Living on the Edge

Phase Transitions in Random Convex Programs

Joel A. Tropp
Michael B. McCoy

Computing + Mathematical Sciences
California Institute of Technology

Joint with Dennis Amelunxen and Martin Lotz (Manchester)
Including work of Samet Oymak and Babak Hassibi (Caltech)

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

...in which we dilate upon the question...

How big is a cone?

(and why you should care)

Statistical Dimension

The Statistical Dimension of a Cone

Definition [Amelunxen, Lotz, McCoy, T 2013].

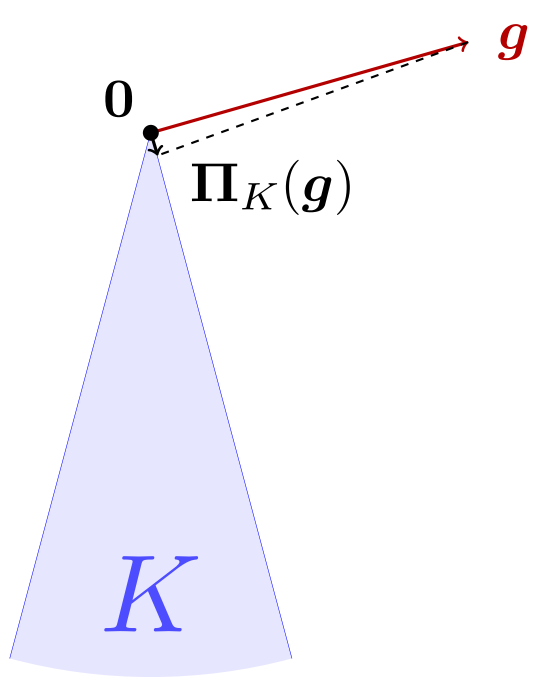
The *statistical dimension* $\delta(K)$ of a closed, convex cone K is the quantity

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

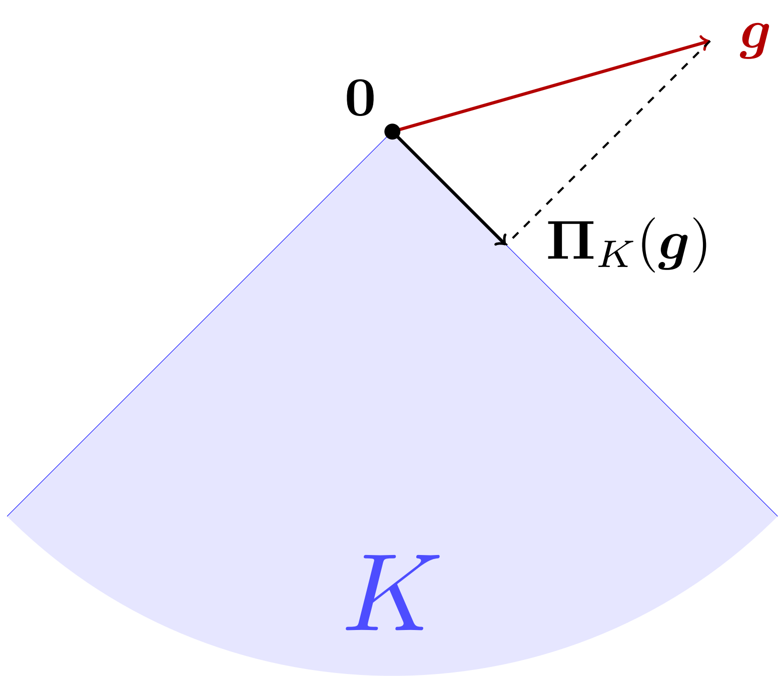
where

- $\mathbf{\Pi}_K$ is the Euclidean projection onto K
- $\boldsymbol{g}$ is a standard normal vector

Statistical Dimension: The Motion Picture



small cone



big cone

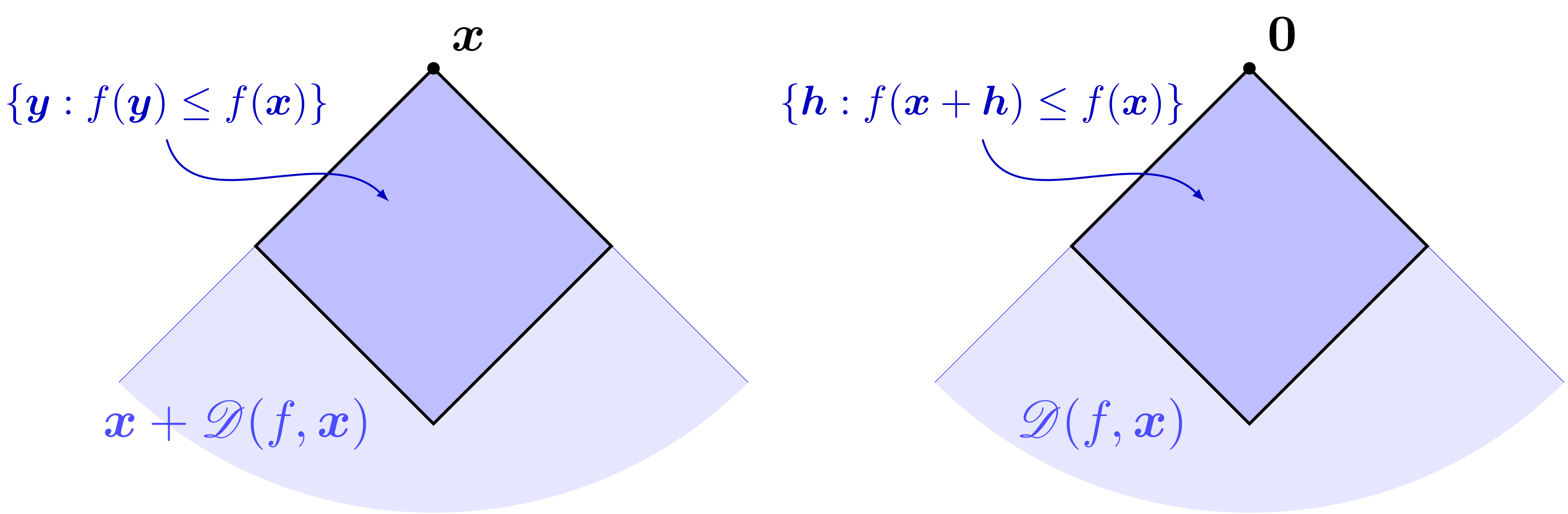
Basic Statistical Dimension Calculations

Cone	Notation	Statistical Dimension
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	$\mathbb{H}^d_+$	$\frac{1}{2}d^2$

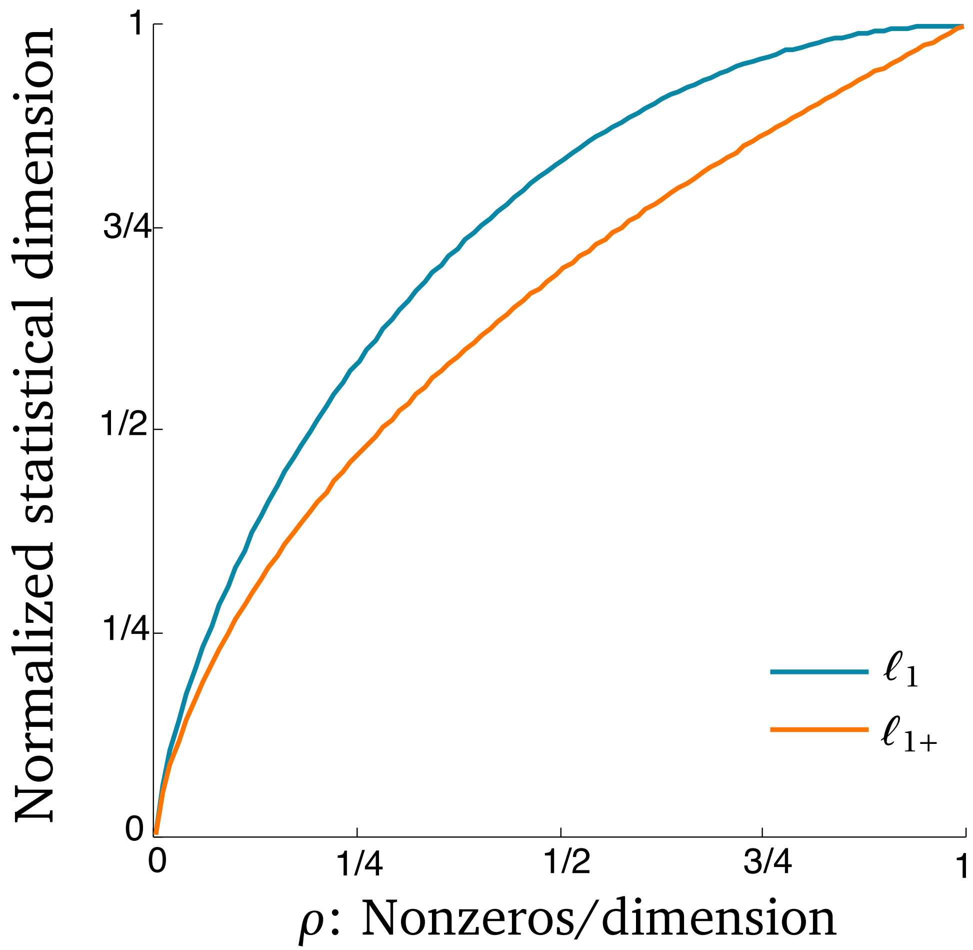
Descent Cones

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

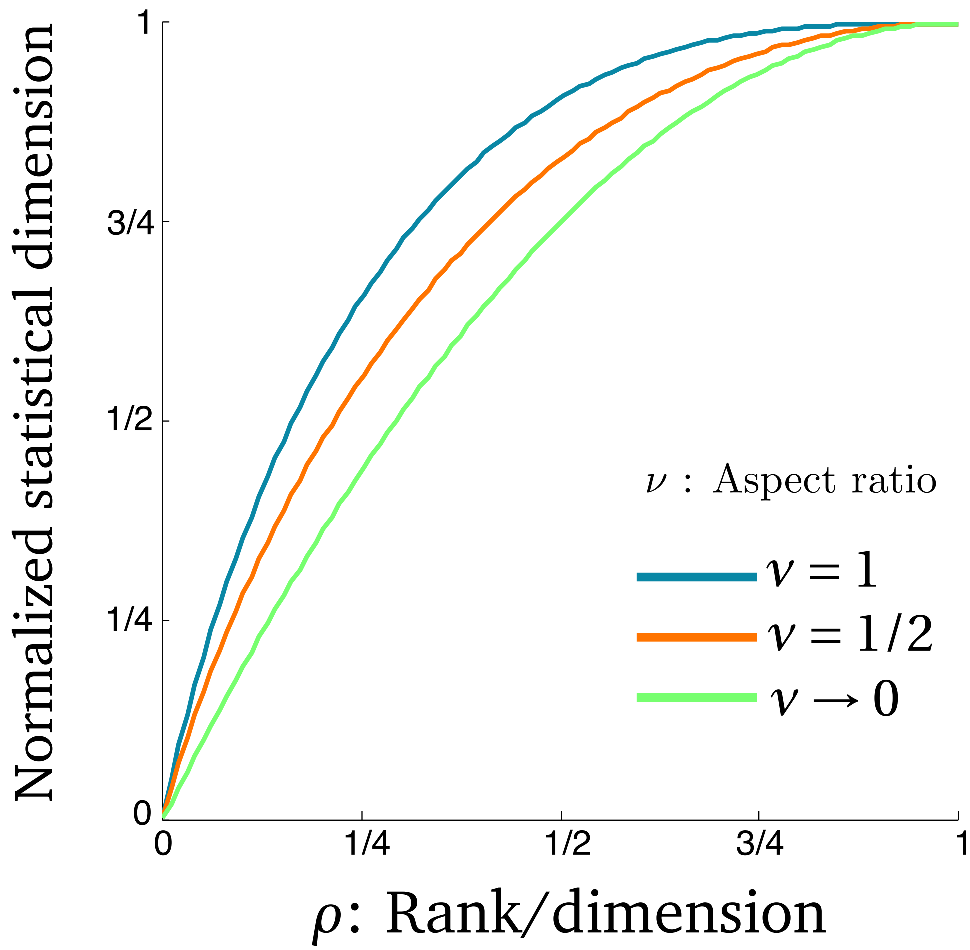
$$\mathscr{D}(f, \boldsymbol{x}) := \{\boldsymbol{h} : f(\boldsymbol{x} + \varepsilon \boldsymbol{h}) \leq f(\boldsymbol{x}) \quad \text{for some } \varepsilon > 0\}$$



Descent Cone of ℓ_1 Norm at Sparse Vector



Descent Cone of S_1 Norm at Low-Rank Matrix



Aside: The Gaussian Width

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

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

- We have the relationship

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

- Statistical dimension is the canonical extension of the linear dimension to the class of convex cones. Gaussian width ain't.

Regularized Denoising

Setup for Regularized Denoising

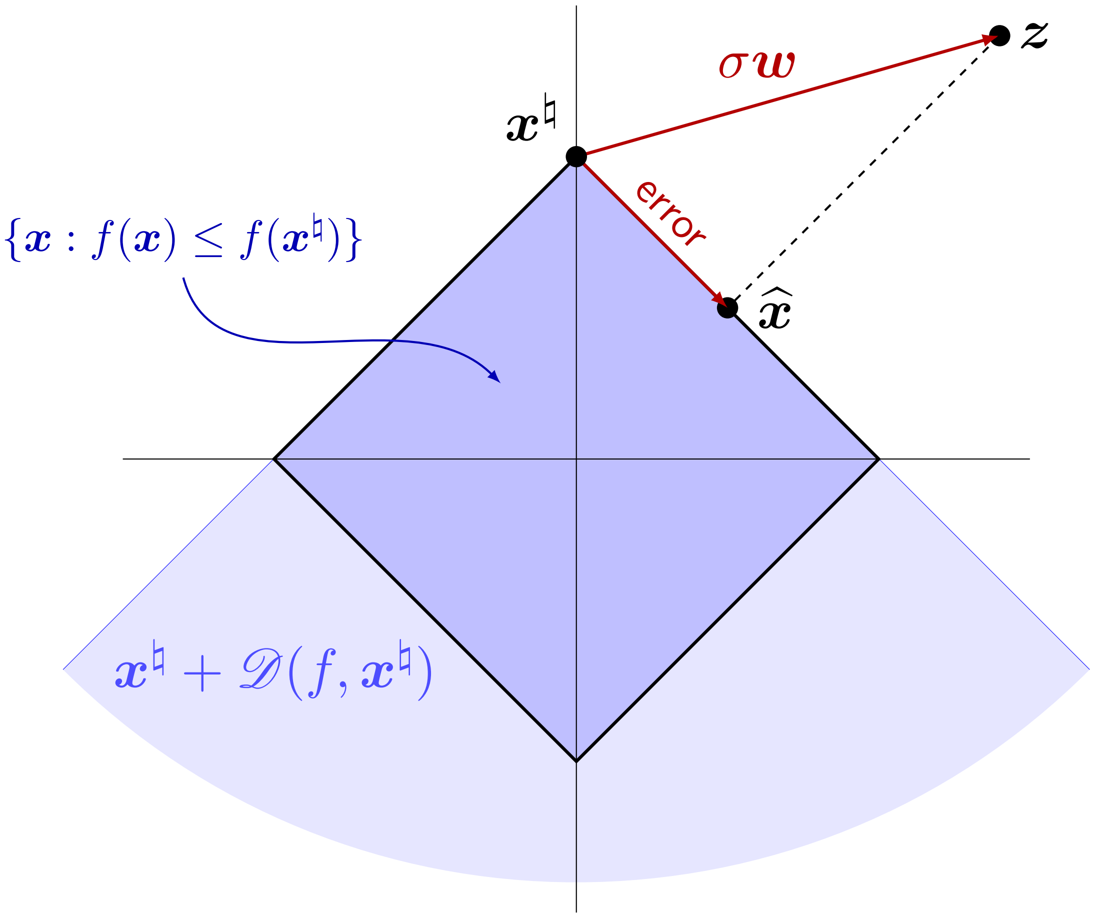
- Let $\boldsymbol{x}^\natural \in \mathbb{R}^d$ be "structured" but unknown
- Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function that reflects "structure"
- Observe $\boldsymbol{z} = \boldsymbol{x}^\natural + \sigma \boldsymbol{w}$ where $\boldsymbol{w} \sim \textsc{normal}(0, \mathbf{I})$
- Remove noise by solving the convex program*

$$\text{minimize} \quad \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{x}\|_2^2 \quad \text{subject to} \quad f(\boldsymbol{x}) \le f(\boldsymbol{x}^\natural)$$

- **Hope:** The minimizer $\widehat{\boldsymbol{x}}$ approximates $\boldsymbol{x}^\natural$

*We assume the side information $f(\boldsymbol{x}^\natural)$ is available. This is equivalent** to knowing the optimal choice of Lagrange multiplier for the constraint.

Geometry of Denoising



The Risk of Regularized Denoising

Theorem 1. [Oymak & Hassibi 2013] **Assume**

- *We observe* $\boldsymbol{z} = \boldsymbol{x}^\natural + \sigma \boldsymbol{w}$ *where* $\boldsymbol{w}$ *is standard normal*
- *The vector* $\widehat{\boldsymbol{x}}$ *solves*

$$\text{minimize} \quad \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{x}\|_2^2 \quad \text{subject to} \quad f(\boldsymbol{x}) \le f(\boldsymbol{x}^\natural)$$

Then

$$\sup_{\sigma > 0} \frac{\mathbb{E}\,\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^\natural\|^2}{\sigma^2} = \delta\big(\mathscr{D}(f, \boldsymbol{x}^\natural)\big)$$

Regularized Linear Inverse Problems

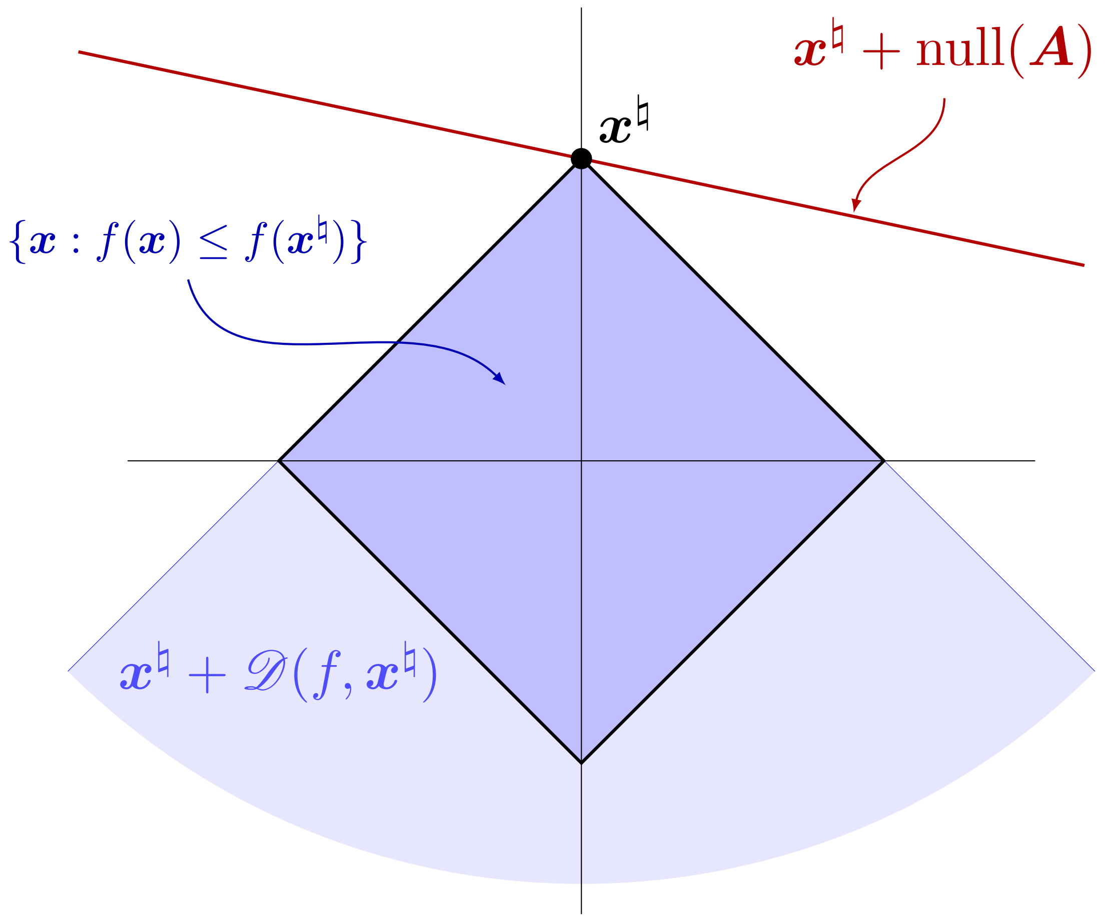
Setup for Linear Inverse Problems

- 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



Linear Inverse Problems with Random Data

Theorem 2. **[Amelunxen, Lotz, McCoy, T 2013]** **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*

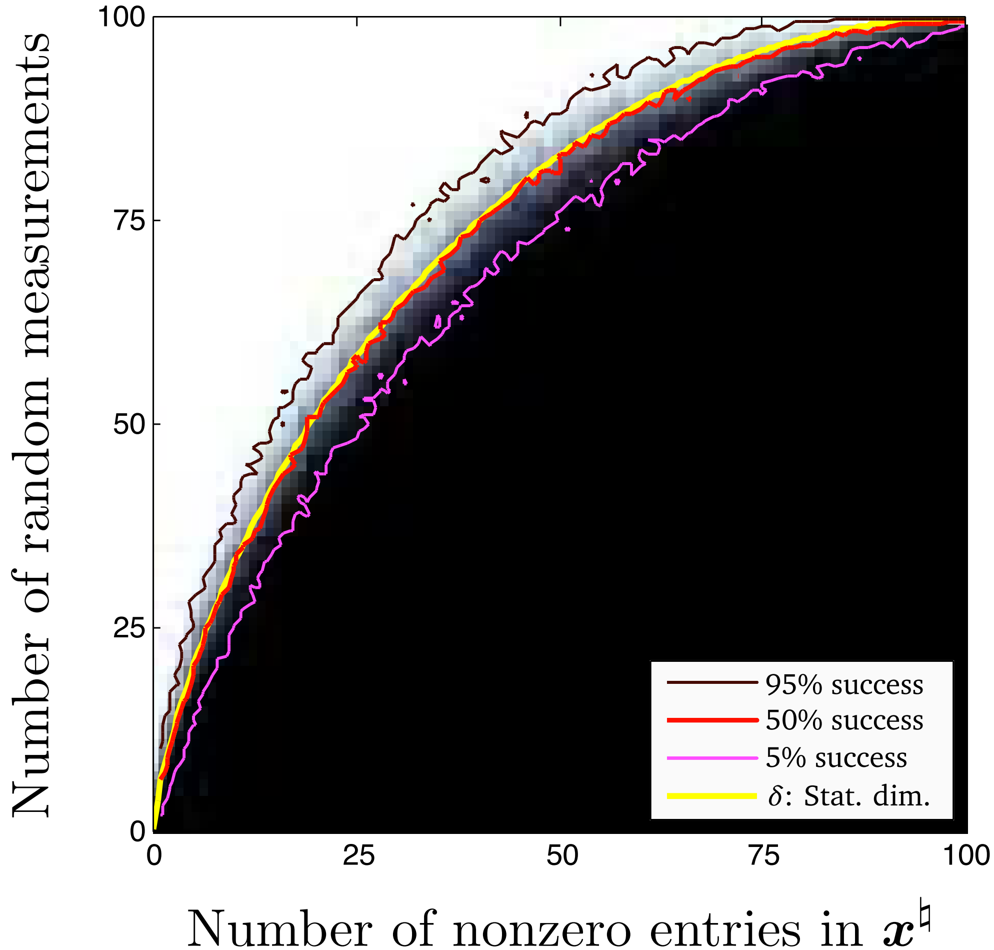
$$\text{minimize} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{z}$$

Then

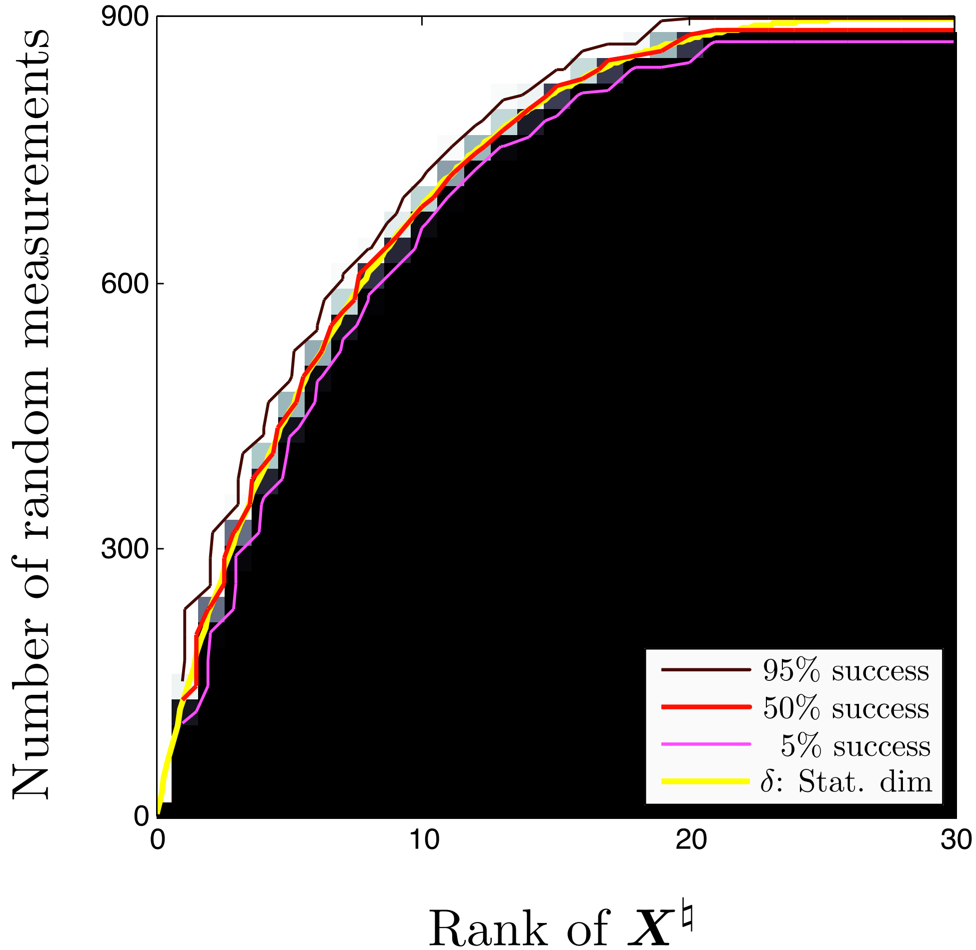
$$m \gtrsim \delta\big(\mathscr{D}(f, \boldsymbol{x}^\natural)\big) \quad \Longrightarrow \quad \widehat{\boldsymbol{x}} = \boldsymbol{x}^\natural \quad \text{whp}$$
$$m \lesssim \delta\big(\mathscr{D}(f, \boldsymbol{x}^\natural)\big) \quad \Longrightarrow \quad \widehat{\boldsymbol{x}} \neq \boldsymbol{x}^\natural \quad \text{whp}$$

Related work: Rudelson–Vershynin 2006, Donoho–Tanner 2008, Stojnic 2009, Chandrasekaran et al. 2010

Sparse Recovery via ℓ_1 Minimization



Low-Rank Recovery via S_1 Minimization



Demixing
Structured Signals

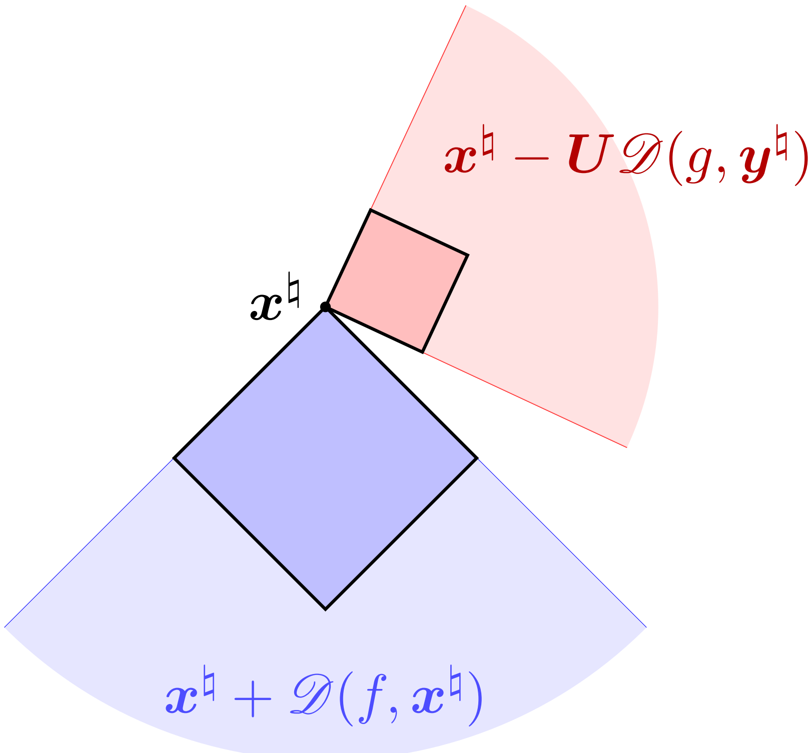
Setup for Demixing Problems

- Let $\boldsymbol{x}^\natural \in \mathbb{R}^d$ and $\boldsymbol{y}^\natural \in \mathbb{R}^d$ be structured, unknown vectors
- Let $f, g : \mathbb{R}^d \to \mathbb{R}$ be convex functions that reflect "structure"
- Let $\boldsymbol{U} \in \mathbb{R}^{d \times d}$ be a known orthogonal matrix
- Observe $\boldsymbol{z} = \boldsymbol{x}^\natural + \boldsymbol{U}\boldsymbol{y}^\natural$
- Demix via convex program

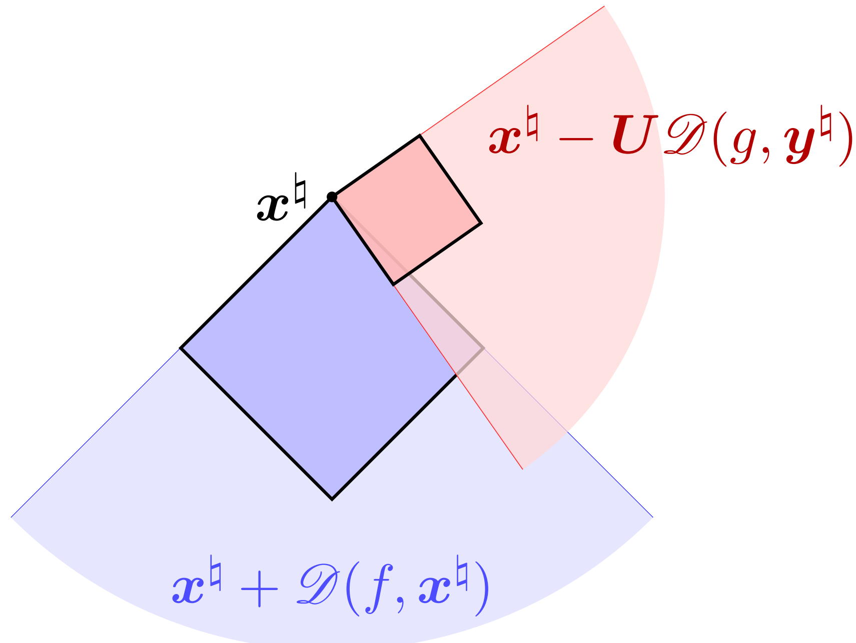
$$\text{minimize} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad g(\boldsymbol{y}) \le g(\boldsymbol{y}^\natural)$$
$$\boldsymbol{x} + \boldsymbol{U}\boldsymbol{y} = \boldsymbol{z}$$

- **Hope:** $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) = (\boldsymbol{x}^\natural, \boldsymbol{y}^\natural)$

Geometry of Demixing Problems



Success!



Failure!

Demixing Problems with Random Incoherence

Theorem 3. **[Amelunxen, Lotz, McCoy, T 2013]** **Assume**

- *The vectors* $\boldsymbol{x}^\natural \in \mathbb{R}^d$ *and* $\boldsymbol{y}^\natural \in \mathbb{R}^d$ *are unknown*
- *The observation* $\boldsymbol{z} = \boldsymbol{x}^\natural + \boldsymbol{Q}\boldsymbol{y}^\natural$ *where* $\boldsymbol{Q}$ *is random orthogonal*
- *The pair* $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}})$ *solves*

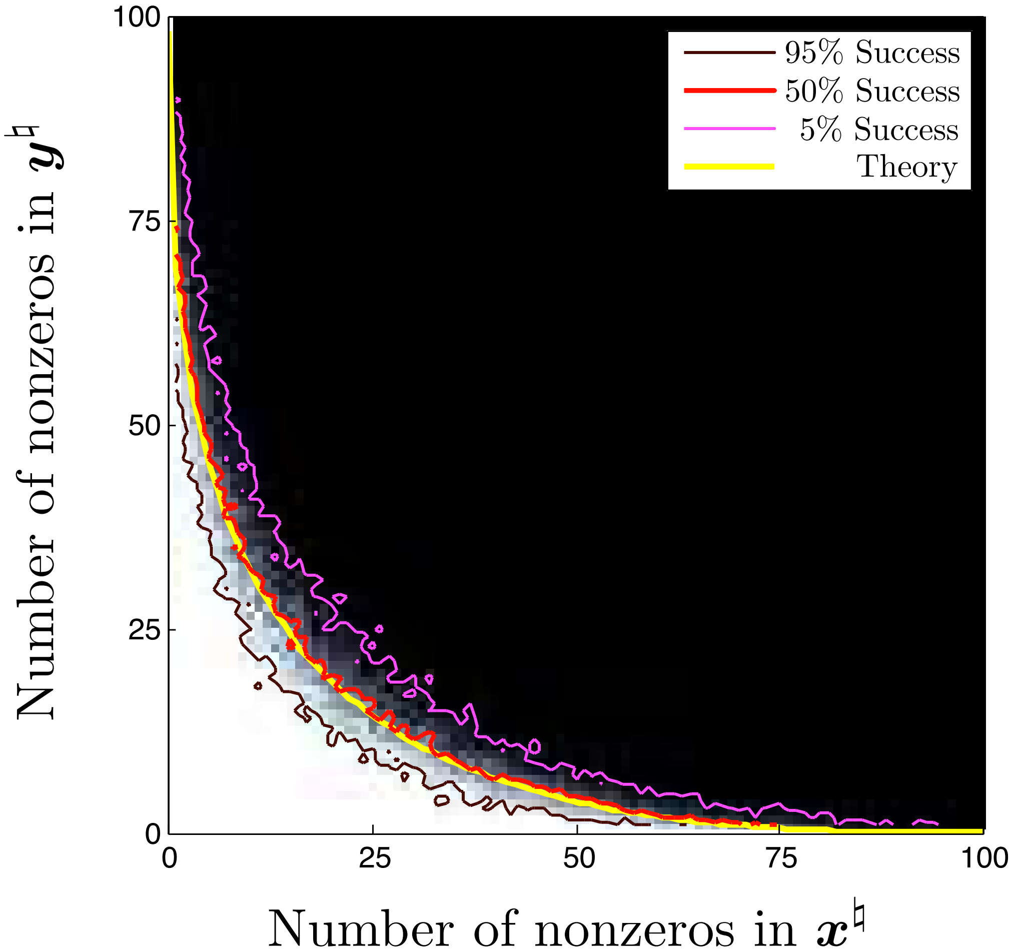
$$\text{minimize} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad g(\boldsymbol{y}) \le g(\boldsymbol{y}^\natural), \quad \boldsymbol{x} + \boldsymbol{Q}\boldsymbol{y} = \boldsymbol{z}$$

Then

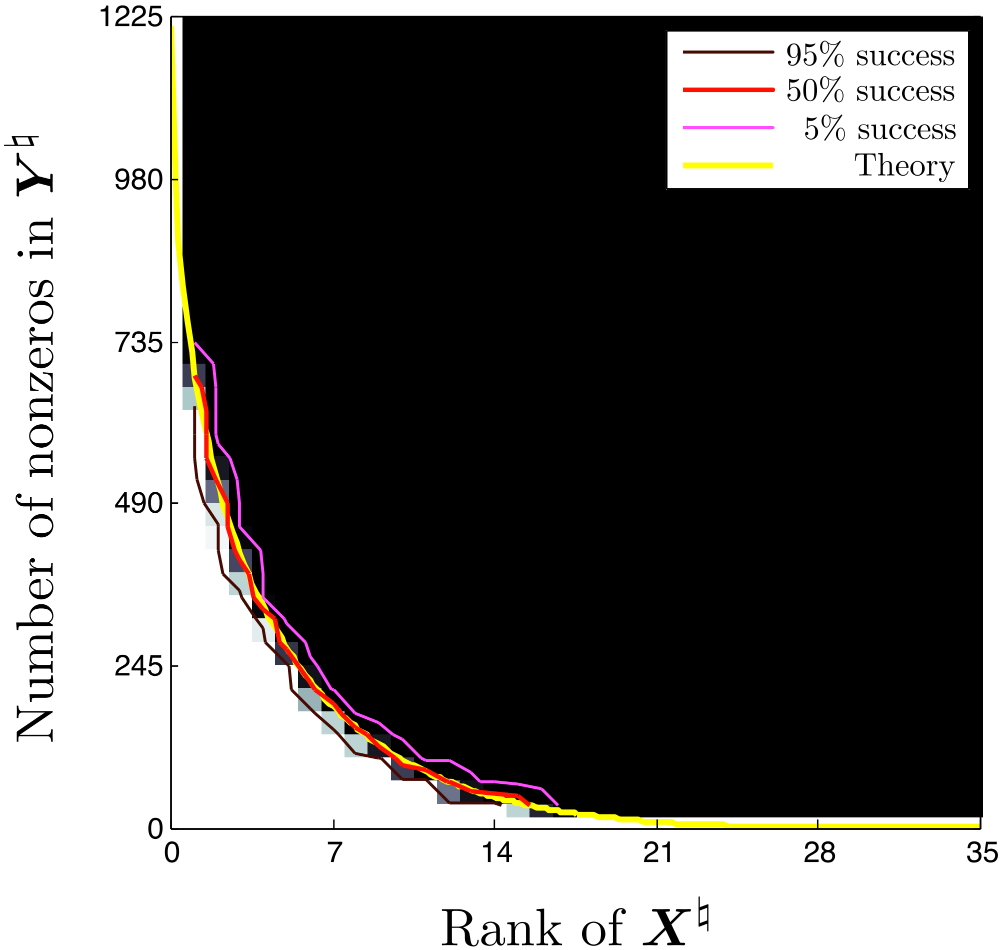
$$\delta\big(\mathscr{D}(f, \boldsymbol{x}^\natural)\big) + \delta\big(\mathscr{D}(g, \boldsymbol{y}^\natural)\big) \lesssim d \quad \Longrightarrow \quad (\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) = (\boldsymbol{x}^\natural, \boldsymbol{y}^\natural) \quad \textit{whp}$$

$$\delta\big(\mathscr{D}(f, \boldsymbol{x}^\natural)\big) + \delta\big(\mathscr{D}(g, \boldsymbol{y}^\natural)\big) \gtrsim d \quad \Longrightarrow \quad (\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) \neq (\boldsymbol{x}^\natural, \boldsymbol{y}^\natural) \quad \textit{whp}$$

Sparse + Sparse via $\ell_1 + \ell_1$ Minimization



Low-Rank + Sparse via $S_1 + \ell_1$ Minimization



To learn more...

E-mail: `mccoy@cms.caltech.edu`
`jtropp@cms.caltech.edu`

Web: `http://users.cms.caltech.edu/~mccoy`
`http://users.cms.caltech.edu/~jtropp`

Papers:

- MT, "Sharp recovery bounds for convex deconvolution, with applications." arXiv cs.IT 1205.1580
- ALMT, "Living on the edge: A geometric theory of phase transitions in convex optimization." arXiv cs.IT 1303.6672
- Oymak & Hassibi, "Asymptotically exact denoising in relation to compressed sensing," arXiv cs.IT 1305.2714
- More to come!

Inference, Control, & Optimization

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

- *A unified view of inference and decision making, emphasizing methods that are computationally efficient and theoretically grounded*
- **Core faculty:** Jim Beck, Venkat Chandrasekaran, John Doyle, Babak Hassibi, Steven Low, Richard Murray, Houman Owhadi, Joel Tropp, Adam Wierman
- **Key Research Areas:** Signal processing, statistics, optimization, control, uncertainty quantification, and their applications in science + engineering