Signal Recovery from Partial Information

Joel A. Tropp
Anna C. Gilbert

{jtropp|annacg}@umich.edu

Department of Mathematics
The University of Michigan

The Signal Recovery Problem

Let $\boldsymbol{s}$ be an m-sparse signal in $\mathbb{R}^d$, for example

$$\boldsymbol{s} = \begin{bmatrix} 0 & -7.3 & 0 & 0 & 0 & 2.7 & 0 & 1.5 & 0 & \ldots \end{bmatrix}^T$$

Use *measurement vectors* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ to collect N *nonadaptive* linear measurements of the signal

$$\langle \boldsymbol{s}, \ \boldsymbol{x}_1 \rangle, \quad \langle \boldsymbol{s}, \ \boldsymbol{x}_2 \rangle, \quad \ldots, \quad \langle \boldsymbol{s}, \ \boldsymbol{x}_N \rangle$$

Q1. How many measurements are necessary to determine the signal?

Q2. How should the measurement vectors be chosen?

Q3. What algorithms can perform the reconstruction task?

Motivations

Medical Imaging

- Tomography provides incomplete, nonadaptive frequency information
- The images typically have a sparse gradient
- Reference: [Candès–Romberg–Tao 2004]

Sensor Networks

- Limited communication favors nonadaptive measurements
- Some types of natural data are approximately sparse
- References: [Haupt–Nowak 2005, Baraniuk et al. 2005]

Q1: How many measurements?

Adaptive measurements

Consider the class of m-sparse signals in $\mathbb{R}^d$ that have 0–1 entries

It is clear that $\log_2 \binom{d}{m}$ bits suffice to distinguish members of this class. By Sterling's approximation,

Storage per signal: $O(m \log(d/m))$ bits

A simple adaptive coding scheme can achieve this rate

Nonadaptive measurements

The naïve approach uses d orthogonal measurement vectors

Storage per signal: $O(d)$ bits

But we can do exponentially better. . .

Q2: What type of measurements?

Idea: Use randomness

Random measurement vectors yield summary statistics that are nonadaptive yet highly informative. Examples:

Bernoulli measurement vectors

Independently draw each $\boldsymbol{x}_n$ uniformly from $\{-1, +1\}^d$

Gaussian measurement vectors

Independently draw each $\boldsymbol{x}_n$ from the distribution

$$\frac{1}{(2\pi)^{d/2}} \, \mathrm{e}^{-\|\boldsymbol{x}\|_2^2/2}$$

Connection with Sparse Approximation

Define the fat $N \times d$ *measurement matrix*

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_N^T \end{bmatrix}$$

The columns of $\boldsymbol{\Phi}$ are denoted $\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_d$

Given an m-sparse signal $\boldsymbol{s}$, form the data vector $\boldsymbol{v} = \boldsymbol{\Phi}\,\boldsymbol{s}$

$$\begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varphi}_1 & \boldsymbol{\varphi}_2 & \boldsymbol{\varphi}_3 & \dots & \boldsymbol{\varphi}_d \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \vdots \\ s_d \end{bmatrix}$$

Note that $\boldsymbol{v}$ is a linear combination of m columns from $\boldsymbol{\Phi}$

Orthogonal Matching Pursuit (OMP)

Input: A measurement matrix $\mathbf{\Phi}$, data vector $\boldsymbol{v}$, and sparsity level m

Initialize the residual $\boldsymbol{r}_0 = \boldsymbol{v}$

For $t = 1, \dots, m$ do

A. Find the column index ω_t that solves

$$\omega_t \; = \; \arg\max_{j=1,\dots,d} \; |\langle \boldsymbol{r}_{t-1}, \; \boldsymbol{\varphi}_j \rangle|$$

B. Calculate the next residual

$$\boldsymbol{r}_t = \boldsymbol{v} - \boldsymbol{P}_t \, \boldsymbol{v}$$

where $\boldsymbol{P}_t$ is the orthogonal projector onto $\operatorname{span}\{\boldsymbol{\varphi}_{\omega_1}, \dots, \boldsymbol{\varphi}_{\omega_t}\}$

Output: An m-sparse estimate $\widehat{\boldsymbol{s}}$ with nonzero entries in components $\omega_1, \dots, \omega_m$. These entries appear in the expansion

$$\boldsymbol{P}_m \, \boldsymbol{v} \quad = \quad \sum_{t=1}^{T} \widehat{s}_{\omega_t} \, \boldsymbol{\varphi}_{\omega_t}$$

Advantages of OMP

We propose OMP as an *effective method for signal recovery* because

- OMP is fast
- OMP is easy to implement
- OMP is surprisingly powerful
- OMP is provably correct

The goal of this lecture is to justify these assertions

Theoretical Performance of OMP

Theorem 1. [T–G 2005] *Choose an error exponent p.*

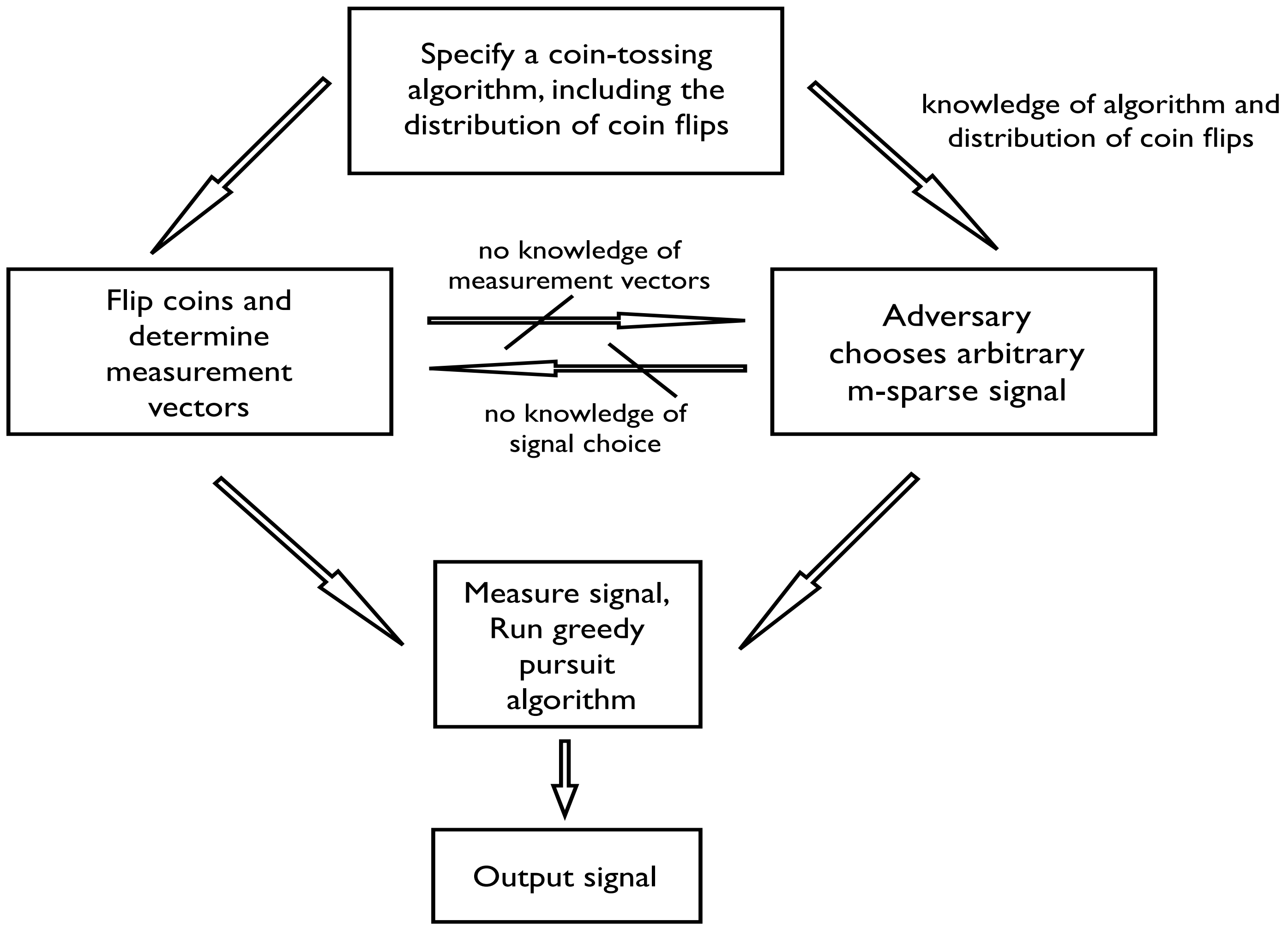
- *Let $\boldsymbol{s}$ be an arbitrary m-sparse signal in $\mathbb{R}^d$*
- *Draw $N = O(p\, m \log d)$ Gaussian or Bernoulli(?) measurements of $\boldsymbol{s}$*
- *Execute OMP with the data vector to obtain an estimate $\widehat{\boldsymbol{s}}$*

The estimate $\widehat{\boldsymbol{s}}$ equals the signal $\boldsymbol{s}$ with probability exceeding $(1 - 2\, d^{-p})$.

To achieve 99% success probability in practice, take

$$\boxed{N \approx 2\, m \ln d}$$

Flowchart for Algorithm



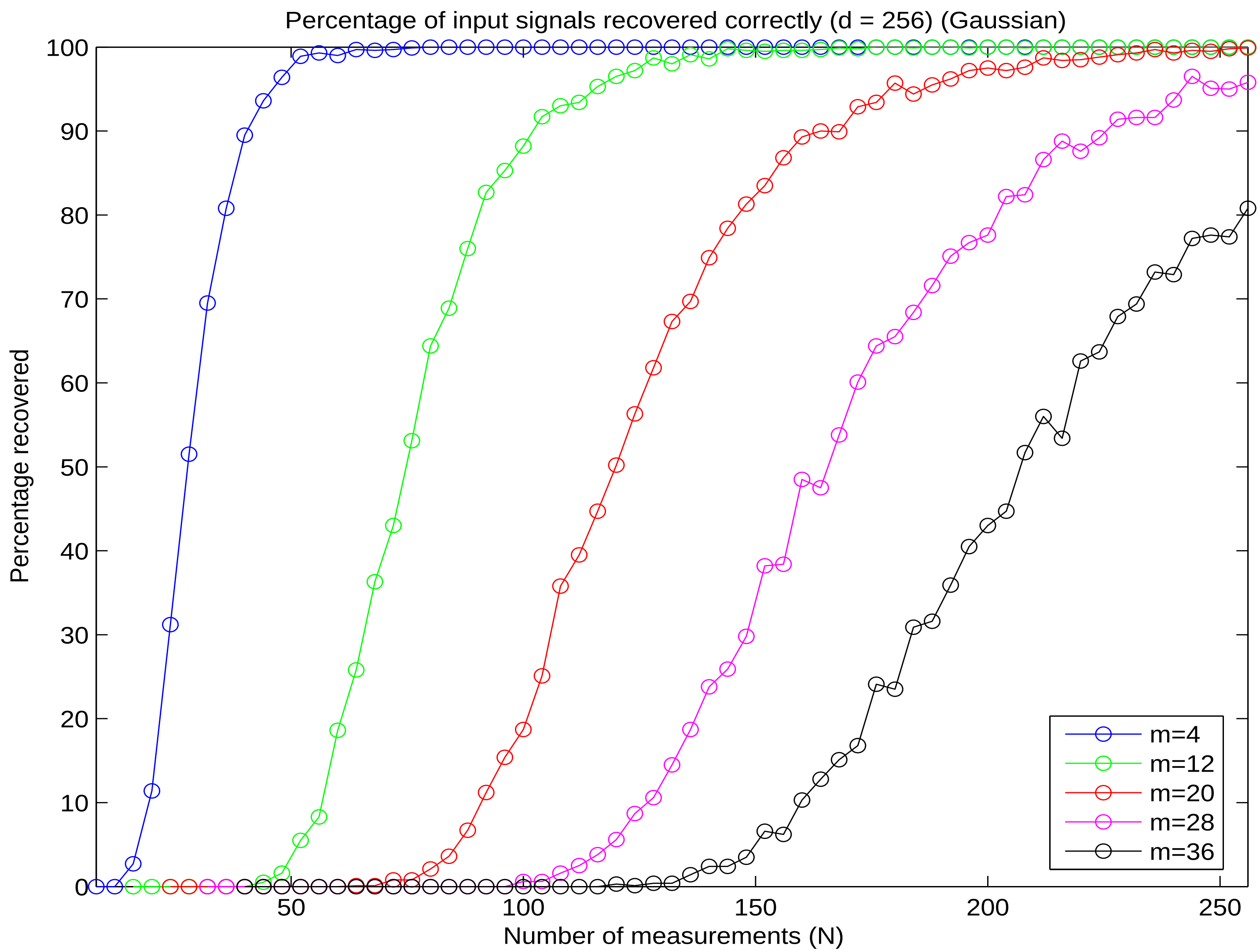
Empirical Results on OMP

For each trial. . .

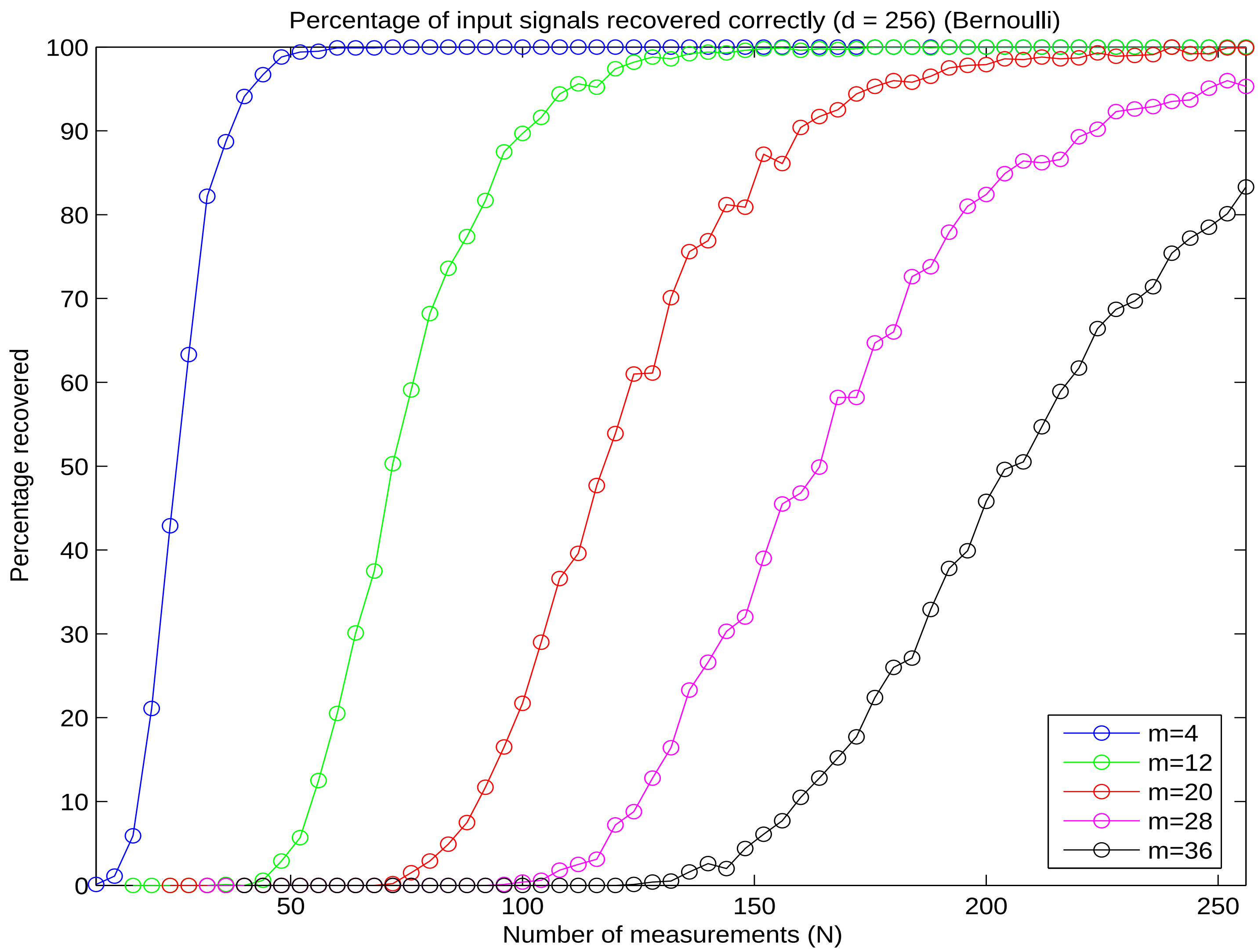
- Generate an m-sparse signal $\boldsymbol{s}$ in $\mathbb{R}^d$ by choosing m components and setting each to one
- Draw N Gaussian measurements of $\boldsymbol{s}$
- Execute OMP to obtain an estimate $\widehat{\boldsymbol{s}}$
- Check whether $\widehat{\boldsymbol{s}} = \boldsymbol{s}$

Perform 1000 independent trials for each triple (m, N, d)

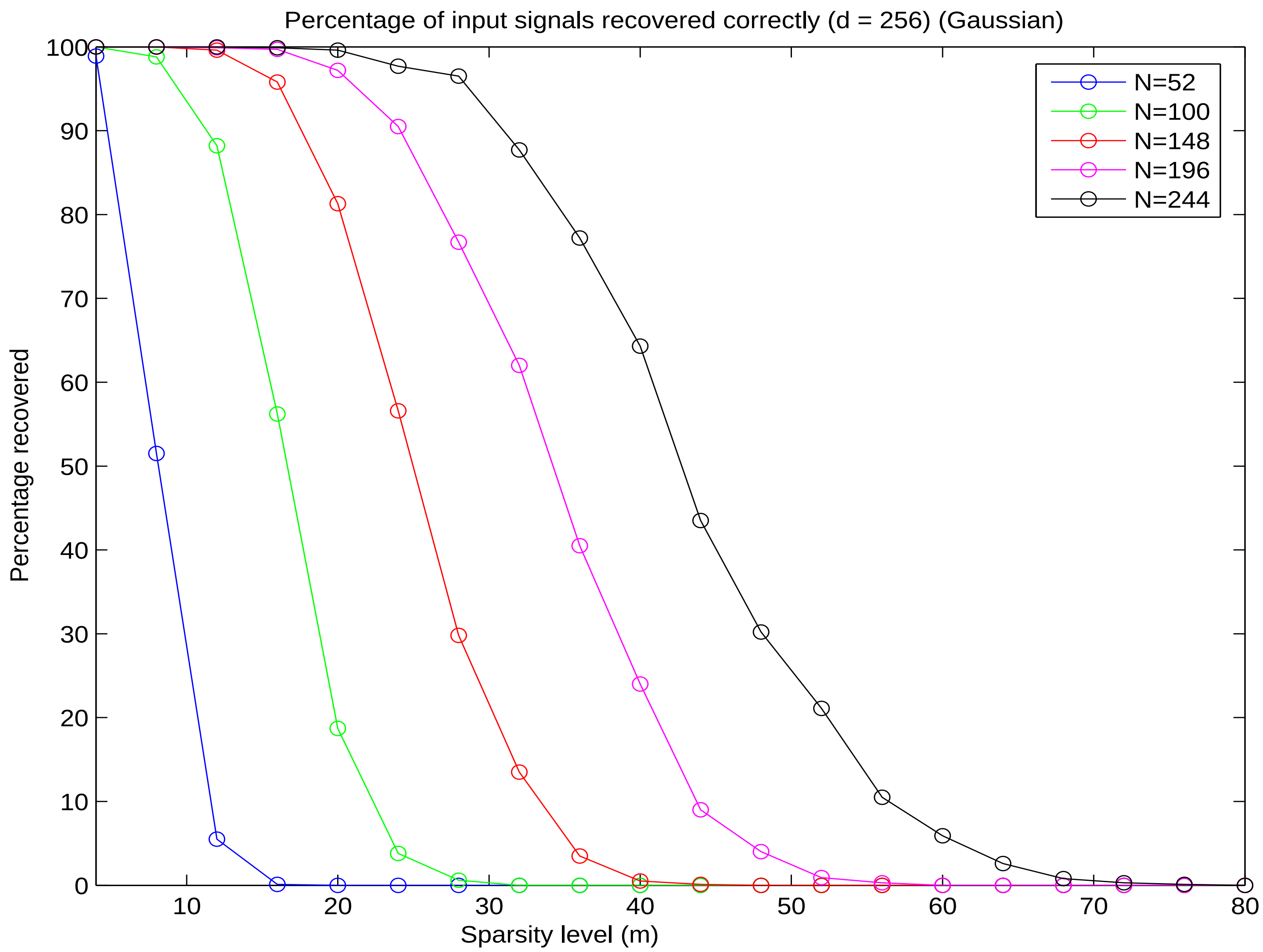
Percentage Recovered vs. Number of Gaussian Measurements



Percentage Recovered vs. Number of Bernoulli Measurements

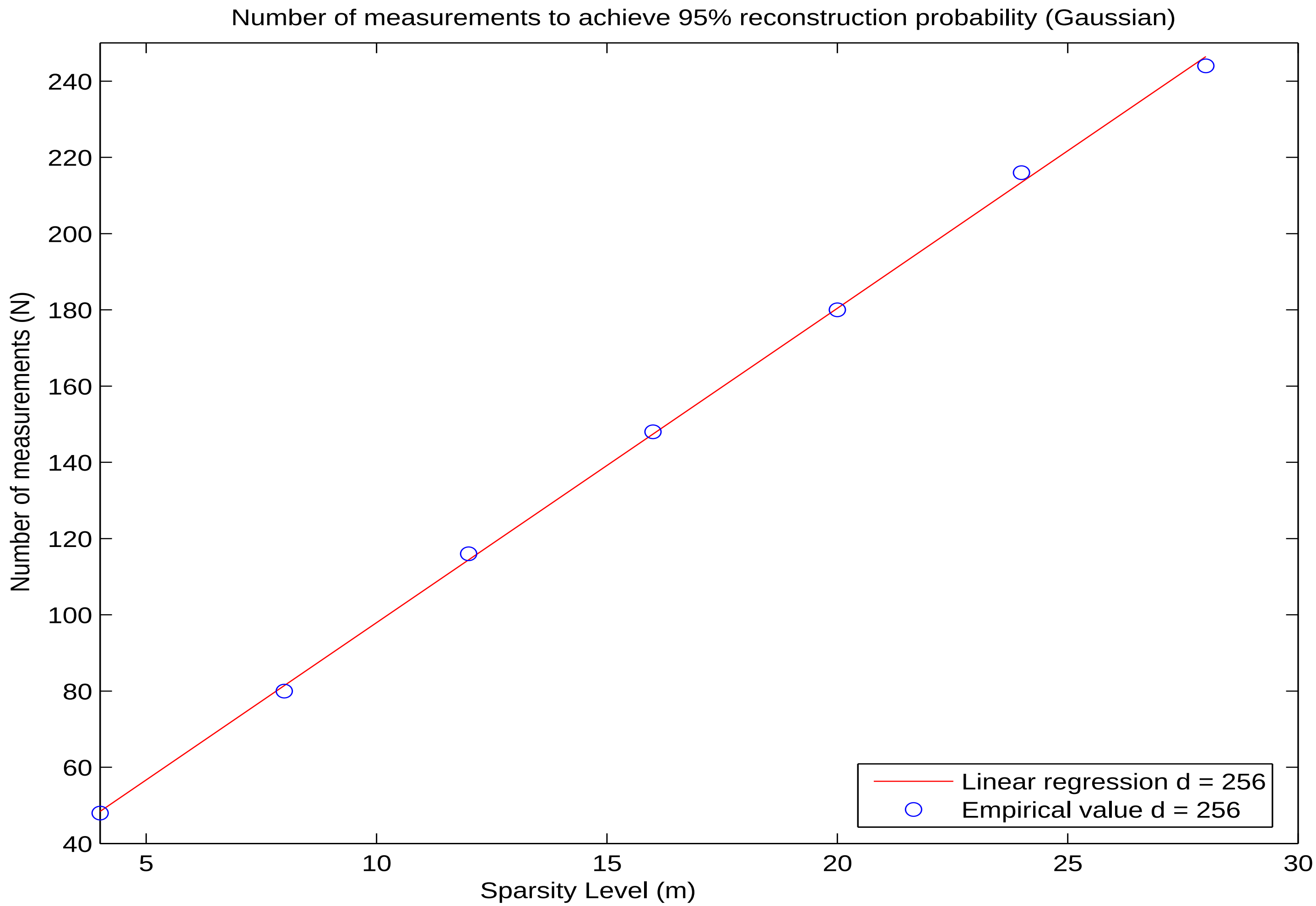


Percentage Recovered vs. Level of Sparsity



Number of Measurements for 95% Recovery

Regression Line: $N = 1.5\, m \ln d + 15.4$



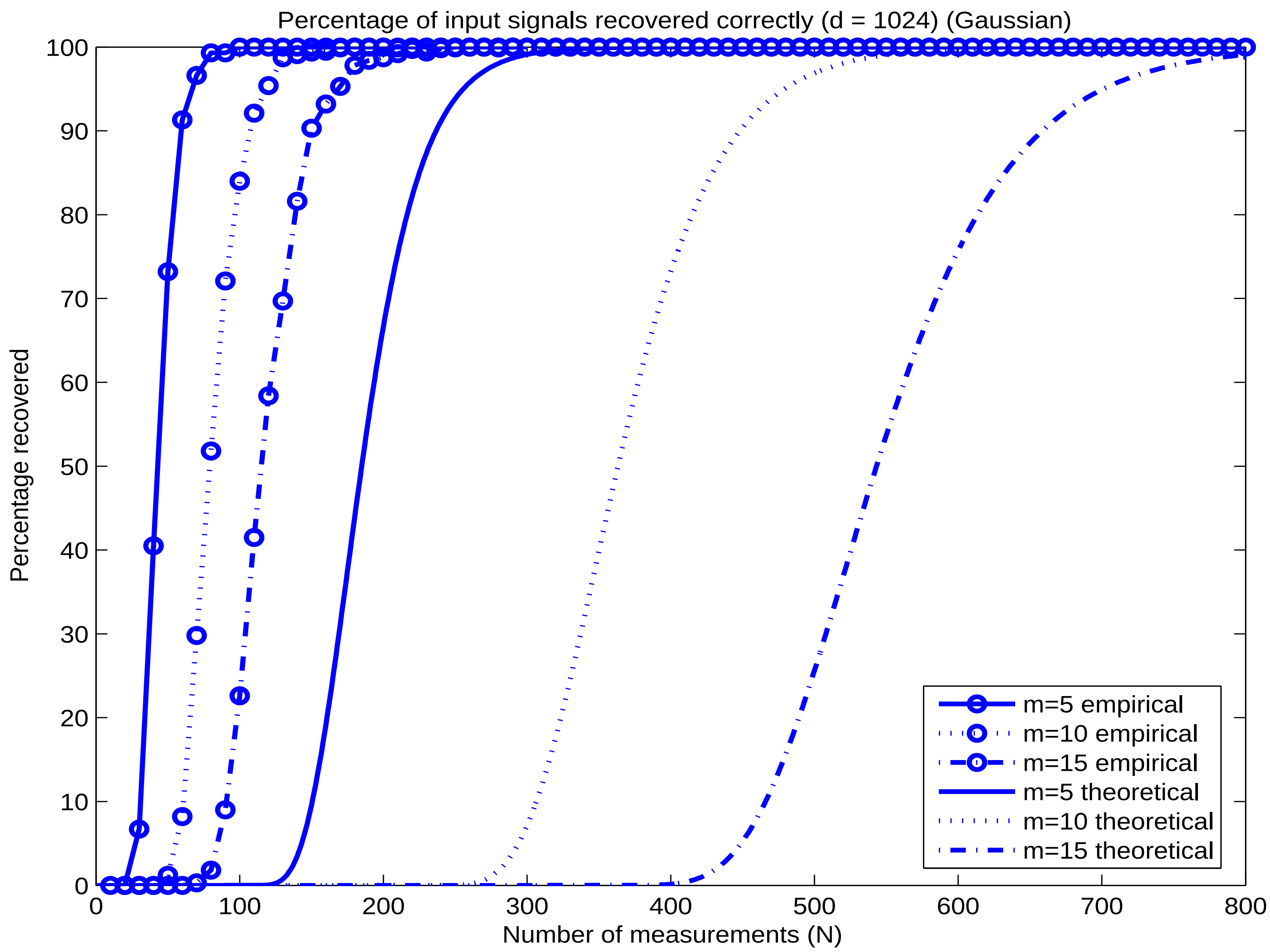
Number of Measurements for 99% Recovery

$d = 256$			$d = 1024$		
m	N	$N/(m \ln d)$	m	N	$N/(m \ln d)$
4	56	2.52	5	80	2.31
8	96	2.16	10	140	2.02
12	136	2.04	15	210	2.02
16	184	2.07			
20	228	2.05			

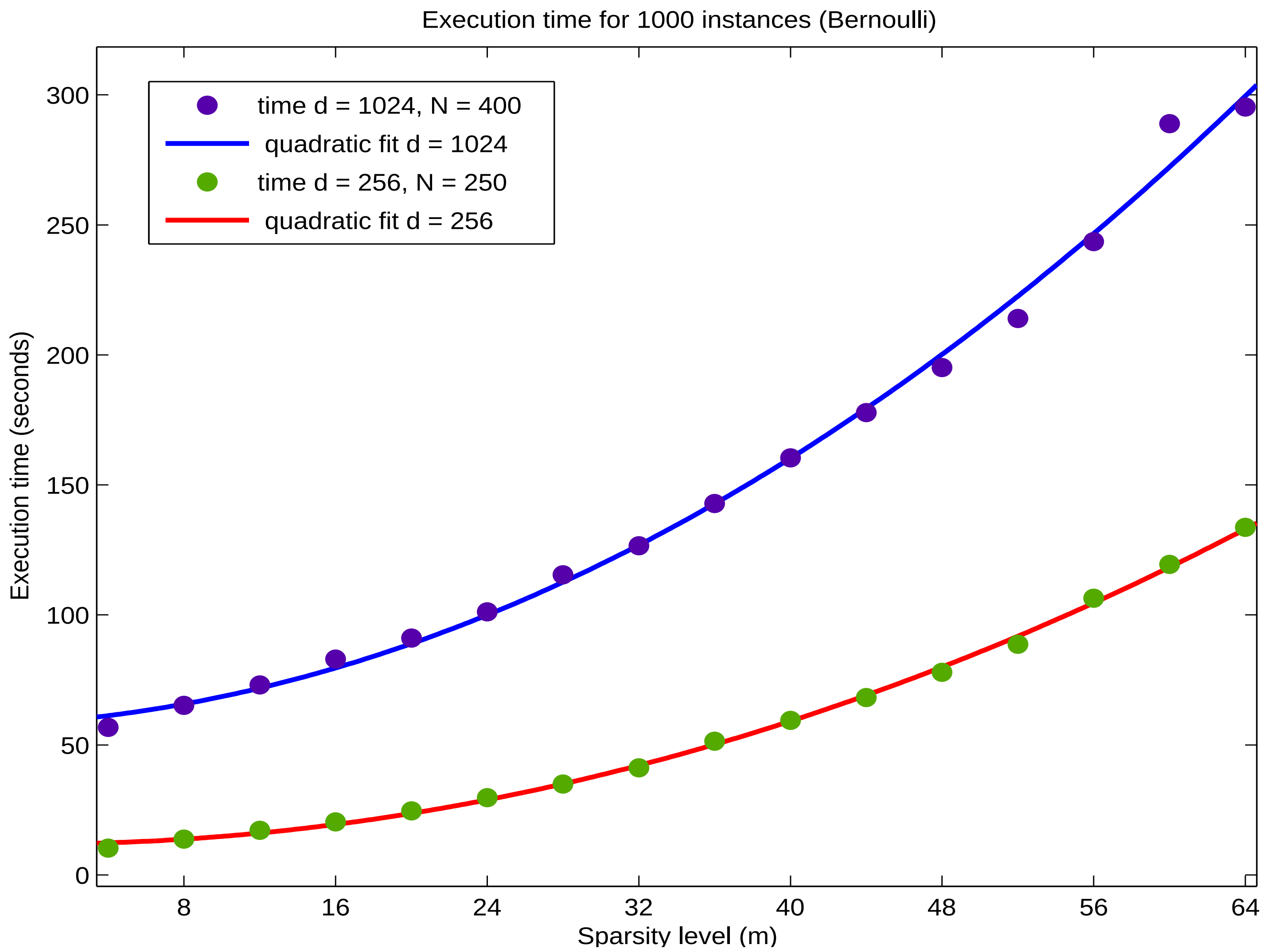
These data justify the rule of thumb

$$\boxed{N \approx 2\, m \ln d}$$

Percentage Recovered: Empirical vs. Theoretical



Execution Time for 1000 Complete Trials



Elements of the Proof I

A Thought Experiment

- Fix an m-sparse signal $\boldsymbol{s}$ and draw a measurement matrix $\mathbf{\Phi}$
- Let $\mathbf{\Phi}_{\rm opt}$ consist of the m correct columns of $\mathbf{\Phi}$
- Imagine we could run OMP with the data vector and the matrix $\mathbf{\Phi}_{\rm opt}$
- It would choose all m columns of $\mathbf{\Phi}_{\rm opt}$ in some order
- If we run OMP with the full matrix $\mathbf{\Phi}$ and it succeeds, then *it must select columns in exactly the same order*

Elements of the Proof II

The Sequence of Residuals

- If OMP succeeds, we know the sequence of residuals $\boldsymbol{r}_1, \dots, \boldsymbol{r}_m$
- Each residual lies in the span of the correct columns of $\boldsymbol{\Phi}$
- Each residual is stochastically independent of the incorrect columns

Elements of the Proof III

The Greedy Selection Ratio

- Suppose that $\boldsymbol{r}$ is the residual in Step A of OMP
- The algorithm picks a correct column of $\boldsymbol{\Phi}$ whenever

$$\rho(\boldsymbol{r}) \;=\; \frac{\max_{\{j \,:\, s_j = 0\}} |\langle \boldsymbol{r},\; \boldsymbol{\varphi}_j \rangle|}{\max_{\{j \,:\, s_j \neq 0\}} |\langle \boldsymbol{r},\; \boldsymbol{\varphi}_j \rangle|} \;<\; 1$$

- The proof shows that $\rho(\boldsymbol{r}_t) < 1$ for all t with high probability

Elements of the Proof IV

Measure Concentration

- The incorrect columns of $\mathbf{\Phi}$ are probably almost orthogonal to $\boldsymbol{r}_t$
- One of the correct columns is probably somewhat correlated with $\boldsymbol{r}_t$

- So the numerator of the greedy selection ratio is probably small

$$\text{Prob}\left\{ \max_{\{j \,:\, s_j = 0\}} |\langle \boldsymbol{r}_t,\ \boldsymbol{\varphi}_j \rangle| \ >\ \varepsilon\, \|\boldsymbol{r}_t\|_2 \right\} \ \lesssim\ d\, \mathrm{e}^{-\varepsilon^2/2}$$

- But the denominator is probably not too small

$$\text{Prob}\left\{ \max_{\{j \,:\, s_j \neq 0\}} |\langle \boldsymbol{r}_t,\ \boldsymbol{\varphi}_j \rangle| \ <\ \left(\sqrt{\frac{N}{m}} - 1 - \varepsilon \right) \|\boldsymbol{r}_t\|_2 \right\} \ \lesssim\ \mathrm{e}^{-\varepsilon^2\, m/2}$$

Another Method: ℓ_1 Minimization

- Suppose $\boldsymbol{s}$ is an m-sparse signal in $\mathbb{R}^d$
- The vector $\boldsymbol{v} = \boldsymbol{\Phi}\,\boldsymbol{s}$ is a linear combination of m columns of $\boldsymbol{\Phi}$
- For Gaussian measurements, this m-term representation is unique

Signal Recovery as a Combinatorial Problem

$$\min_{\widehat{\boldsymbol{s}}} \; \|\widehat{\boldsymbol{s}}\|_0 \quad \text{subject to} \quad \boldsymbol{\Phi}\,\widehat{\boldsymbol{s}} = \boldsymbol{v} \qquad (\ell_0)$$

Relax to a Convex Program

$$\min_{\widehat{\boldsymbol{s}}} \; \|\widehat{\boldsymbol{s}}\|_1 \quad \text{subject to} \quad \boldsymbol{\Phi}\,\widehat{\boldsymbol{s}} = \boldsymbol{v} \qquad (\ell_1)$$

References: [Donoho et al. 1999, 2004] and [Candès et al. 2004]

A Result for ℓ_1 Minimization

Theorem 2. [Rudelson–Vershynin 2005] *Draw* $N = O(m \log(d/m))$ *Gaussian measurement vectors. With probability at least* $(1 - \mathrm{e}^{-d})$*, the following statement holds. For every* m*-sparse signal in* $\mathbb{R}^d$*, the solution to* (ℓ_1) *is identical with the solution to* (ℓ_0)*.*

Notes:

- One set of measurement vectors works for all m-sparse signals
- Related results have been established in [Candès et al. 2004–2005] and in [Donoho et al. 2004–2005]

So, why use OMP?

Ease of implementation and speed

- Writing software to solve (ℓ_1) is difficult
- Even specialized software for solving (ℓ_1) is slow

Sample Execution Times

m	N	d	OMP Time	(ℓ_1) Time
14	175	512	0.02 s	1.5 s
28	500	2048	0.17	14.9
56	1024	8192	2.50	212.6
84	1700	16384	11.94	481.0
112	2400	32768	43.15	1315.6

Randomness

In contrast with ℓ_1, OMP may require randomness during the algorithm

Randomness can be reduced by

- Amortizing over many input signals
- Using a smaller probability space
- Accepting a small failure probability

Research Directions

- (Dis)prove existence of deterministic measurement ensembles
- Extend OMP results to approximately sparse signals
- Applications of signal recovery
- Develop new algorithms

Related Papers and Contact Information

- "Signal recovery from partial information via Orthogonal Matching Pursuit," submitted April 2005
- "Algorithms for simultaneous sparse approximation. Parts I and II," accepted to *EURASIP J. Applied Signal Processing*, April 2005
- "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Info. Theory*, October 2004
- "Just Relax: Convex programming methods for identifying sparse signals," submitted February 2004

All papers available from `http://www.umich.edu/~jtropp`

E-mail: `{jtropp|annacg}@umich.edu`