Living on the Edge

Phase Transitions in Convex Programs with Random Data

Joel A. Tropp

Computing + Mathematical Sciences
California Institute of Technology

Joint with Michael McCoy (CoFacet),
Dennis Amelunxen (CUHK), and Martin Lotz (Manchester)

Research supported in part by ONR, AFOSR, DARPA, and the Sloan Foundation

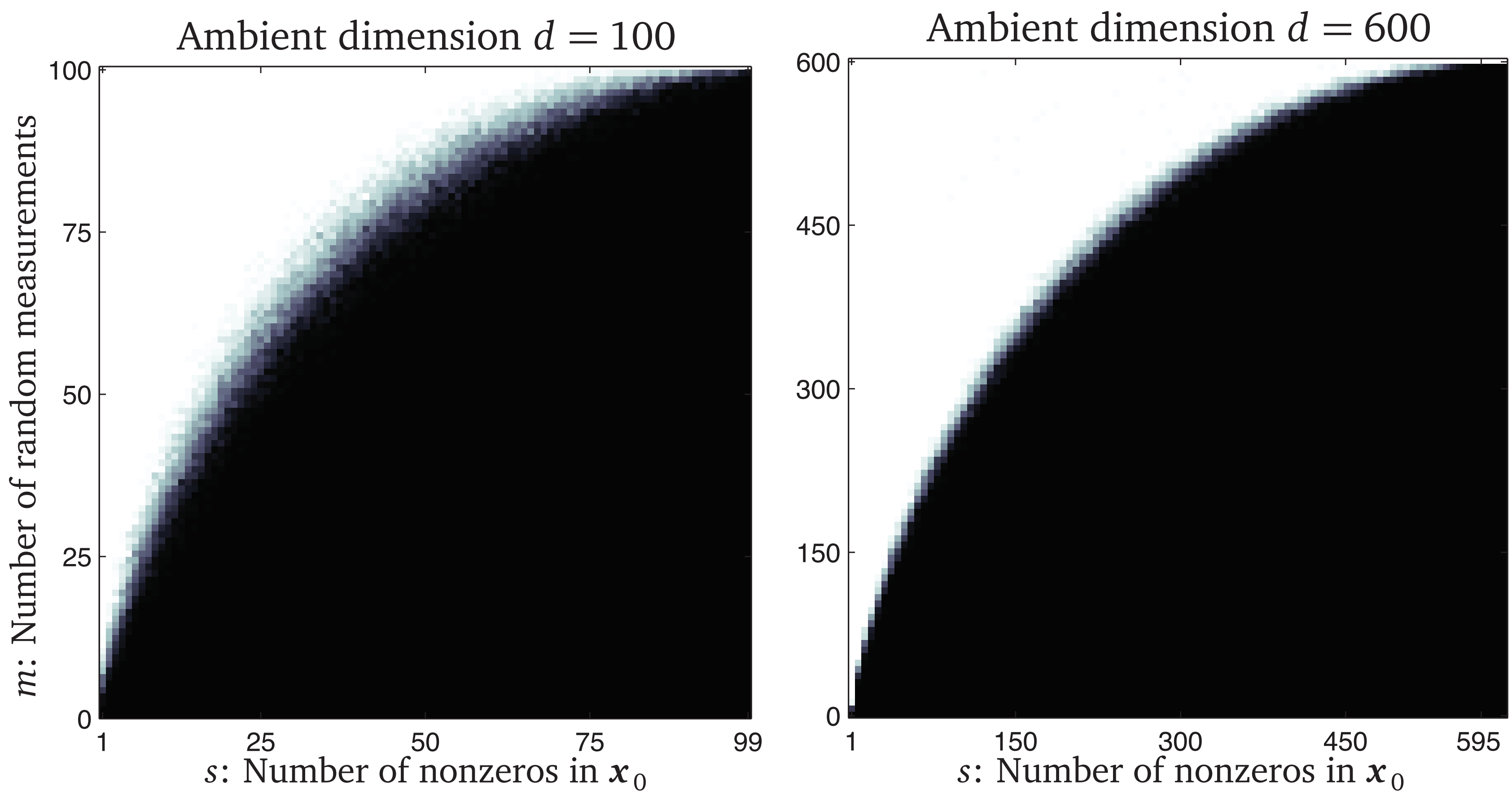
The Compressed Sensing Problem

- Let $\boldsymbol{x}^\natural \in \mathbb{R}^d$ be an unknown vector with s nonzero entries
- Write $\|\cdot\|_1$ for the ℓ_1 norm on $\mathbb{R}^d$
- Let $\boldsymbol{A} \in \mathbb{R}^{m \times d}$ be a Gaussian measurement matrix (iid $N(0,1)$ entries)
- Observe m random measurements: $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}^\natural$
- Produce an estimate $\widehat{\boldsymbol{x}}$ by solving convex program

$$\text{minimize} \quad \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{z}$$

- **Hope:** $\widehat{\boldsymbol{x}} = \boldsymbol{x}^\natural$

A Computer Experiment



Heatmap is probability of success (white = 100%, black = 0%)

Convex Programs with Random Data

Examples...

- **Sensing.** Collect random measurements; reconstruct via optimization
- **Stat and ML.** Random data models; fit model via optimization
- **Coding.** Random channel models; decode via optimization

Motivations...

- **Average-case analysis.** Randomness describes "typical" behavior
- **Fundamental bounds.** Opportunities and limits for convex methods

Research Challenge...

Understand and predict **precise behavior** of random convex programs

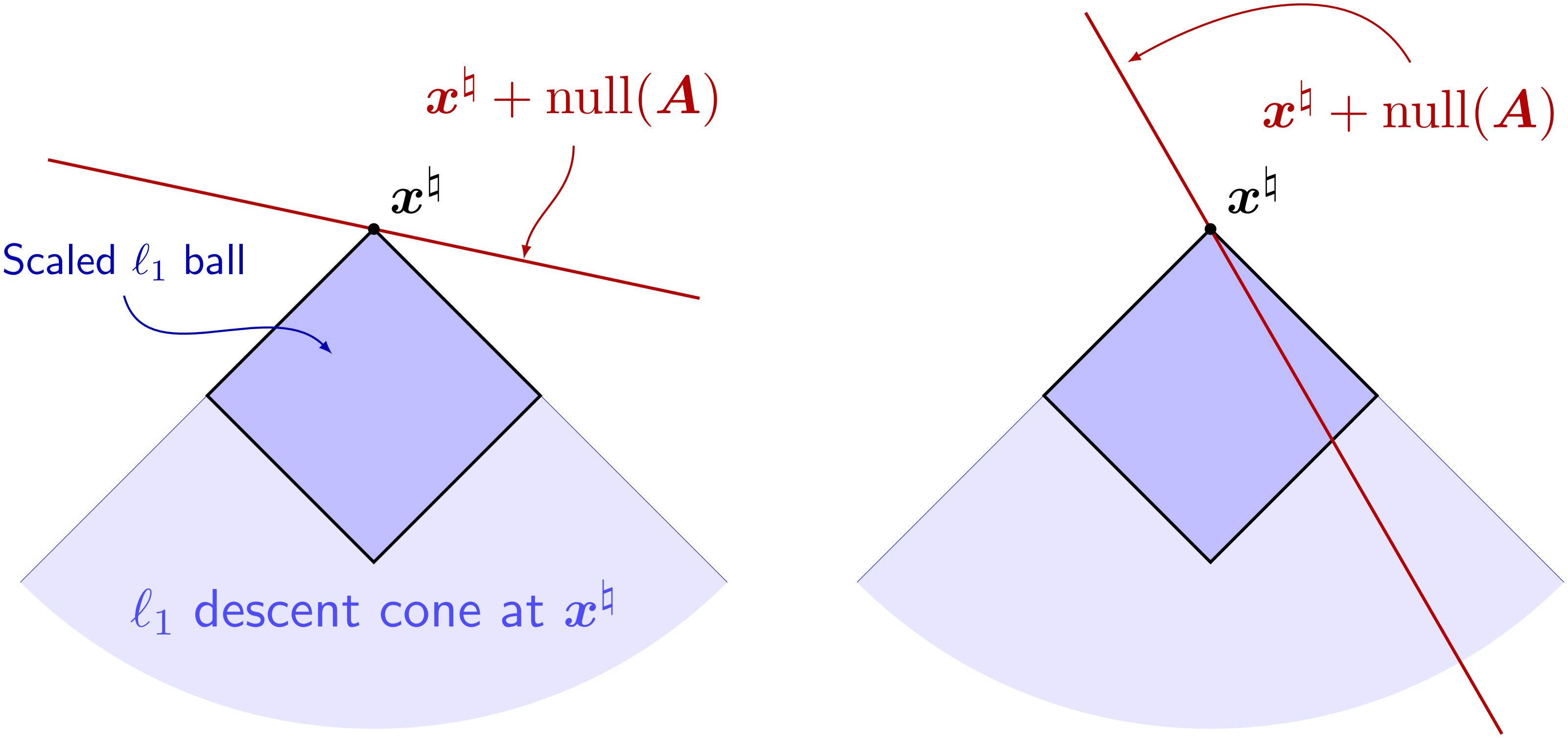
References: Donoho–Maleki–Montanari 2009, Donoho–Johnstone–Montanari 2011, Donoho–Gavish–Montanari 2013

A Theory Emerges...

❧	Vershik & Sporyshev, "An asymptotic estimate for the average number of steps..."	1986
❧	Donoho, "High-dimensional centrally symmetric polytopes..."	2/2005
❧	Rudelson & Vershynin, **"On sparse reconstruction..."**	2/2006
❧	Donoho & Tanner, **"Counting faces of randomly projected polytopes..."**	5/2006
❧	Xu & Hassibi, "Compressed sensing over the Grassmann manifold..."	9/2008
❧	Stojnic, **"Various thresholds for ℓ_1 optimization..."**	7/2009
❧	Bayati & Montanari, "The LASSO risk for gaussian matrices"	8/2010
❧	Oymak & Hassibi, "New null space results and recovery thresholds..."	11/2010
❧	Chandrasekaran, Recht, et al., **"The convex geometry of linear inverse problems"**	12/2010
❧	McCoy & Tropp, "Sharp recovery bounds for convex demixing..."	5/2012
❧	Bayati, Lelarge, & Montanari, "Universality in polytope phase transitions..."	7/2012
❧	Chandrasekaran & Jordan, "Computational & statistical tradeoffs..."	10/2012
❧	Amelunxen, Lotz, McCoy, & Tropp, **"Living on the edge..."**	3/2013
❧	Stojnic, various works	3/2013
❧	Foygel & Mackey, "Corrupted sensing: Novel guarantees..."	5/2013
❧	Oymak & Hassibi, "Asymptotically exact denoising..."	5/2013
❧	McCoy & Tropp, "From Steiner formulas for cones..."	8/2013
❧	McCoy & Tropp, "The achievable performance of convex demixing..."	9/2013
❧	Oymak, Thrampoulidis, & Hassibi, "The squared-error of generalized LASSO..."	11/2013

Ordered by date of release, not date of publication

Geometry of Compressed Sensing Problem



Success! Failure!

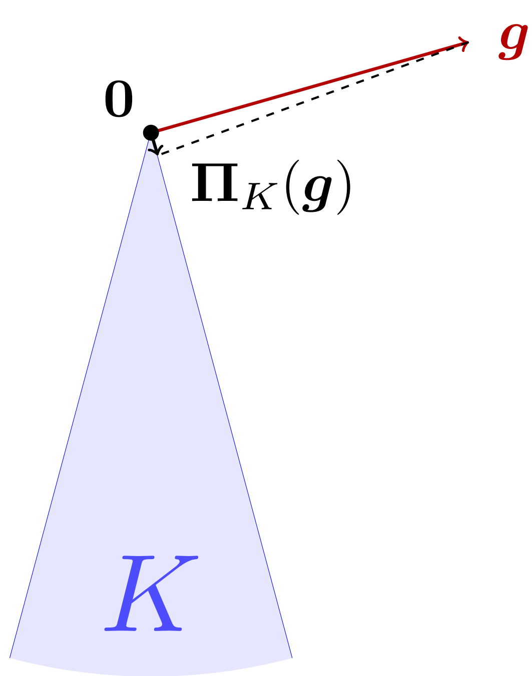
References: Candès–Romberg–Tao 2005, Rudelson–Vershynin 2006, Chandrasekaran et al. 2010, Amelunxen et al. 2013

The Core Question

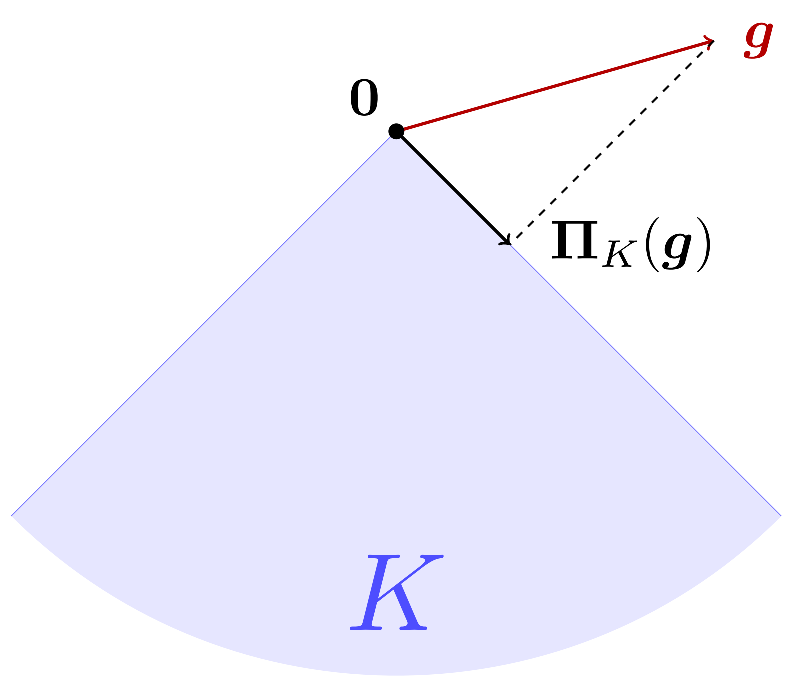
How big is a cone?

Statistical
Dimension

Statistical Dimension: The Motion Picture



small cone



big cone

The Statistical Dimension of a Cone

Definition. The *statistical dimension* $\delta(K)$ of a closed, convex cone K is the quantity

$$\delta(K) := \mathbb{E}\left[\|\mathbf{\Pi}_K(\boldsymbol{g})\|_2^2\right]$$

where

- $\mathbf{\Pi}_K$ is the Euclidean metric projector onto K
- $\boldsymbol{g} \sim \textsc{normal}(\mathbf{0}, \mathbf{I})$ is a standard normal vector

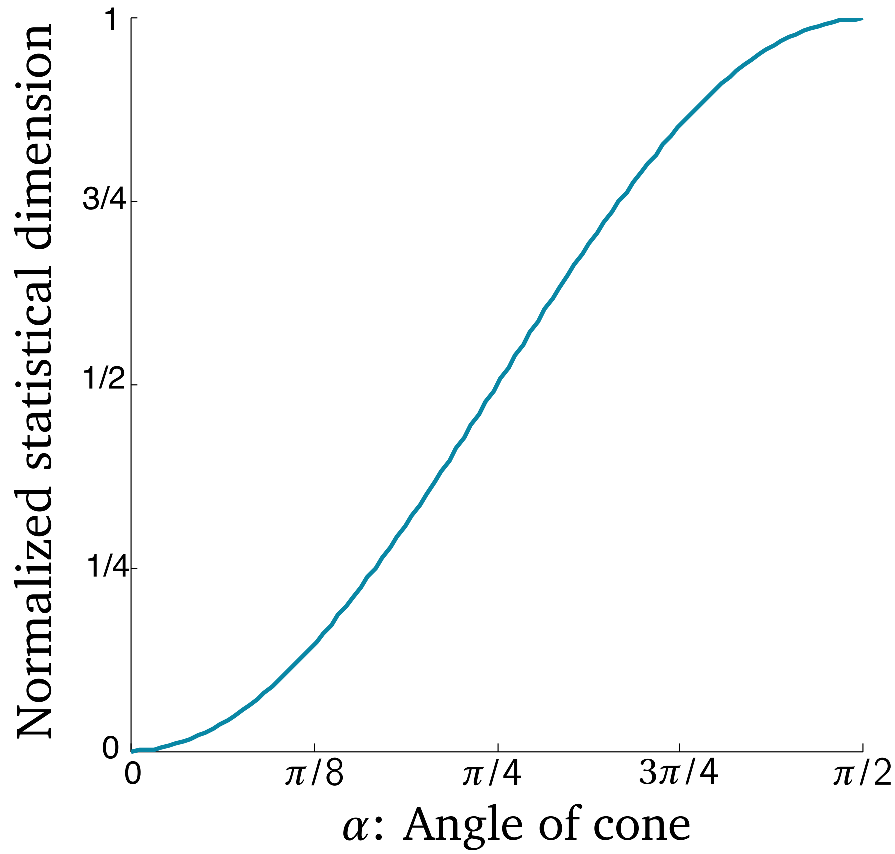
References: Rudelson & Vershynin 2006, Stojnic 2009, Chandrasekaran et al. 2010, Chandrasekaran & Jordan 2012, Amelunxen et al. 2013

Basic Statistical Dimension Calculations

Cone	Notation	Statistical Dimension
j-dim subspace	L_j	j
Nonnegative orthant	$\mathbb{R}_+^d$	$\frac{1}{2}d$
Second-order cone	$\mathbb{L}^{d+1}$	$\frac{1}{2}(d+1)$
Real psd cone	$\mathbb{S}_+^d$	$\frac{1}{4}d(d-1)$
Complex psd cone	$\mathbb{H}_+^d$	$\frac{1}{2}d^2$

References: Chandrasekaran et al. 2010, Amelunxen et al. 2013

Circular Cones

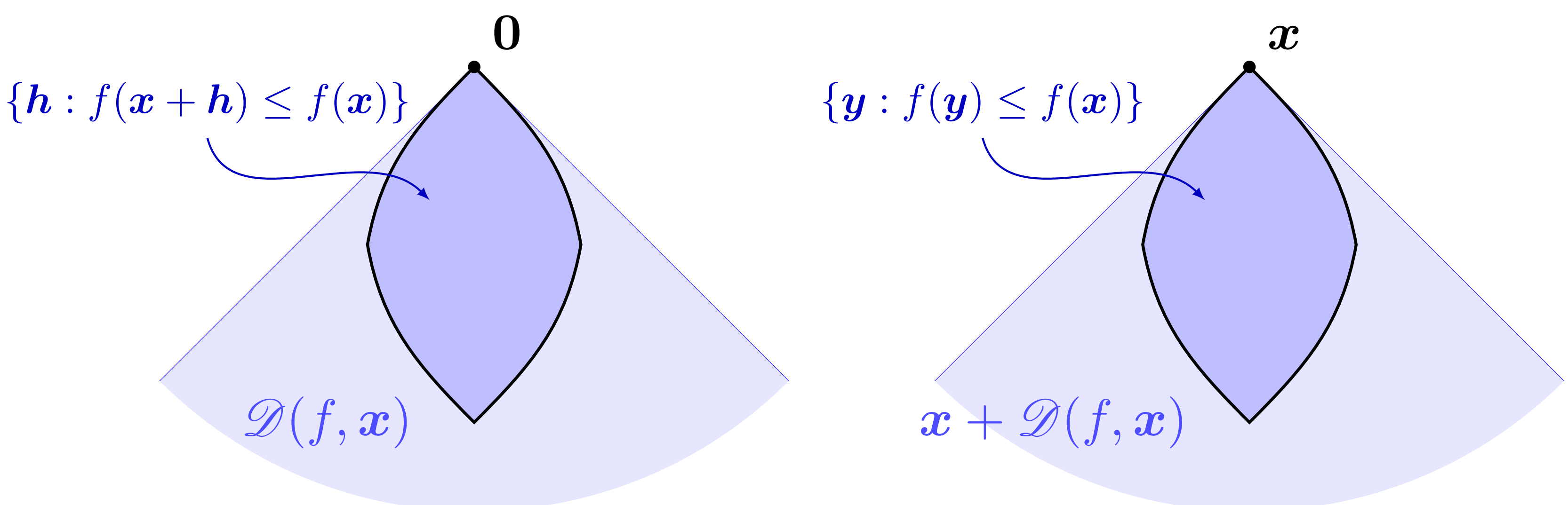


References: Amelunxen et al. 2013, Mu et al. 2013, McCoy & Tropp 2013

Descent Cones

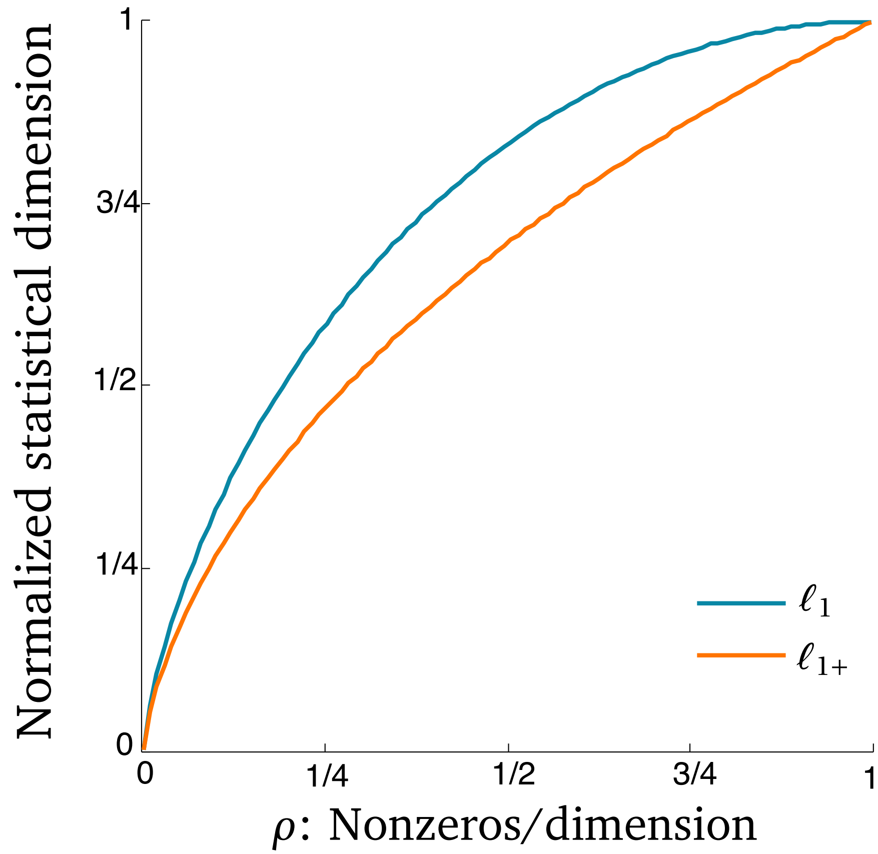
Definition. The *descent cone* of a function f at a point $\boldsymbol{x}$ is

$$\mathscr{D}(f, \boldsymbol{x}) := \{\boldsymbol{h} : f(\boldsymbol{x} + \varepsilon\boldsymbol{h}) \leq f(\boldsymbol{x}) \quad \texttt{for some } \varepsilon > 0\}$$



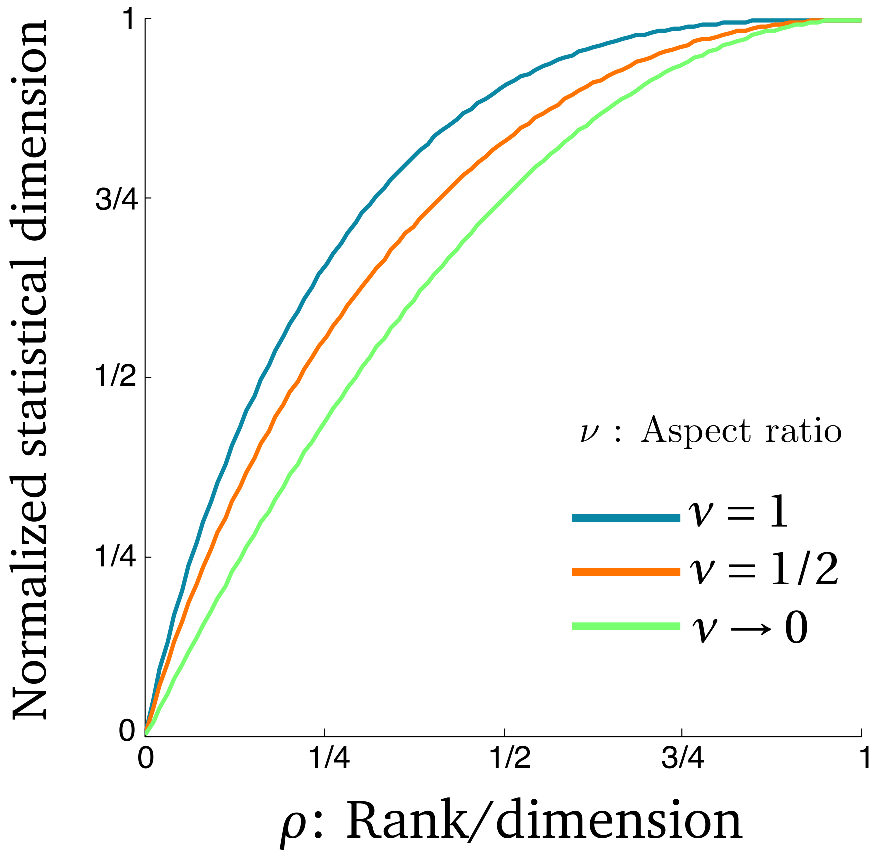
References: Rockafellar 1970, Hiriary-Urruty & Lemaréchal 1996

Descent Cone of ℓ_1 Norm at Sparse Vector



References: Stojnic 2009, Donoho & Tanner 2010, Chandrasekaran et al. 2010, Amelunxen et al. 2013, Mackey & Foygel 2013

Descent Cone of S_1 Norm at Low-Rank Matrix



References: Oymak & Hassibi 2010, Chandrasekaran et al. 2010, Amelunxen et al. 2013, Foygel & Mackey 2013

Statistical Dimension & Phase Transitions

- **Key Question:** When do two randomly oriented cones share a ray?

- **Intuition:** When do two randomly oriented subspaces share a ray?

The Approximate Kinematic Formula

Let C and K be closed convex cones in $\mathbb{R}^d$

$$\delta(C) + \delta(K) \lesssim d \quad \Longrightarrow \quad \mathbb{P}\left\{C \cap \boldsymbol{Q}K = \{\mathbf{0}\}\right\} \approx 1$$

$$\delta(C) + \delta(K) \gtrsim d \quad \Longrightarrow \quad \mathbb{P}\left\{C \cap \boldsymbol{Q}K = \{\mathbf{0}\}\right\} \approx 0$$

where $\boldsymbol{Q}$ is a random orthogonal matrix

References: Amelunxen et al. 2013, McCoy & Tropp 2013

Aside: The Gaussian Width

- The *Gaussian width* $w(K)$ of a convex cone K can be defined as

$$w(K) := \mathbb{E} \sup_{\boldsymbol{x} \in K \cap \mathsf{S}} \langle \boldsymbol{g}, \ \boldsymbol{x} \rangle$$

- The statistical dimension and the Gaussian width are related as

$$w(K)^2 \leq \delta(K) \leq w(K)^2 + 1$$

- Sometimes one parameter is more convenient than the other

- But... **statistical dimension is the canonical extension of the linear dimension to the class of convex cones**

Related: Rudelson–Vershynin 2006, Stojnic 2009, Chandrasekaran et al. 2012

Regularized Linear Inverse Problems

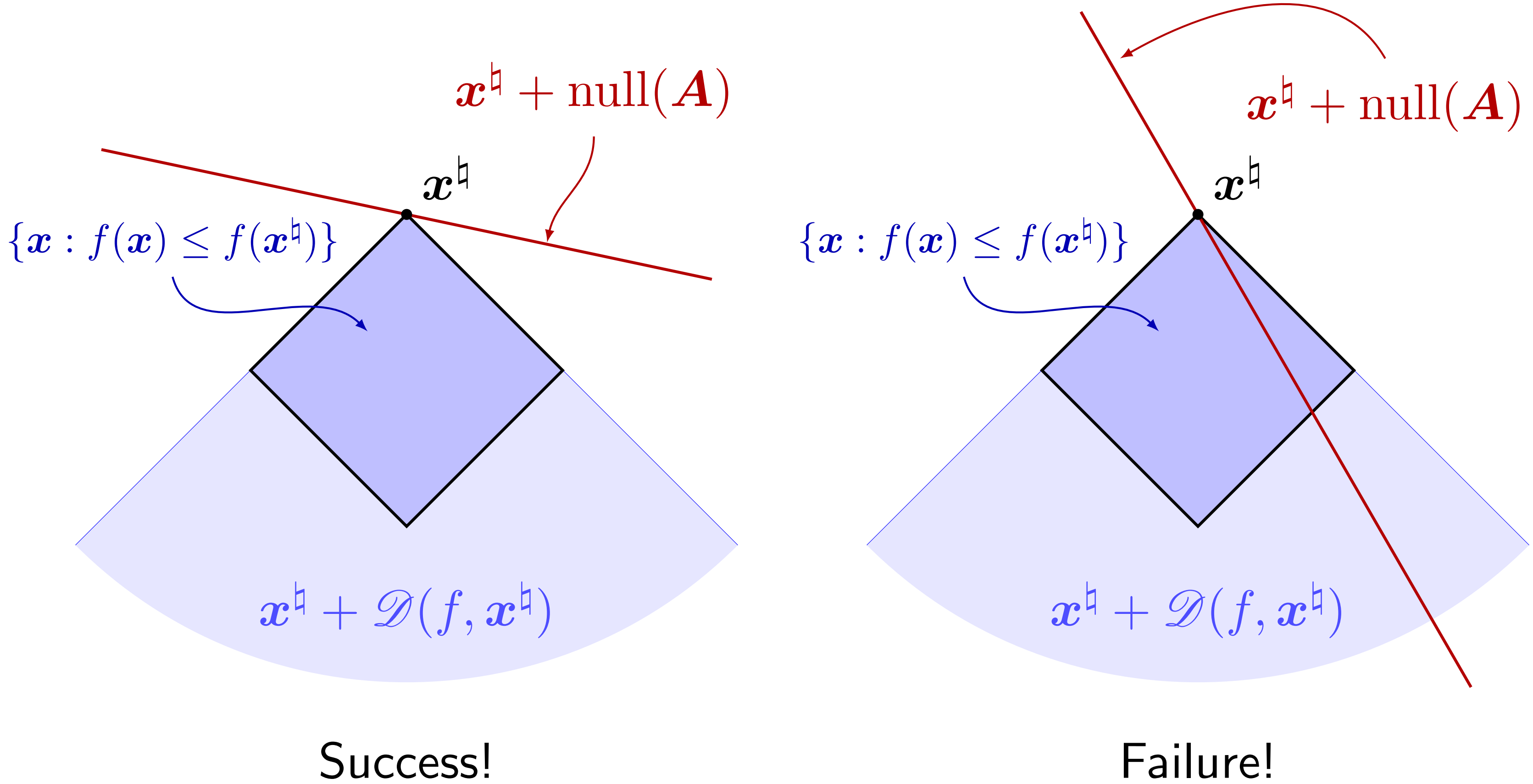
Setup for Linear Inverse Problems

- Let $\boldsymbol{x}^\natural \in \mathbb{R}^d$ be a structured, unknown vector
- Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function that reflects structure
- Let $\boldsymbol{A} \in \mathbb{R}^{m \times d}$ be a measurement operator
- Observe $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}^\natural$
- Find estimate $\widehat{\boldsymbol{x}}$ by solving convex program

$$\text{minimize} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{z}$$

- **Hope:** $\widehat{\boldsymbol{x}} = \boldsymbol{x}^\natural$

Geometry of Linear Inverse Problems



References: Candès–Romberg–Tao 2005, Rudelson–Vershynin 2006, Chandrasekaran et al. 2010, Amelunxen et al. 2013

Linear Inverse Problems with Random Data

Theorem 1. **[CRPW10; ALMT13]** **Assume**

- *The vector* $\boldsymbol{x}^\natural \in \mathbb{R}^d$ *is unknown*
- *The observation* $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}^\natural$ *where* $\boldsymbol{A} \in \mathbb{R}^{m \times d}$ *is standard normal*
- *The vector* $\widehat{\boldsymbol{x}}$ *solves*

$$\text{minimize} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{z}$$

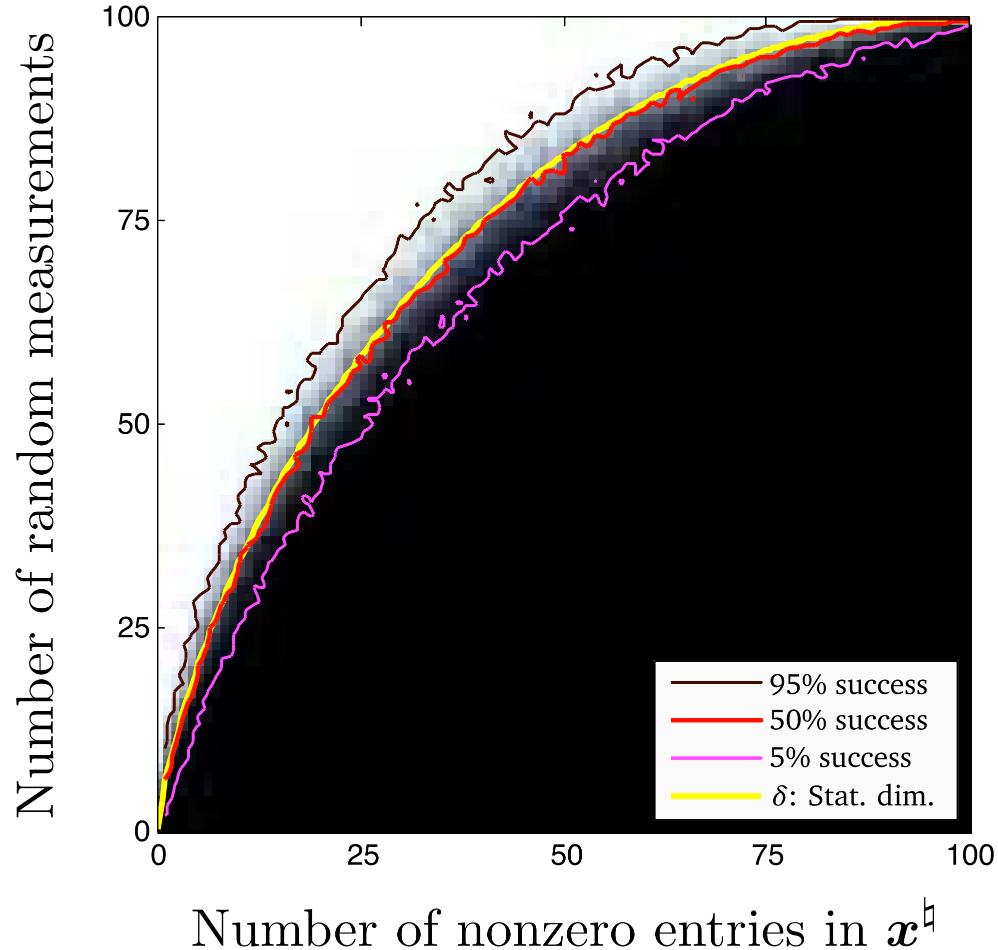
Then *(morally)*

$$m \geq \delta\big(\mathscr{D}(f, \boldsymbol{x}^\natural)\big) \implies \widehat{\boldsymbol{x}} = \boldsymbol{x}^\natural \quad \text{whp} \qquad \textbf{[CRPW10; ALMT13]}$$

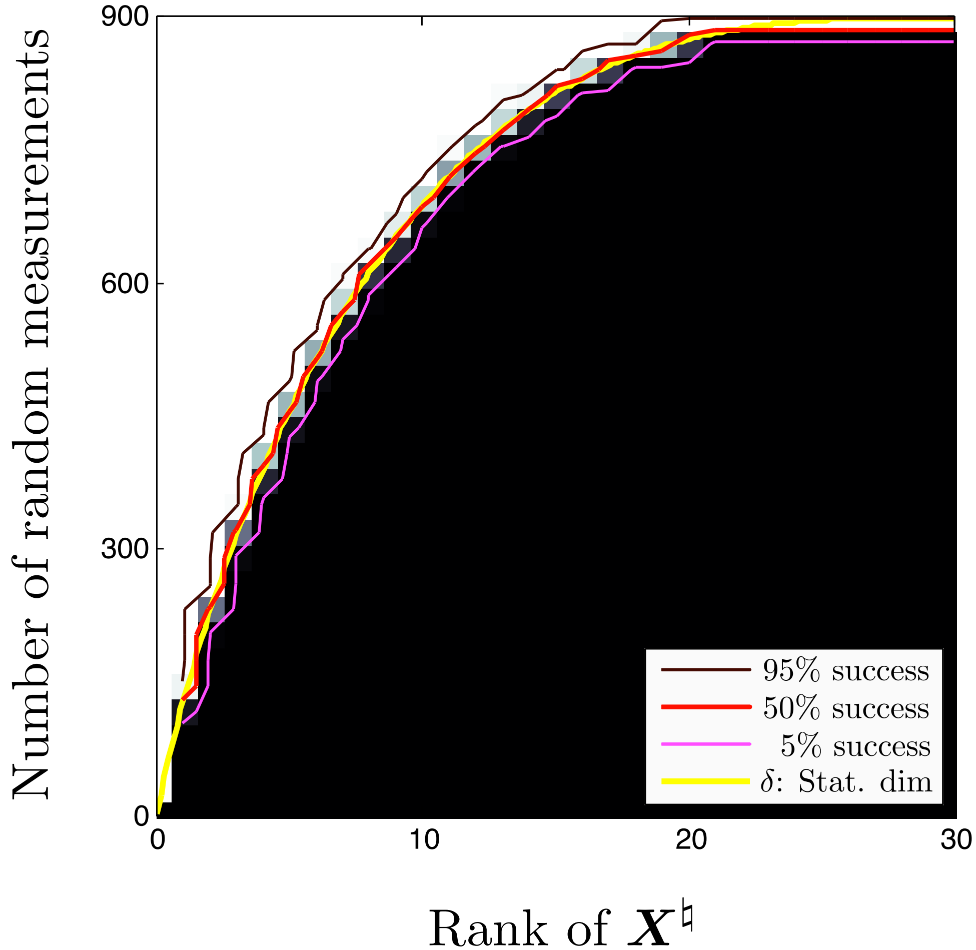
$$m \leq \delta\big(\mathscr{D}(f, \boldsymbol{x}^\natural)\big) \implies \widehat{\boldsymbol{x}} \neq \boldsymbol{x}^\natural \quad \text{whp} \qquad \textbf{[ALMT13]}$$

References: Rudelson–Vershynin 2006, Stojnic 2009, Chandrasekaran et al. 2010, Amelunxen et al. 2013

Sparse Recovery via ℓ_1 Minimization



Low-Rank Recovery via S_1 Minimization



More Examples

Result applies to every (nonsmooth) convex regularizer including...

- ℓ_1 with a nonnegativity constraint
- Simultaneous sparsity and group lasso penalties
- Total variation for signals with sparse gradient
- ℓ_∞ for saturated signals
- The "max norm" for promoting low-rank matrices
- Tensor norms for promoting low-rank tensors
- ...

... but you still have to calculate the statistical dimension!

But My Measurements aren't Gaussian!

There is evidence that the phase transition for many structures is universal (if measurements are centered, isotropic, and incoherent)!

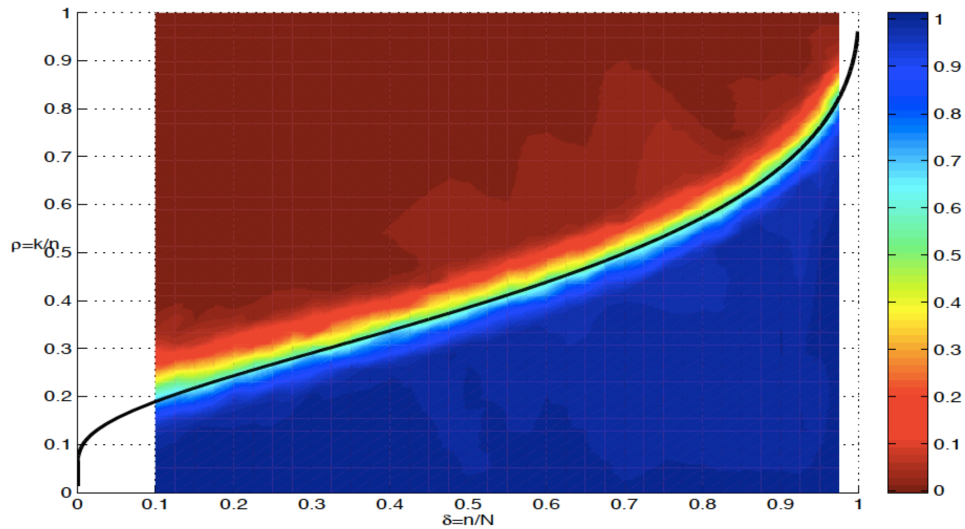


FIGURE 3. *Compressed Sensing from random Fourier measurements.* Shaded attribute: fraction of realizations in which

Theory is currently in progress.

Source: Donoho & Tanner 2009

But My Measurements aren't Isotropic!

Change variables!

- The vector $\boldsymbol{x}^\natural \in \mathbb{R}^d$ is unknown
- The observation $\boldsymbol{z} = \boldsymbol{A\Sigma x}^\natural$ where $\boldsymbol{A} \in \mathbb{R}^{m \times d}$ is standard normal and $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ is nonsingular
- The estimate $\widehat{\boldsymbol{x}}$ solves

$$\text{minimize} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{A\Sigma x} = \boldsymbol{z}$$

- Then the phase transition occurs at $m = \delta\big(\mathscr{D}(f \circ \boldsymbol{\Sigma}^{-1}, \boldsymbol{\Sigma x}^\natural)\big)$

... but you still have to calculate the statistical dimension!

PhD in Computing + Mathematical Sciences

New doctoral program, Department of Computing + Mathematical Sciences at Caltech

- **We are experiencing a convergence of disciplines**
- **A new core emerges:** Optimization, algorithms, statistics, signal processing, networks and markets, mathematical foundations, and applications in science + engineering
- **Web:** `http://www.cms.caltech.edu/academics/grad_cms`

- **Selected faculty:** Shuki Bruck, Venkat Chandrasekaran, Mani Chandy, Mathieu Desbrun, John Doyle, Michelle Effros, Babak Hassibi, Tom Hou, Victoria Kostina, Katrina Ligett, Steven Low, Richard Murray, Houman Owhadi, Pietro Perona, Peter Schröder, Leonard Schulman, Omer Tamuz, Joel Tropp, Chris Umans, PP Vaidyanathan, Thomas Vidick, Adam Wierman, Yisong Yue