)$

Theorem **5** $\forall\, k \in \mathbb{I}_L^0$

- $\widehat{C}_k \subseteq \widehat{R}_\beta$ *if and only if* $\begin{cases} \beta_0 \leqslant \beta < \beta_0^U & \text{for} \quad k = 0 \;; \\ \beta_k \leqslant \beta \leqslant \beta_k^U & \text{for} \quad k \in \{1, \dots, L-1\} \;; \\ \beta_L < \beta \leqslant \beta_L^U & \text{for} \quad k = L \;. \end{cases}$

- $\widehat{C}_k = \widehat{R}_\beta$ *if and only if* $\beta_k < \beta < \beta_k^U$.

Proof based on Theorem 4.

To exploit Theorem 5 we have to clarify the links between (β_k, β_k^U) and c_k

3.3. The effective parameters values

The global minimizers of $\mathcal{F}_\beta$ are always in $\widehat{\mathrm{C}}$ (Theorem 3), so we are interested in the indexes k for which there exist values of β such that $\widehat{\mathrm{C}}_\mathrm{k} \subset \widehat{\mathrm{R}}_\beta$. Their set is obtained from Theorem 5.

Definition **4** *The effective index set* $\mathrm{J} \cup \mathrm{J}^{\mathrm{E}}$*:*

$$\mathrm{J} := \left\{\mathrm{k} \in \mathbb{I}_\mathsf{L}^0 \mid \beta_\mathrm{k} < \beta_\mathrm{k}^{\mathrm{U}}\right\} \quad \text{and} \quad \mathrm{J}^{\mathrm{E}} := \left\{\mathrm{m} \in \mathbb{I}_\mathsf{L}^0 \mid \beta_\mathrm{m} = \beta_\mathrm{m}^{\mathrm{U}}\right\} \ .$$

Ordering: $\mathrm{J} = \{\mathrm{J}_0, \mathrm{J}_1, \ldots, \mathrm{J}_\mathrm{p}\}$ *where* $\mathrm{p} := \sharp\,\mathrm{J} - 1$ *and* $\mathrm{J}_{\mathrm{k}-1} < \mathrm{J}_\mathrm{k} \ \ \forall\, \mathrm{k}$.
Further: $(\mathrm{J}_0 = 0,\ \mathrm{J}_\mathrm{p} = \mathsf{L}) \in \mathrm{J}^2$ *with* $\beta_{\mathrm{J}_{-1}} := \beta_{\mathrm{J}_0}^{\mathrm{U}} \equiv \beta_0^{\mathrm{U}} = +\infty$ *and* $\beta_{\mathrm{J}_\mathrm{p}} \equiv \beta_\mathsf{L} = 0$.

The set J is always nonempty.

Lemma **3** $\widehat{\mathrm{R}} \cap \widehat{\mathrm{C}}_\mathrm{k} = \varnothing$ *if and only if* $\mathrm{k} \in \mathbb{I}_\mathsf{L}^0 \setminus \{\mathrm{J} \cup \mathrm{J}^{\mathrm{E}}\}$.

Definition 3 for reminder:

$$\beta_k := \max\left\{\frac{c_k - c_{k+n}}{n} \mid n \in \{1, \dots, L-k\}\right\} \quad \forall\, k \in \mathbb{I}^0_{L-1} \quad \text{and} \quad \beta_L := 0\ ,$$
$$\beta^U_k := \min\left\{\frac{c_{k-n} - c_k}{n} \mid n \in \{1, \dots, k\}\right\} \quad \forall\, k \in \mathbb{I}_L \quad \text{and} \quad \beta^U_0 := +\infty\ .$$

Simplification of $\{\beta_k,\ \beta^U_k\}_{k\in J\cup J^E}$

Proposition 3 *Let* $\{\beta_k, \beta^U_k\}$ *and* J *be as in Definition 3 and Definition 4, resp. Then*

(a) $\beta_{J_k} < \beta^U_{J_k} = \beta_{J_{k-1}} \qquad \forall\, J_k \in J \setminus \{J_0\} \quad \text{and} \quad \beta_{J^U_0} \equiv \beta_{J_{-1}} = +\infty\ .$

(b) $\beta_{J_k} = \dfrac{c_{J_k} - c_{J_{k+1}}}{J_{k+1} - J_k} \qquad \forall\, J_k \in J \setminus \{J_p\} \quad \text{and} \quad \beta_{J_p} \equiv \beta_L = 0\ .$

(c) $\{\, \beta_m \mid m \in J^E\} \subset \left\{\beta_{J_k} \;\middle|\; J_k \in J \setminus \{J_p\}\right\}\ .$

$\{\beta_k\}_{k\in J}$ is strictly decreasing and its first entry is β_0.

Example **3** *Let* $\{c_k\}_{k=0}^{L}$ *for* $L = 7$ *reads as*

$$c_0 = 48 \quad c_1 = 40 \quad c_2 = 30 \quad c_3 = 22 \quad c_4 = 14 \quad c_5 = 10 \quad c_6 = 4 \quad c_7 = 0\ .$$

By Definition 3 the sequences $\{\beta_k, \beta_k^U\}_{k=0}^{7}$ *are given by*

$$\boldsymbol{\beta_0 = 9} \quad \beta_1 = 10 \quad \boldsymbol{\beta_2 = 8} \quad \beta_3 = 8 \quad \boldsymbol{\beta_4 = 5} \quad \beta_5 = 6 \quad \boldsymbol{\beta_6 = 4} \quad \boldsymbol{\beta_7 = 0}$$

$$\beta_0^U = +\infty \quad \beta_1^U = 8 \quad \beta_2^U = \mathbf{9} \quad \beta_3^U = 8 \quad \beta_4^U = \mathbf{8} \quad \beta_5^U = 4 \quad \beta_6^U = \mathbf{5} \quad \beta_7^U = \mathbf{4}$$

From Definition 4, $J = \{\, J_0 = \mathbf{0},\ J_1 = \mathbf{2},\ J_2 = \mathbf{4},\ J_3 = \mathbf{6},\ J_4 = \mathbf{7}\}$ and $J^E = \{\, \mathbf{3}\,\}$.

- *One has* $\beta_{J_k} = \beta^U_{J_{k+1}}$ *for any* $J_k \in J$ *(Proposition 3(a)).*
- *The formula in Proposition 3(b) holds.*
- $\{\beta_3 \,|\, 3 \in J^E\} \;\Rightarrow\; \beta_3 = \beta_{J_1} = 8 \;\Rightarrow\; \{\beta_3 \,|\, 3 \in J^E\} \subset \{\beta_{J_k} \,|\, J_k \in J \setminus \{J_4\}\}$ *(Proposition 3(c)).*
- $J^E_{\beta_{J_1}} := \{m \in J^E \mid \beta_m = \beta_{J_1}\} = \{3 \in J^E \mid J_1 < 3 < J_2\}$*, see Lemma 4.*
- J *has the smallest indexes so that* $\{\beta_k\}_{k\in J} = \{9, 8, 5, 4, 0\}$ *is the longest strictly decreasing subsequence of* $\{\beta_k\}_{k=0}^{7}$ *containing* β_0 *– see Proposition 4. One has* $\{\beta_k\}_{k\in J} = \{\beta_k\}_{k\in J'}$ *for* $J' := \{0, 3, 4, 6, 7\}$*; however,* $J'_2 > J_2$.

The location of $\{\beta_{\mathrm{m}} \,|\, \mathrm{m} \in \mathrm{J}^{\mathrm{E}}\}$ is given by the (probably empty) subsets

$$\mathrm{J}^{\mathrm{E}}_{\beta_{\mathrm{J_k}}} := \{\mathrm{m} \in \mathrm{J}^{\mathrm{E}} \mid \beta_{\mathrm{m}} = \beta_{\mathrm{J_k}}\} \ .$$

Lemma **4** *The sets* $\mathrm{J}^{\mathrm{E}}_{\beta_{\mathrm{J_k}}}$ *fulfill* $\mathrm{J}^{\mathrm{E}}_{\beta_{\mathrm{J_k}}} = \varnothing$ *for* $\mathrm{k} = \mathrm{p}$ *and for any* $\mathrm{k} \leqslant \mathrm{p} - 1$

$$\mathrm{J}^{\mathrm{E}}_{\beta_{\mathrm{J_k}}} = \{\mathrm{m} \in \mathrm{J}^{\mathrm{E}} \mid \mathrm{J_k} < \mathrm{m} < \mathrm{J_{k+1}}\} \ .$$

J and $\{\beta_{\mathrm{k}}\}_{\mathrm{k} \in \mathrm{J}}$ are characterized next

Proposition **4** *Let* $\{\beta_{\mathrm{k}}\}_{\mathrm{k}=0}^{\mathsf{L}}$ *read as in Definition 3 and* J *as in Definition 4. Then* $0 \in \mathrm{J}$ *and* J *contains the smallest indexes such that* $\{\beta_{\mathrm{k}}\}_{\mathrm{k} \in \mathrm{J}}$ *is the longest strictly decreasing subsequence of* $\{\beta_{\mathrm{k}}\}_{\mathrm{k}=0}^{\mathsf{L}}$ *containing* β_0*.*

In order to find the effective J and $\{\beta_{\mathrm{k}}\}_{\mathrm{k} \in \mathrm{J}}$ we need only $\{\beta_{\mathrm{k}}\}_{\mathrm{k}=0}^{\mathsf{L}}$ in Definition 3.

Outline

1. Two optimization problems involving the ℓ_0 pseudo norm
2. Joint optimality conditions for $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_{\beta})$
3. Parameter values for equality between optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_{\beta})$
4. Equivalences between the global minimizers of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_{\beta})$
   - Partial equivalence
   - Quasi-complete equivalence
5. On the optimal values of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_{\beta})$
6. Cardinality of the optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_{\beta})$
   - Uniqueness of the optimal solutions of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_{\beta})$
7. Numerical tests
8. Conclusions and future directions

4. Equivalence relations between the optimal sets of $(\mathcal{C}_k)$ and $(\mathcal{R}_\beta)$

4.1. Partial equivalence

Theorem **6** *Let* $\{\beta_k\}$ *be as in Definition 3 and* J *as in Definition 4. Then:*

$$\left\{ \widehat{R}_\beta \mid \beta \in \left(\beta_{J_k},\ \beta_{J_{k-1}}\right) \right\} = \widehat{C}_{J_k} \quad \forall\, J_k \in J \ ,$$

$$\left(\bigcup_{n=1}^{p} \left[\beta_{J_k},\ \beta_{J_{k-1}}\right] \right) \cup \left[\beta_{J_0}, \beta_{J_{-1}}\right) = [\,0, +\infty) \ .$$

$\{\beta_{J_0}, \ldots, \beta_{J_{p-1}}\}$ in Definition 4 partition $(0, +\infty)$ into $\sharp J$ proper intervals. For any $\beta \in \left(\beta_{J_k},\ \beta_{J_{k-1}}\right)$ the optimal sets of $(\mathcal{R}_\beta)$ and of $(\mathcal{C}_n)$ for $n = J_k$ coincide.
If $\mathbb{I}_L^0 \setminus J \neq \varnothing$, the optimal sets $(\mathcal{C}_k)$ for $k \in \mathbb{I}_L^0 \setminus J$ cannot be optimal solutions of $(\mathcal{R}_\beta)$ $\forall\, \beta > 0$.

$\Longrightarrow$ partial equivalence.

$\{\beta_{J_k}\}$ is a finite set of isolated values hence $\beta \neq \beta_{J_k}$ generically.

The optimal sets of problem $(\mathcal{R}_\beta)$ for $\beta_{\mathrm{k}},\ \mathrm{k}\in\mathrm{J}$:

Theorem 7 *Let H1 hold. Let $\{\beta_{\mathrm{k}}\}$ be as in Definition 3 and J as in Definition 4. Then*

$$\widehat{\mathrm{R}}_{\beta_{\mathrm{J_k}}} = \widehat{\mathrm{C}}_{\mathrm{J_k}} \cup \widehat{\mathrm{C}}_{\mathrm{J_{k+1}}} \cup \left(\bigcup_{\mathrm{m}\in \mathrm{J}^{\mathrm{E}}_{\beta_{\mathrm{J_k}}}} \widehat{\mathrm{C}}_{\mathrm{m}} \right) \quad \forall\, \mathrm{J_k} \in \mathrm{J}\setminus\{\mathrm{J_p}\}\ ,$$

where $\mathrm{J}^{\mathrm{E}}_{\beta_{\mathrm{J_k}}} = \{\mathrm{m}\in\mathrm{J}^{\mathrm{E}} \mid \mathrm{J_k} < \mathrm{m} < \mathrm{J_{k+1}}\}$ *and* $\widehat{\mathrm{C}}_{\mathrm{k}} \cap \widehat{\mathrm{C}}_{\mathrm{n}} = \varnothing\ \forall (\mathrm{k},\mathrm{n}) \in (\mathrm{J}\cup\mathrm{J}^{\mathrm{E}})^2,\ \mathrm{k}\neq\mathrm{n}.$

Example 4 *[Ex.3, cont.] We had* $\mathrm{J}=\{0,2,4,6,7\}$, $(\beta_{\mathrm{J_0}}=9, \beta_{\mathrm{J_1}}=8, \beta_{\mathrm{J_2}}=5,\ \beta_{\mathrm{J_3}}=4, \beta_{\mathrm{J_4}}=0)$, *and* $\mathrm{J}^{\mathrm{E}}_2 = \{3\}$ *with* $\beta_{\mathrm{J}^{\mathrm{E}}_2}=8$ *and* $\mathrm{J}^{\mathrm{E}}_{\mathrm{k}} = \varnothing$ *otherwise. By Theorems 6 and 7*

$$\{\widehat{\mathrm{R}}_\beta | \beta > 9\} = \widehat{\mathrm{C}}_0 \quad \{\widehat{\mathrm{R}}_\beta | \beta \in (8,9)\} = \widehat{\mathrm{C}}_2 \quad \{\widehat{\mathrm{R}}_\beta | \beta \in (5,8)\} = \widehat{\mathrm{C}}_4 \quad \{\widehat{\mathrm{R}}_\beta | \beta \in (4,5)\} = \widehat{\mathrm{C}}_6 \quad \{\widehat{\mathrm{R}}_\beta | \beta \in (0,4)\} = \widehat{\mathrm{C}}_7$$

and $\widehat{\mathrm{R}}_{\beta=9} = \widehat{\mathrm{C}}_0 \cup \widehat{\mathrm{C}}_2 \quad \widehat{\mathrm{R}}_{\beta=8} = \widehat{\mathrm{C}}_2 \cup \widehat{\mathrm{C}}_3 \cup \widehat{\mathrm{C}}_4 \quad \widehat{\mathrm{R}}_{\beta=5} = \widehat{\mathrm{C}}_4 \cup \widehat{\mathrm{C}}_6 \quad \widehat{\mathrm{R}}_{\beta=4} = \widehat{\mathrm{C}}_6 \cup \widehat{\mathrm{C}}_7$.

A partial equivalence between problems $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$ always exists.

For the $\sharp\,\mathrm{J}-1$ isolated values $\left\{\beta_{\mathrm{k}} \mid \mathrm{k}\in\mathrm{J}\setminus\{\mathsf{L}\}\right\}$ problem $(\mathcal{R}_\beta)$ has normally two optimal sets (Proposition 2).

4.2. Quasi-complete equivalence

Lemma **5** *Let* J *be as in Definition 4. Then the following hold:*

(a) *If the sequence* $\{\beta_{\mathrm{k}}\}_{\mathrm{k}=0}^{\mathsf{L}}$ *in Definition 3 is strictly decreasing, then its entries read as*

$$\boxed{\beta_k = \mathrm{c}_{\mathrm{k}} - \mathrm{c}_{\mathrm{k}+1} \quad \forall\, \mathrm{k} \in \mathbb{I}_{\mathsf{L}-1}^{0} \quad \text{and} \quad \beta_{\mathsf{L}} = 0, \quad \beta_{-1} := \beta_0^{\mathrm{U}} = +\infty\ .} \tag{1}$$

(b) *If the sequence* $\{\beta_{\mathrm{k}}\}_{\mathrm{k}=0}^{\mathsf{L}}$ *in* (1) *is strictly decreasing then* $\mathrm{J} = \mathbb{I}_{\mathsf{L}}^{0}$.

Theorem **8** *Let* $\{\beta_{\mathrm{k}}\}_{\mathrm{k}=0}^{\mathsf{L}}$ *in* (1) *be strictly decreasing. Then*

$$\left\{\, \widehat{\mathrm{R}}_{\beta} \;\middle|\; \beta \in (\,\beta_{\mathrm{k}}, \beta_{\mathrm{k}-1}) \,\right\} = \widehat{\mathrm{C}}_{\mathrm{k}} \quad \forall\, \mathrm{k} \in \mathbb{I}_{\mathsf{L}}^{0}$$

$$\widehat{\mathrm{R}}_{\beta_{\mathrm{k}}} = \widehat{\mathrm{C}}_{\mathrm{k}} \cup \widehat{\mathrm{C}}_{\mathrm{k}+1} \quad \textit{with} \quad \widehat{\mathrm{C}}_{\mathrm{k}} \cap \widehat{\mathrm{C}}_{\mathrm{k}+1} = \varnothing \quad \forall\, \mathrm{k} \in \mathbb{I}_{\mathsf{L}-1}^{0}\ .$$

$\{\beta_{\mathrm{k}}\}_{\mathrm{k}=0}^{\mathsf{L}}$ in (1) strictly decreasing means that $\mathrm{c}_{\mathrm{k}-1} - \mathrm{c}_{\mathrm{k}} > \mathrm{c}_{\mathrm{k}} - \mathrm{c}_{\mathrm{k}+1}$, $\forall\, \mathrm{k} \in \mathbb{I}_{\mathsf{L}-1}$.
Let $\overline{u}$, $\widehat{u}$ and $\widetilde{u}$ be optimal solutions of problems $(\mathcal{C}_{\mathrm{k}-1})$, $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{C}_{\mathrm{k}+1})$, resp.
Denote $\overline{\sigma} := \mathrm{supp}\,(\overline{u})$, $\widehat{u} := \mathrm{supp}\,(\widehat{u})$ and $\widetilde{\sigma} := \mathrm{supp}\,(\widetilde{u})$.
The condition on c_{k}'s reds as $d^{\mathsf{T}}\,(\mathrm{Proj}(A_{\widehat{\sigma}}) - \mathrm{Proj}(A_{\overline{\sigma}}))\,d > d^{\mathsf{T}}\,(\mathrm{Proj}(A_{\widetilde{\sigma}}) - \mathrm{Proj}(A_{\widehat{\sigma}}))\,d > 0$.

Mid-way scenarios...

Outline

1. Two optimization problems involving the ℓ_0 pseudo norm
2. Joint optimality conditions for $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$
3. Parameter values for equality between optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$
4. Equivalences between the global minimizers of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$
   - Partial equivalence
   - Quasi-complete equivalence
5. On the optimal values of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$
6. Cardinality of the optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$
   - Uniqueness of the optimal solutions of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$
7. Numerical tests
8. Conclusions and future directions

5. On the optimal values of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$

$$\Omega_{\mathrm{k}} := \left\{ \omega \subset \mathbb{I}_{\mathsf{N}} \mid \sharp\omega = \mathrm{k} = \mathrm{rank}\,(A_\omega) \right\}.$$

$$\mathrm{E}_{\mathrm{k}} := \bigcup_{\omega\in\Omega_{\mathrm{k}}} \mathrm{range}\,(A_\omega)^\perp \quad \text{and} \quad \mathrm{G}_{\mathrm{k}} := \bigcup_{\omega\in\Omega_{\mathrm{k}}} \mathrm{range}\,(A_\omega)\ .$$

$\mathrm{E}_0 = \mathrm{G}_{\mathsf{M}} = \mathbb{R}^{\mathsf{M}}$ and $\mathrm{E}_{\mathsf{M}} = \mathrm{G}_0 = \{0\}$ by H1.

Proposition **5** *Let* $\mathsf{L}' \leqslant \mathsf{M}$ *be arbitrarily fixed.*

- $\mathrm{c}_{\mathrm{k}} > 0 \quad \forall\, \mathrm{k} \leqslant \mathsf{L}' - 1 \quad \Longleftrightarrow \quad d \in \mathbb{R}^{\mathsf{M}} \setminus \mathrm{G}_{\mathsf{L}'-1}$;
- $d \in \mathbb{R}^{\mathsf{M}} \setminus (\mathrm{E}_2 \cup \mathrm{G}_{\mathsf{L}'-1}) \quad \Longrightarrow \quad \mathrm{c}_{\mathrm{k}-1} > \mathrm{c}_{\mathrm{k}} \quad \forall\, \mathrm{k} \in \mathbb{I}_{\mathsf{L}'}$.

E_2 and $\mathrm{G}_{\mathsf{M}-1}$ are finite unions of vector subspaces of dimensions $\mathsf{M}-2$ and $\mathsf{M}-1$, respectively. Hence, $d \in \mathbb{R}^{\mathsf{M}} \setminus (\mathrm{E}_2 \cup \mathrm{G}_{\mathsf{M}-1})$ is a generic property.

$\Longrightarrow \quad \{\mathrm{c}_{\mathrm{k}}\}_{\mathrm{k}=0}^{\mathsf{M}}$ is strictly decreasing and $\mathsf{L} = \mathsf{M}$ generically.

Proposition 6 $\quad \mathrm{k} \in \mathbb{I}_{\mathsf{L}}^{0} \quad \Longrightarrow \quad \mathcal{F}_\beta(\widehat{u}) = \mathrm{c}_{\mathrm{k}} + \beta\,\mathrm{k} \quad \forall\, \widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{k}}\ .$

By Theorem 3 $\widehat{\mathrm{R}}_\beta \subset \widehat{\mathrm{C}} = \bigcup_{\mathrm{k}=0}^{\mathsf{L}} \widehat{\mathrm{C}}_{\mathrm{k}} \quad \Longrightarrow \quad$ the optimal value of problem $(\mathcal{R}_\beta)$ reads as

$$r_\beta = \min \left\{ \mathrm{c}_{\mathrm{k}} + \beta\,\mathrm{k} \mid \mathrm{k} \in \mathbb{I}_{\mathsf{L}}^{0} \right\}\ .$$

Corollary 2 *The application* $\beta \mapsto r_\beta \ : (0, +\infty) \to \mathbb{R}$ *fulfills*

- $$\begin{cases} r_\beta = \mathrm{c}_{\mathrm{J}_{\mathrm{k}}} + \beta\,\mathrm{J}_{\mathrm{k}} \\ \quad\ = \mathcal{F}_\beta(\widehat{u}) \ \ \forall\, \widehat{u} \in \widehat{\mathrm{C}}_{\mathrm{J}_{\mathrm{k}}} \end{cases} \textit{if and only if } \ \beta \in \begin{cases} [\beta_{\mathrm{J}_0}, +\infty) & \text{for} \quad \mathrm{J}_0 \equiv 0 \\ \left[\beta_{\mathrm{J}_{\mathrm{k}}}, \beta_{\mathrm{J}_{\mathrm{k}-1}}\right] & \text{for} \quad \mathrm{J}_{\mathrm{k}} \in \mathrm{J} \setminus \{0, \mathsf{L}\} \\ \left(0, \beta_{\mathrm{J}_{\mathrm{p}-1}}\right] & \text{for} \quad \mathrm{J}_{\mathrm{p}} \equiv \mathsf{L} \end{cases}$$
- $\beta \mapsto r_\beta$ *is continuous and concave.*
- $r_{\beta_{\mathrm{J}_{\mathrm{k}-1}}} > r_{\beta_{\mathrm{J}_{\mathrm{k}}}} \ \ \forall\, \mathrm{J}_{\mathrm{k}} \in \mathrm{J}\,, \quad r_{\beta_{\mathrm{J}_0}} = \mathrm{c}_{\mathrm{J}_0} = r_\beta \ \ \forall\, \beta \geqslant \beta_{\mathrm{J}_0}$ *and* $r_{\beta_{\mathrm{J}_0}} > r_\beta \ \ \forall\, \beta < \beta_{\mathrm{J}_0}.$

$\beta \mapsto r_\beta$ is affine increasing on each interval $\left(\beta_{\mathrm{J}_{\mathrm{k}}}, \beta_{\mathrm{J}_{\mathrm{k}-1}}\right)$ with upward kinks at $\beta_{\mathrm{J}_{\mathrm{k}}}$ for any $\mathrm{J}_{\mathrm{k}} \in \mathrm{J} \setminus \{\mathsf{L}\}$ and bounded by c_0.

Example **5** [Cont. of Example 3] From Corollary 2, $\beta \mapsto r_\beta$ is given by

$$
\begin{array}{lll}
\beta \in (0,4] & r_\beta = \mathrm{c}_7 + 7\beta = 7\beta & r_{\beta=4} = 28 \\
\beta \in [4,5] & r_\beta = \mathrm{c}_6 + 6\beta = 4 + 6\beta & r_{\beta=5} = 34 \\
\beta \in [5,8] & r_\beta = \mathrm{c}_4 + 4\beta = 14 + 4\beta & r_{\beta=8} = 46 \\
\beta \in [8,9] & r_\beta = \mathrm{c}_2 + 2\beta = 30 + 2\beta & r_{\beta=9} = 48 \\
\beta \in [9,+\infty) & r_\beta = \mathrm{c}_0 + 0\beta = 48 &
\end{array}
$$

affine expressions

Outline

1. Two optimization problems involving the ℓ_0 pseudo norm

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

3. Parameter values for equality between optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$

4. Equivalences between the global minimizers of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$

   – Partial equivalence
   – Quasi-complete equivalence

5. On the optimal values of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$

6. Cardinality of the optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$

   – Uniqueness of the optimal solutions of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$

7. Numerical tests

8. Conclusions and future directions

6. Cardinality of the optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$

For any $\beta > 0$ and $\mathrm{k} \in \mathbb{I}^0_{\mathsf{L}}$ the optimal sets of problems $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$ are composed out of a certain finite number of isolated (hence strict) minimizers.

6.1. Uniqueness of the optimal solutions of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$

$\mathrm{k} \leqslant \min\{\mathsf{L}, \mathsf{M}-1\}$ and $(\widehat{u}, \widetilde{u}) \in (\widehat{\mathrm{C}}_{\mathrm{k}})^2$, $\widehat{u} \neq \widetilde{u}$. Then $\widehat{\sigma} := \mathrm{supp}\,(\widehat{u})$, $\widetilde{\sigma} := \mathrm{supp}\,(\widetilde{u})$ are in $(\Omega_{\mathrm{k}})^2$.

$$\mathrm{c}_{\mathrm{k}} = \|A_{\widehat{\sigma}}\widehat{u}_{\widehat{\sigma}} - d\|^2 = \|A_{\widetilde{\sigma}}\widetilde{u}_{\widetilde{\sigma}} - d\|^2 \quad \text{where} \quad \widehat{\sigma} \neq \widetilde{\sigma} \ .$$

Π_ω the orthogonal projector onto $\mathrm{range}\,(A_\omega)$

$$\|A_{\widehat{\sigma}}\widehat{u}_{\widehat{\sigma}} - d\|^2 - \|A_{\widetilde{\sigma}}\widetilde{u}_{\widetilde{\sigma}} - d\|^2 = d^{\mathsf{T}}\,(\Pi_{\widetilde{\sigma}} - \Pi_{\widehat{\sigma}})\,d = 0 \ .$$

$\mathrm{H}^\star$ For $\mathsf{K} \leqslant \min\{\mathsf{M}-1, \mathsf{L}\}$ fixed, $A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ obeys $\quad \Pi_\omega \neq \Pi_{\overline{\omega}} \ \ \forall(\omega, \overline{\omega}) \in \Omega^2_{\mathrm{k}} \ \ \omega \neq \overline{\omega} \quad \forall\, \mathrm{k} \in \mathbb{I}_{\mathsf{K}}$.

$\mathrm{H}^\star$ is a *generic property* of all matrices in $\mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ [M. N., SIIMS 2013].

$$\Delta_{\mathsf{K}} := \bigcup_{\mathrm{k}=1}^{\mathsf{K}} \bigcup_{(\omega,\overline{\omega})\in(\Omega_{\mathrm{k}})^2} \left\{ g \in \mathbb{R}^{\mathsf{M}} \ \middle|\ \omega \neq \overline{\omega} \ \ \text{and} \ \ g \in \ker\,(\Pi_{\overline{\omega}} - \Pi_\omega) \right\}$$

$\dim(\Delta_{\mathsf{K}}) \leqslant \mathsf{M} - 1$, hence $d \in \mathbb{R}^{\mathsf{M}} \setminus \Delta_{\mathsf{K}}$ *generically*.

$$\mathrm{H}^\star \quad \text{and} \quad d \in \mathbb{R}^{\mathsf{M}} \setminus \Delta_{\mathsf{K}} \quad \Longrightarrow \quad (\mathcal{C}_{\mathrm{k}}) \text{ for } \mathrm{k} \in \mathbb{I}_{\mathsf{K}} \text{ has a unique optimal solution.}$$

$\mathsf{K}' := \max\{\mathrm{k} \in \mathrm{J} \mid \mathrm{k} \leqslant \mathsf{K}\} \quad \Rightarrow \quad (\mathcal{R}_\beta)$ has a unique global minimizer $\forall\beta \in (\beta_{\mathsf{K}'}, +\infty) \setminus \{\beta_{\mathrm{k}}\}_{\mathrm{k}\in\mathrm{J}}$

Outline

1. Two optimization problems involving the ℓ_0 pseudo norm

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

3. Parameter values for equality between optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$

4. Equivalences between the global minimizers of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$

   – Partial equivalence
   – Quasi-complete equivalence

5. On the optimal values of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$

6. Cardinality of the optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$

   – Uniqueness of the optimal solutions of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$

7. Numerical tests

   – Monte Carlo on $\{\beta_{\mathrm{k}}\}$ with 10^5 tests for $(\mathsf{M}, \mathsf{N}) = (5, 10)$

   – Tests on (partial) equivalence with selected matrix data

8. Conclusions and future directions

7. Numerical tests

Two kind of experiments using matrices $A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ for $(\mathsf{M},\mathsf{N}) = (5, 10)$, original vectors $u^{\mathrm{o}} \in \mathbb{R}^{\mathsf{N}}$ and data samples $d = Au^{\mathrm{o}}(+\text{noise})$ with *two different goals*:

- to roughly see the behaviour of the parameters β_{k} in Definition 3 ;
- to verify and illustrate our theoretical findings.

All results were calculated using an exhaustive combinatorial search.

7.1. Monte Carlo experiments on $\{\beta_{\mathrm{k}}\}$ with 10^5 tests for $(\mathsf{M},\mathsf{N}) = (5, 10)$

Two experiments, each one composed of $\mathbf{10^5}$ trials with $A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ for $(\mathsf{M},\mathsf{N}) = (5,\ 10)$

In each trial:

- an "original" $u^{\mathrm{o}} \in \mathbb{R}^{\mathsf{N}}$, random support on $\{1, \ldots, \mathsf{N}\}$ with $\|u^{\mathrm{o}}\|_0 \leqslant \mathsf{M} - 1 = 4$.
- the coefficients of A and $u^{\mathrm{o}}_{\mathrm{supp}\,(u^{\mathrm{o}})}$ – i.i.d.
- $d = Au^{\mathrm{o}} +$ i.i.d. centered Gaussian noise.
- compute the optimal values $\{\mathrm{c}_{\mathrm{k}}\}$ and then compute $(\beta_{\mathrm{k}}, \beta_{\mathrm{k}}^{\mathrm{U}})$ by Definition 3.

Two different distributions for A and $u^{o}_{\mathrm{supp}\,(u^{o})}$

- ◇ **Experiment $\mathcal{N}(0,10)$.** $A(i,j)$ and $u^{o}_{\mathrm{supp}\,(u^{o})} \sim \mathcal{N}(0,10)$. Support length $\sharp\,\mathrm{supp}\,(u^{o}) \in \{1,\ldots,4\}$, mean $= 3.8$. SNR of d in $[10.1,\ 61.1]$, mean $= 33.75$ dB.

- ◇ **Experiment Uni $[0,10]$.** $A(i,j)$ and $u^{o}_{\mathrm{supp}\,(u^{o})} \sim$ uniform on $[0,10]$. $\sharp\,\mathrm{supp}\,(u^{o}) \in \{1,\ldots,4\}$, mean $= 3.8$. SNR of d in $[20,\ 55]$, mean $= 28.95$ dB.

$$N_{\mathrm{k}} := \sharp\,\left\{\mathrm{k} \in \mathbb{I}^0_{\mathsf{M}} \mid \beta_{\mathrm{k}} > \beta_{\mathrm{k}-1}\right\} \ .$$

Table 1: Results on the behaviour of $\{\beta_{\mathrm{k}}\}$ in Definition 3 for two experiments, each one composed of 10^5 random trials. For $\mathrm{k} \geqslant 3$ we have found $N_{\mathrm{k}} = 0$.

	$\beta_{\mathrm{k}} < \beta_{\mathrm{k}-1}, \forall\, \mathrm{k} \in \mathbb{I}^0_{\mathsf{M}}$	$N_{\mathrm{k}} = 1$	$N_{\mathrm{k}} = 2$	mean(SNR)	mean($\|u^{o}\|_0$)
$\mathcal{N}(0,10)$	**93.681 %**	6.254 %	0.065 %	33.75	3.7922
Uni $[0,10]$	**98.783 %**	1.216 %	0.001 %	28.95	3.7936

Observations:

- $\mathsf{L} = \mathsf{M}$ in each trial (Remainder: $\mathsf{L} := \min\left\{\mathrm{k} \in \mathbb{I}_{\mathsf{N}} \mid \mathrm{c}_{\mathrm{k}} = 0\right\}$);
- $\{\mathrm{c}_{\mathrm{k}}\}_{\mathrm{k}=0}^{M}$ was always strictly decreasing (see Proposition 5);
- $\beta_{\mathrm{k}} \neq \beta^{\mathrm{U}}_{\mathrm{k}}$ in each trial (see Proposition 3), so $\mathrm{J}^{\mathrm{E}} = \varnothing$;
- For every A there were d so that $\{\beta_{\mathrm{k}} := \mathrm{c}_{\mathrm{k}} - \mathrm{c}_{\mathrm{k}-1}\}_{\mathrm{k}=0}^{\mathsf{L}}$ was strictly decreasing

6.2. Tests on quasi-equivalence with a selected matrix and selected data

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

$$u^{\mathrm{o}} = \begin{pmatrix} 0 & \mathbf{4} & 0 & 0 & 0 & \mathbf{9} & 0 & 0 & \mathbf{3} & 0 \end{pmatrix}^{\mathsf{T}}.$$

$A(i,j)$ nearly normal distribution with variance 10 and $\mathrm{rank}(A) = \mathsf{M} = 5$

Problem $(\mathcal{C}_{\mathsf{M}})$ has $\sharp\,\Omega_{\mathsf{M}} = 252$ optimal solutions; none of them is shown.

Since $\beta_0 < \beta_0^{\mathrm{U}} = +\infty$, in all cases $\widehat{\mathrm{C}}_0 = \left\{ \widehat{\mathrm{R}}_\beta \,|\, \beta > \beta_0 \right\}$ by Theorem 5.

We selected a couple (A, u^{o}) so that β_{k} are seldom strictly decreasing compared to Tab. 1.

Table 2: 10^5 trials where $d = Au^{\mathrm{o}} +$ i.i.d. centered Gaussian noise.

	$\beta_{\mathrm{k}} < \beta_{\mathrm{k}-1}, \forall\, \mathrm{k} \in \mathbb{I}_{\mathsf{M}}^0$	$N_{\mathrm{k}} = 1$	$N_{\mathrm{k}} = 2$	mean(SNR)	mean($\|u^{\mathrm{o}}\|_0$)
u^{o} in (2)	29.41 %	70.59 %	0 %	36.25	3

Noise-free data

$$d = Au^{\mathrm{o}} = \begin{pmatrix} 64.45 & -171.09 & 114.13 & 153.32 & -38.93 \end{pmatrix}^{\mathsf{T}} .$$

$\widehat{u} = u^{\mathrm{o}}$ is an optimal solution to problems $(\mathcal{C}_{\mathrm{k}})$ with $\mathrm{c_k} = 0$ for $\mathrm{k} \in \{3, 4, 5\}$ and $\mathsf{L} = 3$.

$$\boldsymbol{\beta_3} = 0 < \beta_3^{\mathrm{U}} = \boldsymbol{\beta_1} = 3872.46 < \beta_1^{\mathrm{U}} = \boldsymbol{\beta_0} = 63729 \quad \text{and} \quad \beta_2 = 3968 > \beta_2^{\mathrm{U}} = 3776.82 \ .$$

$\mathrm{J} = \{0, 1, 3\}$

k	$\mathrm{c_k}$	$\widehat{\mathrm{C}}_{\mathrm{k}}$ = the optimal solution of $(\mathcal{C}_{\mathrm{k}})$, singleton										$\widehat{\mathrm{C}}_{\mathrm{k}} = \widehat{\mathrm{R}}_{\beta}$
3	0	0	**4**	0	0	0	**9**	0	0	**3**	0	$\beta \in (\beta_3, \beta_1)$
2	3968	0	**3.25**	0	0	0	**9.29**	0	0	0	0	**no**
1	7745	0	0	0	0	0	**11.76**	0	0	0	0	$\beta \in (\beta_1, \beta_0)$
0	71474	0	0	0	0	0	0	0	0	0	0	$\beta > \beta_0$

Noisy data 1. Nearly normal, centered, i.i.d. noise and SNR$= 32.32$ dB:

$$d = \begin{pmatrix} 69.13 & -171.95 & 113.74 & 150.27 & -36.09 \end{pmatrix}^{\mathsf{T}} .$$

$$\boldsymbol{\beta_5} = 0 < \beta_5^{\mathrm{U}} = \boldsymbol{\beta_4} = 0.068 < \beta_4^{\mathrm{U}} = \boldsymbol{\beta_3} = 36.25 < \beta_3^{\mathrm{U}} = \boldsymbol{\beta_1} = 3987.68 < \beta_1^{\mathrm{U}} = \boldsymbol{\beta_0} = 63154 ,$$

while $\beta_2 = 4002.83 > \beta_2^{\mathrm{U}} = 3972.54$. Hence, $\mathrm{J} = \mathbb{I}_5^0 \setminus \{2\}$

k	$\mathrm{c_k}$	$\widehat{\mathrm{C}}_\mathrm{k} =$ the optimal solution of $(\mathcal{C}_\mathrm{k})$, singleton										$\widehat{\mathrm{C}}_\mathrm{k} = \widehat{\mathrm{R}}_\beta$
4	0.068	0	**4.40**	0	0	0	**8.71**	**0.54**	0	**2.95**	0	$\beta \in (\beta_4, \beta_3)$
3	36.3141	0	**4.09**	0	0	0	**8.88**	0	0	**3.01**	0	$\beta \in (\beta_3, \beta_1)$
2	4039	0	**3.33**	0	0	0	**9.17**	0	0	0	0	**no**
1	8011.68	0	0	0	0	0	**11.71**	0	0	0	0	$\beta \in (\beta_1, \beta_0)$
0	71166	0	0	0	0	0	0	0	0	0	0	$\beta > \beta_0$

Noisy data 2. The noise is nearly normal, centered, i.i.d., SNR$= 25.74$ dB:

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

$$\beta_0 = 60287 \quad \beta_1 = 3825 \quad \beta_2 = 3037.1 \quad \beta_3 = 72.734 \quad \beta_4 = 0.0259 \quad \beta_5 = 0 .$$

$\{\beta_{\mathrm{k}}\}$ is strictly decreasing and hence $\beta_{\mathrm{k}} = \mathrm{c}_{\mathrm{k}} - \mathrm{c}_{\mathrm{k}-1}$

$(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$ are quasi-completely equivalent.

k	c_{k}	$\widehat{\mathrm{C}}_{\mathrm{k}} =$ the optimal solution of $(\mathcal{C}_{\mathrm{k}})$, singleton										$\widehat{\mathrm{C}}_{\mathrm{k}} = \widehat{\mathrm{R}}_\beta$
4	0.0259	0	**8.54**	0	0	**4.59**	**4.90**	**2.73**	0	0	0	$\beta \in (\beta_4, \beta_3)$
3	72.7601	0	**3.93**	0	0	0	**8.70**	0	0	**2.63**	0	$\beta \in (\beta_3, \beta_2)$
2	3109.86	0	**3.27**	0	0	0	**8.95**	0	0	0	0	$\beta \in (\beta_2, \beta_1)$
1	6934.85	0	0	0	0	0	**11.44**	0	0	0	0	$\beta \in (\beta_1, \beta_0)$
0	67222	0	0	0	0	0	0	0	0	0	0	$\beta > \beta_0$

Outline

1. Two optimization problems involving the ℓ_0 pseudo norm
2. Joint optimality conditions for $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$
3. Parameter values for equality between optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$
4. Equivalences between the global minimizers of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$
   - Partial equivalence
   - Quasi-complete equivalence
5. On the optimal values of $(\mathcal{C}_{\mathrm{k}})$ and $(\mathcal{R}_\beta)$
6. Cardinality of the optimal sets of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$
   - Uniqueness of the optimal solutions of $(\mathcal{C}_{\mathrm{k}})$ and of $(\mathcal{R}_\beta)$
7. Numerical tests
8. Conclusions and future directions

8. Conclusions and open questions

- The main equivalence result in a nutshell:

$$\begin{array}{ccccccccccc} \widehat{\mathrm{R}}_\beta = & \widehat{\mathrm{C}}_{\mathsf{L}} & & \widehat{\mathrm{C}}_{\mathrm{J}_{\mathrm{k}+1}} & & \overbrace{\widehat{\mathrm{C}}_{\mathrm{J}_\mathrm{k}} \cup \widehat{\mathrm{C}}_{\mathrm{J}_{\mathrm{k}+1}} \cup \widehat{\mathrm{C}}^0_{\mathrm{J}_\mathrm{k}}}^{} \\ & & & & & \uparrow & & \widehat{\mathrm{C}}_{\mathrm{J}_\mathrm{k}} & & \widehat{\mathrm{C}}_0 \\ 0 = & \beta_{\mathsf{L}} & < & \ldots < \beta_{\mathrm{J}_{\mathrm{k}+1}} & < & \boldsymbol{\beta}_{\mathbf{J}_\mathbf{k}} & < & \beta_{\mathrm{J}_{\mathrm{k}-1}} < \ldots < \beta_{\mathrm{J}_0} & < & \infty \end{array}$$

- The agreement between the optimal sets of problems $(\mathcal{C}_\mathrm{k})$ and $(\mathcal{R}_\beta)$ is driven by the critical parameters $\{\beta_\mathrm{k}\}_{\mathrm{k}=0}^{\mathsf{L}}$ which depend only on the optimal values c_k of problem $(\mathcal{C}_\mathrm{k})$.

- Our comparative results clarify a proper choice between models $(\mathcal{C}_\mathrm{k})$ and $(\mathcal{R}_\beta)$ in applications. If one needs solutions with a fixed number of nonzero entries, $(\mathcal{C}_\mathrm{k})$ is the best choice. If only information on the perturbations is available, $(\mathcal{R}_\beta)$ is a more flexible model.

- If one can solve problem $(\mathcal{C}_\mathrm{k})$ for all k, the global minimizers of problem $(\mathcal{R}_\beta)$ are immediate.

- Our detailed results can give rise to innovative algorithms.

- The degree of partial equivalence depends on the distribution of the coefficients of A and d.

- By specifying a class of matrices A and assumptions on data d, one can infer statistical knowledge on the optimal values $\mathrm{c_k}$ of problems $(\mathcal{C}_\mathrm{k})$ and thus on the critical parameters $\{\beta_\mathrm{k}\}$. Promising theoretical and practical results can be expected.

- A related open question is to know if the optimal solutions of $(\mathcal{R}_\beta)$ are able to eliminate some meaningless solutions of $(\mathcal{C}_\mathrm{k})$.

- Extensions to analysis type penalties $\|Du\|_0$, to low rank matrix recovery, etc., are important.

- Other important results concern algorithms that are known to converge to local minimizers.

Remark **3** *Problem* $(\mathcal{R}_\beta)$ *(for some* $\beta > 0$*) and problems* $(\mathcal{C}_\mathrm{k})$ *for* $\mathrm{k} \in \{0, 1, \ldots, \mathrm{M}\}$ *have the same sets of (strict) local minimizers.*

Thank you for the attention.

Thanks to Zuhair Nashed for the invitation