Phase Transition and Convex Geometry

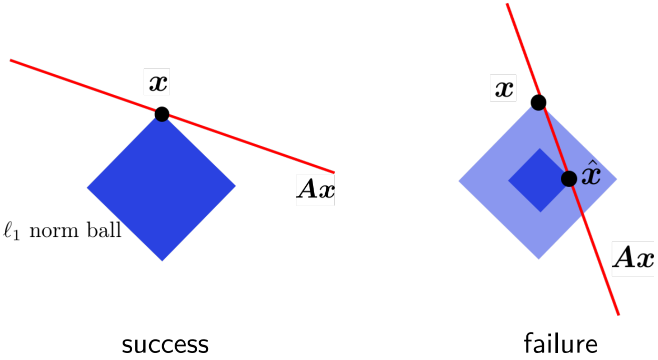
Yuxin Chen

Princeton University, Fall 2020

ℓ_1 minimization for sparse signal recovery

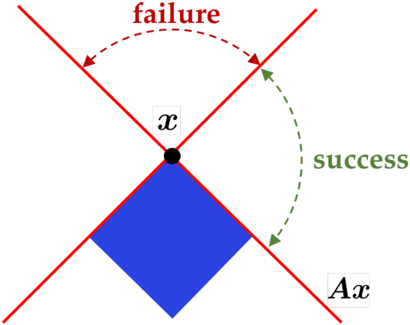
$$\begin{aligned} \text{minimize}_{\boldsymbol{x}\in\mathbb{R}^p} \quad & \|\boldsymbol{x}\|_1 \\ \text{s.t.} \quad & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^n \end{aligned}$$

What is the probability that ℓ_1 minimization succeeds in recovering $\boldsymbol{x}$?



Probability of success of ℓ_1 minimization

Suppose $\boldsymbol{A}$ is randomly generated



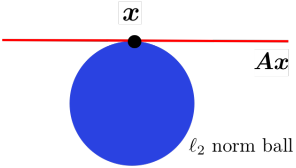
If $\boldsymbol{x} = [0, 1]^\top$ and $\boldsymbol{A} = [a_1, a_2] \in \mathbb{R}^{1\times 2}$ (i.i.d. Gaussian), then

$$\mathbb{P}\{\ell_1\text{-min succeeds}\} = \frac{1}{2} \quad \text{(proved by figure)}$$

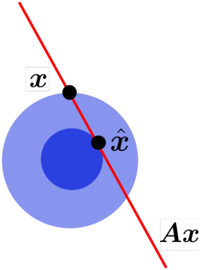
How about ℓ_2 minimization?

Suppose $\boldsymbol{A}$ is randomly generated (e.g. i.i.d. Gaussian), and consider

$$\begin{aligned} \text{minimize}_{\boldsymbol{x}\in\mathbb{R}^p} \quad & \|\boldsymbol{x}\|_2 \\ \text{s.t.} \quad & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^n \end{aligned}$$



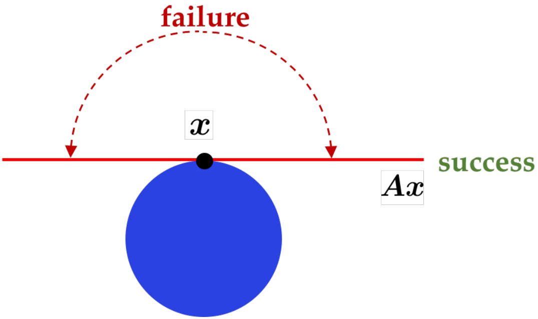
success



failure

Probability of success of ℓ_2 minimization

Suppose $\boldsymbol{A}$ is randomly generated



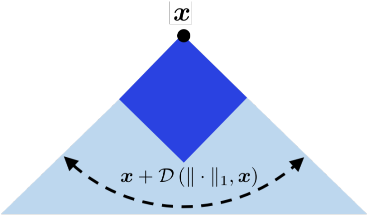
If $\boldsymbol{x} = [0, 1]^\top$ and $\boldsymbol{A} = [a_1, a_2] \in \mathbb{R}^{1\times 2}$ (i.i.d. Gaussian), then

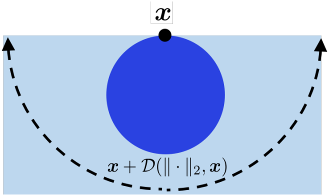
$$\mathbb{P}\{\ell_2\text{-min succeeds}\} = 0 \quad \text{(proved by figure)}$$

Key metric: volume of descent cone

Suppose $\boldsymbol{A}$ is randomly generated, and consider

$$\begin{aligned} \text{minimize}_{\boldsymbol{x}\in\mathbb{R}^p} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^n \end{aligned} \tag{12.1}$$





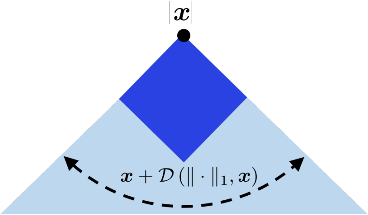
The success probability of (12.1) depends on the volume of the descent cone

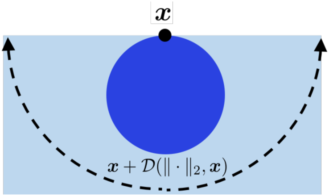
$$\mathcal{D}(f, \boldsymbol{x}) := \{\boldsymbol{h} : \exists\, \epsilon > 0 \text{ s.t. } f(\boldsymbol{x} + \epsilon\boldsymbol{h}) \leq f(\boldsymbol{x})\}$$

Key metric: volume of descent cone

Suppose $\boldsymbol{A}$ is randomly generated, and consider

$$\begin{aligned} \text{minimize}_{\boldsymbol{x}\in\mathbb{R}^p} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^n \end{aligned} \tag{12.1}$$





We need to compute the probability of 2 convex cones sharing a ray:

$$\mathbb{P}\Big\{(12.4) \text{ succeeds}\Big\} = \mathbb{P}\Big\{\mathcal{D}(f, \boldsymbol{x}) \cap \{\boldsymbol{h} : \boldsymbol{A}\boldsymbol{h} = \boldsymbol{0}\} = \{\boldsymbol{0}\}\Big\}$$

Kinematic formula

Lemma 12.1 (Theorem 6.5.6, Schneider & Wolfgang '08)

Let $\mathcal{C}, \mathcal{K} \subseteq \mathbb{R}^d$ be convex cones, and $\boldsymbol{Q}$ a random orthogonal basis:

$$\mathbb{P}\left\{\mathcal{C} \cap \boldsymbol{Q}\mathcal{K} \neq \{\mathbf{0}\}\right\} = \sum_{i=0}^{d} \left(1 + (-1)^{i+1}\right) \sum_{j=i}^{d} \nu_i(\mathcal{C}) \nu_{d+i-j}(\mathcal{K}),$$

where $\nu_k \geq 0$ is called the kth intrinsic volume.

- Exact but not workable formula!
- Calls for a simpler expression

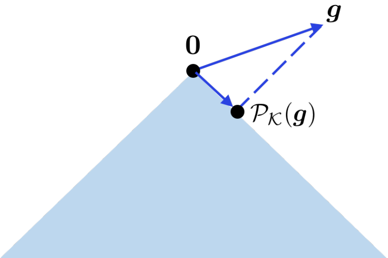
Statistical dimension and Gaussian width

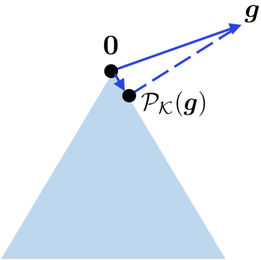
Definition 12.2 (Statistical dimension)

For any convex cone $\mathcal{K}$, its statistical dimension is defined as

$$\text{stat-dim}(\mathcal{K}) := \mathbb{E}[\|\mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\|_2^2]$$

where $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$; $\mathcal{P}_{\mathcal{K}}(\boldsymbol{g}) := \arg\min_{\boldsymbol{z} \in \mathcal{K}} \|\boldsymbol{g} - \boldsymbol{z}\|_2$: Euclidean projection





Statistical dimension and Gaussian width

Definition 12.2 (Statistical dimension)

For any convex cone $\mathcal{K}$, its statistical dimension is defined as

$$\text{stat-dim}(\mathcal{K}) := \mathbb{E}[\|\mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\|_2^2]$$

where $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$; $\mathcal{P}_{\mathcal{K}}(\boldsymbol{g}) := \arg\min_{\boldsymbol{z}\in\mathcal{K}} \|\boldsymbol{g} - \boldsymbol{z}\|_2$: Euclidean projection

- If $\mathcal{K}$ is k-dimensional subspace, then

$$\text{stat-dim}(\mathcal{K}) = k \qquad \text{(so it is indeed a measure of ``dimension'')}$$

Statistical dimension and Gaussian width

Definition 12.2 (Statistical dimension)

For any convex cone $\mathcal{K}$, its statistical dimension is defined as

$$\text{stat-dim}(\mathcal{K}) := \mathbb{E}[\|\mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\|_2^2]$$

where $\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$; $\mathcal{P}_{\mathcal{K}}(\boldsymbol{g}) := \arg\min_{\boldsymbol{z} \in \mathcal{K}} \|\boldsymbol{g} - \boldsymbol{z}\|_2$: Euclidean projection

- A related definition: Gaussian width

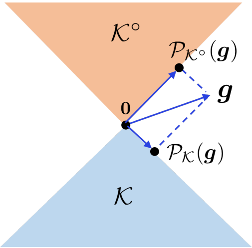
$$w(\mathcal{K}) := \mathbb{E}\left[\sup_{\boldsymbol{z} \in \mathcal{K}, \|\boldsymbol{z}\|_2 = 1} \langle \boldsymbol{z}, \boldsymbol{g} \rangle\right]$$

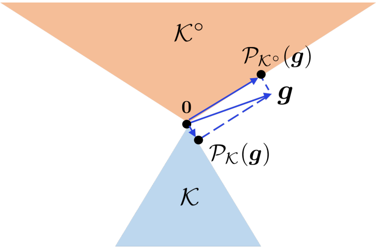
- (Homework) $w^2(\mathcal{K}) \le \text{stat-dim}(\mathcal{K}) \le w^2(\mathcal{K}) + 1$

Polar cone

The polar cone of a convex cone $\mathcal{K}$ is defined as

$$\mathcal{K}^\circ = \{\boldsymbol{y} \mid \boldsymbol{y}^\top \boldsymbol{x} \le 0, \quad \forall \boldsymbol{x} \in \mathcal{K}\}$$





- Moreau's decomposition:

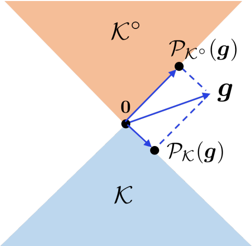
$$\boldsymbol{g} = \mathcal{P}_{\mathcal{K}}(\boldsymbol{g}) + \mathcal{P}_{\mathcal{K}^\circ}(\boldsymbol{g})$$

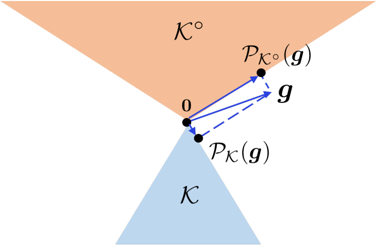
where $\langle \mathcal{P}_{\mathcal{K}}(\boldsymbol{g}), \mathcal{P}_{\mathcal{K}^\circ}(\boldsymbol{g}) \rangle = 0$

Polar cone

The polar cone of a convex cone $\mathcal{K}$ is defined as

$$\mathcal{K}^\circ = \{\boldsymbol{y} \mid \boldsymbol{y}^\top \boldsymbol{x} \leq 0, \quad \forall \boldsymbol{x} \in \mathcal{K}\}$$





- $\mathcal{K}$ is called *self-dual* if

$$\mathcal{K} = \underbrace{-\mathcal{K}^\circ}_{\text{dual cone}}$$

- If $\mathcal{K}$ is self-dual in $\mathbb{R}^d$, then

$$\text{stat-dim}(\mathcal{K}) = d/2 \qquad \text{(by Moreau's decomposition)}$$

Examples

- Nonnegative orthant $\mathbb{R}_+^d$ (self-dual)

$$\text{stat-dim}\big(\mathbb{R}_+^d\big) = \frac{1}{2}d$$

- Second-order cone $\mathbb{L}^d := \left\{ \begin{bmatrix} \boldsymbol{x} \\ \tau \end{bmatrix} \in \mathbb{R}^{d+1} : \|\boldsymbol{x}\|_2 \leq \tau \right\}$ (self-dual)

$$\text{stat-dim}\big(\mathbb{L}^d\big) = \frac{1}{2}(d+1)$$

- Set of symmetric positive semidefinite matrices $\mathbb{S}_+^{d\times d}$ (self-dual)

$$\text{stat-dim}\big(\mathbb{S}_+^{d\times d}\big) = \frac{1}{2} \cdot \frac{d(d+1)}{2}$$

Second-order cone is self-dual

The dual cone of $\mathbb{L}^d$ is

$$\begin{aligned}
\mathcal{C} &= \left\{ \left[\begin{array}{c} \boldsymbol{y} \\ \alpha \end{array} \right] \middle| \, 0 \leq \boldsymbol{y}^\top \boldsymbol{x} + \alpha\tau, \ \forall \left[\begin{array}{c} \boldsymbol{x} \\ \tau \end{array} \right] \in \mathbb{L}^d \right\} \\
&= \left\{ \left[\begin{array}{c} \boldsymbol{y} \\ \alpha \end{array} \right] \middle| \, 0 \leq \inf_{\tau \geq 0} \inf_{\boldsymbol{x}: \|\boldsymbol{x}\|_2 \leq \tau} \left(\boldsymbol{y}^\top \boldsymbol{x} + \alpha\tau \right) \right\} \\
&= \left\{ \left[\begin{array}{c} \boldsymbol{y} \\ \alpha \end{array} \right] \middle| \, 0 \leq \inf_{\tau \geq 0} \inf_{\boldsymbol{x}: \|\boldsymbol{x}\|_2 \leq \tau} \left(-\|\boldsymbol{y}\|_2 \|\boldsymbol{x}\|_2 + \alpha\tau \right) \right\} \\
&= \left\{ \left[\begin{array}{c} \boldsymbol{y} \\ \alpha \end{array} \right] \middle| \, 0 \leq \inf_{\tau \geq 0} \left(\alpha - \|\boldsymbol{y}\|_2 \right) \tau \right\}
\end{aligned}$$

Since $\tau \geq 0$, one has $\left[\begin{array}{c} \boldsymbol{y} \\ \alpha \end{array} \right] \in \mathcal{C}$ iff $\|\boldsymbol{y}\|_2 \leq \alpha$ and hence $\mathcal{C} = \mathbb{L}^d$

Polarity relation

The statistical dimension can be expressed in terms of the polar cone:

$$\text{stat-dim}(\mathcal{K}) = \mathbb{E}\left[\text{dist}^2\left(\boldsymbol{g}, \mathcal{K}^\circ\right)\right] := \mathbb{E}\left[\inf_{\boldsymbol{z}\in\mathcal{K}^\circ} \|\boldsymbol{g}-\boldsymbol{z}\|_2^2\right]$$

where $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$

- A direct consequence:

$$\text{dist}\left(\boldsymbol{g}, \mathcal{K}\right) = \|\boldsymbol{g} - \mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\|_2 \overset{\text{Moreau decomposition}}{=} \|\mathcal{P}_{\mathcal{K}^\circ}(\boldsymbol{g})\|_2$$

Approximate kinematic formula

$\mathcal{C}, \mathcal{K} \in \mathbb{R}^d$: convex cones; $\boldsymbol{Q} \in \mathbb{R}^{d\times d}$: random orthogonal basis

Theorem 12.3 (Amelunxen, Lotz, McCoy & Tropp '13)

$$\begin{aligned} \textit{stat-dim}(\mathcal{C}) + \textit{stat-dim}(\mathcal{K}) \ &\leq \ d - \Theta(\sqrt{d \log d}) \\ \Longrightarrow \quad \boldsymbol{Q}\mathcal{C} \cap \mathcal{K} &= \{\boldsymbol{0}\} \quad \textit{with high prob.} \\ \textit{stat-dim}(\mathcal{C}) + \textit{stat-dim}(\mathcal{K}) \ &\geq \ d + \Theta(\sqrt{d \log d}) \\ \Longrightarrow \quad \boldsymbol{Q}\mathcal{C} \cap \mathcal{K} &\neq \{\boldsymbol{0}\} \quad \textit{with high prob.} \end{aligned}$$

- 2 randomly rotated cones share a ray $\Longleftrightarrow$ their aggregate statistical dimension exceeds ambient dimension
- Sharp concentration: the fluctuation does not exceed $O(\sqrt{d \log d})$

Example: logistic regression

Suppose we obtain n independent *binary* samples:

$$y_i = \begin{cases} 1, & \text{with prob. } \frac{1}{1+\exp(-\boldsymbol{a}_i^\top \boldsymbol{x})} \\ -1, & \text{with prob. } \frac{1}{1+\exp(\boldsymbol{a}_i^\top \boldsymbol{x})} \end{cases} \qquad 1 \le i \le n$$

where $\{\boldsymbol{a}_i \in \mathbb{R}^p\}$: known design vectors; $\boldsymbol{x} \in \mathbb{R}^p$: unknown signal

- the likelihood for each y_i:

$$\begin{aligned} \mathcal{L}(\boldsymbol{x}; y_i) &= \frac{1}{1+\exp(-\boldsymbol{a}_i^\top \boldsymbol{x})} \mathbb{1}\{y_i = 1\} + \frac{1}{1+\exp(\boldsymbol{a}_i^\top \boldsymbol{x})} \mathbb{1}\{y_i = -1\} \\ &= \frac{1}{1+\exp(-y_i \boldsymbol{a}_i^\top \boldsymbol{x})} \end{aligned}$$

Example: logistic regression

Maximum likelihood estimation (logistic regression)

$$\text{minimize}_{\boldsymbol{x}} \;\; -\sum_{i=1}^{n} \log \mathcal{L}(\boldsymbol{x}; y_i) = \sum_{i=1}^{n} \log \left\{ 1 + \exp(-y_i \boldsymbol{a}_i^\top \boldsymbol{x}) \right\}$$

- Consider a simple case

$$\text{true signal } \boldsymbol{x} = \boldsymbol{0} \text{ (global null)}; \qquad \boldsymbol{a}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_p)$$

- We'd naturally hope the MLE $\hat{\boldsymbol{x}}$ to be small (since $\boldsymbol{x} = \boldsymbol{0}$)
- **Question:** is $\|\hat{\boldsymbol{x}}\|_2$ always small under the global null (i.e. $\boldsymbol{x} = \boldsymbol{0}$)?

Example: logistic regression

Fact 12.4 (Cover '65; Sur, Chen, Candes '17)

Suppose $\boldsymbol{x} = \mathbf{0}$. *If* $p > n/2 - \Theta(\sqrt{n \log n})$, *then* $\|\hat{\boldsymbol{x}}\|_2 = \infty$ *w.h.p.*

- $n = 2p$ is indeed a sharp boundary in the sense that $\|\hat{\boldsymbol{x}}\|_2 \lesssim 1$ if $n/p > 2$ (Sur, Chen, Candes '17)

Proof of Fact 12.4

Note that $\log\left\{1+\exp(-y_i \boldsymbol{a}_i^\top \boldsymbol{x})\right\} \geq 0$. Thus, if $\exists\, \hat{\boldsymbol{x}}$ s.t.

$$y_i \boldsymbol{a}_i^\top \hat{\boldsymbol{x}} = +\infty, \qquad 1 \leq i \leq n, \tag{12.2}$$

then $\log\left\{1+\exp(-y_i \boldsymbol{a}_i^\top \boldsymbol{x})\right\} = 0$ for all i, and hence $\hat{\boldsymbol{x}}$ must be the MLE. In this case, $\|\hat{\boldsymbol{x}}\|_2 = \infty$

Proof of Fact 12.4 (cont.)

It remains to check when we can find $\hat{\boldsymbol{x}}$ obeying (12.2), or equivalently, when we have

$$\underbrace{\left\{\boldsymbol{u} \mid u_i = y_i \boldsymbol{a}_i^\top \boldsymbol{x},\ \boldsymbol{x} \in \mathbb{R}^p\right\}}_{\boldsymbol{Q}\mathcal{C}:\ p\text{-dimensional}} \cap \underbrace{\mathbb{R}^n_+}_{\mathcal{K}} \neq \{\boldsymbol{0}\} \tag{12.3}$$

Note that $\boldsymbol{y}$ is independent of $\boldsymbol{A}$ when $\boldsymbol{x} = \boldsymbol{0}$. By Theorem 12.3, if

$$p + \underbrace{\text{stat-dim}(\mathbb{R}^n_+)}_{n/2} > n + \Theta(\sqrt{n \log n}) \qquad (\text{or } p > n/2 + \Theta(\sqrt{n \log n})),$$

then (12.3) holds. This establishes the claim that $\hat{\boldsymbol{x}}$ is unbounded.

Phase transition for inverse problems

Suppose $\boldsymbol{A} \in \mathbb{R}^{n\times p}$ is i.i.d. Gaussian, and consider

$$\begin{aligned} \text{minimize}_{\boldsymbol{x}\in\mathbb{R}^p} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \end{aligned} \tag{12.4}$$

Key: convex geometry

$$(12.4) \text{ succeeds}$$

$$\Updownarrow$$

$$\{\boldsymbol{h} : \boldsymbol{A}\boldsymbol{h} = \boldsymbol{0}\} \cap \mathcal{D}(f, \boldsymbol{x}) = \{\boldsymbol{0}\}$$

$$\Updownarrow$$

$$\underbrace{\text{stat-dim}(\{\boldsymbol{h} : \boldsymbol{A}\boldsymbol{h} = \boldsymbol{0}\})}_{=\,p-n} + \text{stat-dim}(\mathcal{D}(f, \boldsymbol{x})) < p \quad \text{(by Theorem 12.3)}$$

Phase transition for inverse problems

Suppose $\boldsymbol{A} \in \mathbb{R}^{n \times p}$ is i.i.d. Gaussian, and consider

$$\begin{aligned} \text{minimize}_{\boldsymbol{x} \in \mathbb{R}^p} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \end{aligned} \tag{12.4}$$

Theorem 12.5 (Amelunxen, Lotz, McCoy & Tropp '13)

$$n > \textit{stat-dim}(\mathcal{D}(f, \boldsymbol{x})) + \Theta(\sqrt{p \log p})$$

$$\Longrightarrow \qquad (12.4) \textit{ succeeds with high prob.}$$

$$n < \textit{stat-dim}(\mathcal{D}(f, \boldsymbol{x})) - \Theta(\sqrt{p \log p})$$

$$\Longrightarrow \qquad (12.4) \textit{ fails with high prob.}$$

Computing statistical dimension of descent cones?

Example: the decent cone w.r.t. ℓ_∞ norm

$$\text{stat-dim}\left(\mathcal{D}(\|\cdot\|_\infty, \boldsymbol{x})\right) = d - s/2$$

where $s = \#\{i : |x_i| = \|\boldsymbol{x}\|_\infty\}$

Proof: WLOG, suppose $\boldsymbol{x} = [1, \cdots, 1, x_{s+1}, \cdots, x_d]^\top$ with $1 > x_{s+1} \geq \cdots \geq x_d \geq 0$. Then

$$\mathcal{D}(\|\cdot\|_\infty, \boldsymbol{x}) = (\mathbb{R}^s_-) \times \mathbb{R}^{d-s}$$

$$\Longrightarrow \quad \text{stat-dim}\left(\mathcal{D}(\|\cdot\|_\infty, \boldsymbol{x})\right) = \frac{1}{2}s + d - s = d - \frac{1}{2}s$$

Computing statistical dimension of descent cone?

In general, there is a duality between descent cone and $\underbrace{\text{subdifferentials}}_{\text{set of subgradients}}$

$$(\mathcal{D}(f, \boldsymbol{x}))^{\circ} = \mathsf{cone}\left(\partial f(\boldsymbol{x})\right) := \bigcup_{\tau \geq 0} \tau \partial f(\boldsymbol{x})$$

$$\Rightarrow \qquad \mathsf{stat\text{-}dim}\Big(\mathcal{D}\left(f, \boldsymbol{x}\right)\Big) = \mathbb{E}\left[\inf_{\tau \geq 0} \min_{\boldsymbol{u} \in \partial f(\boldsymbol{x})} \|\boldsymbol{g} - \tau \boldsymbol{u}\|_2^2\right]$$

Computing statistical dimension of descent cone?

In general, there is a duality between descent cone and $\underbrace{\text{subdifferentials}}_{\text{set of subgradients}}$

$$(\mathcal{D}(f, \boldsymbol{x}))^{\circ} = \mathsf{cone}\left(\partial f(\boldsymbol{x})\right) := \bigcup_{\tau \geq 0} \tau \partial f(\boldsymbol{x})$$

Lemma 12.6 (informal, Amelunxen, Lotz, McCoy & Tropp '13)

$$\textit{stat-dim}\Big(\mathcal{D}\left(f, \boldsymbol{x}\right)\Big) \;\approx\; \inf_{\tau \geq 0} \mathbb{E}\left[\min_{\boldsymbol{u} \in \partial f(\boldsymbol{x})} \|\boldsymbol{g} - \tau \boldsymbol{u}\|_2^2\right]$$

Example: ℓ_1 minimization

WLOG, suppose $x_1, \cdots, x_k > 0$, $x_{k+1} = \cdots = x_p = 0$.

$$\mathbb{E}\left[\min_{\boldsymbol{u}\in\partial\|\boldsymbol{x}\|_1} \|\boldsymbol{g} - \tau\boldsymbol{u}\|_2^2\right] = \mathbb{E}\left[\sum_{i=1}^{k}(g_i - \tau)^2 + \sum_{i=k+1}^{p}\min_{|u_i|\le 1}(g_i - \tau u_i)^2\right]$$

$$= k\left(1+\tau^2\right) + (p-k)\,\mathbb{E}\left[\min_{|u_i|\le 1}(g_i - \tau u_i)^2\right]$$

$$= k\left(1+\tau^2\right) + (p-k)\,\mathbb{E}\left[(|g_i| - \tau)_+^2\right]$$

By Lemma 12.6,

$$\text{stat-dim}\Big(\mathcal{D}\left(\|\cdot\|_1, \boldsymbol{x}\right)\Big) \;\approx\; \inf_{\tau\ge 0}\mathbb{E}\left[\min_{\boldsymbol{u}\in\partial\|\boldsymbol{x}\|_1}\|\boldsymbol{g} - \tau\boldsymbol{u}\|_2^2\right]$$

$$= \inf_{\tau\ge 0}\left\{k\left(1+\tau^2\right) + (p-k)\sqrt{\frac{2}{\pi}}\int_{\tau}^{\infty}(z-\tau)^2 e^{-z^2}\mathrm{d}z\right\}$$

Numerical phase transition

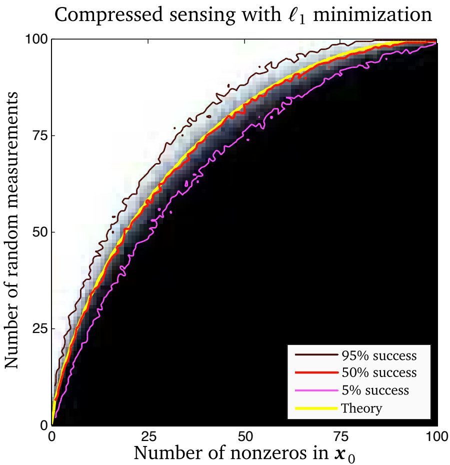


Figure credit: Amelunxen, Lotz, McCoy, & Tropp '13

Reference

- "*Living on the edge: Phase transitions in convex programs with random data*," D. Amelunxen, M. Lotz, M. McCoy, J. Tropp, *Information and Inference*, 2014.
- "*Neighborliness of randomly projected simplices in high dimensions*," D. Donoho and J. Tanner, *PNAS*, 2005.
- "*Stochastic and integral geometry*," R. Schneider and W. Wolfgang, *Springer Science & Business Media*, 2008.
- "*Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition*," T. Cover, *IEEE trans. on electronic computers*, 1965.
- "*The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square*," P. Sur, Y. Chen, and E. Candes, 2017.