The Sparsity Gap

Joel A. Tropp

Computing & Mathematical Sciences
California Institute of Technology
jtropp@acm.caltech.edu
Introduction
We consider linear systems of the form

\[
\begin{bmatrix}
\Phi
\end{bmatrix}
\begin{bmatrix}
x
\end{bmatrix}
= 
\begin{bmatrix}
b
\end{bmatrix}
\]

Assume that

- \(\Phi\) has dimensions \(m \times N\) with \(N > m\)
- \(\Phi\) has full row rank
- The columns of \(\Phi\) have unit \(\ell_2\) norm
The Trichotomy Theorem

**Theorem 1.**  For a linear system $\Phi x = b$, 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:

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

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

- When $$\|x\|_0 \leq s$$, then $$x$$ is called \textit{s-sparse}

- If $$\Phi x = b$$ and $$x$$ is \textit{s-sparse}, then $$x$$ is an \textit{s-sparse representation} of $$b$$

- Since $$\Phi$$ has full row-rank, every $$b$$ has an \textit{m}-sparse representation

\textbf{Question:} What can we say about sparser representations?
Geometry
Key Insight

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

(Column submatrices should not be singular and individual columns should not look like sparse signals)
We call $\Phi$ a \textit{tight frame} when

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

Equivalently, the rows of $\Phi$ form an orthonormal family (up to scaling).

Observe that $N/m$ is the \textit{redundancy} of the frame.

Tight frames have minimal spectral norm among conformal matrices.
Quantifying Niceness II

- The coherence of $\Phi$ is the quantity

$$\mu = \max_{j \neq k} |\langle \varphi_j, \varphi_k \rangle|$$

- 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

Example: Identity + Fourier

A very incoherent tight frame
Uniqueness
Theorem 2. Suppose that a vector $b$ has two representations: $\Phi x = b = \Phi y$. Then
\[ \|x\|_0 + \|y\|_0 > \mu^{-1}. \]

Corollary 3. Suppose that $b = \Phi x$ where
\[ \|x\|_0 < \frac{1}{2} \cdot \mu^{-1}. \]

Then $x$ is the unique solution to (P0).

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

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 $b$ that can be written as either $p$ spikes or $p$ sines

By the Poisson summation formula,

$$b(t) = \sum_{j=0}^{p-1} \delta_{pj}(t) = \frac{1}{\sqrt{m}} \sum_{j=0}^{p-1} e^{-2\pi ipjt/m} \quad \text{for } t = 0, 1, \ldots, m$$

**References:** [Donoho–Stark 1989]
Enter Probability

Insight:
The bad vectors are atypical
**Theorem 4. [T 2010]** Suppose that \( b = \Phi x \) where the nonzero components of \( x \) have a continuous distribution. With probability one, the vector \( b \) has no representation \( b = \Phi y \) where \( \text{supp}(x) \cap \text{supp}(y) = \emptyset \) unless
\[
\|x\|_0 + \|y\|_0 > \mu^{-1} \|x\|_0^{1/2}.
\]

**Corollary 5.** When \( \mu \leq m^{-1/2} \), condition becomes
\[
\|y\|_0 > \|x\|_0 \left[ \sqrt{\frac{m}{\|x\|_0}} - 1 \right]
\]

- 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 $b = \Phi x$**

<table>
<thead>
<tr>
<th>Description</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>The matrix $\Phi$</td>
<td>is a unit-norm tight frame of size $m \times N$ with coherence $\mu \leq c/\log N$.</td>
</tr>
<tr>
<td>The support of $x$</td>
<td>has cardinality $s \leq cm/\log N$ and is uniformly random.</td>
</tr>
<tr>
<td>The nonzero entries of $x$</td>
<td>have a continuous distribution.</td>
</tr>
</tbody>
</table>

Draw a uniformly random set $S$ of $s$ columns from $\Phi$, and define the random column submatrix $A = \Phi_S$. Then

$$\Pr\left\{ \|A^* A - I\| < \frac{1}{2} \right\} \geq 99.72\%$$

and

$$\Pr\left\{ \max_{n \notin S} \|A^* \varphi_n\|_2 < \frac{1}{2} \right\} \geq 99.72\%.$$
The Sparsity Gap

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

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

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

unless

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

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