A New Condition for Exact ℓ_1 Recovery

Charles Dossal[1] Gabriel Peyré[2]

[1]Labag, Université Bordeaux 1

[2]CEREMADE Université Paris Dauphine

2007 AMS Von Neumann Symposium

Sparse Representations and High Dimensional Geometry

The Problem

Problem

- Linear measurements: $y = Ax \in \mathbb{R}^n$, with $x \in \mathbb{R}^p$, $A \in \mathbb{R}^{n \times p}$.
- Problem: recovering x from y when $p > n$, ill-posed inversion.
- Prior: x is sparse $\Leftrightarrow$ $\|x\|_0$ is small.

Applications

- Seismic imaging: x is a collection of diracs, A is a blurring.
- Compressed sensing: x are wavelet coefficients, $A = \Phi U$, where U is the inverse wavelet transform matrix and Φ a random matrix of mesurements.
- ...

ℓ_1 Minimization

Optimization for recovery

- Strict sparsity: $\min \|x\|_0$ subject to $Ax = y$ $\quad P_0(y)$
- Convexification: $\min \|x\|_1$ subject to $Ax = y$ $\quad P_1(y)$
- Relaxation: $\min \frac{1}{2} \|y - Ax\|_2^2 + \gamma \|x\|_1$ $\quad P_1(y, \gamma)$

Identifiability

- x is identifiable $\Leftrightarrow$ x is the unique solution of $P_1(Ax)$
- Conditions on x and A ensuring x is identifiable ?

Notations and Definitions

Notations

- Support: $I(x) = \{i,\, x(i) \neq 0\}$.
- Restriction: $\bar{x} = (x(i))_{i \in I(x)}$.
- Active matrix: $A_I = (a_i)_{i \in I(x)} \Rightarrow Ax = A_I \bar{x}$.
- Pseudo inverse: $B^\dagger = (B^t B)^{-1} B^t$.
- $d_0 = d_0(x) = A_I^{\dagger t} \mathrm{sign}\,(\bar{x})$

Non-degeneracy of A

- $\forall x$ such that A_I has full rank,

$$\forall j \notin I \quad |\langle a_j, d_0(x) \rangle| \neq 1 \qquad \text{(UC)}$$

- Weak condition: almost surely satisfied for most random matrices.

State of the Art (1)

Sparsity criterion: Coherence

- Coherence: $C(A) = \max_{i \neq j} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2}$
- $\|x\|_0 < \frac{1}{2}(1 + \frac{1}{C(A)}) \Rightarrow$ identifiability.

Sparsity criterion: Compressed sensing

- RIH: $\forall I$ with $|I| \leqslant S$,

$$\forall x \in \mathbb{R}^{|I|},\ (1 - \delta_S)\,\|x\|_2^2 \leqslant \|A_I x\|_2^2 \leqslant (1 + \delta_S)\,\|x\|_2^2 \qquad (1)$$

- If $\delta_S + \delta_{2S} + \delta_{3S} < 1$ or if $\delta_{2S} < \sqrt{2} - 1$ then $\|x\|_0 \leqslant S \Rightarrow$ identifiability.
- Robutsness to a bounded noise.
- Some random matrices: $S \leqslant \frac{Cn}{\log_2(p/n)} \Rightarrow A$ satisfies RIH.
- CS bounds are uniform an then are often pessimistic.

State of the Art (2)

Support criterion: Exact Recovery Coefficient (ERC)

- $\mathrm{ERC}(I) = 1 - \max_{j \notin I} \left\| A_I^{\dagger} a_j \right\|_1$.
- $\mathrm{ERC}(I) > 0 \Rightarrow$ then $I(x) \subset I \Rightarrow$ the unique solution $x(\gamma)$ of $\mathrm{P}_1(Ax, \gamma)$ tends to x when $\gamma \to 0$.

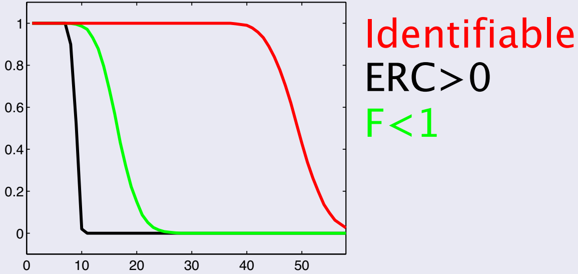
Sign criterion

- $F(x) = \max_{j \notin I} |\langle a_j, d_0 \rangle|$ A_I has full rank and $d_0 = A_I^{\dagger t} \operatorname{sign}(\bar{x})$.
- $\mathcal{F} = \{x \backslash F(x) < 1\}$
- $x \in \mathcal{F} \Rightarrow$ the unique solution $x(\gamma)$ of $\mathrm{P}_1(Ax, \gamma)$ tends to x when $\gamma \to 0$.

Many identifiable vectors

Criterion Efficiency

- Random matrices A with $n = 200$ and $p = 1000$.
- Proportion of vectors satisfying different criteria.

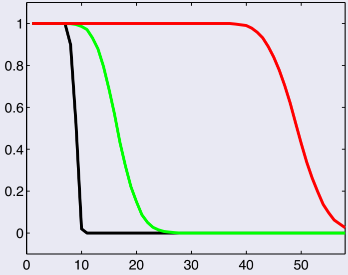


- $ERC(I(x)) > 0$ or $x \in \mathcal{F} \Rightarrow$ x is identifiable.

Many identifiable vectors

Criterion Efficiency

- Random matrices A with $n = 200$ and $p = 1000$.
- Proportion of vectors satisfying different criteria.



Identifiable
ERC>0
F<1

- $ERC(I(x)) > 0$ or $x \in \mathcal{F} \Rightarrow$ x is identifiable.
- We want to caracterize all identifiable vectors

Contributions

A necessary and sufficient condition for identifiability

- If $x \in \bar{\mathcal{F}}$, then x is identifiable.
- If A satisfies condition (UC) then x identifiable implies $x \in \bar{\mathcal{F}}$.

Analysis of the recovery map

If A satifies (UC), the recovery map φ is uniformly Lipschitz

$$\varphi \quad \begin{cases} \mathrm{Im}\,(A) & \longrightarrow & \bar{\mathcal{F}} \\ y & \longmapsto & x \quad \text{solution of problem } \mathrm{P}_1(y) \end{cases}$$

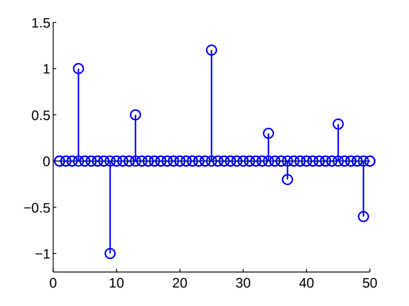
Numerical estimation

- An algorithm to estimate if $x \in \bar{\mathcal{F}}$ or not.
- Numerical simulations.

Sets $\mathcal{F}$ and $\bar{\mathcal{F}}$

$\mathcal{F}$ a union of cones

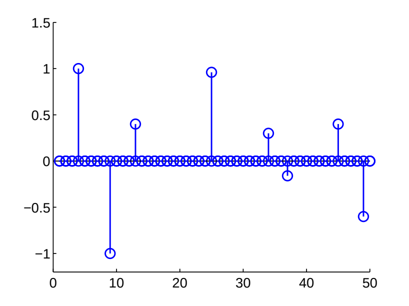
- $\mathcal{F} = \left\{ x \backslash F(x) = \max_{j \notin I} |\langle a_j, A_I^{\dagger t} \mathrm{sign}\,(\bar{x}) \rangle| < 1 \right\}$
- $F(x)$ depends only on the support and sign of x.
- Let's consider $x_1 \in \mathcal{F}$, $F(x_1) = 0.9$.



Sets $\mathcal{F}$ and $\bar{\mathcal{F}}$

$\mathcal{F}$ a union of cones

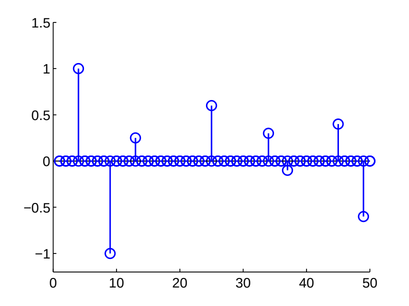
- $\mathcal{F} = \left\{ x \backslash F(x) = \max_{j \notin I} |\langle a_j, A_I^{\dagger t} \mathrm{sign}\,(\bar{x}) \rangle| < 1 \right\}$
- $F(x)$ depends only on the support and sign of x.
- $F(x_2) = F(x_1) = 0.9$ and then $x_2 \in \mathcal{F}$.



Sets $\mathcal{F}$ and $\bar{\mathcal{F}}$

$\mathcal{F}$ a union of cones

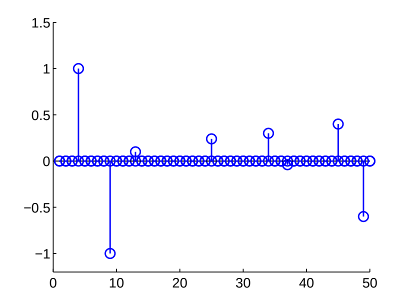
- $\mathcal{F} = \left\{ x \backslash F(x) = \max_{j \notin I} |\langle a_j, A_I^{\dagger t} \mathrm{sign}\,(\bar{x}) \rangle| < 1 \right\}$
- $F(x)$ depends only on the support and sign of x.
- $F(x_3) = F(x_1) = 0.9$ and then $x_2 \in \mathcal{F}$.



Sets $\mathcal{F}$ and $\bar{\mathcal{F}}$

$\mathcal{F}$ a union of cones

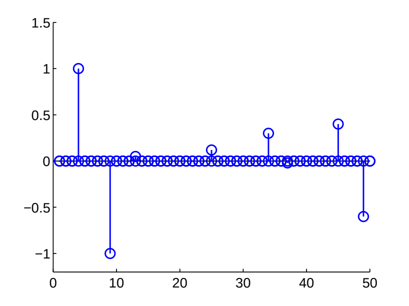
- $\mathcal{F} = \left\{ x \backslash F(x) = \max_{j \notin I} |\langle a_j, A_I^{\dagger t} \mathrm{sign}(\bar{x}) \rangle| < 1 \right\}$
- $F(x)$ depends only on the support and sign of x.
- $F(x_4) = F(x_1) = 0.9$ and then $x_2 \in \mathcal{F}$.



Sets $\mathcal{F}$ and $\bar{\mathcal{F}}$

$\mathcal{F}$ a union of cones

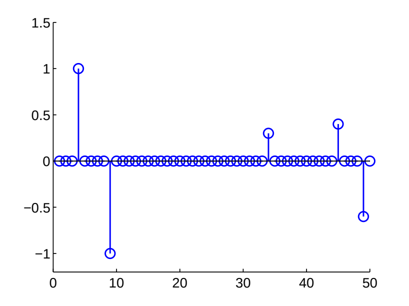
- $\mathcal{F} = \left\{ x \backslash F(x) = \max_{j \notin I} |\langle a_j, A_I^{\dagger t} \mathrm{sign}(\bar{x}) \rangle| < 1 \right\}$
- $F(x)$ depends only on the support and sign of x.
- $F(x_5) = F(x_1) = 0.9$ and then $x_2 \in \mathcal{F}$.



Sets $\mathcal{F}$ and $\bar{\mathcal{F}}$

$\mathcal{F}$ a union of cones

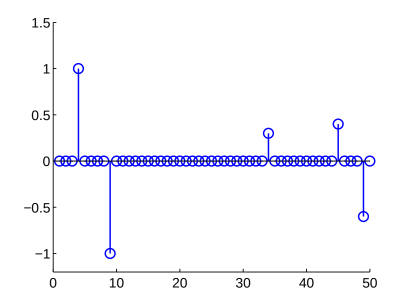
- $\mathcal{F} = \left\{ x \backslash F(x) = \max_{j \notin I} |\langle a_j, A_I^{\dagger t} \text{sign}(\bar{x}) \rangle| < 1 \right\}$
- $F(x)$ depends only on the support and sign of x.
- $I(x_6) \neq I(x_1)$, $F(x_6) = 1.1 > 1$ and then $x_6 \notin \mathcal{F}$ but $x_6 \in \bar{\mathcal{F}}$ and then should be identifiable and it is .



Sets $\mathcal{F}$ and $\bar{\mathcal{F}}$

$\mathcal{F}$ a union of cones

- $\mathcal{F} = \left\{x \backslash F(x) = \max_{j \notin I} |\langle a_j, A_I^{\dagger t} \mathrm{sign}\,(\bar{x})\rangle| < 1\right\}$
- $F(x)$ depends only on the support and sign of x.
- Let's consider now $x \notin \mathcal{F}$, $F(x) = 1.1$.



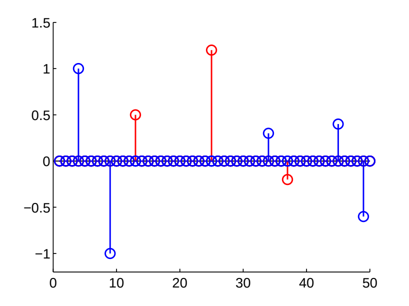
Sets $\mathcal{F}$ and $\bar{\mathcal{F}}$

$\mathcal{F}$ a union of cones

- $\mathcal{F} = \left\{ x \backslash F(x) = \max_{j \notin I} |\langle a_j, A_I^{\dagger t} \mathrm{sign}(\bar{x}) \rangle| < 1 \right\}$
- $F(x)$ depends only on the support and sign of x.

- Let's consider now $x \notin \mathcal{F}$, $F(x) = 1.1$.
- If there is x_1 such that
  - $I(x) \cap I(x_1) = \emptyset$ and
  - $x_2 = x + x_1 \in \mathcal{F}$,

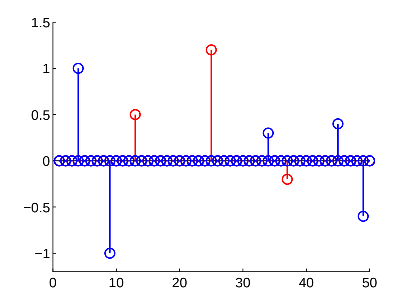
  then $x \in \bar{\mathcal{F}}$,
  $x = \lim_{\lambda \to 0} x + \lambda x_1$



Sets $\mathcal{F}$ and $\bar{\mathcal{F}}$

$\mathcal{F}$ a union of cones

- $\mathcal{F} = \left\{ x \backslash F(x) = \max_{j \notin I} |\langle a_j, A_I^{\dagger t} \mathrm{sign}\,(\bar{x}) \rangle| < 1 \right\}$
- $F(x)$ depends only on the support and sign of x.
- Vectors of $\bar{\mathcal{F}}$ are vectors that can be extented into a vector $x_1 \in \mathcal{F}$.
- $x \in \bar{\mathcal{F}} \Rightarrow x$ identifiable is not a surprise.
- x identifiable$\Rightarrow x \in \bar{\mathcal{F}}$ is more surprising.



The Set $\mathcal{F}$ and the Vector d_0

Definition of d_0

- To ease the explanations: $\|a_i\|_2 = 1$.
- $\mathcal{F} = \left\{x \backslash \max_{j \notin I} |\langle a_j, d_0 \rangle| < 1\right\}$ where $d_0 = A_I^{\dagger t} \mathrm{sign}\,(\bar{x})$
- $d_0 \in \mathrm{Span}(a_i)_{i \in I}$ and $\forall i \in I$, $\langle a_i, d_0 \rangle = \mathrm{sign}\,(x(i))$.

The Set $\mathcal{F}$ and the Vector d_0

Definition of d_0

- To ease the explanations: $\|a_i\|_2 = 1$.
- $\mathcal{F} = \left\{x \backslash \max_{j \notin I} |\langle a_j, d_0 \rangle| < 1\right\}$ where $d_0 = A_I^{\dagger t} \mathrm{sign}\,(\bar{x})$
- $d_0 \in \mathrm{Span}(a_i)_{i \in I}$ and $\forall i \in I$, $\langle a_i, d_0 \rangle = \mathrm{sign}\,(x(i))$.

If $\|x\|_0 = 1$, description of d_0

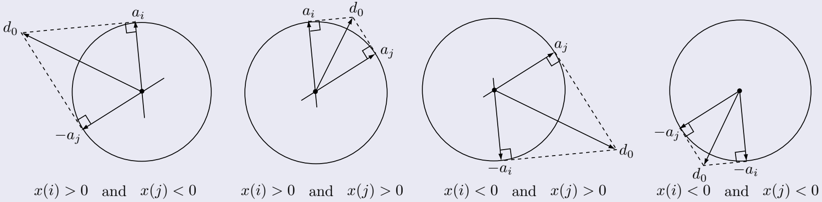
- If $x(i) \neq 0$, $d_0 = \pm a_i$ and $x \in \mathcal{F}$.

The Set $\mathcal{F}$ and the Vector d_0

Definition of d_0

- To ease the explanations: $\|a_i\|_2 = 1$.
- $\mathcal{F} = \left\{x \backslash \max_{j \notin I} |\langle a_j, d_0\rangle| < 1\right\}$ where $d_0 = A_I^{\dagger t}\mathrm{sign}(\bar{x})$
- $d_0 \in \mathrm{Span}(a_i)_{i\in I}$ and $\forall i \in I$, $\langle a_i, d_0\rangle = \mathrm{sign}(x(i))$.

If $\|x\|_0 = 2$, description of d_0

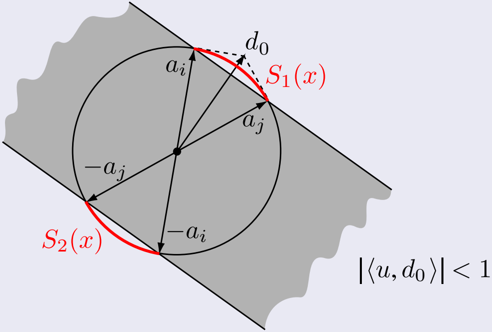


d_0 belongs to the bissector of $\mathrm{sign}(x(i))\, a_i$ and $\mathrm{sign}(x(j))\, a_j$.

The Set $\mathcal{F}$ and the Vector d_0

If $\|x\|_0 = 2$, description of $\mathcal{F}$

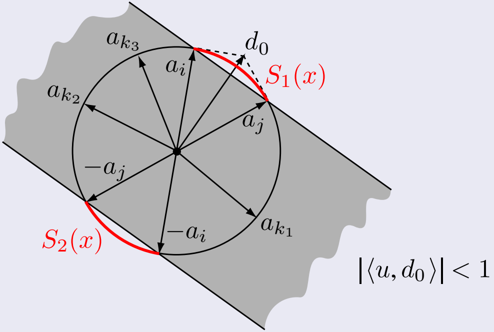
- If $\|x\|_0 = 2$, $x(i) > 0$ and $x(j) > 0$, u with $|\langle u, d_0\rangle| < 1$ belongs to the area in gray.



The Set $\mathcal{F}$ and the Vector d_0

If $\|x\|_0 = 2$, description of $\mathcal{F}$

- If $\|x\|_0 = 2$, $x(i) > 0$ and $x(j) > 0$, u with $|\langle u, d_0 \rangle| < 1$ belongs to the area in gray.

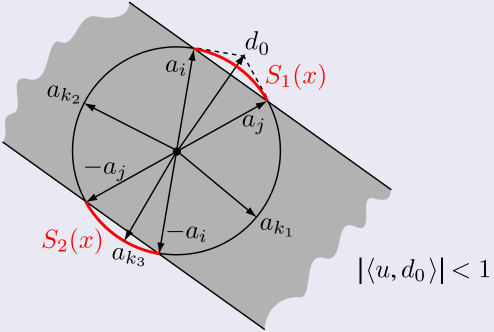


- If no a_k, $k \notin \{i, j\}$ belong to $S_1(x) \cup S_2(x)$, then $F(x) < 1$ and $x \in \mathcal{F}$.

The Set $\mathcal{F}$ and the Vector d_0

If $\|x\|_0 = 2$, description of $\mathcal{F}$

- If $\|x\|_0 = 2$, $x(i) > 0$ and $x(j) > 0$, u with $|\langle u, d_0 \rangle| < 1$ belongs to the area in gray.

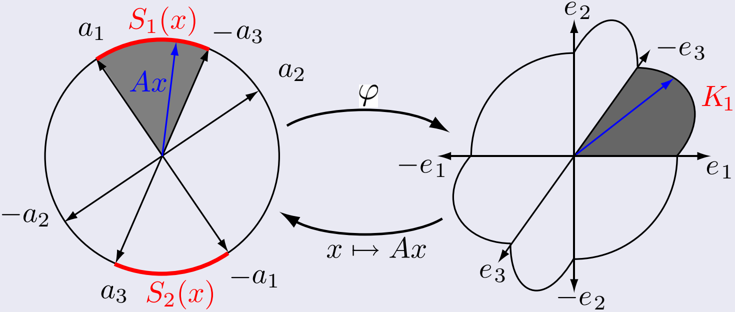


- If there is some a_k, $k \notin \{i, j\}$ in $S_1(x) \cup S_2(x)$, then $F(x) > 1$ and $x \notin \mathcal{F}$.

A Bijection From Cone to Cone

Low dimensional situation: $n = 2$ and $p = 3$

- Matrix $A = (a_1\ a_2\ a_3) \in \mathbb{R}^{2\times 3}$

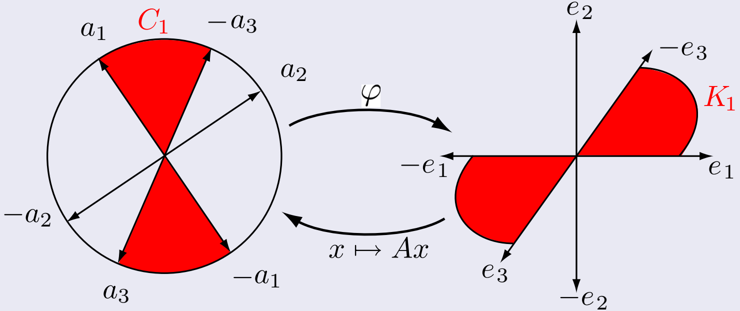


- Let's consider x, $\|x\|_0 = 2$, $x(1) > 0$ and $x(3) < 0$.
- $a_2 \notin S_1(x) \cup S_2(x)$ and then $F(x) < 1$, $x \in \mathcal{F}$

A Bijection From Cone to Cone

Low dimensional situation: $n = 2$ and $p = 3$

- Matrix $A = (a_1\ a_2\ a_3) \in \mathbb{R}^{2\times 3}$

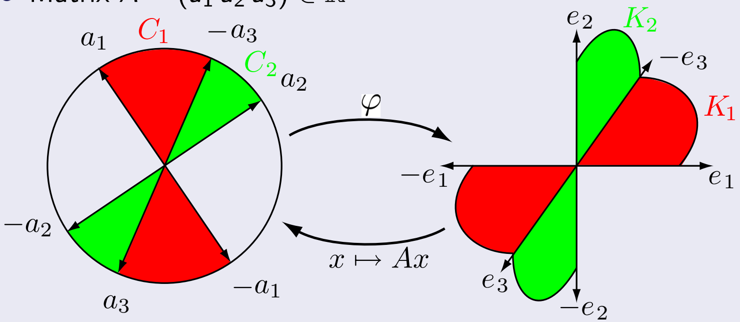


- $C_1 = \{\lambda a_1 + \mu a_3,\ \lambda\mu < 0\} = AK_1$
- $K_1 = \{(\lambda, 0, \mu),\ \lambda\mu < 0\} \subset \mathcal{F}$

A Bijection From Cone to Cone

Low dimensional situation: $n = 2$ and $p = 3$

- Matrix $A = (a_1\ a_2\ a_3) \in \mathbb{R}^{2\times 3}$

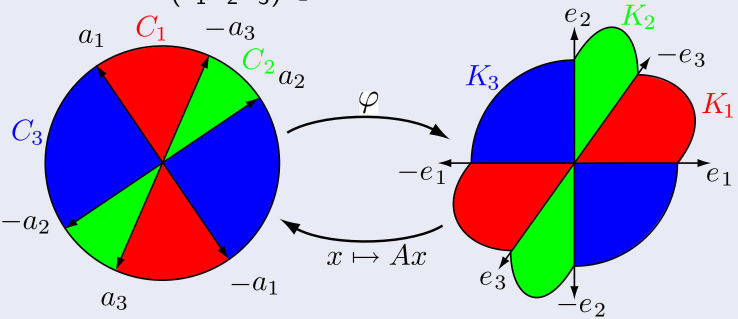


- $C_2 = \{\lambda a_2 + \mu a_3,\ \lambda\mu < 0\} = AK_2$
- $K_2 = \{(0, \lambda, \mu),\ \lambda\mu < 0\} \subset \mathcal{F}$

A Bijection From Cone to Cone

Low dimensional situation: $n = 2$ and $p = 3$

- Matrix $A = (a_1\ a_2\ a_3) \in \mathbb{R}^{2\times 3}$

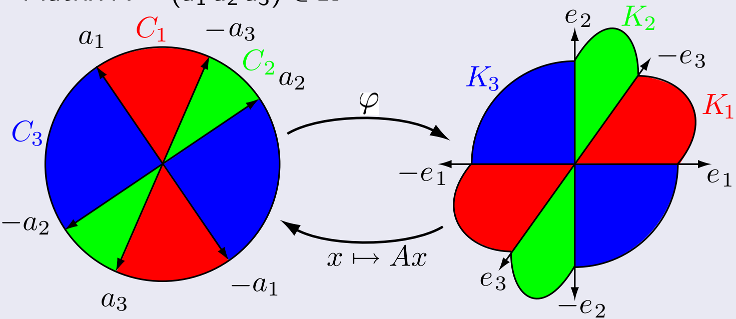


- $C_3 = \{\lambda a_1 + \mu a_2,\ \lambda\mu < 0\} = AK_3$
- $K_3 = \{(\lambda, \mu, 0),\ \lambda\mu < 0\} \subset \mathcal{F}$

A Bijection From Cone to Cone

Low dimensional situation: $n = 2$ and $p = 3$

- Matrix $A = (a_1\ a_2\ a_3) \in \mathbb{R}^{2\times 3}$

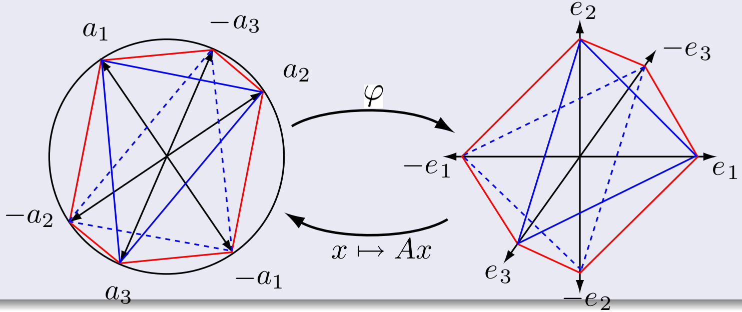


- $\mathbb{R}^2 = \bigcup_{j\leqslant 3} \bar{C}_j \Rightarrow$ no other identifiable vectors.
- $\mathcal{F} = \bigcup_{j\leqslant 3} \bar{K}_j = \bar{\mathcal{F}}$.
- $\varphi : y \in \mathbb{R}^2 \mapsto$ unique solution of $\mathrm{P}_1(y)$ is linear on $C_j \mapsto K_j$.

A Bijection From Cone to Cone

Low dimensional situation: $n = 2$ and $p = 3$

- Matrix $A = (a_1\ a_2\ a_3) \in \mathbb{R}^{2\times 3}$



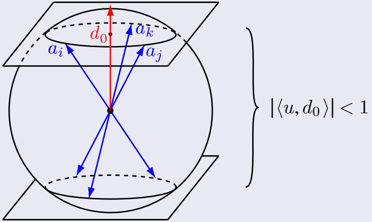
3D Case

Spherical Caps

- $n = 3$, $x \in \mathbb{R}^2$, $\|x\|_0 = 3$, $x(i) > 0, x(j) > 0, x(k) > 0.$
- $d_0 \in \mathrm{Span}(a_i)_{i \in I}$,

$\langle d_0, a_i \rangle = \langle d_0, a_j \rangle = \langle d_0, a_k \rangle = 1.$

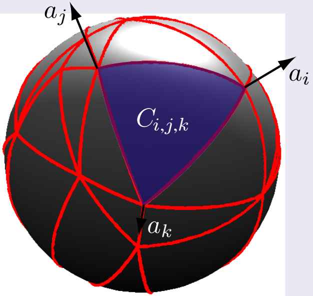
- Caps: $\{u \backslash |\langle u, d_0 \rangle| < 1\}$ delimited by two plans.



Cones in dimensions 3

Delaunay Triangulation

- $\mathcal{A} = (a_i)_i$ points on the sphere.
- There is a triangulation $\mathcal{T}$ such that the caps of all the triangles do not intersect $\mathcal{A}$.
- $\mathcal{T}$ = spher. Delaunay triangulation.
- $\mathcal{T}$ = triangulation of the convex hull of $\mathcal{A}$.



Cones correspondance

- Cones: $(K_j)_{j\in J}$ and $(C_j)_{j\in J}$.
- $K_j \in \mathcal{F}$, $C_j = AK_j$, $\mathbb{R}^3 = \bigcup_j \bar{C}_j$ and $\varphi : C_j \mapsto K_j$ is linear.
- $\exists x \in \bar{K}_j \backslash K_j \subset \bar{\mathcal{F}} \backslash \mathcal{F} \Rightarrow \mathcal{F} \neq \bar{\mathcal{F}}$.

An Algorithm to Estimate $\bar{\mathcal{F}}$

Idea of algorithm

- Algorithm try to extend x into a vector $x_1 \in \mathcal{F}$.
- Components are added to decrease the value of F.

Algorithm (greedy version)

- Initialisation $x \leftarrow x_0$, $I = I(x_0)$ and $d_0 = A_I^{\dagger t} \mathrm{sign}\,(\bar{x}_0)$.
- while $F(x) \geqslant 1$
  - Compute $j_0 = \arg\max_{j \notin I} |\langle a_j, d_0 \rangle|$
  - Set $x(j_0) = \mathrm{sign}\,(\langle a_j, d_0 \rangle)$, update I and d_0.

Efficiency

- 180 000 tests, about 60 000 identifiable vectors and 10 failures.

Conclusion

What's new

- We propose a new sufficient condition ensuring identifiability for any matrix.
- For a large set of matrices this condition is also necessary.
- Condition $x \in \mathcal{F}$ is analytical but $x \in \bar{\mathcal{F}}$ is topological.
- We propose an efficient algorithm to check this new condition.

Thank You!

If you want complete proofs

Preprint can be downloaded (very soon) from

http://www.math.u-bordeaux.fr/~dossal/