The Convex Geometry of Linear Inverse Problems.

Farideh Fazayeli

CSCI 8990
ML at Large Scale and High Dimensions

Mar 3, 2014
V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky
“The Convex Geometry of Linear Inverse Problems.”
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

\[ y = \phi x \quad \phi \in \mathbb{R}^{m \times n} \]

- Given \( y \Rightarrow \) recover \( x \).
- 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 $\ell_1$ norm, yields sparse solution

$$\min_x \| x \|_1$$
$$\text{s.t. } y = \phi x$$

- Minimizing nuclear norm, yields low rank solution
- Both are convex problem, can be solved efficiently
- Can be generalized?
Simple Models from atomic set $A$

$$x = \sum_{i=1}^{r} c_i a_i$$

$\mathbf{Atoms}$

$\mathbf{Weights}$

$\mathbf{Model}$

$\mathbf{Rank}$

Atomic norm induced by convex hull of $A$

$$\| x \|_A = \inf \left\{ t > 0 : x \in t \text{ conv}(A) \right\}$$

$$\| x \|_A = \inf \left\{ \sum_{i=1}^{r} c_i : x = \sum_{i=1}^{r} c_i a_i, c_i \geq 0, a_i \in A \right\}$$
Geometric view - Sparsity

1-sparse vectors of Euclidean norm 1

Convex Hull: $\ell_1$ norm

$$\| x \|_1 = \sum_{i=1}^{n} |x_i|$$
Geometric view - Low rank

2 × 2 rank 1 symmetric matrices (normalized)

Convex Hull: nuclear norm

\[ \| X \|_* = \sum_i \sigma_i(X) \]
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 $x^*$ concise w.r.t to atomic set $A$
- Given linear measurement $y = \phi x^*$, solve

$$\hat{x} = \arg \min_x \| x \|_A$$

s.t. $y = \phi x$

- Recovery condition?
Recovery Condition: Geometric View

- When does $\hat{\mathbf{x}} = \mathbf{x}^*$?
Recovery Condition: Geometric View

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

![Diagram showing a good scenario where $x^* = \hat{x}$ and $y = \phi x$]
Recovery Condition: Geometric View

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

$\bar{x} = \hat{x}$

$y = \phi x$

Good
Recovery Condition: Geometric View

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

- **Good**
  - $x^* = \hat{x}$
  - $y = \phi x$

- **Bad**
  - $y = \phi x$
  - $\hat{x}$
When does $\hat{x} = x^*$?
Recovery Condition

- Tangent Cone at $x$:

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

- Set of descent directions of $\|\cdot\|_A$ at point $x$.

**Proposition 2.1**

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

- Why Atomic Norm?
Recovery from Generic Measurements

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

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

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

**Corollary 3.3**

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

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

- Gordon 1988
proof of Corollary 3.3

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

Theorem 3.2

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

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

- Gordon1988
proof of Corollary 3.3

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

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

- \( \min_{\mathbf{z} \in \Omega} \| \phi \mathbf{z} \|_2 \) is 1-Lipschitz
- \( \mathbb{E} \left[ \min_{\mathbf{z} \in \Omega} \| \phi \mathbf{z} \|_2 \right] \geq \lambda_n - w(\Omega) \)

\[
P\left( \min_{\mathbf{z} \in \Omega} \| \phi \mathbf{z} \|_2 \geq \epsilon \right) \geq 1 - \exp\left(-\frac{1}{2}\left(\lambda_n - w(\Omega) - \sqrt{n}\epsilon\right)^2\right) \geq 0
\]

- Set \( \epsilon = 0 \Rightarrow \lambda_n \geq w(\Omega) \)
- \( w(\Omega) \leq \lambda_n \leq \sqrt{n} \)
Gaussian Width via Dual Cone

- Gaussian width of a cone via the distance to the dual cone
- Polar cone of $C$:

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

**Proposition 3.6**

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

$$w(C) \leq \mathbb{E}_g [\text{dist}(g, C^*)]$$

$$w(C)^2 \leq \mathbb{E}_g [\text{dist}(g, C^*)^2]$$
Proof

- **Gaussian Width:** $w(C \cap S^{p-1}) \leq \mathbb{E}_g \left[ \sup_{z \in C \cap B(0, 1)} g^T z \right]$

- Inside the expected value is the optimal solution to

\[
\max_z g^T z \quad \text{s.t.} \quad z \in C, \quad \| z \|_2^2 \leq 1
\]

- Introducing the Lagrangian:

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

- minimize w.r.t $z$ and $\gamma$

\[
z = \frac{1}{2\gamma} (g - u)
\]

\[
\gamma = \frac{1}{2} \| g - u \|
\]

- Dual Problem:

\[
\min \| g - u \| \quad \text{s.t} \quad u \in C^*
\]
Properties of $w(C)$

- **Lemma 3.7**: $C \subset \mathbb{R}^p$, \[ w(C)^2 + w(C^*)^2 \leq p \]
  
  **proof**: \[ g = \cap_c(g) + \cap_{c^*}(g) \] where \[ \langle \cap_c(g), \cap_{c^*}(g) \rangle = 0 \]

  \[ \text{dist}(g, C) = \| \cap_{c^*}(g) \| \]

  \[ w(C)^2 \leq \mathbb{E}_g[\text{dist}(g, C^*)^2] = \mathbb{E}_g[\| g \|^2 - \| \cap_{c^*}(g) \|^2] = p - \mathbb{E}_g[\text{dist}(g, C)^2] \leq p - w(C^*)^2 \]

- **Corollary 3.8**: Self dual cone $C = -C^*$

  \[ w(C)^2 \leq \frac{p}{2} \]
Special Cases

- Hypercube:
  \[ w(T_A(x^*))^2 \leq \frac{p}{2} \]

- \textit{s}-sparse vector \( x^* \in \mathbb{R}^p \):
  \[ w(T_A(x^*))^2 \leq 2s \log\left(\frac{p}{s}\right) + \frac{5}{4}s \]

- Low-rank matrices \( \in \mathbb{R}^{m_1 \times m_2} \), rank \( r \)
  \[ w(T_A(x^*))^2 \leq 3r(m_1 + m_2 - r) \]
General Cones

Theorem 3.9
- \( C \subseteq \mathbb{R}^p \): close, convex, solid cone
- \( C^* \): has volume of \( \theta \in [0, 1] \)

\[
w(C) \leq 3 \sqrt{\log \frac{4}{\theta}}
\]

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:
  \[ P = \text{conv}\{ z^T z : z \in \{-1, +1\}^m \} \]
- MAX-CUT problem
- Semidefinite relaxation:
  \[ P_1 = \{ \mathcal{M} : \mathcal{M} \text{ Symmetric, } \mathcal{M} \succeq 0, \mathcal{M}_{ii} = 1 \} \]
- Trivial hypercube relaxation:
  \[ P_2 = \{ \mathcal{M} : \mathcal{M} \text{ Symmetric, } \mathcal{M}_{ii} = 1, |\mathcal{M}_{ij}| < 1 \ \forall i \neq j \} \]

- Using \( P \): \( n = O(m) \)
- Using \( P_1 \): \( n = O(m) \)
- Using \( P_2 \): \( n = O(\frac{m^2-m}{4}) \)
Robust Recovery

- Noisy Scenario: \( y = \phi x^* + \omega \quad \| \omega \| \leq \delta \)
  
  \[ \hat{x} = \arg \min_x \| x \|_A \]
  
  s.t. \( \| y - \phi x \| \leq \delta \)

- Robust recovery: \( \| x^* - \hat{x} \| \leq \frac{2\delta}{\epsilon} \) W.H.P. provided by
  
  \[ n \geq \frac{c_0 w(T_A(x^*))^2}{(1 - \epsilon)^2} \]

- \( \| \phi z \| \geq \epsilon \| z \| \)
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?