The Sparsity Gap

Joel A. Tropp

Computing & Mathematical Sciences
California Institute of Technology
jtropp@acm.caltech.edu

Research supported in part by ONR

Introduction

Systems of Linear Equations

We consider linear systems of the form

$$m\left\{\underbrace{\begin{bmatrix} & & \\ & \boldsymbol{\Phi} & \\ & & \end{bmatrix}}_{N}\right. \begin{bmatrix} \\ \boldsymbol{x} \\ \\ \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \end{bmatrix}$$

Assume that

- $\boldsymbol{\Phi}$ has dimensions $m \times N$ with $N > m$
- $\boldsymbol{\Phi}$ has full row rank
- The columns of $\boldsymbol{\Phi}$ have unit ℓ_2 norm

The Trichotomy Theorem

Theorem 1. *For a linear system $\mathbf{\Phi x} = \boldsymbol{b}$, exactly one of the following situations occurs.*

1. *No solution exists.*

2. *The equation has a unique solution.*

3. *The solutions form a linear subspace of positive dimension.*

Regularization via Sparsity

A principled approach to underdetermined systems:

$$\min \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{\Phi x} = \boldsymbol{b} \tag{P0}$$

where $\|\boldsymbol{x}\|_0 = \#\,\mathrm{supp}(\boldsymbol{x}) = \#\{j : x_j \neq 0\}$

- When $\|\boldsymbol{x}\|_0 \leq s$, then $\boldsymbol{x}$ is called *s-sparse*
- If $\boldsymbol{\Phi x} = \boldsymbol{b}$ and $\boldsymbol{x}$ is s-sparse, then $\boldsymbol{x}$ is an *s-sparse representation* of $\boldsymbol{b}$
- Since $\boldsymbol{\Phi}$ has full row-rank, every $\boldsymbol{b}$ has an m-sparse representation
- **Question:** What can we say about sparser representations?

Geometry

Key Insight

Sparse representations are well behaved
when the matrix $\mathbf{\Phi}$ is sufficiently nice

(Column submatrices should not be singular and
individual columns should not look like sparse signals)

Quantifying Niceness I

- We call $\mathbf{\Phi}$ a *tight frame* when

$$\mathbf{\Phi}\mathbf{\Phi}^* = \frac{N}{m} \cdot \mathbf{I}$$

- Equivalently, the rows of $\mathbf{\Phi}$ form an orthonormal family (up to scaling)
- Observe that N/m is the *redundancy* of the frame
- Tight frames have minimal spectral norm among conformal matrices

Quantifying Niceness II

- The *coherence* of $\mathbf{\Phi}$ is the quantity

$$\mu = \max_{j \neq k} |\langle \boldsymbol{\varphi}_j, \ \boldsymbol{\varphi}_k \rangle|$$

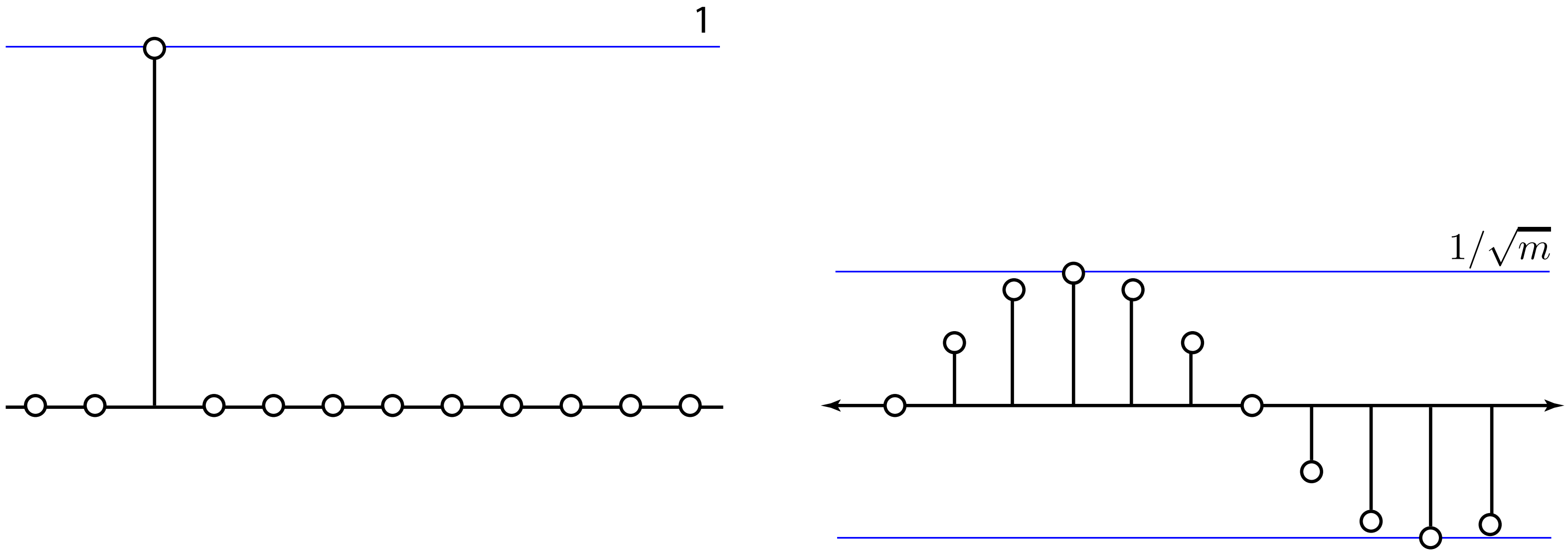
- Measures the angle between columns
- When $N \geq 2m$, the coherence satisfies

$$\mu \gtrsim \frac{1}{\sqrt{m}}$$

- Incoherent matrices appear often in signal processing applications

References: [Welch 1974, Mallat–Zhang 1993, Donoho–Huo 2001, Gribonval–Nielsen 2003, Strohmer–Heath 2003]

Example: Identity + Fourier



Impulses Complex Exponentials

A very incoherent tight frame

Uniqueness

Uncertainty implies Uniqueness

Theorem 2. *Suppose that a vector* $\boldsymbol{b}$ *has two representations:* $\boldsymbol{\Phi x} = \boldsymbol{b} = \boldsymbol{\Phi y}$. *Then*

$$\|\boldsymbol{x}\|_0 + \|\boldsymbol{y}\|_0 > \mu^{-1}.$$

Corollary 3. *Suppose that* $\boldsymbol{b} = \boldsymbol{\Phi x}$ *where*

$$\|\boldsymbol{x}\|_0 < \frac{1}{2} \cdot \mu^{-1}.$$

Then $\boldsymbol{x}$ *is the unique solution to* (P0).

❧ Very strict requirement since $\mu^{-1} \lesssim \sqrt{m}$

References: [Donoho–Stark 1989, Donoho–Huo 2001, Gribonval–Nielsen 2003, Donoho–Elad 2003]

The Square-Root Threshold

- Sparse representations are not necessarily unique past the $\sqrt{m}$ threshold

Example: The Dirac Comb

- Consider the Identity + Fourier matrix with $m = p^2$
- There is a vector $\boldsymbol{b}$ that can be written as either p spikes or p sines
- By the Poisson summation formula,

$$\boldsymbol{b}(t) = \sum_{j=0}^{p-1} \delta_{pj}(t) = \frac{1}{\sqrt{m}} \sum_{j=0}^{p-1} \mathrm{e}^{-2\pi \mathrm{i} pjt/m} \qquad \text{for } t = 0, 1, \dots, m$$

References: [Donoho–Stark 1989]

Enter Probability

Insight:
The bad vectors are atypical

An Uncertainty Principle for Generic Signals

Theorem 4. [T 2010] *Suppose that $\boldsymbol{b} = \boldsymbol{\Phi x}$ where the nonzero components of $\boldsymbol{x}$ have a continuous distribution. With probability one, the vector $\boldsymbol{b}$ has no representation $\boldsymbol{b} = \boldsymbol{\Phi y}$ where $\operatorname{supp}(\boldsymbol{x}) \cap \operatorname{supp}(\boldsymbol{y}) = \emptyset$ unless*

$$\|\boldsymbol{x}\|_0 + \|\boldsymbol{y}\|_0 > \mu^{-1} \|\boldsymbol{x}\|_0^{1/2}.$$

Corollary 5. *When $\mu \leq m^{-1/2}$, condition becomes*

$$\|\boldsymbol{y}\|_0 > \|\boldsymbol{x}\|_0 \left[\sqrt{\frac{m}{\|\boldsymbol{x}\|_0}} - 1 \right]$$

- Even with refinements, this approach does not yield uniqueness!
- **Problem:** Some *supports* could be bad

References: [*The Sparsity Gap*]

Enter More Probability

Insight:
The bad supports are atypical

A Simple Model for Random Sparse Vectors

Model (M0) for $\boldsymbol{b} = \boldsymbol{\Phi x}$	
The matrix $\boldsymbol{\Phi}$	is a unit-norm tight frame of size $m \times N$ with coherence $\mu \leq \mathrm{c}/\log N$.
The support of $\boldsymbol{x}$	has cardinality $s \leq cm/\log N$ and is uniformly random.
The nonzero entries of $\boldsymbol{x}$	have a continuous distribution.

Random Submatrices & Sparse Representation

Theorem 6. [T 2006, 2008] *Assume all parameters satisfy Model* (M0). *Draw a uniformly random set S of s columns from $\mathbf{\Phi}$, and define the random column submatrix $\boldsymbol{A} = \mathbf{\Phi}_S$. Then*

$$\text{Prob}\left\{\|\boldsymbol{A}^*\boldsymbol{A} - \mathbf{I}\| < \frac{1}{2}\right\} \;\geq\; 99.72\%$$

and

$$\text{Prob}\left\{\max_{n \notin S} \|\boldsymbol{A}^*\boldsymbol{\varphi}_n\|_2 < \frac{1}{2}\right\} \;\geq\; 99.72\%.$$

References: [*Random Subdictionaries*, *Random Submatrices*]

The Sparsity Gap

Theorem 7. [T 2008, 2010] *Let* $\boldsymbol{b} = \boldsymbol{\Phi x}$ *be a vector drawn from Model* (M0). *With probability at least 99.44%, the following statements hold.*

1. *The vector* $\boldsymbol{x}$ *is the unique solution to* (P0).
2. *Furthermore, there is no disjoint representation*

$$\boldsymbol{b} = \boldsymbol{\Phi y} \quad \textit{where} \quad \operatorname{supp}(\boldsymbol{x}) \cap \operatorname{supp}(\boldsymbol{y}) = \emptyset$$

   unless

$$\|\boldsymbol{y}\|_0 > \|\boldsymbol{x}\|_0 \left(1 + 2 \cdot \frac{m}{N}\right).$$

References: [Candès–Romberg 2006, *Random Subdictionaries*, *Random Submatrices*, *The Sparsity Gap*]

To learn more...

Web: `http://www.acm.caltech.edu/~jtropp`

E-mail: `jtropp@acm.caltech.edu`

Relevant papers:

- "Conditioning of random subdictionaries," *ACHA*, 2008
- "Norms of random submatrices," *CRAS*, 2008
- "Spikes and sines," *JFAA*, 2008
- "The sparsity gap," *Proc. CISS*, 2010