Structured Signal Processing

Joel A. Tropp

Computing + Mathematical Sciences

California Institute of Technology

jtropp@cms.caltech.edu

Research supported by ONR, AFOSR, NSF, DARPA, Sloan, and Moore.

Amuse Bouche

The Compressed Sensing Problem

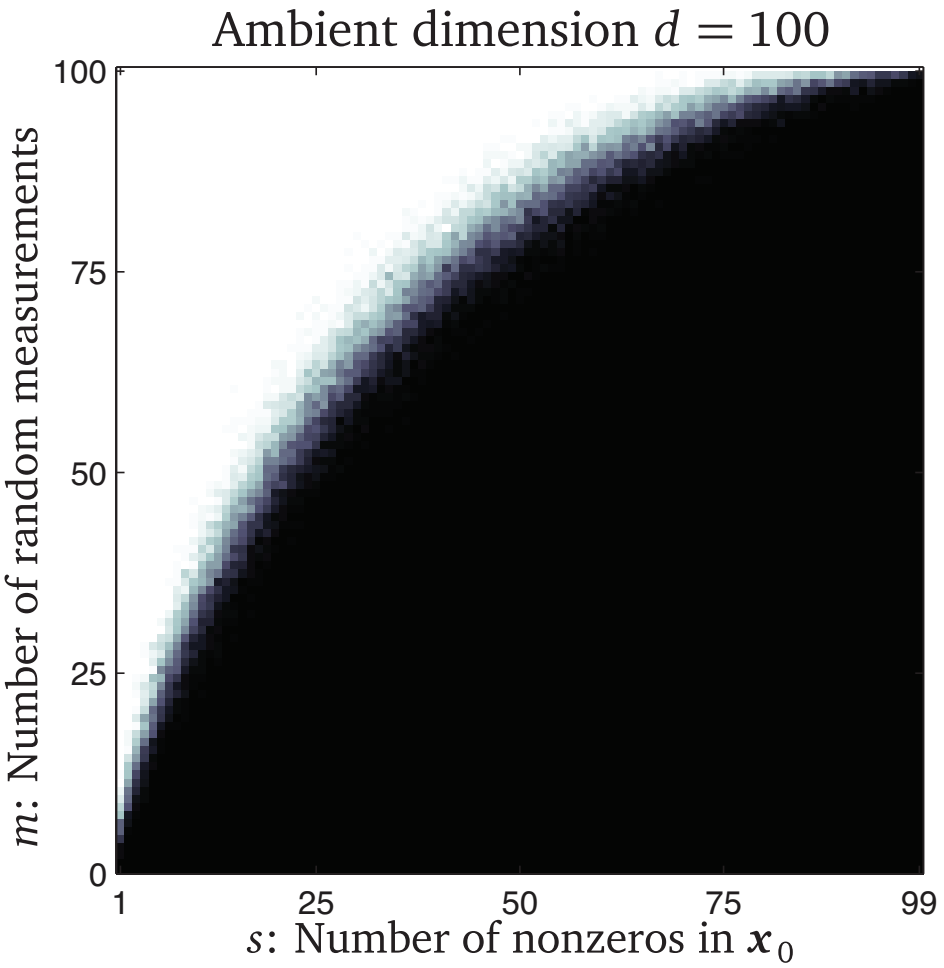
- Suppose $\boldsymbol{x}_\natural \in \mathbb{R}^d$ has s nonzero entries
- Let $\boldsymbol{\Gamma} \in \mathbb{R}^{m \times d}$ be a standard normal matrix
- Observe $\boldsymbol{z} = \boldsymbol{\Gamma} \boldsymbol{x}_\natural \in \mathbb{R}^m$
- Find estimate $\widehat{\boldsymbol{x}}$ by solving convex program

$$\text{minimize} \quad \|\boldsymbol{x}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Gamma} \boldsymbol{x} = \boldsymbol{z}$$

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

Sources: Donoho 2006; Candès & Tao 2006.

Empirical Performance of Compressed Sensing

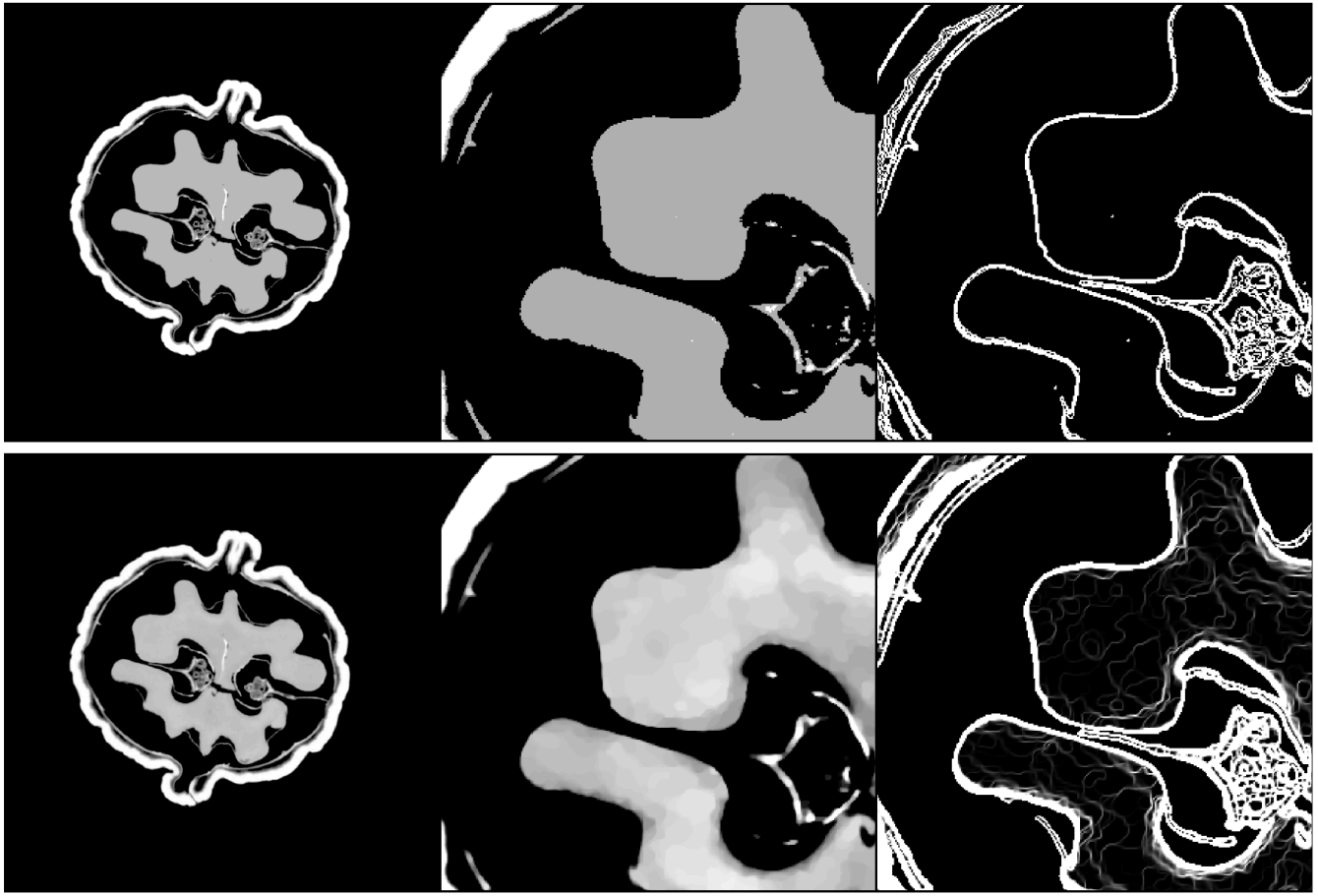


What's Going on Here?!?

- What is the probability of success as a function of (s, m, d)?
- Does a phase transition exist?
- Can we locate the phase transition?
- How wide is the transition region?
- Is there a geometric explanation for this phenomenon?
- Can we export this reasoning to understand other problems?
- Who cares?

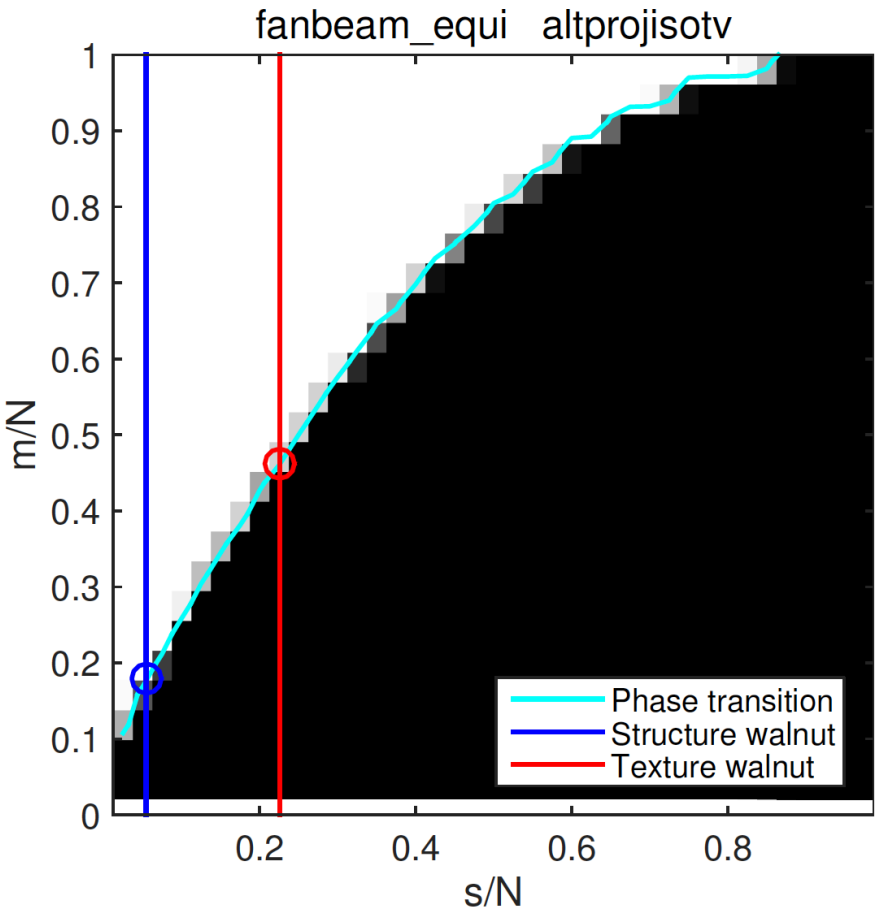
Sources: Donoho 2006; Donoho & Tanner 2009; Stojnic 2009, 2013; McCoy & Tropp 2013, 2014; Thrampoulidis et al. 2013–2016; Amelunxen et al. 2014; Foygel & Mackey 2014; Goldstein et al. 2016.

Case Study: Walnut Phantoms



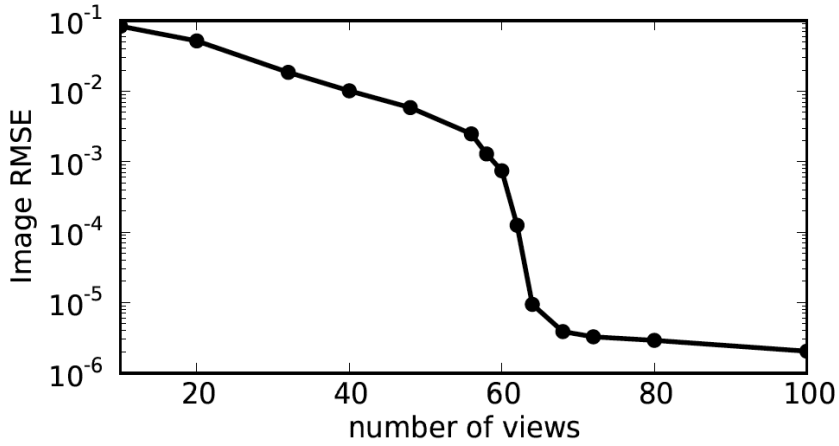
Source: Jørgensen & Sidky 2014.

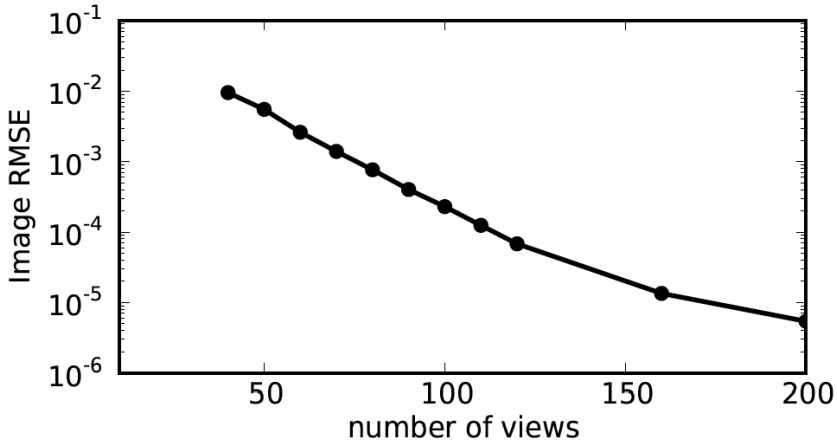
Case Study: Walnut Phantoms



Source: Jørgensen & Sidky 2014.

Case Study: Walnut Phantoms



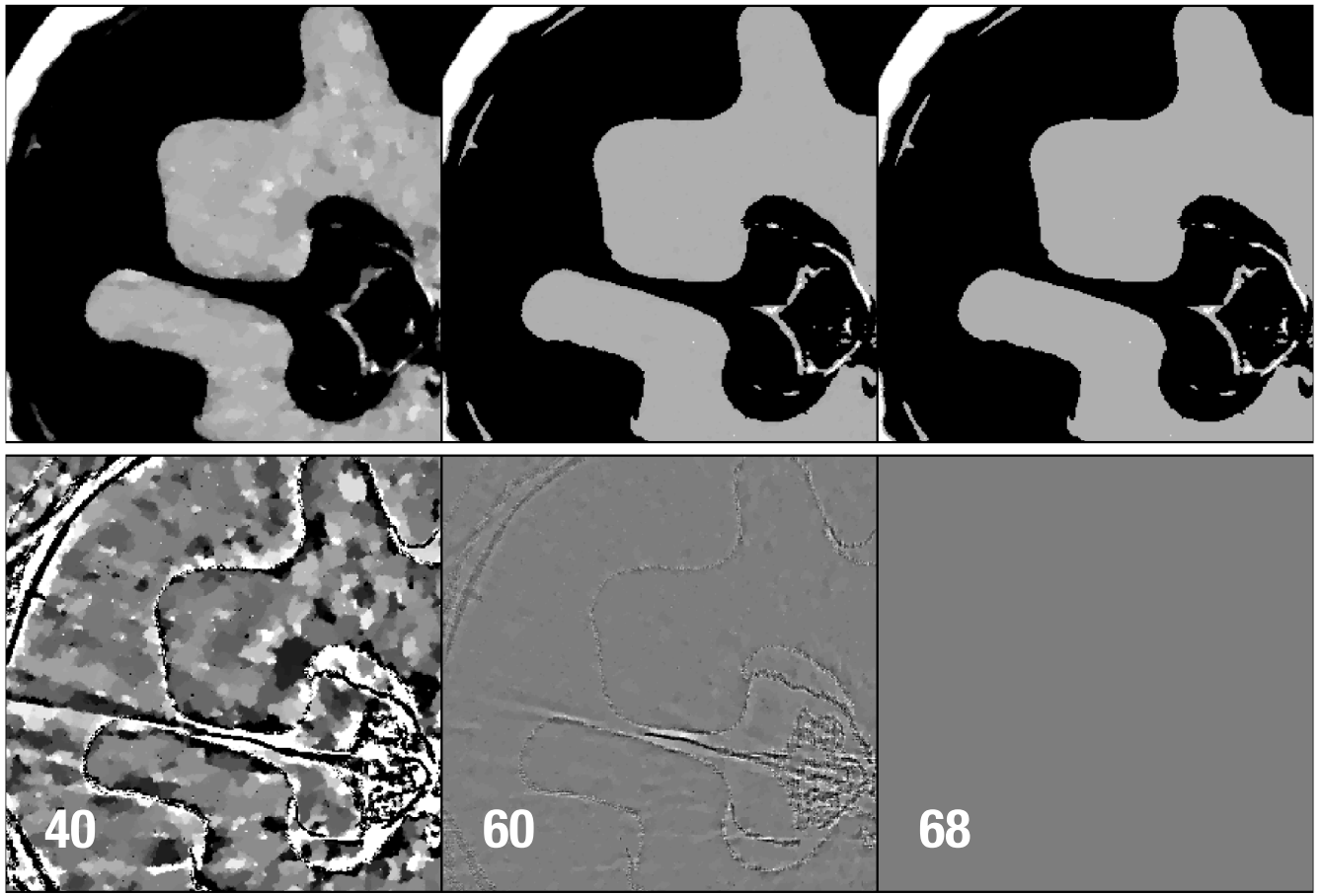


Walnut image	Gradient sparsity	Recovered at	DT prediction	ALMT prediction
Structure	$45,074$	68	69.3	71.7
Texture	$186,306$	?	188.7	185.8

Table 1: Walnut test images with gradient-domain sparsity levels, number of projections at which recovery is observed, and DT and ALMT phase-diagram predictions of critical sampling levels. A reference point of full sampling is $N_v \geq 403$ projections, where the system matrix has more rows than columns.

Source: Jørgensen & Sidky 2014.

Case Study: Walnut Phantoms



Source: Jørgensen & Sidky 2014.

Atomic Decomposition

Atoms and Dictionaries

Definition 1. Consider a compact collection $\mathcal{A}$ of vectors:

$$\mathcal{A} = \{\boldsymbol{a}_\xi : \xi \in \Xi\} \subset \mathbb{R}^d$$

The collection $\mathcal{A}$ is called a dictionary, and the elements $\boldsymbol{a}_\xi$ are called atoms.

- Atoms are "elementary structures" that compose signals of interest
- Closely related to definition from nonlinear approximation (1990s)
- Terminology motivated by atomic decomposition in harmonic analysis (1970s)
- Generalizes the concept of a frame in signal processing (1980s)

Sources: Duffin & Schaeffer 1952; Coifman 1974; Daubechies et al. 1986; Mallat & Zhang 1993; Davis et al. 1994; DeVore & Temlyakov 1996; Chen et al. 1997, 2001; Temlyakov 2002; Donoho 2005; Fuchs 2005; Chandrasekaran et al. 2012.

Example: Astronomical Image

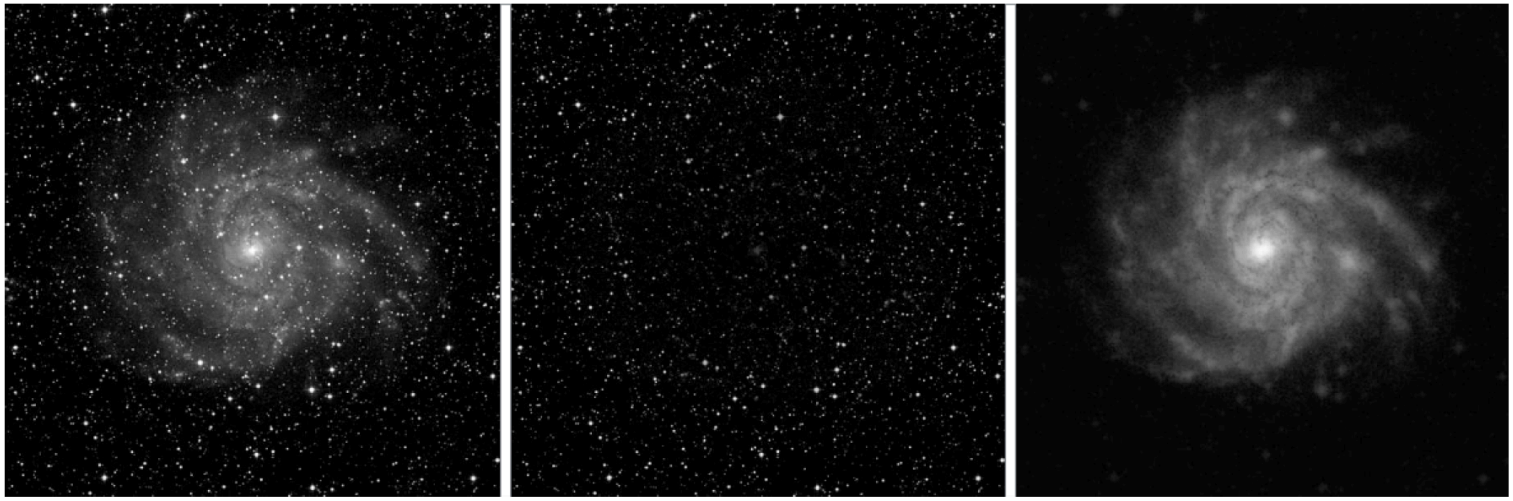


Image credit: NASA

Observation z_0 Sparse component x_0 DCT-sparse component y_0

Source: Starck et al. 2003; McCoy & Tropp 2013; McCoy et al. 2013; Amelunxen et al. 2014.

Sparsity with Respect to a Dictionary

Definition 2. For a natural number s, define the set of vectors of the form

$$\mathcal{A}_s = \left\{ \boldsymbol{x} = \sum_{\omega \in \Omega} c_\omega \boldsymbol{a}_\omega : c_\omega \geq 0 \text{ and } |\Omega| \leq s \right\}$$

The members of $\mathcal{A}_s$ are said to be s-sparse with respect to the dictionary

- $\mathcal{A}_s$ are increasing approximation classes
- Sparsity s parameterizes complexity of signals
- A "semiparametric" model with wide application

Sources: Stechkin 1955; Miller 1989; Jones 1992; Barron 1993; Mallat & Zhang 1993; Davis et al. 1994; DeVore & Temlyakov 1996; Chen et al. 1997, 2001; Donoho & Huo 2001; Temlyakov 2002.

Sparse Approximation Problems

- Given a vector $\boldsymbol{x}_\natural$ and a dictionary $\mathcal{A}$...
- Minimize the error subject to a bound on the sparsity:

$$\text{minimize} \quad \|\boldsymbol{x} - \boldsymbol{x}_\natural\|^2 \quad \text{subject to} \quad \boldsymbol{x} \in \mathcal{A}_s$$

- Minimize the sparsity subject to a bound on the error:

$$\text{minimize} \quad s \quad \text{subject to} \quad \boldsymbol{x} \in \mathcal{A}_s \quad \text{and} \quad \|\boldsymbol{x} - \boldsymbol{x}_\natural\|^2 \leq \varepsilon$$

- **Fact** (Natarajan 1995)**:** In general, sparse approximation is NP-hard

$\|\cdot\|$ always denotes the Euclidean norm

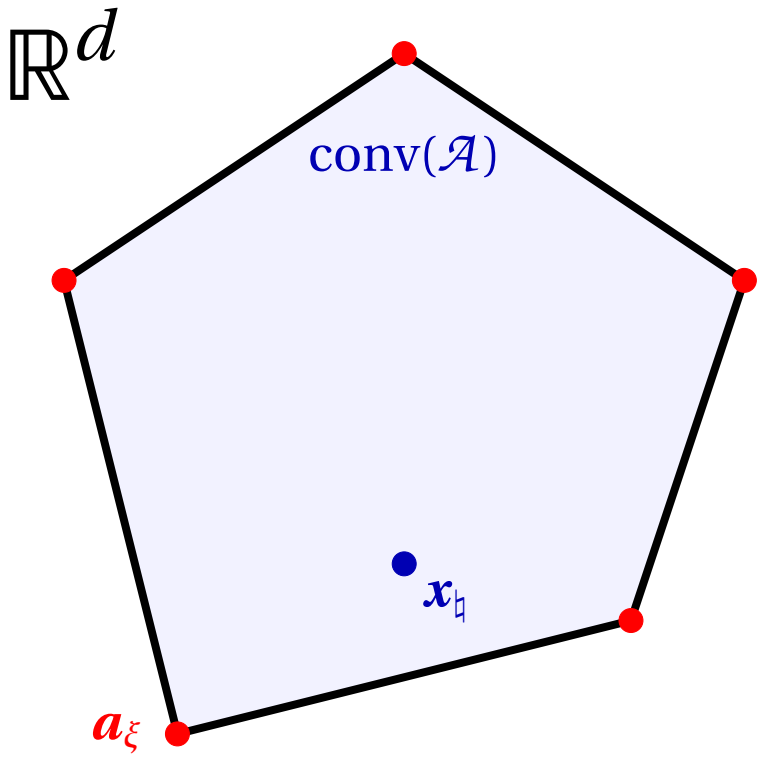
Sources: Schmidt 1908; Stechkin 1955; Friedman & Stuetzle 1981; Miller 1989; Jones 1992; Barron 1993; Mallat & Zhang 1993; Davis et al. 1994; Natarajan 1995; DeVore & Temlyakov 1996; Chen et al. 1997, 2001; Donoho & Huo 2001; Temlyakov 2002; ....

Advice for Research & Life

If at first you don't succeed…

Lower your standards!

Relax: Approximation in Convex Hulls



- Carathéodory: Can write $\boldsymbol{x}_\natural$ as a convex combination of $d + 1$ atoms
- How well can we approximate $\boldsymbol{x}_\natural$ with s atoms?

Approximate Carathéodory

Theorem 3 (Maurey 1970s). ***Assume***

- $\mathcal{A}$ *is a dictionary*
- $\boldsymbol{x}_\natural \in \mathrm{conv}(\mathcal{A})$

Then *there exists an* s*-sparse vector* $\boldsymbol{x}_s \in \mathcal{A}_s$ *with*

$$\|\boldsymbol{x}_s - \boldsymbol{x}_\natural\| \leq \mathrm{diam}(\mathcal{A}) \cdot \frac{1}{\sqrt{s}}$$

where $\mathrm{diam}(\mathcal{A}) = \max_{\omega,\xi \in \Xi} \|\boldsymbol{a}_\omega - \boldsymbol{a}_\xi\|$

Sources: Maurey 1970s; Pisier 1980; Carl 1985; DeVore & Temlyakov 1996.

Proof of Maurey's Theorem: The Empirical Method

- Carathéodory implies

$$\boldsymbol{x}_\natural = \sum_{\omega\in\Omega} p_\omega \boldsymbol{a}_\omega \quad \text{where} \quad p_\omega \geq 0, \quad \sum_{\omega\in\Omega} p_\omega = 1, \quad |\Omega| \leq d+1$$

- Define random vector $\boldsymbol{z}$ that takes values $\boldsymbol{z} = \boldsymbol{a}_\omega$ with probability p_ω
- Observe: $\mathbb{E}\,\boldsymbol{z} = \boldsymbol{x}_\natural$
- Set $\bar{\boldsymbol{z}} = \frac{1}{s}\sum_{i=1}^{s} \boldsymbol{z}_i \in \mathcal{A}_s$ where $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_s$ are iid copies of $\boldsymbol{z}$
- Using orthogonality,

$$\mathbb{E}\,\|\bar{\boldsymbol{z}} - \boldsymbol{x}_\natural\|^2 = \frac{1}{s^2}\mathbb{E}\left\|\sum_{i=1}^{s}(\boldsymbol{z}_i - \mathbb{E}\,\boldsymbol{z}_i)\right\|^2 = \frac{1}{s^2}\sum_{i=1}^{s}\mathbb{E}\,\|\boldsymbol{z}_i - \mathbb{E}\,\boldsymbol{z}_i\|^2 = \frac{1}{s}\mathbb{E}\,\|\boldsymbol{z} - \boldsymbol{x}_\natural\|^2 \leq \frac{\operatorname{diam}^2(\mathcal{A})}{s}$$

- The probabilistic method yields the desired $\boldsymbol{x}_s \in \mathcal{A}_s$

Atomic Gauges

Definition 4. Let $\mathcal{A}$ be a dictionary. The atomic gauge is defined as

$$\|\boldsymbol{x}\|_{\mathcal{A}} = \min\left\{t \geq 0 : \boldsymbol{x} \in t \cdot \operatorname{conv}(\mathcal{A})\right\}$$

$$= \min\left\{\sum_{\xi \in \Xi} c_\xi : \boldsymbol{x} = \sum_{\xi \in \Xi} c_\xi \boldsymbol{a}_\xi \text{ and } c_\xi \geq 0\right\}$$

The gauge is nonnegative, positively homogeneous, and convex.

- Gauges are norm-like functions
- They are linear on rays from the origin
- They can take the values zero or $+\infty$ on an entire ray

Sources: Stechkin 1955; DeVore & Temlyakov 1996; Chen et al. 1997, 2001; Temlyakov 2002; Donoho 2005; Fuchs 2005; Chandrasekaran et al. 2012.

Atomic Gauges and Sparse Approximation

Corollary 5. *For any vector $\boldsymbol{x}_\natural$, there is an s-sparse vector $\boldsymbol{x}_s \in \mathcal{A}_s$ that achieves*

$$\|\boldsymbol{x}_s - \boldsymbol{x}_\natural\| \leq \operatorname{diam}(\mathcal{A}) \cdot \frac{\|\boldsymbol{x}_\natural\|_{\mathcal{A}}}{\sqrt{s}}$$

- Atomic gauge always controls the quality of sparse approximations
- Evidence that atomic gauge is a reasonable proxy for complexity with respect to dictionary
- Bound is optimal for worst-case $\boldsymbol{x}_\natural$
- Very poor bound for exactly sparse $\boldsymbol{x}_\natural$

Sources: Stechkin 1955; Jones 1992; DeVore & Temlyakov 1996; Chen et al. 1997, 2001; Temlyakov 2002; Jaggi 2013.

Examples of Dictionaries and Atomic Gauges

Signal Type	Dictionary ($\mathcal{A}$)	Atomic Gauge ($\|\cdot\|_{\mathcal{A}}$)
Sparse vector	$\{\pm \mathbf{e}_i\}$	$\|\cdot\|_{\ell_1}$
Frequency-sparse	$\{\pm \mathbf{f}_i\}$	$\|\mathcal{F}(\cdot)\|_{\ell_1}$
Spikes + sines	$\{\pm \mathbf{e}_i\} \cup \{\pm \mathbf{f}_i\}$	$\|\cdot\|_{\ell_1} \square \|\mathcal{F}(\cdot)\|_{\ell_1}$
Sparse gradient	$\{\pm(\mathbf{e}_{i+1} - \mathbf{e}_i)\}$	$\|\cdot\|_{\mathrm{TV}}$
Sparse + nonnegative	$\{\mathbf{e}_i\}$	$\|\cdot\|_{\ell_1}$ if nn, else $+\infty$
Saturated	$\{\pm 1\}^d$	$\|\cdot\|_{\ell_\infty}$
Row-sparse matrix	$\{\mathbf{e}_i \boldsymbol{u}^* : \|\boldsymbol{u}\| = 1\}$	$\sum_i \|(\cdot)_{i:}\|$
Low-rank matrix	$\{\boldsymbol{u}\boldsymbol{v}^* : \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1\}$	$\|\cdot\|_{S_1}$
Low rank + psd	$\{\boldsymbol{u}\boldsymbol{u}^* : \|\boldsymbol{u}\| = 1\}$	$\|\cdot\|_{S_1}$ if psd, else $+\infty$
Orthogonal	$\{\boldsymbol{U} : \boldsymbol{U}\boldsymbol{U}^* = \boldsymbol{U}^*\boldsymbol{U} = \mathbf{I}\}$	$\|\cdot\|_{S_\infty}$

Sources: Rudin et al. 1992; Mallat & Zhang 1993; DeVore & Temlyakov 1996; Chen et al. 1997, 2001; Donoho & Huo 2001; Temlyakov 2002; Fazel 2002; Tropp 2006; Chandrasekaran et al. 2012; Jaggi 2011, 2013; ....

Atomic Regularization for Sparse Approximation

- Given a vector $\boldsymbol{x}_\natural$ and a dictionary $\mathcal{A}$...
- Minimize the error subject to a bound on the atomic gauge:

$$\text{minimize} \quad \|\boldsymbol{x} - \boldsymbol{x}_\natural\|^2 \quad \text{subject to} \quad \|\boldsymbol{x}\|_{\mathcal{A}} \leq \alpha$$

- Minimize the atomic gauge subject to a bound on the error:

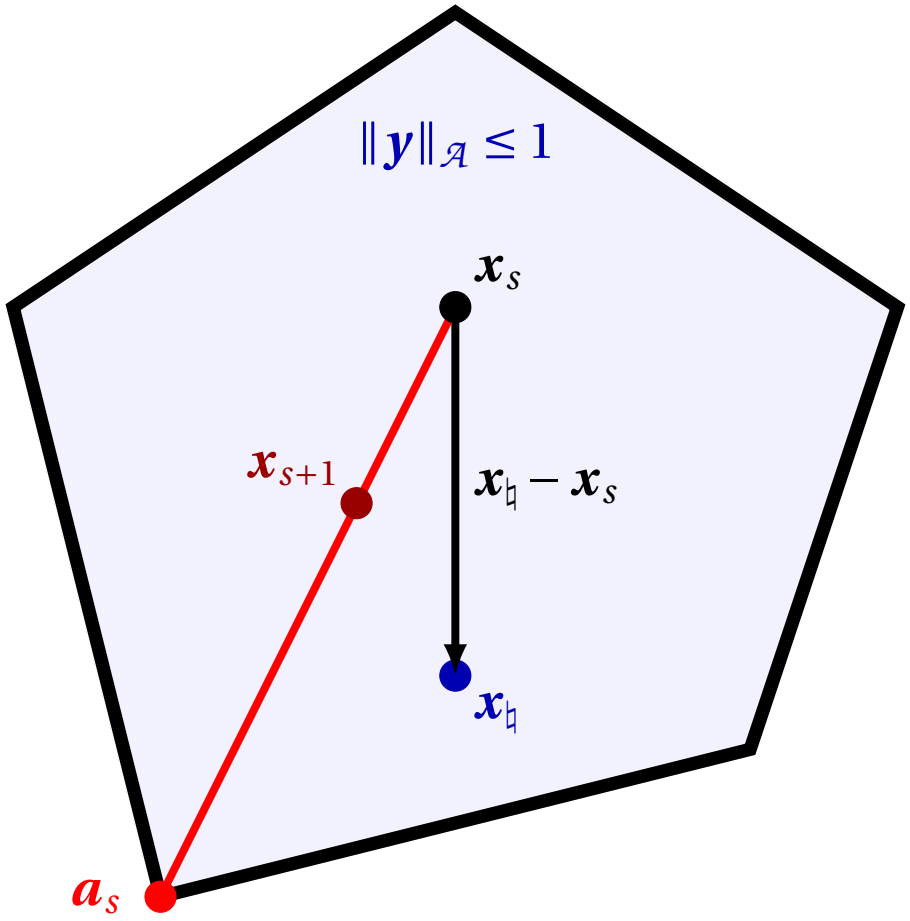
$$\text{minimize} \quad \|\boldsymbol{x}\|_{\mathcal{A}} \quad \text{subject to} \quad \|\boldsymbol{x} - \boldsymbol{x}_\natural\|^2 \leq \varepsilon$$

- Tradeoff the value of the error and the atomic gauge:

$$\text{minimize} \quad \|\boldsymbol{x} - \boldsymbol{x}_\natural\|^2 + \lambda \|\boldsymbol{x}\|_{\mathcal{A}}$$

Sources: Chen et al. 1997, 2001; Donoho & Huo 2001; Gribonval & Nielson 2002; Donoho & Elad 2004; Tropp 2006; Fuchs 2006; Donoho et al. 2006; Chandrasekaran et al. 2012; ....

Conditional Gradient Method (CGM)



Initialize $\boldsymbol{x}_0 = \boldsymbol{0}$; $s = 0$

$$\boldsymbol{a}_s = \arg\max_{\|\boldsymbol{y}\|_{\mathcal{A}} \leq 1} \langle \boldsymbol{y},\ \boldsymbol{x}_\natural - \boldsymbol{x}_s \rangle$$

$$\boldsymbol{x}_{s+1} = (1 - \theta_s)\boldsymbol{x}_s + \theta_s \boldsymbol{a}_s$$

$$\theta_s = (s+1)^{-1}$$

$$\min_{\|\boldsymbol{x}\|_{\mathcal{A}} \leq 1} \|\boldsymbol{x} - \boldsymbol{x}_\natural\|^2$$

Sparse Approximation via CGM

Theorem 6 (CGM Convergence). ***Assume***

- $\mathcal{A}$ *is a dictionary*
- $\boldsymbol{x}_\natural$ *is a vector with* $\|\boldsymbol{x}_\natural\|_{\mathcal{A}} \leq \alpha$, *where* α *is known*

Then*, after* s *iterations, CGM produces an* s*-sparse vector* $\boldsymbol{x}_s$ *with*

$$\|\boldsymbol{x}_s - \boldsymbol{x}_\natural\| \leq \operatorname{diam}(\mathcal{A}) \cdot \frac{\alpha}{\sqrt{s+1}}$$

- If α is misspecified, CGM converges to best approximation of $\boldsymbol{x}_\natural$ with such α
- CGM also known as Frank–Wolfe, Relaxed Greedy Algorithm, or Relaxed Matching Pursuit

Sources: Frank & Wolfe 1956; Levitin & Poljak 1967; Jones 1992; DeVore & Temlyakov 1996; Temlyakov 2002; Hazan 2008; Clarkson 2010; Jaggi 2011, 2013.

Proof of CGM Convergence

- Without loss, assume $\alpha = 1$
- Since $\boldsymbol{x}_{s+1} = (1-\theta_s)\boldsymbol{x}_s + \theta_s \boldsymbol{a}_s$,

$$\|\boldsymbol{x}_{s+1} - \boldsymbol{x}_\natural\|^2 = \|\boldsymbol{x}_s - \boldsymbol{x}_\natural\|^2 - 2\theta_s \langle \boldsymbol{x}_\natural - \boldsymbol{x}_s,\ \boldsymbol{a}_s - \boldsymbol{x}_s\rangle + \theta_s^2 \|\boldsymbol{a}_s - \boldsymbol{x}_s\|^2$$

- By construction of $\boldsymbol{a}_s$,

$$\langle \boldsymbol{x}_\natural - \boldsymbol{x}_s,\ \boldsymbol{a}_s - \boldsymbol{x}_s\rangle = \max_{\|\boldsymbol{y}\|_{\mathcal{A}} \le 1} \langle \boldsymbol{x}_\natural - \boldsymbol{x}_s,\ \boldsymbol{y} - \boldsymbol{x}_s\rangle \ge \langle \boldsymbol{x}_\natural - \boldsymbol{x}_s,\ \boldsymbol{x}_\natural - \boldsymbol{x}_s\rangle = \|\boldsymbol{x}_s - \boldsymbol{x}_\natural\|^2$$

- Therefore,

$$\|\boldsymbol{x}_{s+1} - \boldsymbol{x}_\natural\|^2 \le (1-2\theta_s)\|\boldsymbol{x}_s - \boldsymbol{x}_\natural\|^2 + \theta_s^2 \operatorname{diam}^2(\mathcal{A})$$

- Since $\theta_s(1-\theta_s) < \theta_{s+1}$, induction yields

$$\|\boldsymbol{x}_s - \boldsymbol{x}_\natural\|^2 \le \theta_s \operatorname{diam}^2(\mathcal{A})$$

Exact Recovery Theorems

Theorem 7 (Donoho & Elad 2004; Fuchs 2004; Tropp 2004). ***Assume***

- $\mathcal{A}$ *is a standardized dictionary* $(\|\boldsymbol{a}_\xi\| = 1)$
- *the coherence* $\mu = \max_{\xi \neq \omega} |\langle \boldsymbol{a}_\xi, \boldsymbol{a}_\omega \rangle|$
- $\boldsymbol{x}_\natural \in \mathcal{A}_s$ *where* $s \leq \frac{1}{2}\sqrt{\mu^{-1} + 1}$

Then *we can obtain an s-sparse representation of* $\boldsymbol{x}_\natural$ *by solving the linear program*

$$\text{minimize} \quad \sum_{\xi \in \Xi} c_\xi \quad \text{subject to} \quad \boldsymbol{x}_\natural = \sum_{\xi \in \Xi} c_\xi \boldsymbol{a}_\xi \quad \text{and} \quad c_\xi \geq 0$$

- Can sometimes obtain optimal error bounds for sparse approximation!

Sources: Donoho & Huo 2001; Gribonval & Nielsen 2002; Gilbert et al. 2003; Fuchs 2004–2006; Gribonval & Vandergheynst 2004; Donoho & Elad 2004; Tropp 2004, 2006.

Applications of Sparse Models...

- Denoising and estimation
- Signal recovery, regression, and compressed sensing
- Simultaneous sparse approximation and group sparsity
- Demixing and morphological component analysis
- Matrix completion and phase retrieval
- Superresolution and line spectral estimation
- Blind deconvolution and self-calibration
- ...

Sources: Donoho & Johnstone 1992; Mallat & Zhang 1993; Chen et al. 1997, 2001; Fazel 2002; Starck et al. 2003; Tropp et al. 2006; Recht et al. 2009, 2010; Bodmann et al. 2009; Jaggi 2011, 2013; Bhaskar et al. 2012; Fernandez-Granda 2013; Romberg et al. 2013; McCoy & Tropp 2013; Amelunxen et al. 2014; Ling & Strohmer 2015; ....

Statistical Dimension

Regularized Denoising

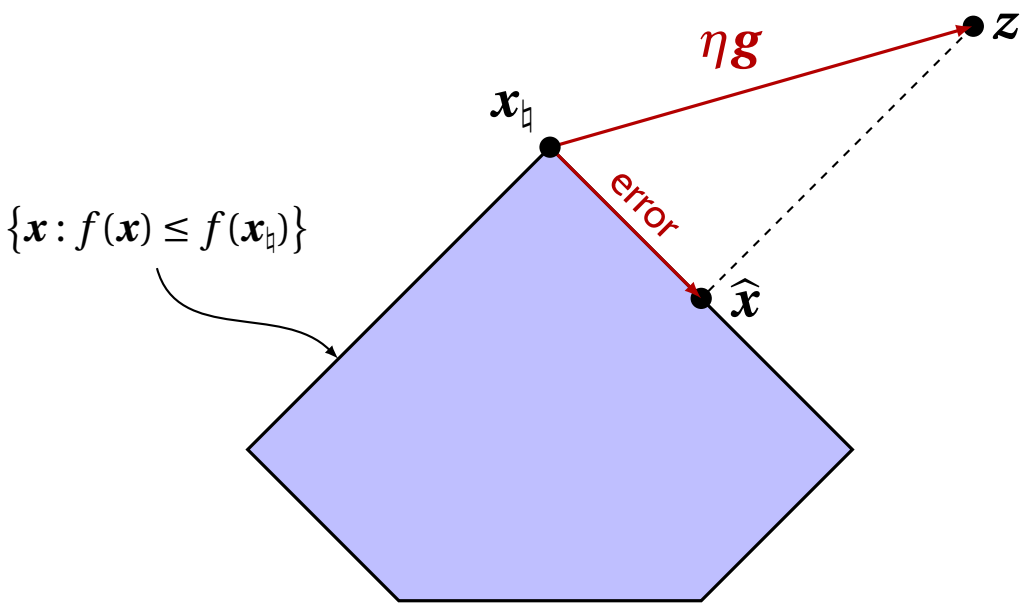
- Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex "structural" penalty (e.g., an atomic gauge)
- Let $\boldsymbol{x}_\natural$ be "structured" but unknown
- Observe $\boldsymbol{z} = \boldsymbol{x}_\natural + \eta \boldsymbol{g}$ where $\boldsymbol{g} \sim \text{NORMAL}(0, \mathbf{I})$
- Remove noise by solving the convex program

$$\text{minimize} \quad \|\boldsymbol{z} - \boldsymbol{x}\|^2 \qquad \text{subject to} \qquad f(\boldsymbol{x}) \leq f(\boldsymbol{x}_\natural)$$

- **Hope:** The minimizer $\widehat{\boldsymbol{x}}$ approximates $\boldsymbol{x}_\natural$
- **Remark:** Other formulations more practical, but this is easier to analyze

Sources: Donoho et al. 2009, 2013; Bhaskar et al. 2012; Chandrasekaran & Jordan 2013; Oymak & Hassibi 2013; Amelunxen et al. 2014.

Geometry of Atomic Denoising I



Cones and Projections

Definition 8. A convex cone is a convex set K that satisfies $K \subseteq \tau K$ for $\tau \geq 0$.

Definition 9. Let K be a convex cone. The polar is the closed convex cone $K^\circ = \{\boldsymbol{y} : \langle \boldsymbol{y}, \boldsymbol{x} \rangle \leq 0 \text{ for all } \boldsymbol{x} \in K\}$.

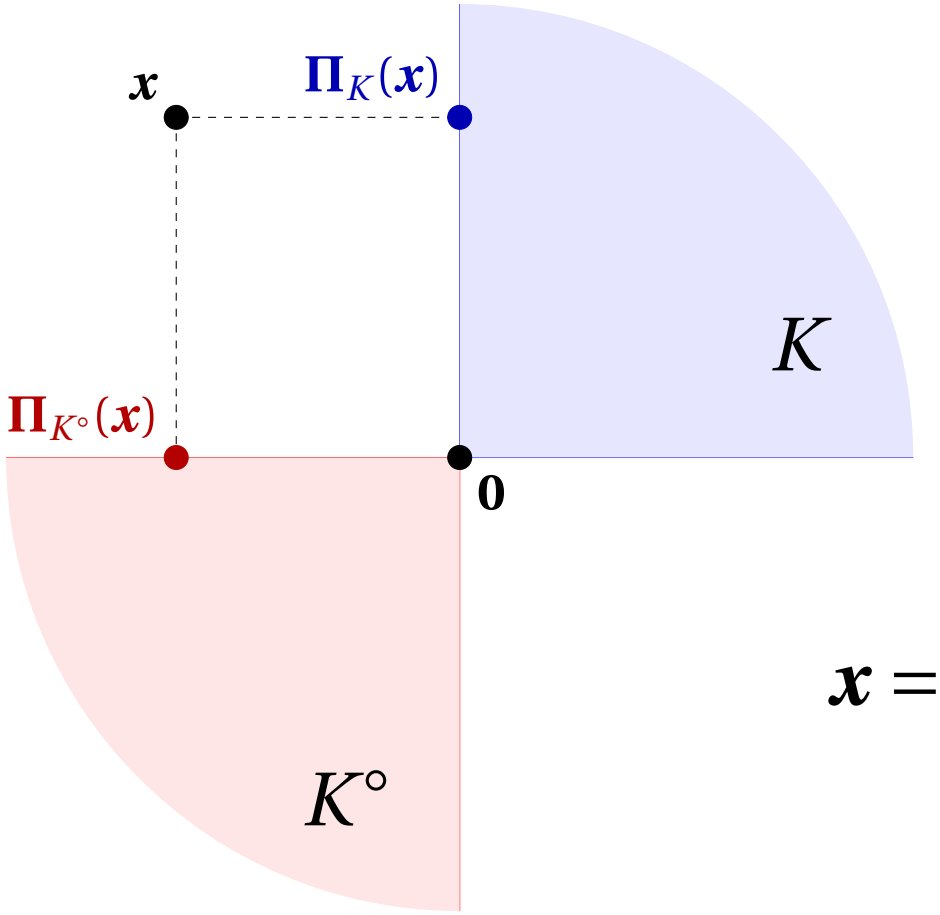
Definition 10. Let K be a closed convex cone. The Euclidean projection onto K is

$$\boldsymbol{\Pi}_K(\boldsymbol{x}) = \arg\min_{\boldsymbol{y} \in K} \|\boldsymbol{y} - \boldsymbol{x}\|^2$$

For a general convex cone C, define $\boldsymbol{\Pi}_C = \boldsymbol{\Pi}_{\text{closure}(C)}$

Sources: Rockafellar 1970; Rockafellar & Wets 1997; Hiriart–Urruty & Lemaréchal 2002.

Moreau's Theorem



$$x = \Pi_K(x) + \Pi_{K^\circ}(x)$$

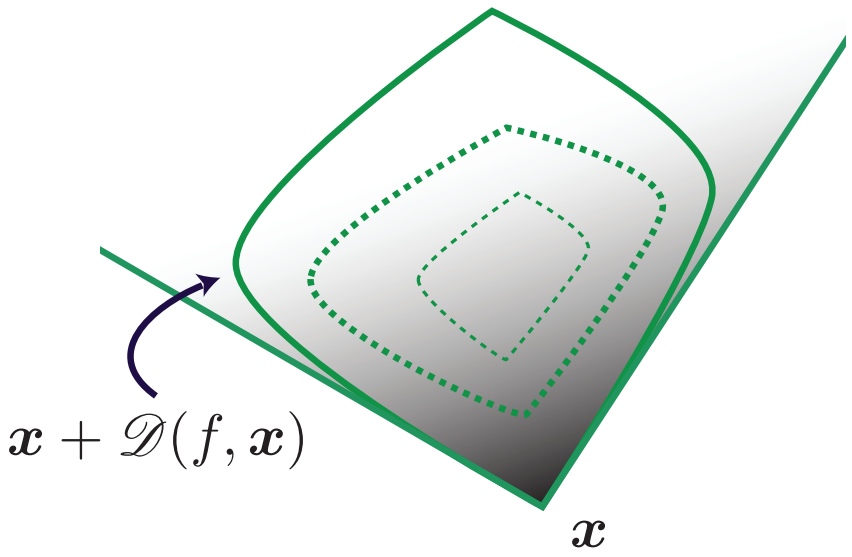
$$\Pi_K(x) \perp \Pi_{K^\circ}(x)$$

Sources: Moreau 1965; Rockafellar 1970; Rockafellar & Wets 1997; Hiriart-Urruty & Lemaréchal 2002.

Descent Cones

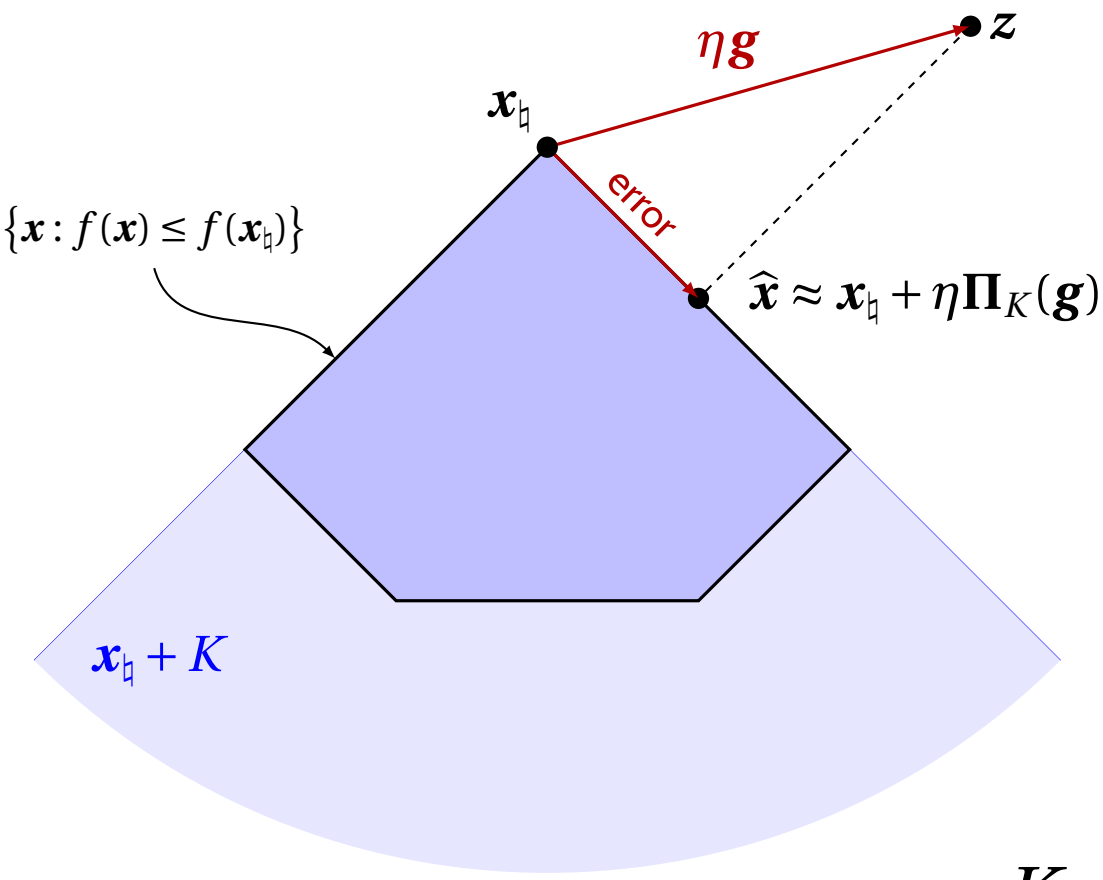
Definition 11. For convex f and a point $\boldsymbol{x}$, the descent cone is the convex cone

$$\mathcal{D}(f, \boldsymbol{x}) = \left\{ \boldsymbol{u} : f(\boldsymbol{x} + \varepsilon \boldsymbol{u}) \leq f(\boldsymbol{x}) \text{ for some } \varepsilon > 0 \right\}$$



Sources: Rockafellar 1970; Rockafellar & Wets 1997; Hiriart–Urruty & Lemaréchal 2002.

Geometry of Regularized Denoising II



$$K = \mathcal{D}(f, x_\natural)$$

Analysis of Regularized Denoising

Theorem 12 (Oymak & Hassibi 2013). ***Assume***

- f *is a convex function and* $\boldsymbol{x}_\natural \in \operatorname{dom}(f)$
- *Observe* $\boldsymbol{z} = \boldsymbol{x}_\natural + \eta \boldsymbol{g}$ *where* $\boldsymbol{g} \sim \textsc{normal}(\mathbf{0}, \mathbf{I})$
- *The vector* $\widehat{\boldsymbol{x}}$ *solves*

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

Then

$$\sup_{\eta > 0} \frac{\mathbb{E}\,\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_\natural\|^2}{\eta^2} \quad = \quad \lim_{\eta \downarrow 0} \frac{\mathbb{E}\,\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_\natural\|^2}{\eta^2} \quad = \quad \mathbb{E}\,\|\boldsymbol{\Pi}_K(\boldsymbol{g})\|^2$$

where $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$

Sources: Donoho et al. 2009, 2013; Oymak & Hassibi 2013; Chandrasekaran & Jordan 2013; Amelunxen et al. 2014.

Regularized Denoising: Upper Bound

- Change variables: $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{x}_\natural$

$$\text{minimize} \quad \|\boldsymbol{u} - \eta \boldsymbol{g}\|^2 \quad \text{subject to} \quad f(\boldsymbol{x}_\natural + \boldsymbol{u}) \le f(\boldsymbol{x}_\natural)$$

- First-order optimality for $\widehat{\boldsymbol{u}} = \widehat{\boldsymbol{x}} - \boldsymbol{x}_\natural$:

$$\langle \boldsymbol{v} - \widehat{\boldsymbol{u}},\ \widehat{\boldsymbol{u}} - \eta \boldsymbol{g} \rangle \ge 0 \quad \text{for all feasible } \boldsymbol{v}$$

- Choose $\boldsymbol{v} = \boldsymbol{0}$ and use the fact $\widehat{\boldsymbol{u}} \in \mathcal{D}(f, \boldsymbol{x}_\natural) = K$:

$$\|\widehat{\boldsymbol{u}}\|^2 \le \langle \eta \boldsymbol{g},\ \widehat{\boldsymbol{u}} \rangle = \eta \left\langle \boldsymbol{\Pi}_K(\boldsymbol{g}) + \boldsymbol{\Pi}_{K^\circ}(\boldsymbol{g}),\ \widehat{\boldsymbol{u}} \right\rangle \le \eta \langle \boldsymbol{\Pi}_K(\boldsymbol{g}),\ \widehat{\boldsymbol{u}} \rangle \le \eta \|\widehat{\boldsymbol{u}}\| \, \|\boldsymbol{\Pi}_K(\boldsymbol{g})\|$$

- Rearrange and take expectation:

$$\eta^{-2} \mathbb{E} \|\widehat{\boldsymbol{u}}\|^2 \quad \le \quad \mathbb{E} \|\boldsymbol{\Pi}_K(\boldsymbol{g})\|^2$$

- The lower bound is technical

Statistical Dimension

Definition 13. Let K be a convex cone in $\mathbb{R}^d$. The statistical dimension is

$$\delta(K) = \mathbb{E}\,\|\mathbf{\Pi}_K(\boldsymbol{g})\|^2 \quad \text{where} \quad \boldsymbol{g} \sim \text{NORMAL}(\mathbf{0}, \mathbf{I})$$

An extension of linear dimension to cones:

- $0 \le \delta(K) \le d$
- $C \subset K$ implies $\delta(C) \le \delta(K)$
- If L is a subspace, $\delta(L) = \dim(L)$

Sources: Amelunxen et al. 2014; McCoy & Tropp 2014.

Properties of Statistical Dimension

- Let C, K be convex cones in $\mathbb{R}^d$
- Gaussian formulation: $\delta(K) = \mathbb{E} \|\boldsymbol{\Pi}_K(\boldsymbol{g})\|^2$
  - Nonnegativity: $\delta(K) \geq 0$
  - Subspaces: $\delta(L) = \dim(L)$ for a subspace L
  - Rotational invariance: $\delta(K) = \delta(\boldsymbol{Q}K)$ for any orthogonal $\boldsymbol{Q}$
- Complementarity: $\delta(K) + \delta(K^\circ) = d$
  - Upper bound: $\delta(K) \leq d$
- Polar formulation: $\delta(K) = \mathbb{E} \operatorname{dist}^2(\boldsymbol{g}, K^\circ)$
- Mean-squared-width formulation: $\delta(K) = \mathbb{E} \left(\sup_{\|\boldsymbol{u}\| \leq 1, \boldsymbol{u} \in K} \langle \boldsymbol{g}, \boldsymbol{u} \rangle \right)^2$
  - Monotonicity: $C \subset K$ implies $\delta(C) \leq \delta(K)$

Sources: Chandrasekaran et al. 2012; Amelunxen et al. 2014; McCoy & Tropp 2014.

Proof of Statistical Dimension Properties

- Moreau's Theorem + basic facts about standard normal vectors
- Complementarity:

$$d = \mathbb{E}\,\|\boldsymbol{g}\|^2 = \mathbb{E}\,\|\boldsymbol{\Pi}_K(\boldsymbol{g})\|^2 + \mathbb{E}\,\|\boldsymbol{\Pi}_{K^\circ}(\boldsymbol{g})\|^2 = \delta(K) + \delta(K^\circ)$$

- Polar formulation:

$$\delta(K) = \mathbb{E}\,\|\boldsymbol{\Pi}_K(\boldsymbol{g})\|^2 = \mathbb{E}\,\|\boldsymbol{g} - \boldsymbol{\Pi}_{K^\circ}(\boldsymbol{g})\|^2 = \mathbb{E}\,\mathrm{dist}^2(\boldsymbol{g}, K^\circ)$$

- Mean-squared-width formulation:

$$\sup_{\substack{\|\boldsymbol{u}\| \le 1 \\ \boldsymbol{u} \in K}} \langle \boldsymbol{g},\, \boldsymbol{u} \rangle = \sup_{\substack{\|\boldsymbol{u}\| \le 1 \\ \boldsymbol{u} \in K}} \left\langle \boldsymbol{\Pi}_K(\boldsymbol{g}) + \boldsymbol{\Pi}_{K^\circ}(\boldsymbol{g}),\, \boldsymbol{u} \right\rangle \le \sup_{\substack{\|\boldsymbol{u}\| \le 1 \\ \boldsymbol{u} \in K}} \langle \boldsymbol{\Pi}_K(\boldsymbol{g}),\, \boldsymbol{u} \rangle = \|\boldsymbol{\Pi}_K(\boldsymbol{g})\|$$

$$\sup_{\substack{\|\boldsymbol{u}\| \le 1 \\ \boldsymbol{u} \in K}} \langle \boldsymbol{g},\, \boldsymbol{u} \rangle \ge \langle \boldsymbol{g},\, \boldsymbol{\Pi}_K(\boldsymbol{g}) / \|\boldsymbol{\Pi}_K(\boldsymbol{g})\| \rangle = \|\boldsymbol{\Pi}_K(\boldsymbol{g})\|$$

Gaussian Width and Statistical Dimension

Proposition 14 (Amelunxen et al. 2014). *Assume K is a convex cone in $\mathbb{R}^d$ and $\boldsymbol{g} \in \mathbb{R}^d$ is standard normal. Then*

$$\delta(K) - 1 \quad \leq \quad \left(\mathbb{E} \sup_{\|\boldsymbol{u}\|=1, \boldsymbol{u} \in K} \langle \boldsymbol{g}, \boldsymbol{u} \rangle \right)^2 \quad \leq \quad \delta(K)$$

- Upper bound (Jensen):

$$\left(\mathbb{E} \sup_{\|\boldsymbol{u}\|=1, \boldsymbol{u} \in K} \langle \boldsymbol{g}, \boldsymbol{u} \rangle \right)^2 \leq \mathbb{E} \left(\sup_{\|\boldsymbol{u}\|=1, \boldsymbol{u} \in K} \langle \boldsymbol{g}, \boldsymbol{u} \rangle \right)^2 \leq \mathbb{E} \left(\sup_{\|\boldsymbol{u}\| \leq 1, \boldsymbol{u} \in K} \langle \boldsymbol{g}, \boldsymbol{u} \rangle \right)^2 = \delta(K)$$

- Lower bound (Poincaré):

$$\begin{aligned} 1 + \left(\mathbb{E} \sup_{\|\boldsymbol{u}\|=1, \boldsymbol{u} \in K} \langle \boldsymbol{g}, \boldsymbol{u} \rangle \right)^2 &\geq \mathbb{E} \left(\sup_{\|\boldsymbol{u}\|=1, \boldsymbol{u} \in K} \langle \boldsymbol{g}, \boldsymbol{u} \rangle \right)^2 \\ &\geq \mathbb{E} \left[\left(\sup_{\|\boldsymbol{u}\|=1, \boldsymbol{u} \in K} \langle \boldsymbol{g}, \boldsymbol{u} \rangle \right)^2 \cdot \mathbb{1}_{\boldsymbol{g} \in \mathbb{R}^d \setminus K^\circ} \right] = \mathbb{E} \left[\left(\sup_{\|\boldsymbol{u}\| \leq 1, \boldsymbol{u} \in K} \langle \boldsymbol{g}, \boldsymbol{u} \rangle \right)^2 \right] = \delta(K) \end{aligned}$$

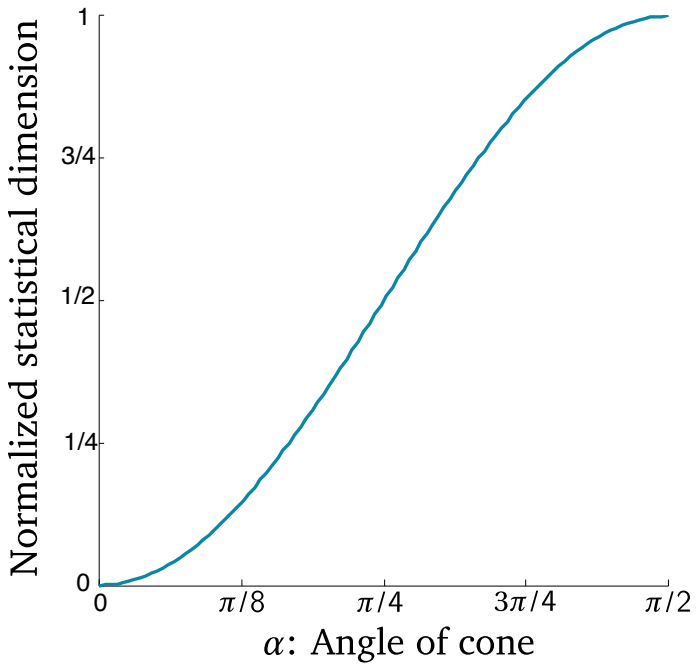
Source: Amelunxen et al. 2014.

Basic Examples of Statistical Dimension

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	$\mathbb{H}_+^d$	$\frac{1}{2}d^2$

Sources: Chandrasekaran et al. 2012; Amelunxen et al. 2014; McCoy & Tropp 2014.

Circular Cones



- For $\alpha \in (0, \pi/2)$, define $\mathrm{Circ}_d(\alpha) = \{\boldsymbol{x} \in \mathbb{R}^d : x_1 \geq \|\boldsymbol{x}\| \cos(\alpha)\}$
- $\delta\big(\mathrm{Circ}_d(\alpha)\big) = d \sin^2(\alpha) + \cos(2\alpha) + o(1)$

Sources: Amelunxen et al. 2014; McCoy & Tropp 2014.

Polar Form of a Descent Cone

Definition 15. The subdifferential ∂f of a convex function f at a point $\boldsymbol{x}$ is

$$\partial f(\boldsymbol{x}) = \left\{ \boldsymbol{u} : f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{u},\ \boldsymbol{y} - \boldsymbol{x} \rangle \text{ for all } \boldsymbol{y} \right\}$$

Fact 16. *Assume that $\partial f(\boldsymbol{x})$ is nonempty, compact, and does not contain the origin. Then*

$$\mathcal{D}(f, \boldsymbol{x})^\circ = \bigcup_{\tau \geq 0} \tau \cdot \partial f(\boldsymbol{x})$$

Remark: Compactness is not essential.

Sources: Rockafellar 1970; Rockafellar & Wets 1998; Hiriart-Urruty & Lemaréchal 2002.

The Descent Cone Recipe

Proposition 17 (Amelunxen et al. 2014)**.** *Assume that f is a convex function whose subdifferential $\partial f(\boldsymbol{x})$ is nonempty and does not contain the origin. Then*

$$\delta\big(\mathcal{D}(f,\boldsymbol{x})\big) \quad \leq \quad \inf_{\tau \geq 0} \mathbb{E}\, \operatorname{dist}^2\big(\boldsymbol{g}, \tau \cdot \partial f(\boldsymbol{x})\big) \quad = \quad \inf_{\tau \geq 0} J(\tau)$$

where $\boldsymbol{g}$ is standard normal

- Assume $\partial f(\boldsymbol{x})$ is compact for simplicity
- Calculate:

$$\begin{aligned}\delta\big(\mathcal{D}(f,\boldsymbol{x})\big) &= \mathbb{E}\, \operatorname{dist}^2\big(\boldsymbol{g}, \mathcal{D}(f,\boldsymbol{x})^\circ\big) = \mathbb{E}\, \operatorname{dist}^2\Big(\boldsymbol{g}, \bigcup_{\tau \geq 0} \tau \cdot \partial f(\boldsymbol{x})\Big) \\ &= \mathbb{E} \inf_{\tau \geq 0} \operatorname{dist}^2\big(\boldsymbol{g}, \tau \cdot \partial f(\boldsymbol{x})\big) \leq \inf_{\tau \geq 0} \mathbb{E}\, \operatorname{dist}^2\big(\boldsymbol{g}, \tau \cdot \partial f(\boldsymbol{x})\big)\end{aligned}$$

Sources: Stojnic 2009, 2013; Chandrasekaran et al. 2012; Amelunxen et al. 2014; Foygel & Mackey 2014.

The Descent Cone Recipe: Error Estimate

Theorem 18 (Amelunxen et al. 2014)**.** *Assume that f is a norm. For nonzero $\boldsymbol{x}$,*

$$0 \quad \leq \quad \inf_{\tau \geq 0} J(\tau) - \delta\big(\mathcal{D}(f, \boldsymbol{x})\big) \quad \leq \quad \frac{2 \max\big\{ \|\boldsymbol{u}\| : \boldsymbol{u} \in \partial f(\boldsymbol{x}) \big\}}{f(\boldsymbol{x} / \|\boldsymbol{x}\|)}$$

- **Idea:** Linearize $\tau \mapsto \operatorname{dist}^2\big(\boldsymbol{g}, \tau \cdot \partial f(\boldsymbol{x})\big)$ at minimizer $\tau_\star$ of J
- Related result of Foygel & Mackey based on other ideas
- Still room for improvement

Sources: Amelunxen et al. 2014; Foygel & Mackey 2014.

Example: ℓ_1 Statistical Dimension I

- **Goal:** Compute $\delta\big(\mathcal{D}(\|\cdot\|_{\ell_1}, \boldsymbol{x}_s)\big)$ for vector $\boldsymbol{x}_s \in \mathbb{R}^d$ with s nonzero entries
- By symmetry, can assume $\boldsymbol{x}_s = (\mathbf{1}_s, \mathbf{0}_{d-s})$
- Subdifferential: $\partial f(\boldsymbol{x}_s) = \big\{(\mathbf{1}_s, \boldsymbol{v}) \in \mathbb{R}^d : \|\boldsymbol{v}\|_\infty \le 1\big\}$
- Distance to scaled subdifferential:

$$J_s(\tau) = \mathbb{E}\,\mathrm{dist}^2\big(\boldsymbol{g}, \tau \cdot \partial \|\boldsymbol{x}_s\|_{\ell_1}\big) = \sum_{i=1}^{s} \mathbb{E}(g_i - \tau)^2 + \sum_{i=s+1}^{d} \mathbb{E} \min_{|v_i| \le 1} (g_i - \tau v_i)^2$$

$$= s \cdot \mathbb{E}(g - \tau)^2 + (d-s) \cdot \mathbb{E}(|g| - \tau)_+^2 \le s \cdot (1 + \tau^2) + 2(d-s) \cdot \mathrm{e}^{-\tau^2/2}$$

- Choose $\tau^2 = 2\log(d/s)$ to reach

$$\delta\big(\mathcal{D}(\|\cdot\|_{\ell_1}, \boldsymbol{x}_s)\big) \quad \le \quad \inf_{\tau \ge 0} J_s(\tau) \quad \le \quad 2s(1 + \log(d/s))$$

Sources: Donoho 2006; Donoho & Tanner 2009; Stojnic 2009, 2013; Chandrasekaran et al. 2012; Amelunxen et al. 2014; Foygel & Mackey 2014; Goldstein et al. 2014.

Example: ℓ_1 Statistical Dimension II

- $\boldsymbol{x}_s \in \mathbb{R}^d$ has s nonzero entries
- Observe $\boldsymbol{z} = \boldsymbol{x}_s + \eta \boldsymbol{g}$
- $\widehat{\boldsymbol{x}}$ solves

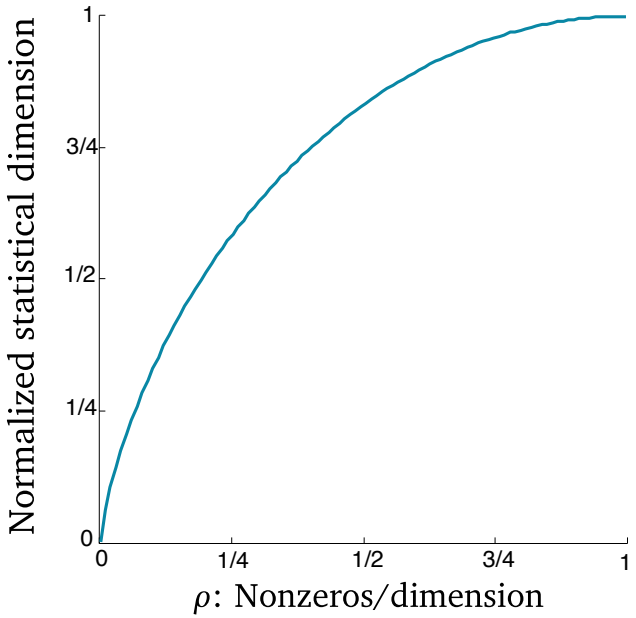
$$\text{minimize} \quad \|\boldsymbol{x} - \boldsymbol{z}\|^2 \quad \text{subject to} \quad \|\boldsymbol{x}\|_{\ell_1} \le \|\boldsymbol{x}_s\|_{\ell_1}$$

- Then

$$\mathbb{E}\,\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_\natural\|^2 \quad \le \quad 2s(1 + \log(d/s)) \cdot \eta^2$$

- Almost achieve same MSE as if we knew $\operatorname{supp}(\boldsymbol{x}_s)$!

Example: ℓ_1 Statistical Dimension III

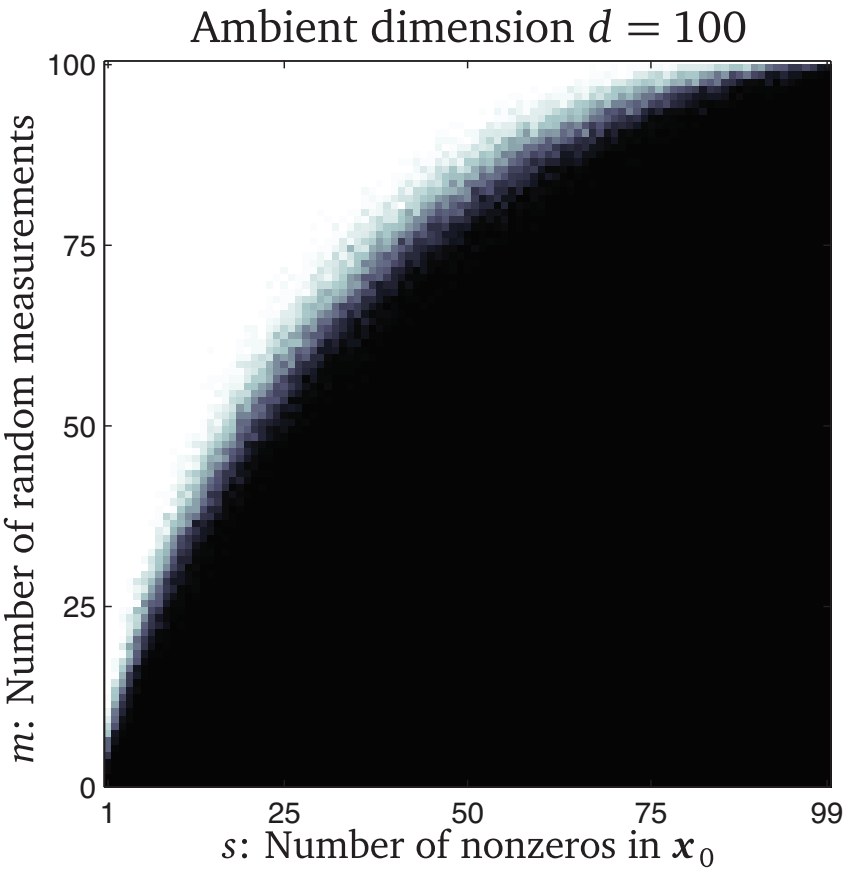


- With $\rho = s/d \in [0, 1]$,

$$\frac{\delta\big(\mathcal{D}(f, \boldsymbol{x}_s)\big)}{d} \quad \leq \quad \inf_{\tau \geq 0} \left[\rho(1+\tau^2) + (1-\rho)\,\mathbb{E}(|g| - \tau)_+^2\right] \quad \leq \quad \frac{\delta\big(\mathcal{D}(f, \boldsymbol{x}_s)\big)}{d} + \frac{1}{d\sqrt{\rho}}$$

Sources: Affentranger & Schneider 1992; Betke & Henk 1993; Böröczky & Henk 1999; Donoho 2006; Donoho & Tanner 2009; Stojnic 2009, 2013; McCoy & Tropp 2013; Amelunxen et al. 2014; Foygel & Mackey 2014; Goldstein et al. 2014.

Compressed Sensing: Hmmm...



Sources: Donoho et al. 2009–2013.

Example: S_1 Statistical Dimension I

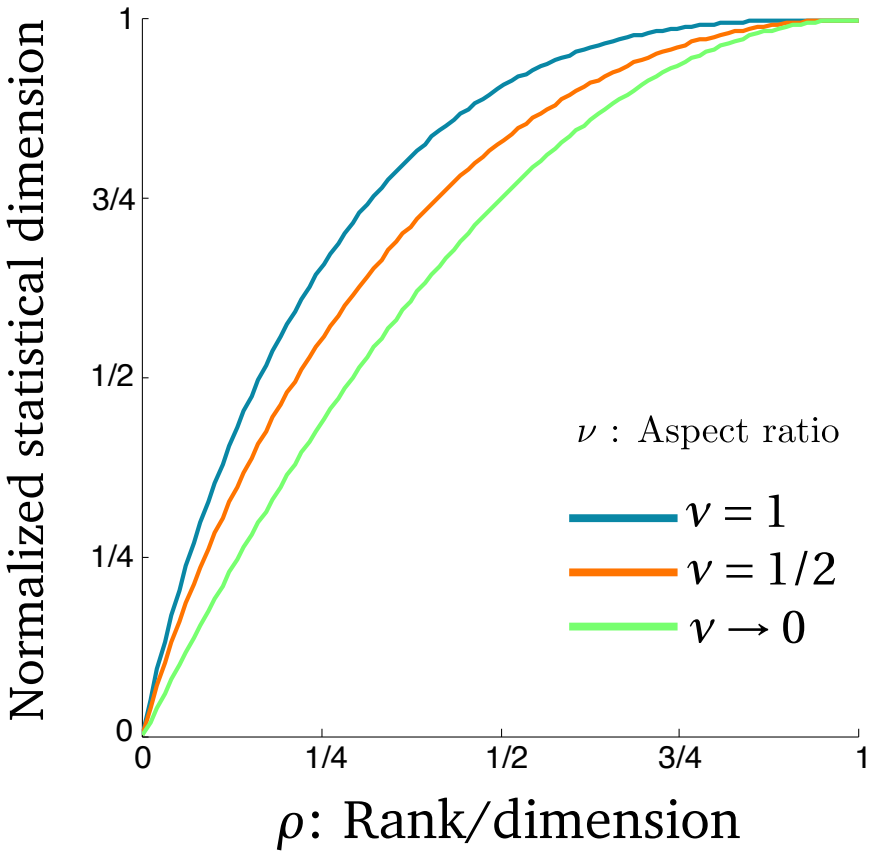
- $\boldsymbol{X}_r \in \mathbb{R}^{d_1 \times d_2}$ has rank r
- Observe $\boldsymbol{Z} = \boldsymbol{X}_r + \eta \boldsymbol{G}$
- $\widehat{\boldsymbol{X}}$ solves

$$\text{minimize} \quad \|\boldsymbol{X} - \boldsymbol{Z}\|_{\mathrm{F}}^2 \quad \text{subject to} \quad \|\boldsymbol{X}\|_{S_1} \leq \|\boldsymbol{X}_r\|_{S_1}$$

- Then

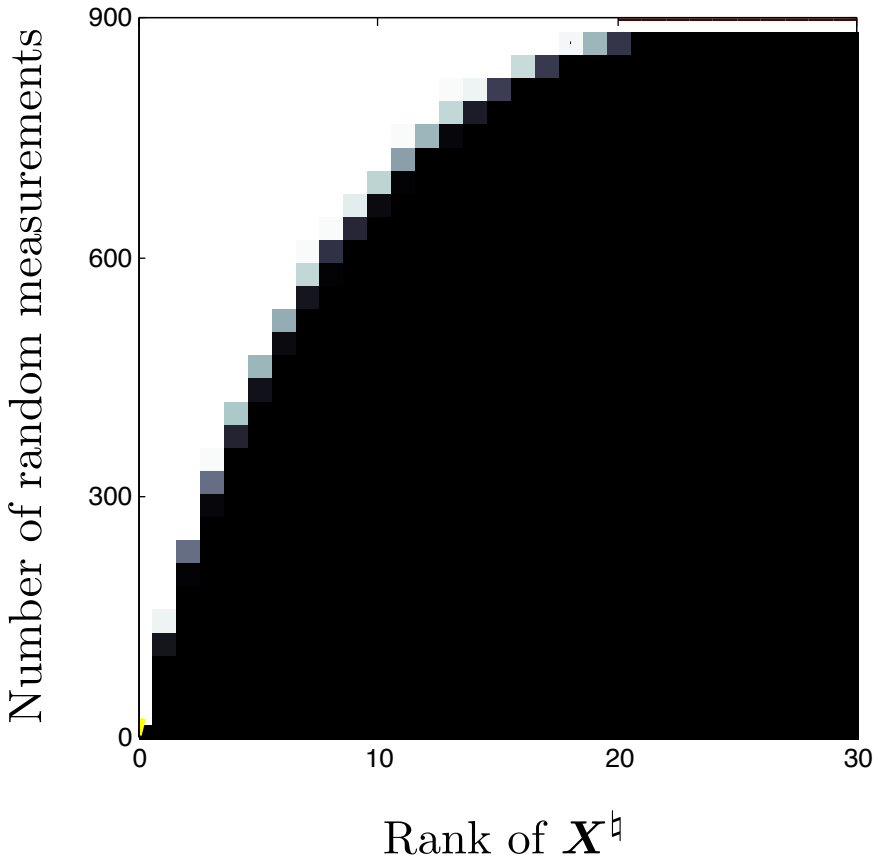
$$\mathbb{E}\,\|\widehat{\boldsymbol{X}} - \boldsymbol{X}_r\|_{\mathrm{F}}^2 \quad \leq \quad \delta\big(\mathcal{D}(\|\cdot\|_{S_1}, \boldsymbol{X}_r)\big) \cdot \eta^2 \quad \leq \quad 3r(d_1 + d_2 - r) \cdot \eta^2$$

Example: S_1 Statistical Dimension II



Sources: Oymak et al. 2010; Chandrasekaran et al. 2012; Amelunxen et al. 2014; McCoy & Tropp 2013, 2014; Goldstein et al. 2014.

Matrix Compressed Sensing: Hmmm...



Sources: Recht et al. 2010; Oymak et al. 2010; Donoho et al. 2013.

A Conjecture...

PNAS

The phase transition of matrix recovery from Gaussian measurements matches the minimax MSE of matrix denoising

David L. Donoho[a,1], Matan Gavish[a], and Andrea Montanari[a,b]

Departments of [a]Statistics and [b]Electrical Engineering, Stanford University, Stanford, CA 94305

Contributed by David L. Donoho, April 3, 2013 (sent for review February 5, 2013)

Let X_0 be an unknown M by N matrix. In matrix recovery, one takes $n < MN$ linear measurements $y_1,\ldots,y_n$ of X_0, where $y_i = \mathrm{Tr}(A_i^T X_0)$ and each A_i is an M by N matrix. A popular approach for matrix recovery is nuclear norm minimization (NNM): solving the convex optimization problem min $\|X\|_*$ subject to $y_i = \mathrm{Tr}(A_i^T X)$ for all $1 \le i \le n$, where $\|\cdot\|_*$ denotes the nuclear norm, namely, the sum of singular values. Empirical work reveals a phase transition curve, stated in terms of the undersampling fraction $\delta(n,\mathrm{M},\mathrm{N}) = n/(MN)$, rank fraction $\rho = \mathit{rank}(X_0)/\min\{M,\mathrm{N}\}$, and aspect ratio $\beta = M/N$. Specifically when the measurement matrices A_i have independent standard Gaussian random entries, a curve $\delta^*(\rho) = \delta^*(\rho;\beta)$ exists such that, if $\delta > \delta^*(\rho)$, NNM typically succeeds for large M,N, whereas if $\delta < \delta^*(\rho)$, it typically fails. An apparently quite different problem is matrix denoising in Gaussian noise, in which an unknown M by N matrix X_0 is to be estimated based on direct noisy measurements $Y = X_0 + Z$, where the matrix Z has independent and identically distributed Gaussian entries. A popular

measurement operator $\mathcal{A}$, then the solution $X_1 = X_1(y)$ to (P_{nuc}) is precisely X_0. Such incoherence can be obtained by letting $\mathcal{A}$ be random, for instance if $\mathcal{A}(X_0)_i = \mathrm{Tr}(A_i^T X_0)$ with $A_i \in \mathbb{R}^{m\times n}$ having independent and identically distributed (i.i.d) Gaussian entries. In this case we speak of "matrix recovery from Gaussian measurements" (1).

A key phrase from the previous paragraph is "if X_0 is sufficiently low rank." Clearly, there must be a quantitative tradeoff between the rank of X_0 and the number of measurements required, such that higher-rank matrices require more measurements. In the Gaussian measurements model, with N, M sufficiently large, empirical work by Recht et al. (1, 7, 8), Tanner and Wei (9), and Oymak and Hassibi (10) documents a phase transition phenomenon. For matrices of a given rank, there is a fairly precise number of required samples, in the sense that a transition from nonrecovery to complete recovery takes place sharply as the number of samples varies across this value. For example, in Table S1 we present data from

APPLIED MATHEMATICS

Gaussian Comparisons

Slepian's Lemma

Theorem 19 (Slepian 1962). ***Assume*** *$\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{y} \in \mathbb{R}^N$ be centered, jointly Gaussian vectors whose covariance structures satisfy*

$$\begin{cases} \mathbb{E}\, X_i X_j \le \mathbb{E}\, Y_i Y_j & \text{for all } i \neq j \\ \mathbb{E}\, X_i^2 = \mathbb{E}\, Y_i^2 & \text{for all } i \end{cases}$$

Then*, for all choices of $\lambda_i \in \mathbb{R}$,*

$$\mathbb{P}\left(\bigcup_{i=1}^N \{Y_i > \lambda_i\}\right) \le \mathbb{P}\left(\bigcup_{i=1}^N \{X_i > \lambda_i\}\right)$$

In particular,

$$\mathbb{E} \max_i Y_i \le \mathbb{E} \max_i X_i$$

Sources: Slepian 1962; Sudakov 1969, 1971; Marcus & Shepp 1970, 1972; Fernique 1974; Joag-Dev et al. 1983; Kahane 1986; Ledoux & Talagrand (Cor. 3.12).

Chevet's Theorem

Corollary 20 (Chevet 1977, Gordon 1985). ***Assume***

- $U \subset \mathbb{R}^m$ *and* $V \subset \mathbb{R}^n$ *are compact subsets of the unit sphere*
- $\mathbf{\Gamma} \in \mathbb{R}^{n \times m}$ *and* $\boldsymbol{g} \in \mathbb{R}^m$ *and* $\boldsymbol{h} \in \mathbb{R}^n$ *are independent standard normal*

Then

$$\mathbb{E} \max_{\boldsymbol{u} \in U, \boldsymbol{v} \in V} \langle \mathbf{\Gamma} \boldsymbol{u}, \boldsymbol{v} \rangle \quad \leq \quad \mathbb{E} \max_{\boldsymbol{u} \in U, \boldsymbol{v} \in V} \big[\langle \boldsymbol{g}, \boldsymbol{u} \rangle + \langle \boldsymbol{h}, \boldsymbol{v} \rangle \big]$$

Sources: Chevet 1977; Gordon 1985, 1988; Ledoux & Talagrand (Thm. 3.20); Davidson & Szarek 2001; Stojnic 2013; Amelunxen et al. 2014; Thrampoulidis et al. 2015; Amelunxen & Lotz 2015.

Example: Spectral Norm of a Gaussian Matrix

- Goal: For standard normal $\boldsymbol{\Gamma} \in \mathbb{R}^{n \times m}$, bound expectation of

$$\|\boldsymbol{\Gamma}\| = \max_{\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1} \langle \boldsymbol{\Gamma} \boldsymbol{u}, \boldsymbol{v} \rangle$$

- Apply Chevet's Theorem:

$$\begin{aligned}
\mathbb{E}\,\|\boldsymbol{\Gamma}\| &= \mathbb{E} \max_{\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1} \langle \boldsymbol{\Gamma} \boldsymbol{u}, \boldsymbol{v} \rangle \\
&\leq \mathbb{E} \max_{\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1} \left[\langle \boldsymbol{g}, \boldsymbol{u} \rangle + \langle \boldsymbol{h}, \boldsymbol{v} \rangle \right] \\
&= \mathbb{E}\,\|\boldsymbol{g}\| + \mathbb{E}\,\|\boldsymbol{h}\| \\
&\leq \sqrt{m} + \sqrt{n}
\end{aligned}$$

- Result is sharp, including constants!

Sources: Marchenko & Pastur 1967; Chevet 1977; Gordon 1985, 1988; Yin et al. 1988; Bai et al. 1988; Edelman 1988; Ledoux & Talagrand (Thm. 3.20); Davidson & Szarek 2001.

Proof of Chevet's Theorem

- Let $\gamma \in \mathbb{R}$ be an independent standard normal
- Define independent Gaussian processes on $\{(\boldsymbol{u}, \boldsymbol{v}) : \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1\}$:

$$Y_{\boldsymbol{uv}} = \langle \boldsymbol{\Gamma u},\ \boldsymbol{v} \rangle + \gamma \quad \text{and} \quad X_{\boldsymbol{uv}} = \langle \boldsymbol{g},\ \boldsymbol{u} \rangle + \langle \boldsymbol{h},\ \boldsymbol{v} \rangle$$

- Compute the covariances:

$$\mathbb{E}\, Y_{\boldsymbol{uv}} Y_{\boldsymbol{u'v'}} = \langle \boldsymbol{u},\ \boldsymbol{u'} \rangle \langle \boldsymbol{v},\ \boldsymbol{v'} \rangle + 1 \quad \text{and} \quad \mathbb{E}\, X_{\boldsymbol{uv}} X_{\boldsymbol{u'v'}} = \langle \boldsymbol{u},\ \boldsymbol{u'} \rangle + \langle \boldsymbol{v},\ \boldsymbol{v'} \rangle$$

- Comparison:

$$\mathbb{E}\, Y_{\boldsymbol{uv}} Y_{\boldsymbol{u'v'}} - \mathbb{E}\, X_{\boldsymbol{uv}} X_{\boldsymbol{u'v'}} = \big(1 - \langle \boldsymbol{u},\ \boldsymbol{u'} \rangle\big)\big(1 - \langle \boldsymbol{v},\ \boldsymbol{v'} \rangle\big) \geq 0$$

$$\mathbb{E}\, Y_{\boldsymbol{uv}}^2 - \mathbb{E}\, X_{\boldsymbol{uv}}^2 = 0$$

- Apply Slepian's Lemma (on finite subsets of U, V):

$$\mathbb{E} \max_{\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1} \langle \boldsymbol{\Gamma u},\ \boldsymbol{v} \rangle = \mathbb{E} \max_{\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1} Y_{\boldsymbol{uv}} \leq \mathbb{E} \max_{\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1} X_{\boldsymbol{uv}} = \mathbb{E} \max_{\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1} \big[\, \langle \boldsymbol{g},\ \boldsymbol{u} \rangle + \langle \boldsymbol{h},\ \boldsymbol{v} \rangle \,\big]$$

Gordon's Theorem

Theorem 21 (Gordon 1985). ***Assume*** *$\boldsymbol{X} \in \mathbb{R}^{M \times N}$ and $\boldsymbol{Y} \in \mathbb{R}^{M \times N}$ are centered, jointly Gaussian matrices whose covariance structures satisfy*

$$\begin{cases} \mathbb{E}\, X_{ij} X_{k\ell} \le \mathbb{E}\, Y_{ij} Y_{k\ell} & \text{for all } i \neq k \text{ and all } j, \ell \\ \mathbb{E}\, X_{ij} X_{i\ell} \ge \mathbb{E}\, Y_{ij} Y_{i\ell} & \text{for all } i \text{ and } j \neq \ell \\ \mathbb{E}\, X_{ij}^2 = \mathbb{E}\, Y_{ij}^2 & \text{for all } i, j \end{cases}$$

Then*, for all choices of $\lambda_{ij} \in \mathbb{R}$,*

$$\mathbb{P}\left(\bigcap_{i=1}^{M} \bigcup_{j=1}^{N} \{ Y_{ij} > \lambda_{ij} \} \right) \ge \mathbb{P}\left(\bigcap_{i=1}^{M} \bigcup_{j=1}^{N} \{ X_{ij} > \lambda_{ij} \} \right)$$

In particular,

$$\mathbb{E} \min_i \max_j Y_{ij} \ge \mathbb{E} \min_i \max_j X_{ij}$$

Sources: Joag-Dev et al. 1983; Gordon 1985, 1988; Kahane 1986; Ledoux & Talagrand (Cor. 3.13).

The Gaussian Minimax Theorem

Corollary 22 (Gordon 1985). ***Assume***

- $U \subset \mathbb{R}^m$ *and* $V \subset \mathbb{R}^n$ *are compact subsets of the unit sphere*
- $\boldsymbol{\Gamma} \in \mathbb{R}^{n \times m}$ *and* $\boldsymbol{g} \in \mathbb{R}^m$ *and* $\boldsymbol{h} \in \mathbb{R}^n$ *are independent standard normal*

Then

$$\mathbb{E} \min_{\boldsymbol{u} \in U} \max_{\boldsymbol{v} \in V} \langle \boldsymbol{\Gamma} \boldsymbol{u}, \boldsymbol{v} \rangle \quad \geq \quad \mathbb{E} \min_{\boldsymbol{u} \in U} \max_{\boldsymbol{v} \in V} \big[\langle \boldsymbol{g}, \boldsymbol{u} \rangle + \langle \boldsymbol{h}, \boldsymbol{v} \rangle \big]$$

Sources: Chevet 1977; Gordon 1985, 1988; Ledoux & Talagrand (Thm. 3.20); Davidson & Szarek 2001; Stojnic 2013; Amelunxen et al. 2014; Thrampoulidis et al. 2015; Amelunxen & Lotz 2015.

Example: Minimum Singular Value of a Gaussian Matrix

- Goal: For standard normal $\mathbf{\Gamma} \in \mathbb{R}^{n \times m}$, bound

$$\mathbb{E}\,\sigma_{\min}(\mathbf{\Gamma}) = \min_{\|\boldsymbol{u}\|=1} \max_{\|\boldsymbol{v}\|=1} \langle \mathbf{\Gamma}\boldsymbol{u},\ \boldsymbol{v} \rangle$$

- Gaussian Minimax Theorem:

$$\begin{aligned}
\mathbb{E}\,\sigma_{\min}(\mathbf{\Gamma}) &= \min_{\|\boldsymbol{u}\|=1} \max_{\|\boldsymbol{v}\|=1} \langle \mathbf{\Gamma}\boldsymbol{u},\ \boldsymbol{v} \rangle \\
&\geq \mathbb{E} \min_{\|\boldsymbol{u}\|=1} \max_{\|\boldsymbol{v}\|=1} \big[\langle \boldsymbol{g},\ \boldsymbol{u} \rangle + \langle \boldsymbol{h},\ \boldsymbol{v} \rangle \big] \\
&= \mathbb{E}\,\|\boldsymbol{h}\| - \mathbb{E}\,\|\boldsymbol{g}\| \\
&\geq \sqrt{n-1} - \sqrt{m}
\end{aligned}$$

- Result is sharp, including constants!
- **Remark:** Can replace $\sqrt{n-1}$ with $\sqrt{n}$ if you work hard enough

Sources: Marchenko & Pastur 1967; Chevet 1977; Silverstein 1985; Gordon 1985, 1988; Ledoux & Talagrand (Thm. 3.20); Szarek 1991; Bai & Yin 1993; Davidson & Szarek 2001.

Proof of Gaussian Minimax Theorem

- Define independent Gaussian processes on $\{(\boldsymbol{u}, \boldsymbol{v}) : \boldsymbol{u} \in U, \boldsymbol{v} \in V\}$:

$$Y_{\boldsymbol{uv}} = \langle \boldsymbol{\Gamma u},\ \boldsymbol{v} \rangle + \gamma \quad \text{and} \quad X_{\boldsymbol{uv}} = \langle \boldsymbol{g},\ \boldsymbol{u} \rangle + \langle \boldsymbol{h},\ \boldsymbol{v} \rangle$$

- Compute the covariances:

$$\mathbb{E}\, Y_{\boldsymbol{uv}} Y_{\boldsymbol{u'v'}} = \langle \boldsymbol{u},\ \boldsymbol{u'} \rangle \langle \boldsymbol{v},\ \boldsymbol{v'} \rangle + 1 \quad \text{and} \quad \mathbb{E}\, X_{\boldsymbol{uv}} X_{\boldsymbol{u'v'}} = \langle \boldsymbol{u},\ \boldsymbol{u'} \rangle + \langle \boldsymbol{v},\ \boldsymbol{v'} \rangle$$

- Comparison:

$$\mathbb{E}\, Y_{\boldsymbol{uv}} Y_{\boldsymbol{u'v'}} - \mathbb{E}\, X_{\boldsymbol{uv}} X_{\boldsymbol{u'v'}} = \big(1 - \langle \boldsymbol{u},\ \boldsymbol{u'} \rangle\big)\big(1 - \langle \boldsymbol{v},\ \boldsymbol{v'} \rangle\big) \geq 0$$

$$\mathbb{E}\, Y_{\boldsymbol{uv}} Y_{\boldsymbol{uv'}} - \mathbb{E}\, X_{\boldsymbol{uv}} X_{\boldsymbol{uv'}} = 0$$

$$\mathbb{E}\, Y_{\boldsymbol{uv}}^2 - \mathbb{E}\, X_{\boldsymbol{uv}}^2 = 0$$

- Apply Gordon's Theorem (on finite subsets of U, V):

$$\begin{aligned} \mathbb{E} \min_{\boldsymbol{u} \in U} \max_{\boldsymbol{v} \in V} \langle \boldsymbol{\Gamma u},\ \boldsymbol{v} \rangle &= \mathbb{E} \min_{\boldsymbol{u} \in U} \max_{\boldsymbol{v} \in V} Y_{\boldsymbol{uv}} \\ &\geq \mathbb{E} \min_{\boldsymbol{u} \in U} \max_{\boldsymbol{v} \in V} X_{\boldsymbol{uv}} = \mathbb{E} \min_{\boldsymbol{u} \in U} \max_{\boldsymbol{v} \in V} \big[\langle \boldsymbol{g},\ \boldsymbol{u} \rangle + \langle \boldsymbol{h},\ \boldsymbol{v} \rangle \big] \end{aligned}$$

Kahane's Approach to Gaussian Comparison

Theorem 23 (Kahane 1986). ***Assume*** *$\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{y} \in \mathbb{R}^N$ are centered, jointly Gaussian vectors. For sets $A, B \subset \{1, \dots, N\}^2$, assume that the covariance structures satisfy*

$$\begin{cases} \mathbb{E}\, X_i X_j \le \mathbb{E}\, Y_i Y_j & \text{for } (i,j) \in A \\ \mathbb{E}\, X_i X_j \ge \mathbb{E}\, Y_i Y_j & \text{for } (i,j) \in B \\ \mathbb{E}\, X_i X_j = \mathbb{E}\, Y_i Y_j & \text{for } (i,j) \notin A \cup B \end{cases}$$

Let $f : \mathbb{R}^N \to \mathbb{R}$ be a function whose second (distributional) derivative satisfies

$$\begin{cases} \partial_{ij} f \ge 0 & \text{for } (i,j) \in A \\ \partial_{ij} f \le 0 & \text{for } (i,j) \in B \end{cases}$$

Then

$$\mathbb{E}\, f(\boldsymbol{x}) \le \mathbb{E}\, f(\boldsymbol{y})$$

Sources: Joag-Dev et al. 1983; Kahane 1986; Ledoux & Talagrand (Thm. 3.11).

Gaussian Integration by Parts I

Lemma 24 (Univariate Gaussian IBP)**.** *Let $\gamma \in \mathbb{R}$ be a standard normal random variable. For "any" function $f : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}\left[\gamma f(\gamma)\right] = \mathbb{E}\left[f'(\gamma)\right]$$

- Calculate:

$$\mathbb{E}\left[\gamma f(\gamma)\right] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u f(u) \mathrm{e}^{-u^2/2} \,\mathrm{d}u = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f'(u) \mathrm{e}^{-u^2/2} \,\mathrm{d}u = \mathbb{E}\left[f'(\gamma)\right]$$

- Sufficient that f absolutely continuous and $f' \in L_1(\mathrm{d}\gamma)$

Gaussian Integration by Parts II

Lemma 25 (Gaussian IBP). *Let $\boldsymbol{x} \in \mathbb{R}^n$ be a centered, jointly Gaussian vector with covariance $\boldsymbol{\Sigma}$. For "any" function $f : \mathbb{R}^n \to \mathbb{R}$,*

$$\mathbb{E}\left[X_i f(\boldsymbol{x})\right] = \sum_{j=1}^{n} (\boldsymbol{\Sigma})_{ij} \mathbb{E}\left[(\partial_j f)(\boldsymbol{x})\right]$$

- Since $\boldsymbol{x} \sim \boldsymbol{\Sigma}^{1/2} \boldsymbol{z}$ for standard normal $\boldsymbol{z} \in \mathbb{R}^n$,

$$\mathbb{E}\left[X_i f(\boldsymbol{x})\right] = \sum_{k=1}^{n} \left(\boldsymbol{\Sigma}^{1/2}\right)_{ik} \mathbb{E}\left[Z_k f\left(\boldsymbol{\Sigma}^{1/2} \boldsymbol{z}\right)\right] = \sum_{k=1}^{n} \left(\boldsymbol{\Sigma}^{1/2}\right)_{ik} \mathbb{E}\left[Z_k g(\boldsymbol{z})\right]$$

- By univariate Gaussian IBP,

$$\mathbb{E}\left[Z_k g(\boldsymbol{z})\right] = \mathbb{E}\left[(\partial_k g)(\boldsymbol{z})\right] = \sum_{j=1}^{n} \left(\boldsymbol{\Sigma}^{1/2}\right)_{kj} \mathbb{E}\left[(\partial_j f)\left(\boldsymbol{\Sigma}^{1/2} \boldsymbol{z}\right)\right] = \sum_{j=1}^{n} \left(\boldsymbol{\Sigma}^{1/2}\right)_{kj} \mathbb{E}\left[(\partial_j f)(\boldsymbol{x})\right]$$

- Therefore,

$$\mathbb{E}\left[X_i f(\boldsymbol{x})\right] = \sum_{j,k=1}^{n} \left(\boldsymbol{\Sigma}^{1/2}\right)_{ik} \left(\boldsymbol{\Sigma}^{1/2}\right)_{kj} \mathbb{E}\left[(\partial_j f)(\boldsymbol{x})\right] = \sum_{j=1}^{n} \boldsymbol{\Sigma}_{ij} \mathbb{E}\left[(\partial_j f)(\boldsymbol{x})\right]$$

Gaussian Interpolation

Lemma 26 (Gaussian Interpolation). *Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^n$ be independent, centered, jointly Gaussian vectors with covariances $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{y}}$. Define*

$$\boldsymbol{z}(t) = \sqrt{t}\,\boldsymbol{x} + \sqrt{1-t}\,\boldsymbol{y} \quad \text{for } t \in [0,1]$$

For "any" function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[f(\boldsymbol{z}(t))\right] = \frac{1}{2}\sum_{i,j=1}^{n}\left((\boldsymbol{\Sigma}_{\boldsymbol{x}})_{ij} - (\boldsymbol{\Sigma}_{\boldsymbol{y}})_{ij}\right)\mathbb{E}\left[(\partial_{ij}f)(\boldsymbol{z}(t))\right]$$

- Calculate:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[f(\boldsymbol{z}(t))\right] = \sum_{i=1}^{n}\mathbb{E}\left[(\partial_i f)(\boldsymbol{z}(t))(\partial_i \boldsymbol{z})(t)\right] = \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[(\partial_i f)(\boldsymbol{z}(t))\left(\frac{1}{\sqrt{t}}X_i - \frac{1}{\sqrt{1-t}}Y_i\right)\right]$$

- Apply Gaussian IBP to each term; e.g.,

$$\frac{1}{\sqrt{t}}\mathbb{E}\left[(\partial_i f)(\boldsymbol{z}(t))X_i\right] = \sum_{j=1}^{n}(\boldsymbol{\Sigma}_{\boldsymbol{x}})_{ij}\mathbb{E}\left[(\partial_{ij}f)(\boldsymbol{z}(t))\right]$$

Proof of Kahane's Theorem

- Observe that $f(\boldsymbol{z}(0)) = f(\boldsymbol{y})$ and $f(\boldsymbol{z}(1)) = f(\boldsymbol{x})$
- By Gaussian interpolation,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[f(\boldsymbol{z}(t))\right] = \frac{1}{2} \sum_{i,j=1}^{n} \left((\boldsymbol{\Sigma}_{\boldsymbol{x}})_{ij} - (\boldsymbol{\Sigma}_{\boldsymbol{y}})_{ij}\right) \mathbb{E}\left[(\partial_{ij} f)(\boldsymbol{z}(t))\right]$$

- By hypothesis,

$$\begin{aligned} \text{for } (i,j) \in A, \quad & (\boldsymbol{\Sigma}_{\boldsymbol{x}})_{ij} \le (\boldsymbol{\Sigma}_{\boldsymbol{y}})_{ij} \quad \text{and} \quad \partial_{ij} f \ge 0 \\ \text{for } (i,j) \in B, \quad & (\boldsymbol{\Sigma}_{\boldsymbol{x}})_{ij} \ge (\boldsymbol{\Sigma}_{\boldsymbol{y}})_{ij} \quad \text{and} \quad \partial_{ij} f \le 0 \\ \text{otherwise}, \quad & (\boldsymbol{\Sigma}_{\boldsymbol{x}})_{ij} = (\boldsymbol{\Sigma}_{\boldsymbol{y}})_{ij} \end{aligned}$$

- Thus,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[f(\boldsymbol{z}(t))\right] \le 0$$

- Conclude: $\mathbb{E}\, f(\boldsymbol{x}) \le \mathbb{E}\, f(\boldsymbol{y})$

Proof of Slepian's Lemma

- Let $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{y} \in \mathbb{R}^N$ be centered, jointly Gaussian vectors with

$$\begin{cases} \mathbb{E}\, X_i X_j \le \mathbb{E}\, Y_i Y_j & \text{for all } i \neq j \\ \mathbb{E}\, X_i^2 = \mathbb{E}\, Y_i^2 & \text{for all } i \end{cases}$$

- Set $A = \{(i,j) : i \neq j\}$ and $B = \emptyset$. Define the function

$$f(\boldsymbol{w}) = \prod_{i=1}^{N} \mathbb{1}\{w_i \le \lambda_i\}$$

- For $(i,j) \in A$, compute second derivative:

$$(\partial_i f)(\boldsymbol{w}) = -\mathbb{1}\{w_i = \lambda_i\} \prod_{j \neq i} \mathbb{1}\{w_j \le \lambda_j\}$$

$$(\partial_{ij} f)(\boldsymbol{w}) = \mathbb{1}\{w_i = \lambda_i, w_j = \lambda_j\} \prod_{k \notin \{i,j\}} \mathbb{1}\{w_k \le \lambda_k\} \ge 0$$

- Apply Kahane's Theorem:

$$\mathbb{P}\left(\bigcap_{i=1}^N \{X_i \le \lambda_i\}\right) = \mathbb{E}\, f(\boldsymbol{x}) \le \mathbb{E}\, f(\boldsymbol{y}) = \mathbb{P}\left(\bigcap_{i=1}^N \{Y_i \le \lambda_i\}\right)$$

Proof of Gordon's Theorem I

- Let $\boldsymbol{X} \in \mathbb{R}^{M \times N}$ and $\boldsymbol{Y} \in \mathbb{R}^{M \times N}$ be centered, jointly Gaussian matrices with

$$\begin{cases} \mathbb{E}\, X_{ij} X_{k\ell} \le \mathbb{E}\, Y_{ij} Y_{k\ell} & \text{for all } i \neq k \text{ and all } j, \ell \\ \mathbb{E}\, X_{ij} X_{i\ell} \ge \mathbb{E}\, Y_{ij} Y_{i\ell} & \text{for all } i \text{ and } j \neq \ell \\ \mathbb{E}\, X_{ij}^2 = \mathbb{E}\, Y_{ij}^2 & \text{for all } i, j \end{cases}$$

- Set $A = \{((i,j),(k,\ell)) : i \neq k\}$ and $B = \{((i,j),(k,\ell)) : i = k, j \neq \ell\}$. Define

$$f(\boldsymbol{W}) = \prod_{i=1}^{N} \left[1 - \prod_{j=1}^{M} \mathbb{1}\{w_{ij} \le \lambda_{ij}\} \right]$$

- Compute first derivative:

$$(\partial_{(i,j)} f)(\boldsymbol{W}) = \mathbb{1}\{w_{ij} = \lambda_{ij}\} \prod_{j' \neq j} \mathbb{1}\{w_{ij'} \le \lambda_{ij'}\} \prod_{i' \neq i} \left[1 - \prod_{j'} \mathbb{1}\{w_{i'j'} \le \lambda_{i'j'}\} \right]$$

Proof of Gordon's Theorem II

- For $((i,j),(k,\ell)) \in B$, compute second derivative:

$$(\partial_{(i,j),(k,\ell)} f)(\boldsymbol{W}) = -\mathbb{1}\{w_{ij} = \lambda_{ij}, w_{i\ell} = \lambda_{i\ell}\} \times \prod_{j' \notin \{j,\ell\}} \mathbb{1}\{w_{ij'} \le \lambda_{ij'}\} \prod_{i' \neq i} \left[1 - \prod_{j'} \mathbb{1}\{w_{i'j'} \le \lambda_{i'j'}\} \right] \le 0$$

- For $((i,j),(k,\ell)) \in A$, compute second derivative:

$$(\partial_{(i,j),(k,\ell)} f)(\boldsymbol{W}) = \mathbb{1}\{w_{ij} = \lambda_{ij}, w_{k\ell} = \lambda_{k\ell}\} \times \prod_{j' \neq j} \mathbb{1}\{w_{ij'} \le \lambda_{ij'}\} \prod_{j' \neq \ell} \mathbb{1}\{w_{kj'} \le \lambda_{kj'}\} \prod_{i' \notin \{i,k\}} \left[1 - \prod_{j'} \mathbb{1}\{w_{i'j'} \le \lambda_{i'j'}\} \right] \ge 0$$

- Apply Kahane's Theorem:

$$\mathbb{P}\left(\bigcap_{i=1}^{N} \left(\bigcap_{j=1}^{M} \{X_{ij} \le \lambda_{ij}\} \right)^{\complement} \right) = \mathbb{E} f(\boldsymbol{x}) \le \mathbb{E} f(\boldsymbol{y}) = \mathbb{P}\left(\bigcap_{i=1}^{N} \left(\bigcap_{j=1}^{M} \{Y_{ij} \le \lambda_{ij}\} \right)^{\complement} \right)$$

Gaussian Concentration

Theorem 27. ***Assume***

- $f : \mathbb{R}^d \to \mathbb{R}$ *is a Lipschitz function:* $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L \cdot \|\boldsymbol{x} - \boldsymbol{y}\|$ *for all* $\boldsymbol{x}, \boldsymbol{y}$
- $Z = f(\boldsymbol{g})$ *where* $\boldsymbol{g}$ *is standard normal*

Then

$$\operatorname{Var}[Z] = \mathbb{E}(Z - \mathbb{E}\, Z)^2 \quad \leq \quad L^2 \qquad \textit{(Poincaré)}$$

$$\mathbb{P}\{|Z - \mathbb{E}\, Z| \geq t\} \quad \leq \quad \mathrm{e}^{-t^2/(2L^2)} \quad \textit{for all } t \geq 0 \qquad \textit{(concentration)}$$

Sources: Ledoux & Talagrand (Sec. 1.1); Bogachev (Sec. 1.7); Boucheron et al. (Sec. 3.7, 5.4).

Pisier's Approach to Gaussian Concentration

Theorem 28 (Pisier 1986). ***Assume***

- $f : \mathbb{R}^d \to \mathbb{R}$ *is an* L*-Lipschitz function*
- $\boldsymbol{x} \in \mathbb{R}^d$ *and* $\gamma \in \mathbb{R}$ *are standard normal*
- $\Phi : \mathbb{R} \to \mathbb{R}$ *is convex*
- $t \mapsto \Phi(t) + \Phi(-t)$ *is increasing on* $\mathbb{R}_+$

Then

$$\mathbb{E}\,\Phi\big(f(\boldsymbol{x}) - \mathbb{E}\,f(\boldsymbol{x})\big) \quad \leq \quad \mathbb{E}\,\Phi\left(\tfrac{1}{2}\pi L\gamma\right)$$

In particular,

$$\Phi(t) = t^2: \qquad \operatorname{Var}[f(\boldsymbol{x})] \leq \tfrac{1}{4}\pi^2 L^2$$

$$\Phi(t) = \mathrm{e}^{\theta t}: \qquad \log \mathbb{E}\,\mathrm{e}^{\theta(f(\boldsymbol{x}) - \mathbb{E}\,f(\boldsymbol{x}))} \leq \tfrac{1}{8}\pi^2 L^2 \theta^2 \quad \text{for } \theta \in \mathbb{R}$$

Sources: Pisier 1986; Ledoux & Talagrand (Eqn. (1.5)).

Proof of Pisier's Theorem

- Let $\boldsymbol{x}, \boldsymbol{y}$ be independent standard normal variables
- Define $\boldsymbol{z}(\alpha) = \cos(\alpha)\boldsymbol{x} + \sin(\alpha)\boldsymbol{y}$ and $\boldsymbol{z}'(\alpha) = -\sin(\alpha)\boldsymbol{x} + \cos(\alpha)\boldsymbol{y}$
- For fixed α, the vectors $\boldsymbol{z}(\alpha)$ and $\boldsymbol{z}'(\alpha)$ are independent standard normal
- Calculate:

$$\begin{aligned}
\mathbb{E}\,\Phi\big(f(\boldsymbol{x}) - \mathbb{E}\,f(\boldsymbol{y})\big) &\le \mathbb{E}\,\Phi\big(f(\boldsymbol{x}) - f(\boldsymbol{y})\big) \\
&= \mathbb{E}\,\Phi\left(\int_0^{\pi/2} \big\langle \nabla f(\boldsymbol{z}(\alpha)),\, \boldsymbol{z}'(\alpha)\big\rangle \,\mathrm{d}\alpha\right) \\
&\le \frac{2}{\pi}\int_0^{\pi/2} \mathbb{E}\,\Phi\left(\frac{\pi}{2}\big\langle \nabla f(\boldsymbol{z}(\alpha)),\, \boldsymbol{z}'(\alpha)\big\rangle\right) \mathrm{d}\alpha \\
&= \mathbb{E}\,\Phi\left(\frac{\pi}{2}\big\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{y}\big\rangle\right) \\
&= \mathbb{E}\,\Phi\left(\frac{\pi}{2}\,\|\nabla f(\boldsymbol{x})\| \cdot \gamma\right) \\
&\le \mathbb{E}\,\Phi\left(\frac{\pi}{2} L \cdot \gamma\right)
\end{aligned}$$

Signal Recovery

Convex Signal Recovery

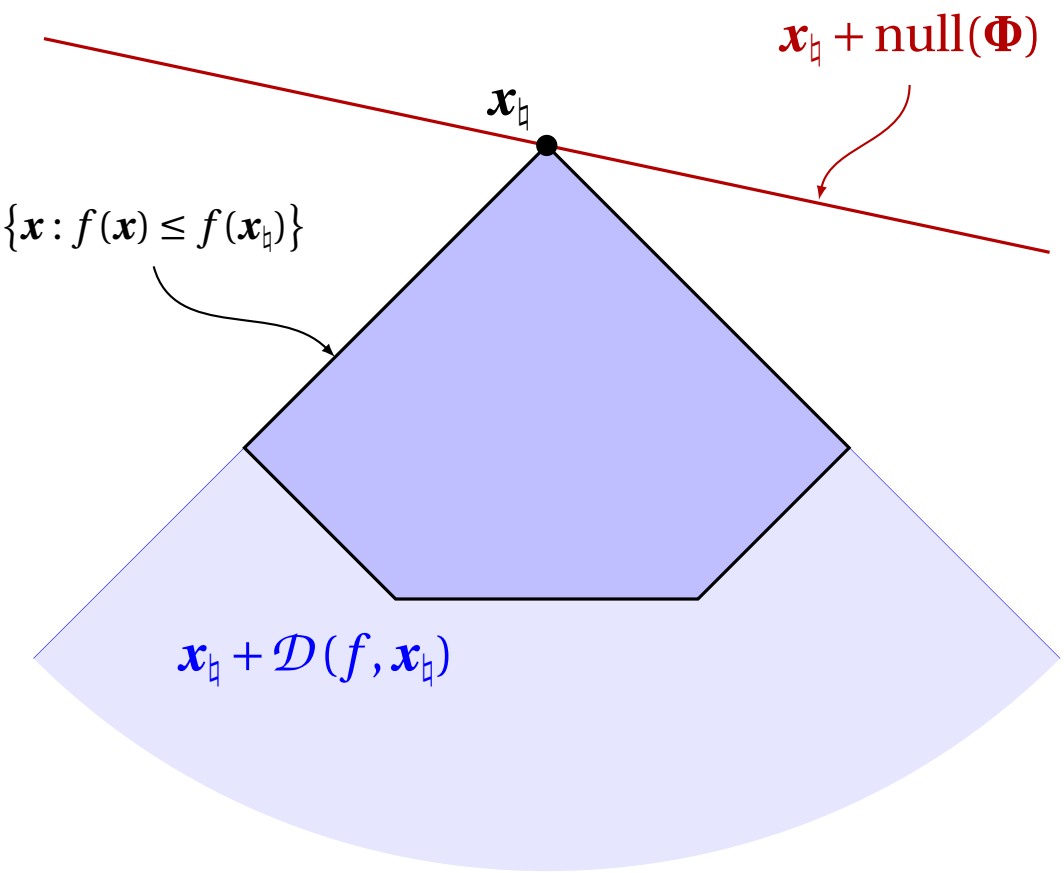
- Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex structural penalty (e.g., an atomic gauge)
- Let $\boldsymbol{x}_\natural \in \mathbb{R}^d$ be "structured" but unknown
- Let $\boldsymbol{\Phi} \in \mathbb{R}^{m \times d}$ be a known measurement matrix
- Observe $\boldsymbol{z} = \boldsymbol{\Phi} \boldsymbol{x}_\natural \in \mathbb{R}^m$
- Find estimate $\widehat{\boldsymbol{x}}$ by solving convex program

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

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

Sources: Chen et al. 1997, 2001; Chandrasekaran et al. 2012, McCoy & Tropp 2013; Oymak et al. 2013; Amelunxen et al. 2014; Thrampoulidis et al. 2014–2016; Tropp 2015.

Geometry of Convex Signal Recovery



Analysis of Convex Signal Recovery: Success

Proposition 29 (Geometric Formulation)**.** *Convex signal recovery succeeds* ($\widehat{\boldsymbol{x}} = \boldsymbol{x}_\natural$) *if and only if*

$$\mathcal{D}(f, \boldsymbol{x}_\natural) \cap \operatorname{null}(\boldsymbol{\Phi}) = \{\boldsymbol{0}\}$$

Proposition 30 (Analytic Condition for Success)**.** *Convex signal recovery succeeds if*

$$\sigma_{\min}(\boldsymbol{\Phi}; K) = \inf_{\substack{\|\boldsymbol{x}\|=1 \\ \boldsymbol{x} \in K}} \|\boldsymbol{\Phi}\boldsymbol{x}\| > 0 \qquad \textit{("Minimum conic singular value")}$$

where $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$

Sources: Rudelson & Vershynin 2006; Stojnic 2009, 2013; Oymak et al. 2010; Chandrasekaran et al. 2012; McCoy & Tropp 2013; Oymak et al. 2013; Amelunxen et al. 2014; Thrampoulidis et al. 2014–2016; Tropp 2015.

Gaussian Measurements

Proposition 31 (Gaussian Measurements). ***Assume*** $\mathbf{\Gamma} \in \mathbb{R}^{m \times d}$ *is a standard normal matrix.* ***Then*** $\operatorname{null}(\mathbf{\Gamma})$ *is a uniformly distributed subspace of* $\mathbb{R}^d$ *with codimension* $m \wedge d$*, almost surely.*

Sources: Donoho 2006; Candès & Tao 2006; Rudelson & Vershynin 2006; Stojnic 2009, 2013; Donoho & Tanner 2009; Recht et al. 2010; Oymak et al. 2010; Chandrasekaran et al. 2012; McCoy & Tropp 2013; Oymak et al. 2013; Amelunxen et al. 2014; Goldstein et al. 2014; Thrampoulidis et al. 2014–2016; Tropp 2015.

Minimum Conic Singular Value of a Gaussian Matrix

Proposition 32. ***Assume***

- K *is a convex cone in* $\mathbb{R}^d$
- $\mathbf{\Gamma} \in \mathbb{R}^{m \times d}$ *is standard normal*

Then

$$\mathbb{E}\,\sigma_{\min}(\mathbf{\Gamma}; K) \quad \geq \quad \sqrt{m-1} - \sqrt{\delta(K)}$$

- Write $\sigma_{\min}(\mathbf{\Gamma}; K) = \inf_{\|\boldsymbol{u}\|=1, \boldsymbol{u} \in K} \max_{\|\boldsymbol{v}\|=1} \langle \mathbf{\Gamma} \boldsymbol{u},\ \boldsymbol{v} \rangle$
- Gaussian Minimax Theorem:

$$\begin{aligned} \mathbb{E}\,\sigma_{\min}(\mathbf{\Gamma}; K) \quad &\geq \quad \mathbb{E} \inf_{\|\boldsymbol{u}\|=1, \boldsymbol{u} \in K} \max_{\|\boldsymbol{v}\|=1} \left[\langle \boldsymbol{g},\ \boldsymbol{u} \rangle + \langle \boldsymbol{h},\ \boldsymbol{v} \rangle \right] \\ &\geq \quad \sqrt{m-1} - \mathbb{E} \sup_{\|\boldsymbol{u}\|=1, \boldsymbol{u} \in K} \langle \boldsymbol{g},\ \boldsymbol{u} \rangle \quad \geq \quad \sqrt{m-1} - \sqrt{\delta(K)} \end{aligned}$$

Sources: Gordon 1985, 1988; Rudelson & Vershynin 2006; Stojnic 2009, 2013; Oymak et al. 2010; Chandrasekaran et al. 2012; Oymak et al. 2013; Amelunxen et al. 2014, Thrampoulidis et al. 2014–2016; Tropp 2015.

Concentration of Minimum Conic Singular Value

Proposition 33. ***Assume*** *K is a convex cone, and let $\mathbf{\Gamma}$ be standard normal.* ***Then***

$$\mathbb{P}\{\sigma_{\min}(\mathbf{\Gamma}; K) < \mathbb{E}\,\sigma_{\min}(\mathbf{\Gamma}; K) - t\} \quad \leq \quad \mathrm{e}^{-t^2/2}$$

- Bound the Lipschitz constant of $\sigma_{\min}(\cdot; K)$:

$$\inf_{\|\boldsymbol{u}\|=1, \boldsymbol{u}\in K} \|\mathbf{\Gamma}\boldsymbol{u}\| - \inf_{\|\boldsymbol{u}\|=1, \boldsymbol{u}\in K} \|\mathbf{\Gamma}'\boldsymbol{u}\| \quad \leq \quad \|\mathbf{\Gamma}\boldsymbol{u}'\| - \|\mathbf{\Gamma}'\boldsymbol{u}'\| \quad \leq \quad \|(\mathbf{\Gamma} - \mathbf{\Gamma}')\boldsymbol{u}'\| \quad \leq \quad \|\mathbf{\Gamma} - \mathbf{\Gamma}'\|_{\mathrm{F}}$$

  where $\boldsymbol{u}'$ is a near-minimizer of the second term

- Apply Gaussian concentration with $L = 1$

Sources: Rudelson & Vershynin 2006; Stojnic 2009, 2013; Oymak et al. 2010; Chandrasekaran et al. 2012; Oymak et al. 2013; Amelunxen et al. 2014, Thrampoulidis et al. 2013–2016; Tropp 2015.

Gaussian Measurements: Success

Theorem 34 (Chandrasekaran et al. 2012). ***Assume***

- $f : \mathbb{R}^d \to \mathbb{R}$ *is a convex function and* $\boldsymbol{x}_\natural \in \operatorname{dom}(f)$
- *Observe* $\boldsymbol{z} = \boldsymbol{\Gamma} \boldsymbol{x}_\natural$ *where* $\boldsymbol{\Gamma} \in \mathbb{R}^{m \times d}$ *is standard normal*
- *The vector* $\widehat{\boldsymbol{x}}$ *solves*

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

Then

$$m \geq \delta(K) + \mathrm{C}\sqrt{d} \quad \text{implies} \quad \widehat{\boldsymbol{x}} = \boldsymbol{x}_\natural \quad \text{with high probability}$$

where $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$

Sources: Rudelson & Vershynin 2006; Stojnic 2009, 2013; Chandrasekaran et al. 2012; Amelunxen et al. 2014; Thrampoulidis et al. 2014–2016; Tropp 2015.

Gaussian Measurements: Success Proof

- Let $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$
- Combine last two results:

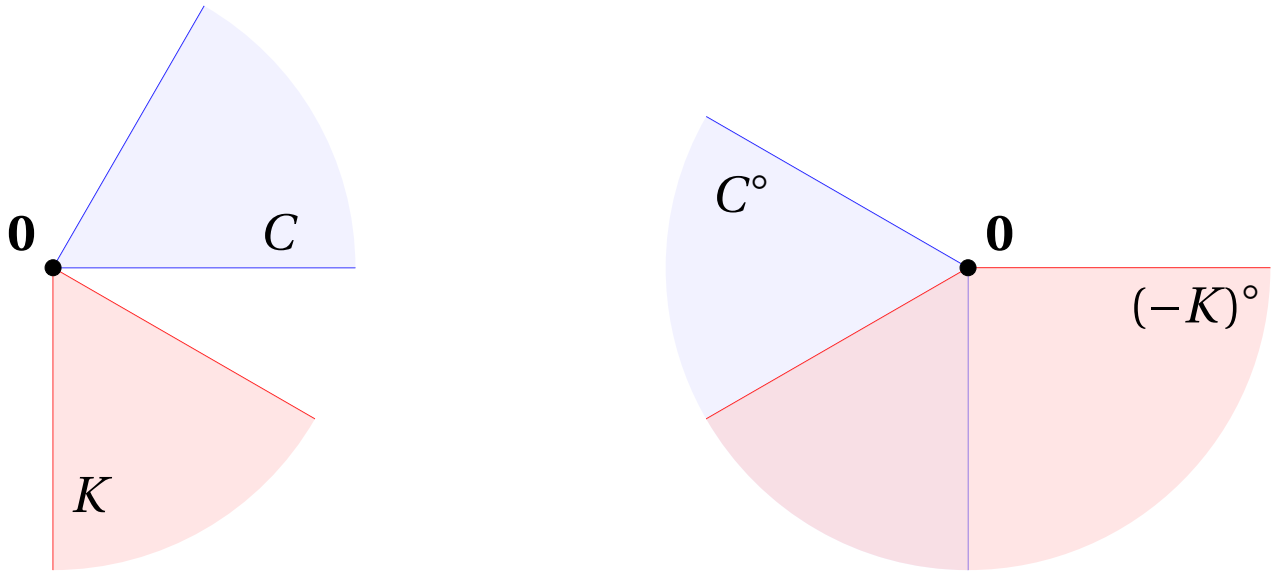
$$\begin{aligned} \mathrm{e}^{-t^2/2} &\geq \mathbb{P}\left\{\sigma_{\min}(\boldsymbol{\Gamma}; K) < \mathbb{E}\,\sigma_{\min}(\boldsymbol{\Gamma}; K) - t\right\} \\ &\geq \mathbb{P}\left\{\sigma_{\min}(\boldsymbol{\Gamma}; K) < \sqrt{m-1} - \sqrt{\delta(K)} - t\right\} \end{aligned}$$

- Set $t = 3$ to achieve probability less than 2%
- Success $(\sigma_{\min}(\boldsymbol{\Gamma}; K) > 0)$: $\sqrt{m-1} - \sqrt{\delta(K)} - 3 > 0$
- Equivalently, $m > \delta(K) + 6\sqrt{\delta(K)} + 10$
- Use upper bound $\delta(K) \leq d$

Separation of Convex Cones

Theorem 35 (Klee 1955). ***Assume*** *C and K are convex cones in $\mathbb{R}^d$, one of which is not a subspace.* ***Then***

$$C \cap K = \{\mathbf{0}\} \quad \textit{implies} \quad C^\circ \cap (-K)^\circ \neq \{\mathbf{0}\}$$



Source: Klee 1955.

Analysis of Convex Signal Recovery: Failure

Proposition 36 (Geometric Formulation)**.** *Convex signal recovery fails (i.e., $\boldsymbol{x}_\natural$ is not the unique solution) if and only if*

$$\mathcal{D}(f, \boldsymbol{x}_\natural) \cap \operatorname{null}(\boldsymbol{\Phi}) \neq \{\boldsymbol{0}\}$$

If $\mathcal{D}(f, \boldsymbol{x}_\natural)$ is not a subspace, a sufficient condition for failure is

$$\big(\mathcal{D}(f, \boldsymbol{x}_\natural)\big)^\circ \cap \operatorname{null}(\boldsymbol{\Phi})^\circ = \{\boldsymbol{0}\}$$

Proposition 37 (Analytic Condition for Failure)**.** *Let $\boldsymbol{\Psi}$ be a matrix with $\operatorname{null}(\boldsymbol{\Psi}) = \operatorname{null}(\boldsymbol{\Phi})^\circ$. Convex signal recovery fails if*

$$\sigma_{\min}(\boldsymbol{\Psi}; K^\circ) = \min_{\substack{\|\boldsymbol{x}\|=1 \\ \boldsymbol{x} \in K^\circ}} \|\boldsymbol{\Psi}\boldsymbol{x}\| > 0$$

where $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$

Sources: Stojnic 2013; McCoy & Tropp 2013; Amelunxen et al. 2014; Thrampoulidis et al. 2014–2016; Tropp 2015.

Gaussian Measurements: Failure

Theorem 38 (Amelunxen et al. 2014). ***Assume***

- $f : \mathbb{R}^d \to \mathbb{R}$ *is a convex function and* $\boldsymbol{x}_\natural \in \operatorname{dom}(f)$
- *Observe* $\boldsymbol{z} = \boldsymbol{\Gamma} \boldsymbol{x}_\natural$ *where* $\boldsymbol{\Gamma} \in \mathbb{R}^{m \times d}$ *is standard normal*
- *The vector* $\widehat{\boldsymbol{x}}$ *solves*

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

Then

$$m \leq \delta(K) - \mathrm{C}\sqrt{d} \quad \text{implies} \quad \boldsymbol{x}_\natural \text{ is not the unique solution} \quad \text{with high prob.}$$

where $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$

Sources: Stojnic 2013; Oymak et al. 2013; Amelunxen et al. 2014; Foygel & Mackey 2014; Thrampoulidis et al. 2014–2016; Tropp 2015.

Gaussian Measurements: Failure Proof

- Can assume $m < d$ or else failure with probability zero
- If $\mathcal{D}(f, \boldsymbol{x}_\natural)$ is a subspace, just count dimensions
- Let $\boldsymbol{\Gamma}^\circ \in \mathbb{R}^{(d-m)\times d}$ be standard normal
- $\mathrm{null}(\boldsymbol{\Gamma}^\circ)$ has same distribution as $\mathrm{null}(\boldsymbol{\Gamma})^\circ$ (a unif. rdm subspace, codim m)
- Let $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$. As before,

$$\begin{aligned} \mathrm{e}^{-t^2/2} &\geq \mathbb{P}\left\{\sigma_{\min}(\boldsymbol{\Gamma}^\circ; K^\circ) < \mathbb{E}\,\sigma_{\min}(\boldsymbol{\Gamma}^\circ; K^\circ) - t\right\} \\ &\geq \mathbb{P}\left\{\sigma_{\min}(\boldsymbol{\Gamma}^\circ; K^\circ) < \sqrt{d-m-1} - \sqrt{\delta(K^\circ)} - t\right\} \end{aligned}$$

- Failure $(\sigma_{\min}(\boldsymbol{\Gamma}^\circ; K^\circ) > 0)$: $\quad \sqrt{d-m-1} - \sqrt{\delta(K^\circ)} - 3 > 0$
- Equivalently, $\quad d - m > \delta(K^\circ) + 6\sqrt{\delta(K^\circ)} + 10$
- Use facts $\quad \delta(K^\circ) = d - \delta(K) \leq d$

Gaussian Measurements: Summary

Theorem 39 (Chandrasekaran et al. 2012; Amelunxen et al. 2014). ***Assume***

- $f : \mathbb{R}^d \to \mathbb{R}$ *is a convex function and* $\boldsymbol{x}_\natural \in \operatorname{dom}(f)$
- *Observe* $\boldsymbol{z} = \boldsymbol{\Gamma} \boldsymbol{x}_\natural$ *where* $\boldsymbol{\Gamma} \in \mathbb{R}^{m \times d}$ *is standard normal*
- *The vector* $\widehat{\boldsymbol{x}}$ *solves*

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

Then

$$m \le \delta(K) - \mathrm{C}\sqrt{d} \quad \text{implies} \quad \text{success} \quad \text{with high probability}$$

$$m \ge \delta(K) + \mathrm{C}\sqrt{d} \quad \text{implies} \quad \text{failure} \quad \text{with high probability}$$

where $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$

Sources: Rudelson & Vershynin 2006; Stojnic 2009, 2013; Oymak et al. 2010, 2013; Chandrasekaran et al. 2012; Amelunxen et al. 2014; Foygel & Mackey 2014; Thrampoulidis et al. 2014–2016; Tropp 2015.

Gaussian Measurements: Improved

Theorem 40 (Amelunxen et al. 2014). ***Assume***

- $f : \mathbb{R}^d \to \mathbb{R}$ *is a convex function and* $\boldsymbol{x}_\natural \in \operatorname{dom}(f)$
- *Observe* $\boldsymbol{z} = \boldsymbol{\Gamma}\boldsymbol{x}_\natural$ *where* $\boldsymbol{\Gamma} \in \mathbb{R}^{m \times d}$ *is standard normal*
- *The vector* $\widehat{\boldsymbol{x}}$ *solves*

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

Then

$$m \le \delta(K) - \mathrm{C}\sqrt{\delta(K) \wedge \delta(K^\circ)} \quad \text{implies} \quad \text{success} \quad \text{with high prob.}$$

$$m \ge \delta(K) + \mathrm{C}\sqrt{\delta(K) \wedge \delta(K^\circ)} \quad \text{implies} \quad \text{failure} \quad \text{with high prob.}$$

where $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$

Sources: Amelunxen et al. 2014; Thrampoulidis et al. 2014–2016; Goldstein et al. 2017.

Example: ℓ_1 Minimization

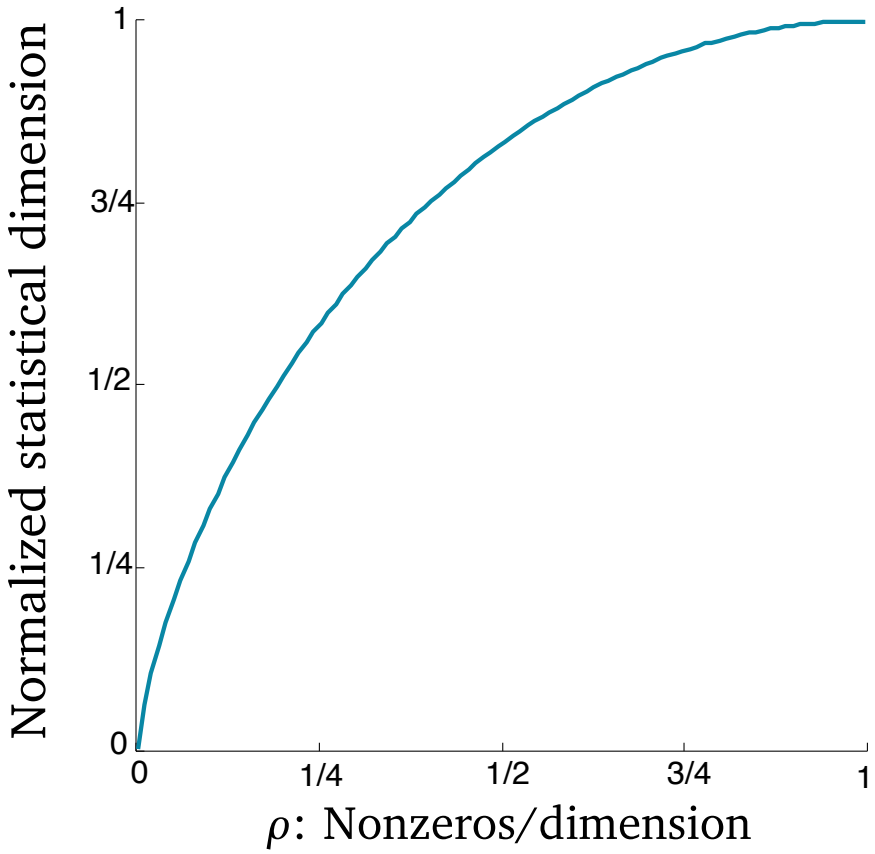
- Suppose $\boldsymbol{x}_\natural \in \mathbb{R}^d$ has s nonzero entries
- Let $\boldsymbol{\Gamma} \in \mathbb{R}^{m \times d}$ be a standard normal matrix
- Observe $\boldsymbol{z} = \boldsymbol{\Gamma} \boldsymbol{x}_\natural$
- Find estimate $\widehat{\boldsymbol{x}}$ by solving convex program

$$\text{minimize} \quad \|\boldsymbol{x}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Gamma} \boldsymbol{x} = \boldsymbol{z}$$

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

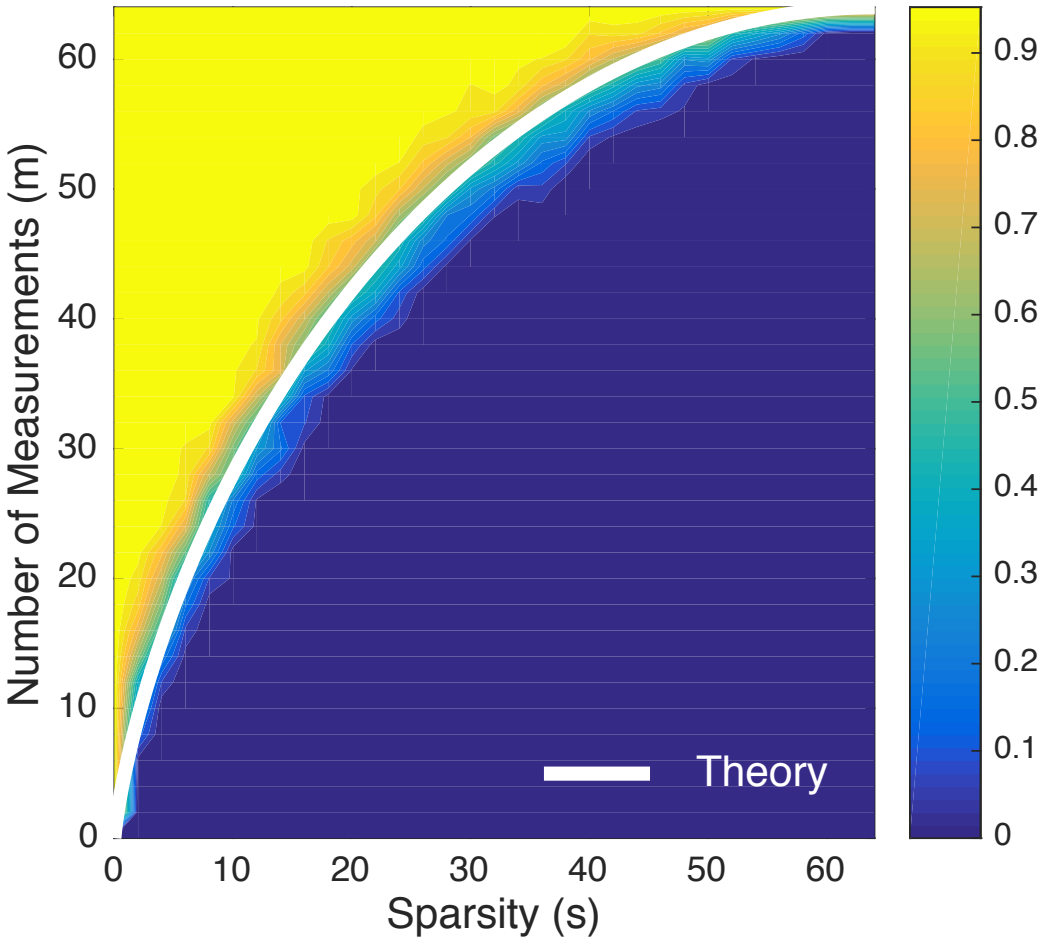
Sources: Donoho 2004, 2006; Candès & Tao 2006; Rudelson & Vershynin 2006; Donoho & Tanner 2009; Stojnic 2009, 2013; Chandrasekaran et al. 2012; Amelunxen et al. 2014; Foygel & Mackey 2014; Goldstein et al. 2017.

ℓ_1 Statistical Dimension Curve

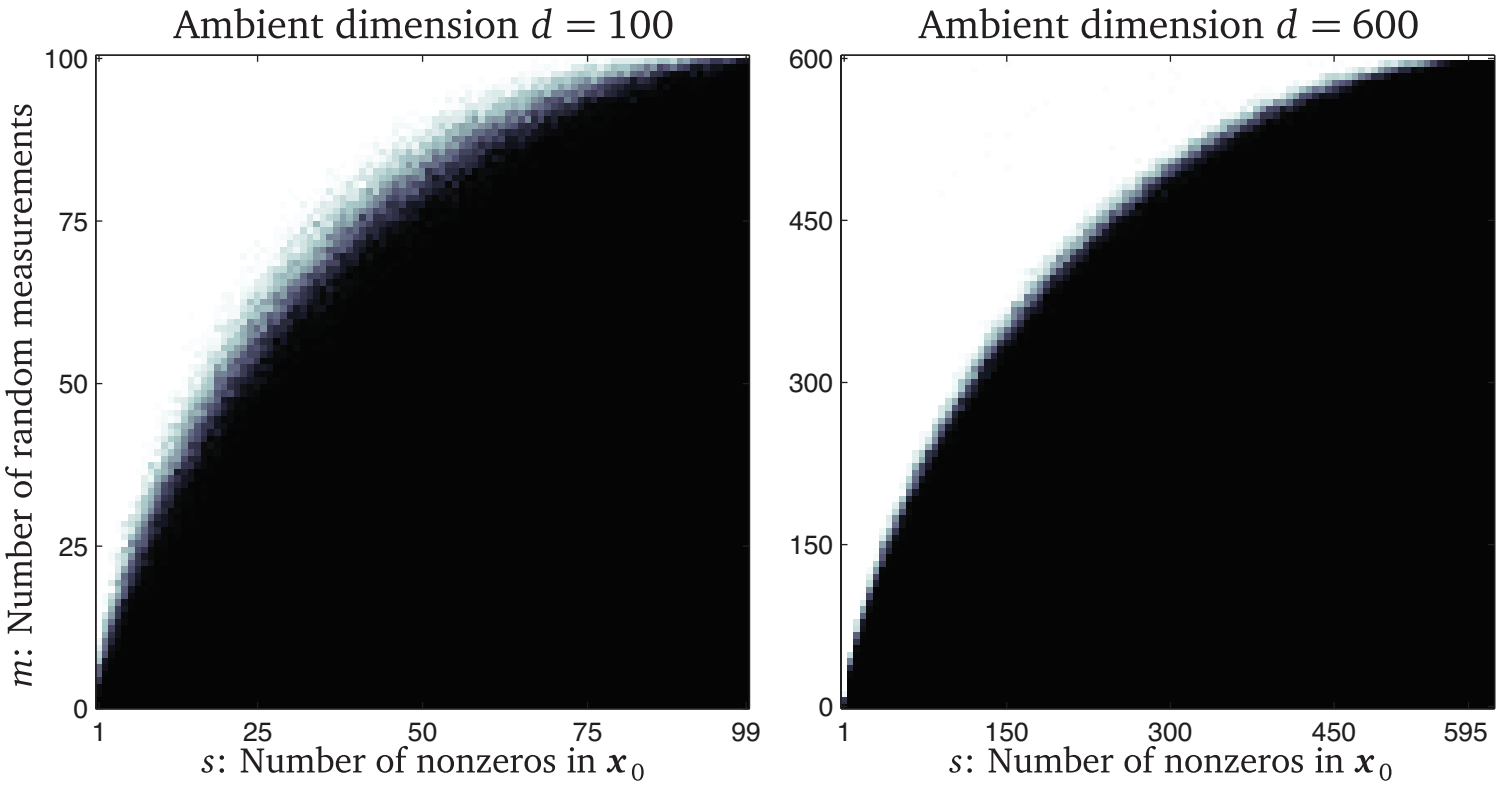


Sources: Affentranger & Schneider 1992; Betke & Henk 1993; Böröczky & Henk 1999; Donoho 2006; Donoho & Tanner 2009; Stojnic 2009, 2013; Chandrasekaran et al. 2012; McCoy & Tropp 2013; Amelunxen et al. 2014; Foygel & Mackey 2014.

Example: Performance of ℓ_1 Minimization



Example: Emergence of ℓ_1 Phase Transition



Example: S_1 Minimization

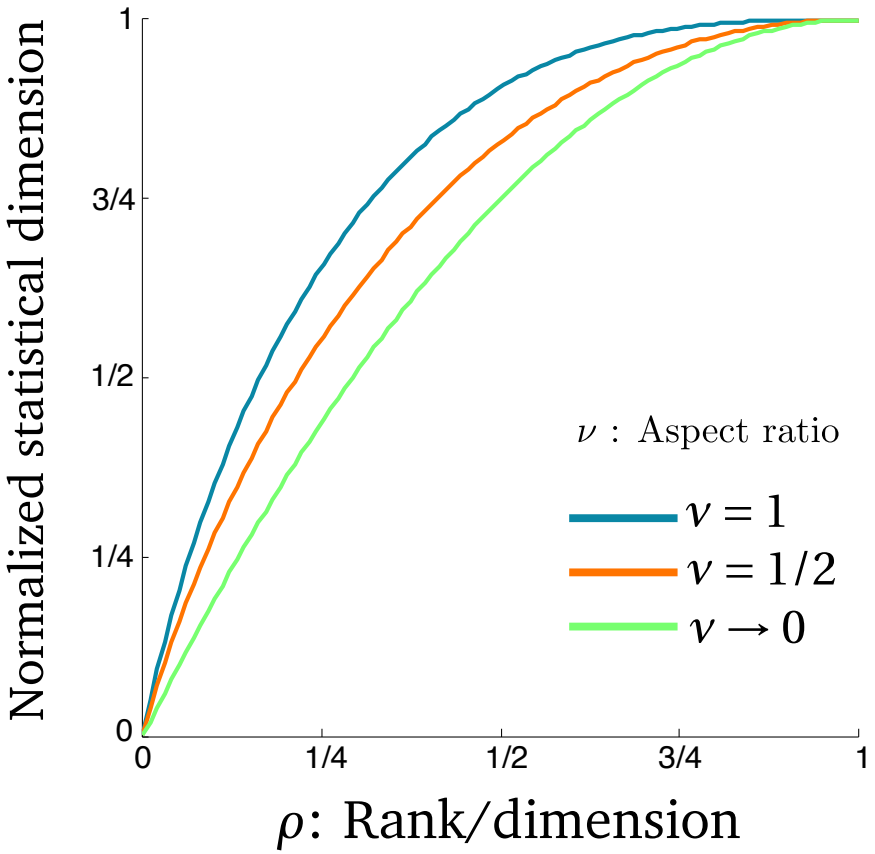
- Suppose $\boldsymbol{X}_\natural \in \mathbb{R}^{d_1 \times d_2}$ has rank r
- Let $\boldsymbol{\Gamma} \in \mathbb{R}^{m \times (d_1 \times d_2)}$ be a standard normal matrix
- Observe $\boldsymbol{z} = \boldsymbol{\Gamma}(\operatorname{vec} \boldsymbol{X}_\natural)$
- Find estimate $\widehat{\boldsymbol{X}}$ by solving convex program

$$\text{minimize} \quad \|\boldsymbol{X}\|_{S_1} \quad \text{subject to} \quad \boldsymbol{\Gamma}(\operatorname{vec} \boldsymbol{X}) = \boldsymbol{z}$$

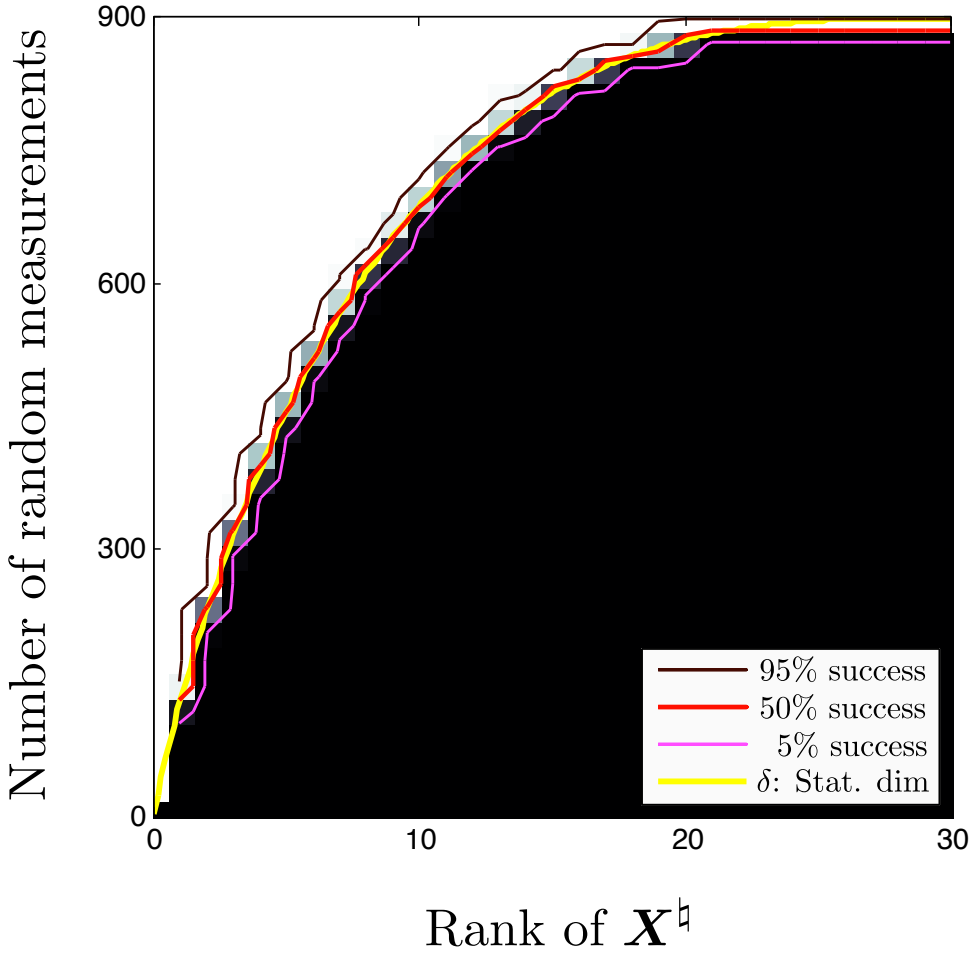
- **Hope:** $\widehat{\boldsymbol{X}} = \boldsymbol{X}_\natural$

Sources: Fazel 2002; Recht et al. 2010; Oymak et al. 2010, 2013; Chandrasekaran et al. 2012; Amelunxen et al. 2014; Thrampoulidis et al. 2014–2016; Tropp 2015; Goldstein et al. 2017.

S_1 Statistical Dimension Curve



Example: Performance of S_1 Minimization



Desserts + Digestifs

Complements

- Universality
- Signal recovery with noise
- Sharp analysis for Gaussian signal recovery with Gaussian noise
- Non-Gaussian measurements
- Demixing
- ...

Universality I

Theorem 41 (Oymak & Tropp 2015). ***Assume***

- $f : \mathbb{R}^d \to \mathbb{R}$ *is a convex function and* $\boldsymbol{x}_\natural \in \operatorname{dom}(f)$
- $\boldsymbol{\Phi} \in \mathbb{R}^{m \times d}$ *has iid standardized, symmetric entries with* $4+$ *moments*
- *Observe* $\boldsymbol{z} = \boldsymbol{\Phi} \boldsymbol{x}_\natural$
- *The vector* $\widehat{\boldsymbol{x}}$ *solves*

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

Then

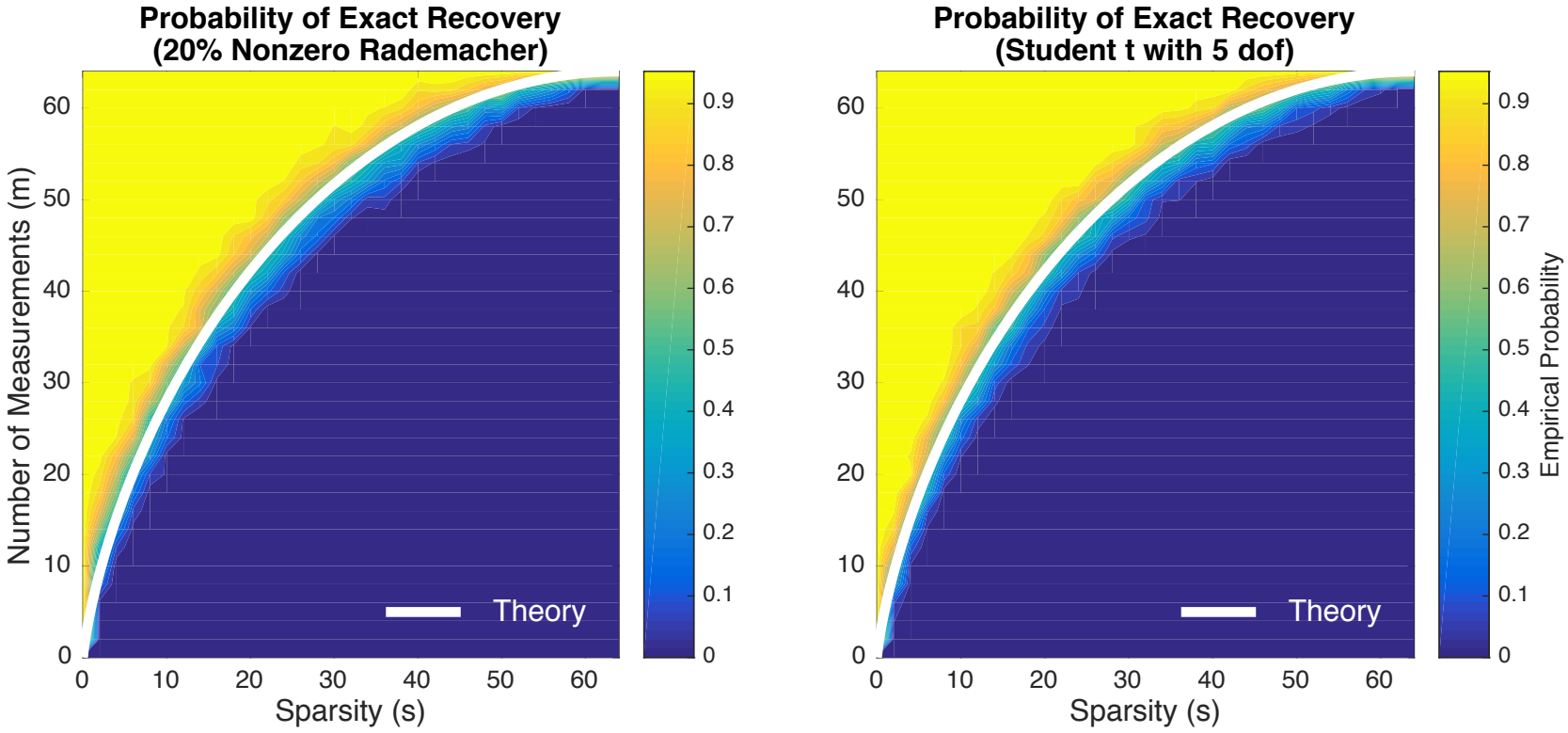
$$m \le \delta(K) - o(d) \quad \text{implies} \quad \text{success} \quad \text{with high probability}$$

$$m \ge \delta(K) + o(d) \quad \text{implies} \quad \text{failure} \quad \text{with high probability}$$

where $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$

Sources: Donoho & Tanner 2009; Bayati et al. 2015; Oymak & Tropp 2015.

Universality II



minimize $\|\boldsymbol{x}\|_{\ell_1}$ subject to $\boldsymbol{\Phi x} = \boldsymbol{\Phi x}_s$

Signal Recovery with Noise I

Theorem 42. ***Assume***

- $f : \mathbb{R}^d \to \mathbb{R}$ *is a convex function and* $\boldsymbol{x}_\natural \in \operatorname{dom}(f)$
- *The matrix* $\boldsymbol{\Phi} \in \mathbb{R}^{m \times d}$
- *Observe* $\boldsymbol{z} = \boldsymbol{\Phi}\boldsymbol{x}_\natural + \boldsymbol{e}$ *where* $\|\boldsymbol{e}\| \leq \eta$
- *The vector* $\widehat{\boldsymbol{x}}$ *solves*

$$\text{minimize} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad \|\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{z}\| \leq \eta$$

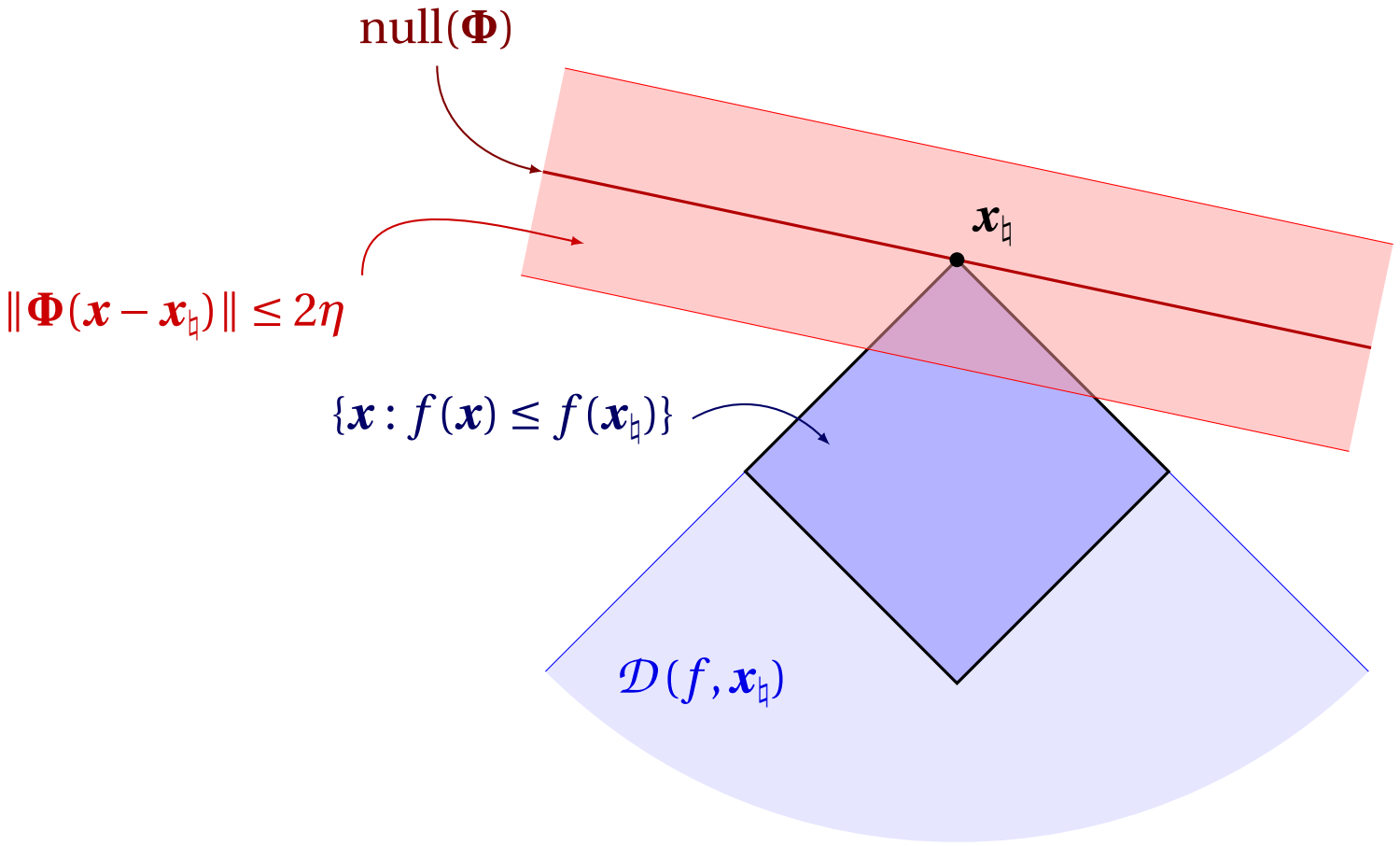
Then

$$\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_\natural\| \leq \frac{2\eta}{\sigma_{\min}(\boldsymbol{\Phi}; K)}$$

where $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$

Sources: Candès et al. 2006; Chandrasekaran et al. 2012; Tropp 2015.

Signal Recovery with Noise II



Signal Recovery with Noise III

Theorem 43 (Oymak et al. 2013). ***Assume***

- $f : \mathbb{R}^d \to \mathbb{R}$ *is a convex function and* $\boldsymbol{x}_\natural \in \operatorname{dom}(f)$
- *Let* $\boldsymbol{\Gamma} \in \mathbb{R}^{m \times d}$ *is standard normal*
- *Observe* $\boldsymbol{z} = \boldsymbol{\Phi}\boldsymbol{x}_\natural + \eta \boldsymbol{g}$ *where* $\boldsymbol{g}$ *is standard normal*
- *The vector* $\widehat{\boldsymbol{x}}$ *solves*

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

Then *(roughly)*

$$\sup_{\eta > 0} \frac{\mathbb{E}\,\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_\natural\|^2}{\eta^2} = \lim_{\eta \downarrow 0} \frac{\mathbb{E}\,\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_\natural\|^2}{\eta^2} = \frac{m}{m - \delta(K)}$$

where $K = \mathcal{D}(f, \boldsymbol{x}_\natural)$

Sources: Oymak et al. 2013; Thrampoulidis et al. 2014–2016.

Non-Gaussian Measurements

Proposition 44 (Mendelson 2013). ***Assume***

- *The rows of* $\boldsymbol{\Phi} \in \mathbb{R}^{m \times d}$ *are iid copies of* $\boldsymbol{\varphi} \in \mathbb{R}^d$
- K *is a convex cone*
- *Define the small ball probability*

$$Q = \inf_{\|\boldsymbol{u}\| \le 1, \boldsymbol{u} \in K} \mathbb{P}\{|\langle \boldsymbol{u}, \boldsymbol{\varphi} \rangle| \ge 1/6\}$$

- *For independent Rademacher variables* $\{\varepsilon_i\}$, *define the mean empirical width*

$$W_m = \mathbb{E} \sup_{\|\boldsymbol{u}\| = 1, \boldsymbol{u} \in K} \left\langle \boldsymbol{u}, \frac{1}{\sqrt{m}} \sum_{i=1}^m \varepsilon_i \boldsymbol{\varphi}_i \right\rangle$$

Then*, for all* $t > 0$,

$$\sigma_{\min}(\boldsymbol{\Phi}; K) \ge \tfrac{1}{3}\sqrt{m}\, Q - 2W_m - t \quad \textit{with prob.} \quad \ge 1 - \mathrm{e}^{-18t^2}$$

Sources: Mendelson et al. 2013–2016; Tropp 2015.

Demixing I

Theorem 45 (Amelunxen et al. 2014). ***Assume***

- $f, g : \mathbb{R}^d \to \mathbb{R}$ *are convex functions*
- $\boldsymbol{x}_\natural \in \operatorname{dom}(f)$ *and* $\boldsymbol{y}_\natural \in \operatorname{dom}(g)$
- *Observe* $\boldsymbol{z} = \boldsymbol{x}_\natural + \boldsymbol{Q}\boldsymbol{y}_\natural$ *where* $\boldsymbol{Q} \in \mathbb{R}^{d \times d}$ *is random orthogonal*
- *The pair* $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}})$ *solves*

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

Then

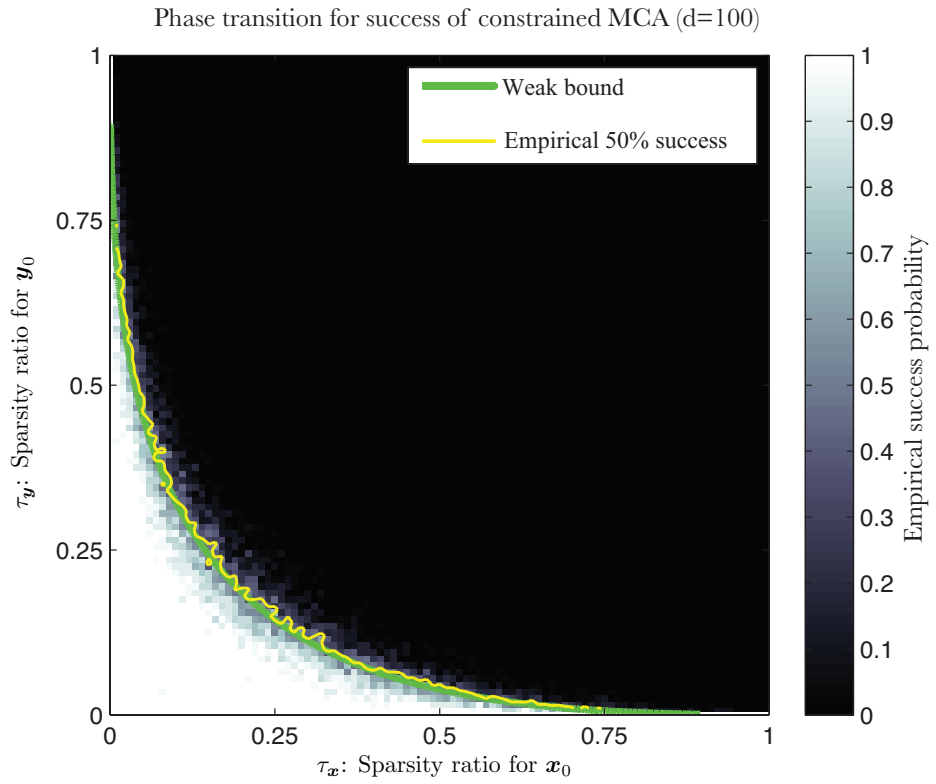
$$\delta(C) + \delta(K) \le d - O(\sqrt{d}) \quad \text{implies} \quad (\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) = (\boldsymbol{x}_\natural, \boldsymbol{y}_\natural) \quad \text{with high prob.}$$

$$\delta(C) + \delta(K) \ge d + O(\sqrt{d}) \quad \text{implies} \quad (\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) \ne (\boldsymbol{x}_\natural, \boldsymbol{y}_\natural) \quad \text{with high prob.}$$

where $C = \mathcal{D}(f, \boldsymbol{x}_\natural)$ *and* $K = \mathcal{D}(g, \boldsymbol{y}_\natural)$

Sources: Amelunxen et al. 2014.

Demixing II



minimize $\|\boldsymbol{x}\|_{\ell_1}$ subject to $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{Q}\boldsymbol{y}, \quad \|\boldsymbol{y}\|_{\ell_1} \leq \|\boldsymbol{y}_\natural\|_{\ell_1}$

Source: Starck et al. 2003; McCoy & Tropp 2013; McCoy et al. 2013; Amelunxen 2014.

Demixing III

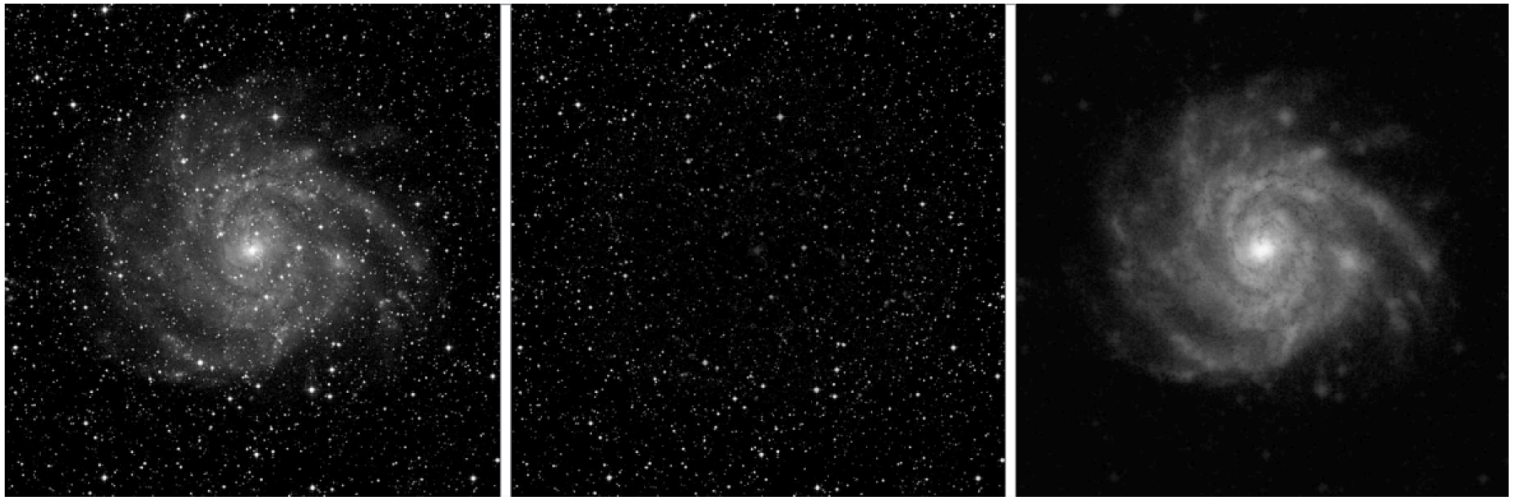


Image credit: NASA

Observation z_0 Sparse component x_0 DCT-sparse component y_0

Source: Starck et al. 2003; McCoy & Tropp 2013; McCoy et al. 2013; Amelunxen 2014.

To learn more...

E-mail: jtropp@cms.caltech.edu

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

Some Sources:

- Ledoux & Talagrand, "Probability in Banach spaces," Springer, 1991
- Chen et al., "Atomic decomposition by Basis Pursuit," *SIAM Rev.*, 2001
- Tropp, "Topics in sparse approximation," PhD Thesis, UT-Austin, 2004
- Mallat & Péyre, "A wavelet tour of signal processing," 3rd ed., Academic Press, 2009
- Jaggi, "Sparse convex optimization methods for machine learning," PhD Thesis, ETH, 2011
- Chandrasekaran et al., "The convex geometry of linear inverse problems," *Found. Comput. Math.*, 2012
- Foucart & Rauhut, "A mathematical introduction to compressive sampling," Birkhäuser, 2013
- McCoy, "A geometric analysis of convex demixing," PhD Thesis, Caltech, 2013
- McCoy & Tropp, "The achievable performance of convex demixing," ACM Report 2017-02, Caltech, 2013
- Amelunxen et al., "Living on the edge...," *Inform. Inference*, 2014
- McCoy & Tropp, "From Steiner formulas for cones...," *Discrete Comput. Geom*, 2014
- Oymak, "Convex relaxation for low-dimensional representation...," PhD Thesis, Caltech, 2014
- Tropp, "Convex recovery of structured signals...," in *Sampling Theory: A Renaissance*, Birkhäuser, 2015
- Thrampoulidis, "Recovering structured signals in high dimensions...," PhD Thesis, Caltech, 2016
- Goldstein et al., "Gaussian phase transitions and conic intrinsic volumes...," *Ann. Appl. Probab.*, 2017
- Amelunxen & Lotz, "Intrinsic volumes of polyhedral cones," *Discrete Comput. Geom.*, 2017
- Oymak & Tropp, "Universality laws for randomized dimension reduction...," arXiv:1511.09433, in revision