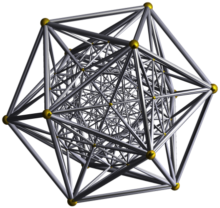
Computational Principles for High-dim Data Analysis

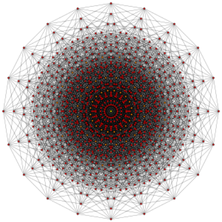
(Lecture Eight)

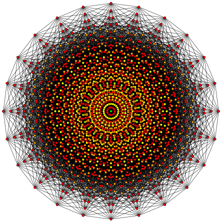
Yi Ma

EECS Department, UC Berkeley

September 23, 2021







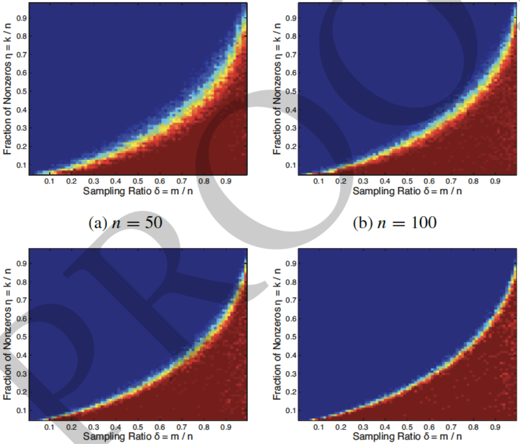
Convex Methods for Sparse Signal Recovery
(Phase Transition in Sparse Recovery)

"Algebra is but written geometry; geometry is but drawn algebra."
– Sophie Germain

Phase Transition Phenomenon

Success probability of the ℓ^1 minimization:

$$\min \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$



(a) $n = 50$ (b) $n = 100$

(c) $n = 200$ (d) $n = 400$

Phase Transition Phenomenon

Conjecture: measurement ratio $\delta = m/n$ exceeds a certain function $\psi(\eta)$ of the sparsity ratio $\eta = k/n$. That is, the precise number of measurements needed for success of ℓ^1 minimization:

$$m^\star \geq \psi(k/n)n.$$

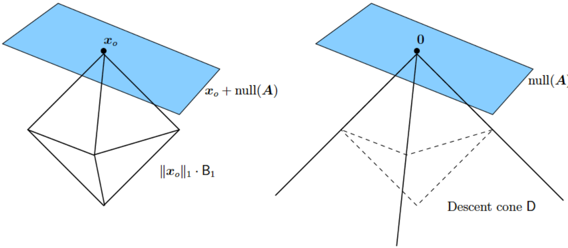
When do we expect this to happen? (compared to RIP)

- From a deterministic to a *random* matrix $\boldsymbol{A} \sim_{iid} \mathcal{N}(0, \frac{1}{m})$.
- From recovery of all sparse to a *fixed* sparse $\boldsymbol{x}_o$.

A More Rigorous (and Weaker) Statement: For a given, *fixed* $\boldsymbol{x}_o$, with high probability in the *random* matrix $\boldsymbol{A}$, ℓ^1 minimization recovers that particular $\boldsymbol{x}_o$ from the measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_o$.

Phase Transition: Geometric Intuition

In the Coefficient Space $\boldsymbol{x} \in \mathbb{R}^n$:



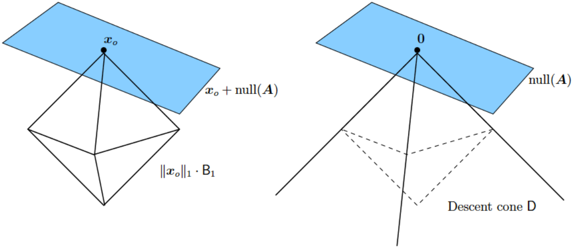
Necessary and Sufficient Condition: x_o is the only intersection between the affine subspace:

$$\mathsf{S}: \quad \{\boldsymbol{x} \mid \boldsymbol{x} \in \boldsymbol{x}_o + \mathrm{null}(\boldsymbol{A})\} \tag{1}$$

of feasible solutions and the scaled ℓ^1 ball:

$$\|\boldsymbol{x}_o\|_1 \cdot \mathsf{B}_1 = \{\boldsymbol{x} \mid \|\boldsymbol{x}\|_1 \leq \|\boldsymbol{x}_o\|_1\}. \tag{2}$$

Phase Transition: Coefficient Space



Lemma

Suppose that $y = Ax_o$. *Then* x_o *is the unique optimal solution to the* ℓ^1 *minimization problem if and only if* $\mathsf{D} \cap \mathrm{null}(A) = \{\mathbf{0}\}$, *where* D *is the* ℓ^1 *descent cone:*

$$\mathsf{D} = \{v \mid \|x_o + tv\|_1 \le \|x_o\|_1 \text{ for some } t > 0\}. \quad (3)$$

Phase Transition: Example of Two Random Subspaces

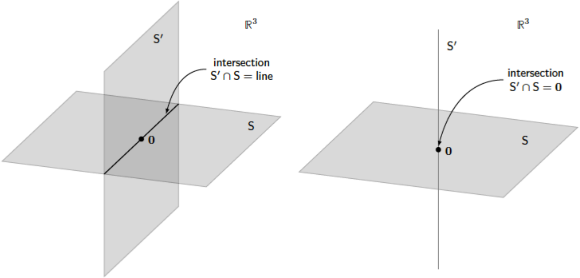
When does a randomly chosen subspace S intersect another subspace S′?

Example (Intersection of Two Linear Subspaces)

Let S′ be any linear subspace of $\mathbb{R}^n$, and let S be a uniform random subspace. Then

$$\mathbb{P}\left[\mathsf{S} \cap \mathsf{S}' = \{\mathbf{0}\}\right] = 0, \quad \dim(\mathsf{S}) + \dim(\mathsf{S}') > n; \tag{4}$$

$$\mathbb{P}\left[\mathsf{S} \cap \mathsf{S}' = \{\mathbf{0}\}\right] = 1, \quad \dim(\mathsf{S}) + \dim(\mathsf{S}') \leq n. \tag{5}$$



Phase Transition: Example of Two Random Cones

When does a cone C_1 intersect another randomly chosen cone C_2?

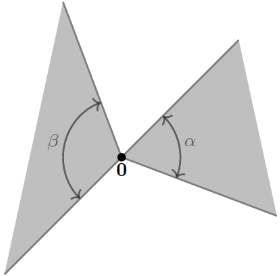
Example (Two Cones in $\mathbb{R}^2$)

Notice that if we have two convex cones C_1 and C_2 in $\mathbb{R}^2$, with angle α and β respectively. Let C_1 be fixed and we rotate C_2 by a rotation $\boldsymbol{R}$ uniformly chosen from $\mathbb{S}^1$. Then we have

$$\mathbb{P}[\mathsf{C}_1 \cap \boldsymbol{R}(\mathsf{C}_2) \neq \{\mathbf{0}\}] = \min\{1, (\alpha+\beta)/2\pi\}. \quad (6)$$

How to generalize these special cases to general convex cones?

the notion of
dimension or size of angle for cones in $\mathbb{R}^2$
to cones in spaces of higher dimension?...



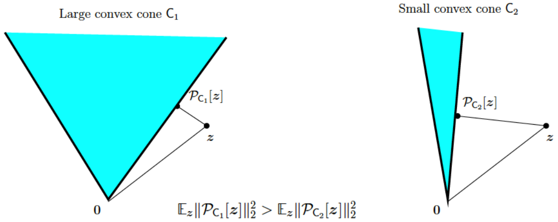
Phase Transition: Geometric & Statistical Dimension

Consider a Gaussian vector, $\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, projected onto the subspace S:

$$\mathcal{P}_{\mathsf{S}}[\boldsymbol{g}] \doteq \arg\min_{\boldsymbol{x} \in \mathsf{S}} \|\boldsymbol{x} - \boldsymbol{g}\|_2^2 . \tag{7}$$

Then, an equivalent definition of dimension of S:

$$d = \dim(\mathsf{S}) = \mathbb{E}_{\boldsymbol{g}}\left[\|\mathcal{P}_{\mathsf{S}}[\boldsymbol{g}]\|_2^2\right] . \tag{8}$$



Phase Transition: Coefficient Space

Definition (Statistical Dimension)

Given C is a closed convex cone in $\mathbb{R}^n$, then its statistical dimension, denoted as $\delta(\mathsf{C})$, is given by:

$$\delta(\mathsf{C}) \doteq \mathbb{E}_{\boldsymbol{g}}\left[\|\mathcal{P}_{\mathsf{C}}[\boldsymbol{g}]\|_2^2\right], \quad \text{with } \boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}). \tag{9}$$

Fact: if S is a random subspace of $\mathbb{R}^n$, and C a closed convex cone, then we have:

$$\delta(\mathsf{S}) + \delta(\mathsf{C}) \gg n \quad \Longrightarrow \quad \mathsf{S} \cap \mathsf{C} \neq \{\mathbf{0}\} \text{ with high probability;}$$
$$\delta(\mathsf{S}) + \delta(\mathsf{C}) \ll n \quad \Longrightarrow \quad \mathsf{S} \cap \mathsf{C} = \{\mathbf{0}\} \text{ with high probability.}$$

For a more precise statement see Chapter 6 or a proof.[1]

[1] *Living on the edge: Phase transitions in convex programs with random data.* D. Amelunxen, M. Lotz, M. McCoy, and J. Tropp, *Information and Inference*, 2014

Phase Transition: Coefficient Space

Proposition (Phase Transition for ℓ^1 Minimization – Qualitative)

Suppose that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_o$ with $\boldsymbol{x}_o$ sparse. Let D *denote the descent cone of the ℓ^1 norm $\|\cdot\|_1$ at $\boldsymbol{x}_o$. Since* $\dim[\mathrm{null}(\boldsymbol{A})] = n - m$*, then:*

$$\mathbb{P}[\ell^1 \textit{ recovers } \boldsymbol{x}_o] \;\le\; C \exp\left(-c\frac{(\delta(\mathsf{D}) - m)^2}{n}\right), \quad m \le \delta(\mathsf{D});$$

$$\mathbb{P}[\ell^1 \textit{ recovers } \boldsymbol{x}_o] \;\ge\; 1 - C \exp\left(-c\frac{(m - \delta(\mathsf{D}))^2}{n}\right), \quad m \ge \delta(\mathsf{D}).$$

Methods and results can be generalized to:

- Any other atomic norms $\|\cdot\|_{\mathcal{D}}$ (e.g. nuclear norm for matrices).
- Intersections between two general convex cones (e.g. RPCA).

Phase Transition: Coefficient Space

Proposition (Phase Transition for ℓ^1 Minimization – Quantitative)

Let D *be the descent cone of the* ℓ^1 *norm at any* $\boldsymbol{x}_o \in \mathbb{R}^n$ *satisfying* $\|\boldsymbol{x}_o\|_0 = k$. *Then*

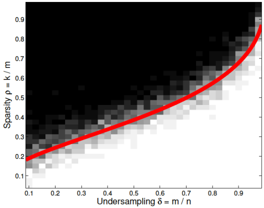
$$n\psi\left(\frac{k}{n}\right) - 4\sqrt{n/k} \le \delta(\mathrm{D}) \le n\psi\left(\frac{k}{n}\right), \tag{10}$$

where

$$\psi(\eta) = \min_{t\ge 0}\left\{\eta(1+t^2) + (1-\eta)\sqrt{\frac{2}{\pi}}\int_t^\infty (s-t)^2 \exp\left(-\frac{s^2}{2}\right) ds\right\}. \tag{11}$$

See Chapter 6 for derivation...
Phase transition for ℓ^1 minimization takes place at:

$$m^\star = \psi\left(\frac{k}{n}\right) n. \tag{12}$$



Phase Transition: Observation Space

The unit ℓ^1 ball in $\mathbb{R}^n$:

$$\mathsf{B}_1 \doteq \{\boldsymbol{x} \mid \|\boldsymbol{x}\|_1 \le 1\}$$

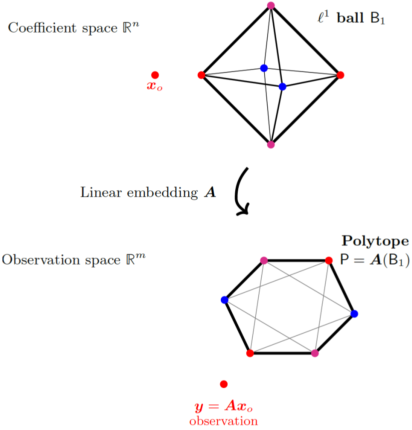
and its projection into $\mathbb{R}^m$:

$$\mathsf{P} \doteq \boldsymbol{A}(\mathsf{B}_1) = \{\boldsymbol{A}\boldsymbol{x} \mid \|\boldsymbol{x}\|_1 \le 1\}.$$

ℓ^1 minimization uniquely recovers any $\boldsymbol{x}$ with support I and signs $\boldsymbol{\sigma}$ if and only if

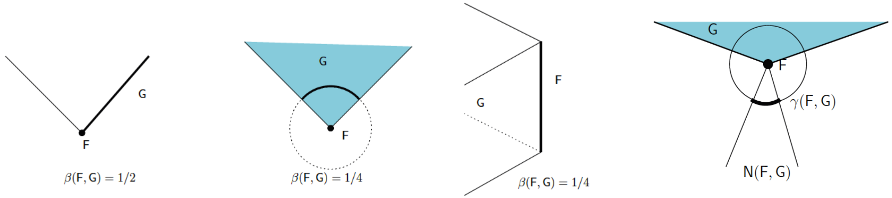
$$\mathsf{F} \doteq \mathrm{conv}(\{\sigma_i \boldsymbol{a}_i \mid i \in \mathsf{I}\}) \qquad (13)$$

forms a face of the polytope P.



Phase Transition: Observation Space

Internal angle and **external angle** of a face F on a polytope G.



Fact[2]**:** for an $m \times n$ Gaussian matrix $\boldsymbol{A}$,

$$\mathbb{E}_{\boldsymbol{A}}[f_k(\boldsymbol{A}\mathsf{P})] = f_k(\mathsf{P}) - 2 \underbrace{\sum_{\ell=m+1,m+3,\ldots} \sum_{\mathsf{F}\in\mathsf{F}_k(\mathsf{P})} \sum_{\mathsf{G}\in\mathsf{F}_\ell(\mathsf{P})} \beta(\mathsf{F},\mathsf{G})\gamma(\mathsf{G},\mathsf{P})}_{\Delta=\text{Expected number of faces lost}}.$$

When Δ is substantially smaller than one, w.h.p., we have

$$f_k(\boldsymbol{A}(\mathsf{P})) = f_k(\mathsf{P}).$$

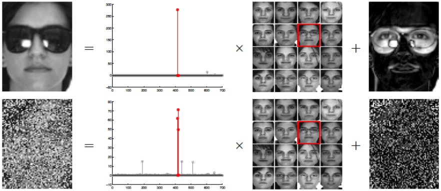
[2] Counting faces of randomly projected polytopes when the projection radically lowers dimension, D. Donoho and J. Tanner, 2009.

Phase Transition for Recovering Support

Recall

face identification problem: From noisy observations $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_o + \boldsymbol{z}$, estimate the signed support:

$$\boldsymbol{\sigma}_o = \mathrm{sign}(\boldsymbol{x}_o), \qquad (14)$$



by solving the Lasso problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{1}{2}\left\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right\|_2^2 + \lambda\left\|\boldsymbol{x}\right\|_1 .$$

Two scenarios:

- **Partial support recovery**: $\mathrm{supp}(\hat{\boldsymbol{x}}) \subseteq \mathrm{supp}(\boldsymbol{x}_o)$. The estimator exhibits no "false positives".
- **Signed support recovery**: $\mathrm{sign}(\hat{\boldsymbol{x}}) = \boldsymbol{\sigma}_o$. The estimator correctly determines all nonzero entries of $\boldsymbol{x}_o$ and their signs – difficult!.

Phase Transition for Recovering Support

Theorem (Phase Transition in Partial Support Recovery)

Suppose that $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with entries iid $\mathcal{N}(0, \frac{1}{m})$ random variables, and let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_o + \boldsymbol{z}$, with $\boldsymbol{x}_o$ a k-sparse vector and $\boldsymbol{z} \sim_{\text{iid}} \mathcal{N}\left(0, \frac{\sigma^2}{m}\right)$. If

$$m \geq \left(1 + \frac{\sigma^2}{\lambda^2 k} + \epsilon\right) 2k \log(n-k), \tag{15}$$

then with probability at least $1 - Cn^{-\epsilon}$, any solution $\hat{\boldsymbol{x}}$ to the Lasso problem satisfies $\operatorname{supp}(\hat{\boldsymbol{x}}) \subseteq \operatorname{supp}(\boldsymbol{x}_o)$. Conversely, if

$$m < \left(1 + \frac{\sigma^2}{\lambda^2 k} - \epsilon\right) 2k \log(n-k), \tag{16}$$

then the probability that there exists a solution $\hat{\boldsymbol{x}}$ of the Lasso which satisfies $\operatorname{sign}(\hat{\boldsymbol{x}}) = \operatorname{sign}(\boldsymbol{x}_o)$ is at most $Cn^{-\epsilon}$.

Note: we can have $\operatorname{sign}(\hat{\boldsymbol{x}}) = \operatorname{sign}(\boldsymbol{x}_o)$ w.h.p. only if $\min_{i \in \mathsf{I}} |\boldsymbol{x}_{oi}| > C\lambda$.

Recovering Support

Proof ideas: $\hat{\boldsymbol{x}}$ is optimal if and only if **(looking familiar?)**

$$\boldsymbol{A}^* \left(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}\right) \in \lambda \partial \left\|\cdot\right\|_1 \left(\hat{\boldsymbol{x}}\right). \tag{17}$$

Let $\mathsf{J} = \mathrm{supp}(\hat{\boldsymbol{x}})$ and the condition (17) decomposes into two conditions:

$$\boldsymbol{A}_{\mathsf{J}}^* \left(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}\right) = \lambda \, \mathrm{sign}(\hat{\boldsymbol{x}}_{\mathsf{J}}), \tag{18}$$

$$\left\|\boldsymbol{A}_{\mathsf{J}^c}^* \left(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}\right)\right\|_\infty \leq \lambda. \tag{19}$$

Construct a guess solution vector $\boldsymbol{x}_\star$ by solving a *restricted* Lasso problem:

$$\boldsymbol{x}_\star \in \arg\min_{\mathrm{supp}(\boldsymbol{x}) \subseteq \mathsf{I}} \left\{ \tfrac{1}{2} \left\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\right\|_2^2 + \lambda \left\|\boldsymbol{x}\right\|_1 \right\}, \tag{20}$$

which satisfies the equality constraint (18). We will then be left to check the inequality constraints (19).

Conclusions (of Chapter 3)

Conditions when ℓ^1 minimization find the correct k-sparse solution:

$$\min \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

- **Mutual Coherence:**

$$m = O(k^2).$$

- **Restricted Isometry:**

$$m = O\big(k \log(n/k)\big).$$

- **Phase Transition:**

$$m^\star = \psi\left(\frac{k}{n}\right) n.$$

Recovery is also stable w.r.t. to noise and approximate sparsity.

Assignments

- Reading: Section 3.6 and 3.7 of Chapter 3.
- Written Homework #2.
- Advanced Reading: Section 6.2 of Chapter 6.