The Convex Geometry of Linear Inverse Problems.

Farideh Fazayeli

CSCI 8990
ML at Large Scale and High Dimensions

Mar 3, 2014

Reference

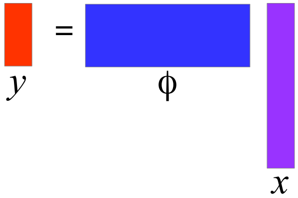
V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky
"The Convex Geometry of Linear Inverse Problems."
Foundations of Computational Mathematics, 12, 805-849, 2012.

1. Introduction
2. Unified Convex optimization Framework
3. Recovery Condition
   - Number of required measurements for unique true recovery
4. Computational Issues
5. Noisy Scenario

Inverse Problem

- Find the solution of

$$\mathbf{y} = \phi \mathbf{x} \quad \phi \in \mathbb{R}^{m \times n}$$

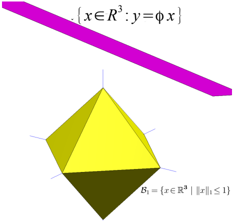


- Given $\mathbf{y} \Rightarrow$ recover $\mathbf{x}$.
- Limited Linear Measurements: ill posed problem
- Infinite solution, which one to choose?
- Examples
  - Sparse vectors: signal processing, statistics
  - Low-rank matrices: control, statistics, collaborative filtering
  - Sums of a few permutation matrices: ranked elections, multiobject tracking
  - Low-rank tensors: computer vision, neuroscience
  - Orthogonal matrices: machine learning

Review: Sparsity and Low-rank

- Minimizing ℓ_1 norm, yields sparse solution

$$\min_{\mathbf{x}} \| x \|_1 \quad s.t. \quad \mathbf{y} = \phi \mathbf{x}$$



- Minimizing nuclear norm, yields low rank solution
- Both are convex problem, can be solved efficiently
- Can be generalized?

Convex Optimization Framework

- Simple Models from atomic set A

Rank

$$x = \sum_{i=1}^{r} c_i a_i \qquad a_i \in A, \quad c_i \geq 0$$

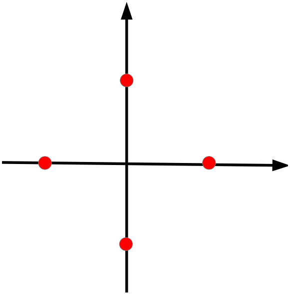
Model — Weights — Atoms

- Atomic norm induced by convex hull of A

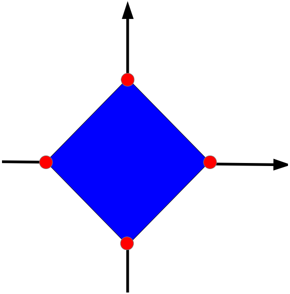
$$\| x \|_A = \inf\{t > 0 : x \in t\,\mathrm{conv}(A)\}$$

$$\| x \|_A = \inf\left\{\sum_i c_i : x = \sum_{i=1}^{r} c_i \mathbf{a}_i, c_i \geq 0, \mathbf{a}_i \in A\right\}$$

Geometric view - Sparsity



$\Rightarrow$

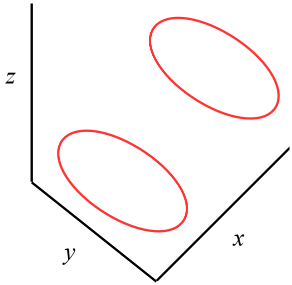


1-sparse vectors of
Euclidean norm 1

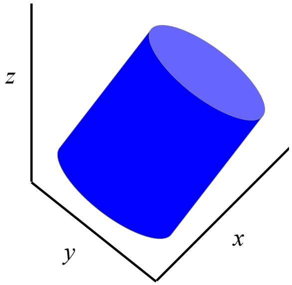
Convex Hull: ℓ_1 norm

$$\| \mathbf{x} \|_1 = \sum_{i=1}^{n} |x_i|$$

Geometric view - Low rank



$$\begin{pmatrix} x & y \\ y & z \end{pmatrix} \Rightarrow$$



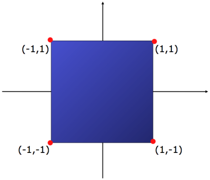
2×2 rank 1 symmetric matrices (normalized)

Convex Hull: nuclear norm

$$\| \mathbf{X} \|_* = \sum_i \sigma_i(X)$$

Other Convex Hulls

Sign Vectors

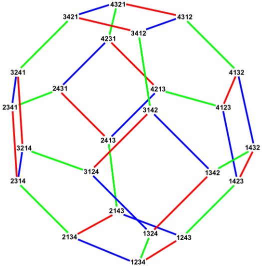


- Hypercube polytope
- Integer Programming

Cut Matrices

- Cut polytope
- Atoms: rank-1 sign matrices

Permutation Matrices



- Birkhoff polytope
- Permutahedra
- Ranking context
- Object tracking context

Convex Optimization Framework

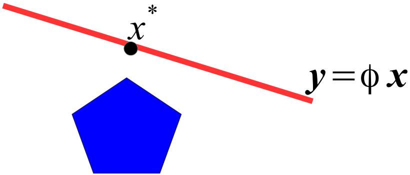
- Consider true $\mathbf{x}^*$ concise w.r.t to atomic set A
- Given linear measurement $\mathbf{y} = \phi \mathbf{x}^*$, solve

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \| x \|_A \quad \text{s.t.} \quad \mathbf{y} = \phi \mathbf{x}$$

- Recovery condition?

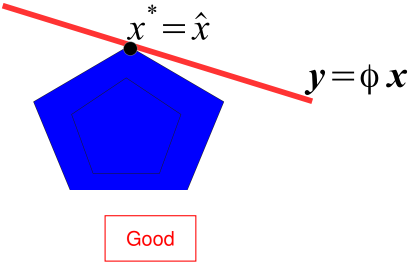
Recovery Condition: Geometric View

- When does $\hat{\mathbf{x}} = \mathbf{x}^*$?



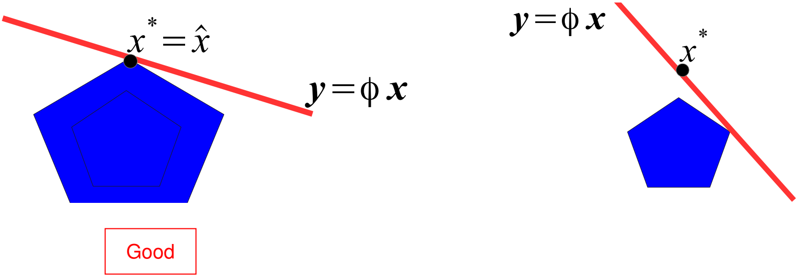
Recovery Condition: Geometric View

- When does $\hat{\mathbf{x}} = \mathbf{x}^*$?



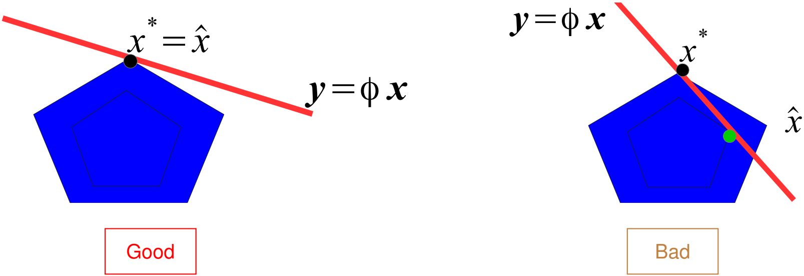
Recovery Condition: Geometric View

- When does $\hat{\mathbf{x}} = \mathbf{x}^*$?



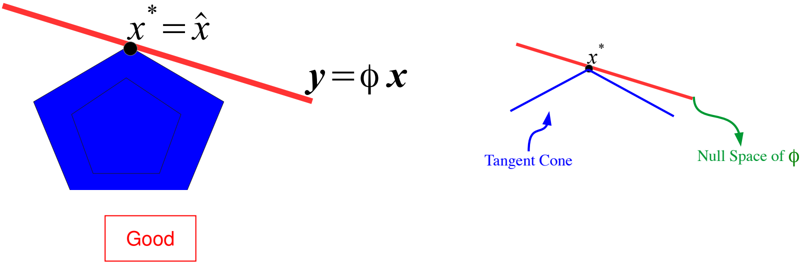
Recovery Condition: Geometric View

- When does $\hat{\mathbf{x}} = \mathbf{x}^*$?



Recovery Condition: Geometric View

- When does $\hat{\mathbf{x}} = \mathbf{x}^*$?



Recovery Condition

- Tangent Cone at x:

$$T_A(\mathbf{x}) = \{\mathbf{z} - \mathbf{x} : \| z \|_A \leq \| x \|_A\}$$

- Set of descent directions of $\| \, . \, \|_A$ at point $\mathbf{x}$.

Proposition 2.1

$$\hat{\mathbf{x}} = \mathbf{x}^* \iff null(\phi) \cap T_A(\mathbf{x}^*) = \{0\}$$

- Why Atomic Norm?

Recovery from Generic Measurements

- Number of measurements n for exact recovery?
- Gaussian Width:

$$w(S) := \mathbb{E}_{\mathbf{g}} \left[\sup_{\mathbf{z} \in S \cap \mathcal{B}(0,1)} \mathbf{g}^T \mathbf{z} \right]$$

- $\mathbf{g} \sim \mathcal{N}(0, I)$
- $\mathcal{B}(0, 1)$: Unit Euclidean ball.

Corollary 3.3

- $\mathbf{y} = \phi \mathbf{x}^*$
- $\phi : \mathbb{R}^p \to \mathbb{R}^n$ i.i.d. zero-mean Gaussian entries
- $\hat{\mathbf{x}} = \mathbf{x}^*$ W.H.P. if

$$n \geq w(T_A(\mathbf{x}^*))^2 + 1$$

- Gordon 1988

proof of Corollary 3.3

- $\| \phi \mathbf{z} \|$ Minimum gain of the ϕ restricted to $T_A(\mathbf{x}^*)$
- Bounding $\| \phi \mathbf{z} \|$ away from zero $\quad \mathbf{z} \in T_A(\mathbf{x}^*)$

Theorem 3.2

- Restricted minimum singular values
- λ_n: expected length of a n-dimensional Gaussian vector
- $\frac{n}{\sqrt{n+1}} \leq \lambda_n \leq \sqrt{n}$
- Ω: Closed subset of unit sphere $\mathbb{S}^{p-1}$
- $\phi : \mathbb{R}^p \leftarrow \mathbb{R}^n$: random map with i.i.d Gaussian entries

$$\mathbb{E}\left[\min_{\mathbf{z}\in\Omega} \| \phi \mathbf{z} \|_2\right] \geq \lambda_n - w(\Omega)$$

- Gordon1988

proof of Corollary 3.3

- $\mathbf{g} \sim \mathcal{N}(0, I)$
- f be Lipschitz constant L

$$P(f(\mathbf{g}) \geq \mathbb{E}[f] - t) \geq 1 - \exp(-\frac{t^2}{2L^2})$$

- $\min_{\mathbf{z}\in\Omega} \parallel \phi\mathbf{z} \parallel_2$ is 1-Lipschitz
- $\mathbb{E}\left[\min_{\mathbf{z}\in\Omega} \parallel \phi\mathbf{z} \parallel_2\right] \geq \lambda_n - w(\Omega)$

$$P(\min_{\mathbf{z}\in\Omega} \parallel \phi\mathbf{z} \parallel_2 \geq \epsilon) \geq 1 - \exp(-\frac{1}{2}(\lambda_n - w(\Omega) - \sqrt{n}\epsilon)^2)$$

$$\geq 0$$

- Set $\epsilon = 0 \quad \Rightarrow \quad \lambda_n \geq w(\Omega)$
- $w(\Omega) \leq \lambda_n \leq \sqrt{n}$

Gaussian Width via Dual Cone

- Gaussian width of a cone via the distance to the dual cone
- Polar cone of $\mathcal{C}$:

$$\mathcal{C}^* = \{\mathbf{x} \in \mathbb{R}^p : \langle \mathbf{x}, \mathbf{z} \rangle \leq 0 \ \ \forall \mathbf{z} \in \mathcal{C}\}$$

Proposition 3.6

- $\mathbf{g} \sim \mathcal{N}(0, I)$
- dist: Euclidean distance of a point to a set

$$w(\mathcal{C}) \leq \mathbb{E}_{\mathbf{g}} \left[\mathrm{dist}(\mathbf{g}, \mathcal{C}^*)\right]$$

$$w(\mathcal{C})^2 \leq \mathbb{E}_{\mathbf{g}} \left[\mathrm{dist}(\mathbf{g}, \mathcal{C}^*)^2\right]$$

Proof

- Gaussian Width: $w(\mathcal{C} \cap \mathbb{S}^{p-1}) \leq \mathbb{E}_{\mathbf{g}} \left[\sup_{\mathbf{z} \in \mathcal{C} \cap \mathcal{B}(0,1)} \mathbf{g}^T \mathbf{z} \right]$
- Inside the expected value is the optimal solution to

$$\max_{z} \mathbf{g}^T \mathbf{z} \quad \text{s.t.} \quad \mathbf{z} \in \mathcal{C}, \quad \| z \|^2 \leq 1$$

- Introducing the Lagrangian:

$$\mathcal{L}(\mathbf{z}, \mathbf{u}, \gamma) = \mathbf{g}^T \mathbf{z} + \gamma(1 - \mathbf{z}^T \mathbf{z}) - \mathbf{u}^T \mathbf{z}$$

- minimize w.r.t $\mathbf{z}$ and γ

$$\mathbf{z} = \frac{1}{2\gamma}(\mathbf{g} - \mathbf{u}) \qquad \gamma = \frac{1}{2} \| \mathbf{g} - \mathbf{u} \|$$

- Dual Problem:

$$\min \| \mathbf{g} - \mathbf{u} \| \quad \text{s.t} \quad \mathbf{u} \in \mathcal{C}^*$$

Properties of $w(\mathcal{C})$

- **Lemma 3.7**: $\mathcal{C} \subset \mathbb{R}^p$, $\quad w(\mathcal{C})^2 + w(\mathcal{C}^*)^2 \leq p$
- proof:
- $\mathbf{g} = \sqcap_{\mathcal{C}}(\mathbf{g}) + \sqcap_{\mathcal{C}^*}(\mathbf{g})$ where $\langle \sqcap_{\mathcal{C}}(\mathbf{g}), \sqcap_{\mathcal{C}^*}(\mathbf{g}) \rangle = 0$
- $\text{dist}(\mathbf{g}, \mathcal{C}) = \| \sqcap_{\mathcal{C}^*}(\mathbf{g}) \|$

$$
\begin{aligned}
w(\mathcal{C})^2 &\leq \mathbb{E}_{\mathbf{g}}[\text{dist}(\mathbf{g}, \mathcal{C}^*)^2] \\
&= \mathbb{E}_{\mathbf{g}}[\| \mathbf{g} \|^2 - \| \sqcap_{\mathcal{C}^*}(\mathbf{g}) \|^2] = p - \mathbb{E}_{\mathbf{g}}[\text{dist}(\mathbf{g}, \mathcal{C})^2] \\
&\leq p - w(\mathcal{C}^*)^2
\end{aligned}
$$

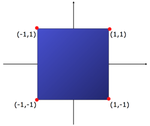
- **Corollary 3.8**: Self dual cone $\mathcal{C} = -\mathcal{C}^*$

$$
w(\mathcal{C})^2 \leq \frac{p}{2}
$$

Special Cases

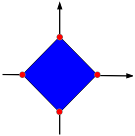
- Hypercube:

$$w(T_A(\mathbf{x}^*))^2 \leq \frac{p}{2}$$



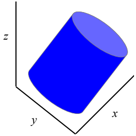
- s-sparse vector $\mathbf{x}^* \in \mathbb{R}^p$:

$$w(T_A(\mathbf{x}^*))^2 \leq 2s\log(\frac{p}{s}) + \frac{5}{4}s$$



- Low-rank matrices $\in \mathbb{R}^{m_1 \times m_2}$, rank r

$$w(T_A(\mathbf{x}^*))^2 \leq 3r(m_1 + m_2 - r)$$



General Cones

Theorem 3.9

- $\mathcal{C} \subseteq \mathbb{R}^p$: close, convex, solid cone
- $\mathcal{C}^*$: has volume of $\theta \in [0, 1]$

$$w(\mathcal{C}) \leq 3\sqrt{\log \frac{4}{\theta}}$$

- **Corollary 3.14**
  For a symmetric polytope with m vertices

$$n \geq O(\log m)$$

Approximations

- Atomic set A are Algebraic variety
- Well-approximated in a constructive manner by
  - linear matrix inequality constraints
- Semidefinite representations are intractable?
  - Hierarchy of tractable semidefinite relaxations

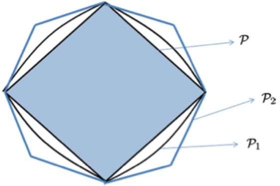
Complexity vs Number of Measurements

- Intractable to compute norm induced by cut Polytope:

$$\mathcal{P} = \text{conv}\{\mathbf{z}^T\mathbf{z} : \mathbf{z} \in \{-1, +1\}^m\}$$

- MAX-CUT problem
- Semidefinite relaxation:
  $\mathcal{P}_1 = \{\mathcal{M} : \mathcal{M} \ \text{Symmetric}, \ \ \mathcal{M} \succcurlyeq 0, \mathcal{M}_{ii} = 1\}$
- Trivial hypercube relaxation:
  $\mathcal{P}_2 = \{\mathcal{M} : \mathcal{M} \ \text{Symmetric}, \ \ \mathcal{M}_{ii} = 1, \ \ |\mathcal{M}_{ij}| < 1 \ \ \forall i \neq j\}$

- Using $\mathcal{P} : n = O(m)$
- Using $\mathcal{P}_1 : n = O(m)$
- Using $\mathcal{P}_2 : n = O(\frac{m^2 - m}{4})$



Robust Recovery

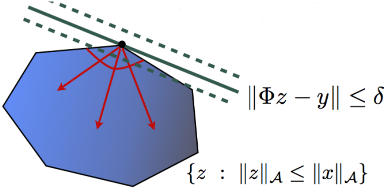
- Noisy Scenario: $\mathbf{y} = \phi \mathbf{x}^* + \omega \qquad \| \omega \| \leq \delta$

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \| \mathbf{x} \|_A$$
$$\text{s.t.} \quad \| \mathbf{y} - \phi \mathbf{x} \| \leq \delta$$

- Robust recovery: $\| \mathbf{x}^* - \hat{\mathbf{x}} \| \leq \frac{2\delta}{\epsilon}$ W.H.P. provided by

$$n \geq \frac{c_0 w(T_A(\mathbf{x}^*))^2}{(1-\epsilon)^2}$$

- $\| \phi \mathbf{z} \| \geq \epsilon \| \mathbf{z} \|$



Conclusion

- Providing a unified convex optimization framework for Inverse problem
- Recovery condition
    - Noiseless scenario
    - Noisy scenario
- Number of measurements for true unique recovery
- Tradeoff: complexity and number of measurements

Questions?