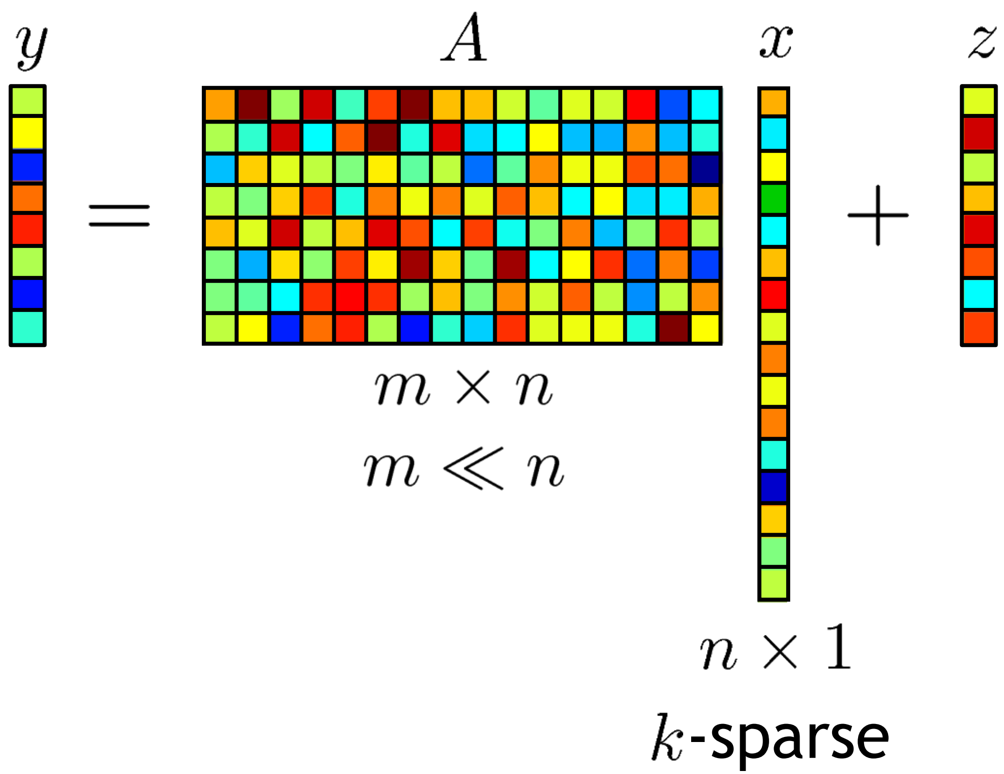
Adaptive sensing of sparse signals in noise

Mark A. Davenport

Georgia Institute of Technology

School of Electrical and Computer Engineering

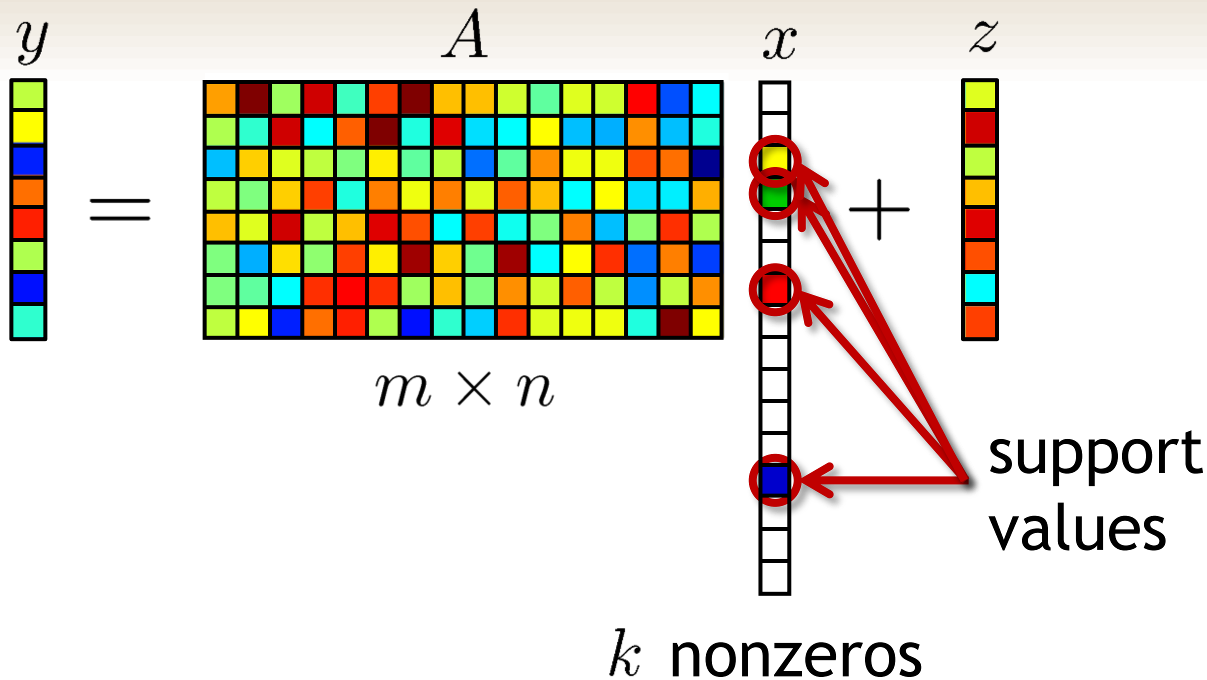
Sparse recovery



When (and how well) can we estimate x from the measurements y?

Review of Nonadaptive Sparse Recovery

Sparse recovery



- How should we design A to ensure that y contains as much information about x as possible?
- What algorithms do we have for recovering x from y?

How to design A?

Prototypical sensing model:

$$y = Ax + z \qquad\qquad z \sim \mathcal{N}(0, \sigma^2 I)$$

- Constrain A to have fixed Frobenius norm
- Pick A at *random*
  - i.i.d. sub-Gaussian entries
  - random rows from a unitary matrix
- As long as $m \geq Ck \log(n/k)$, with high probability a random A will satisfy nice properties
  - "restricted isometry property"
  - deep connections with *Johnson-Lindenstrauss Lemma*

How to recover x ?

Lots and lots of algorithms

- ℓ_1-minimization
- greedy algorithms (matching pursuit, CoSaMP, IHT)
- ....

Suppose that $y = Ax + z$ with $z \sim \mathcal{N}(0, \sigma^2 I)$.
If we choose A at random, then for any x with $\|x\|_0 \leq k$ we have that

$$\widehat{x} = \arg\min_{x' \in \mathbb{R}^n} \|x'\|_1 \quad \text{s.t.} \quad \|A^*(y - Ax')\|_\infty \leq c$$

satisfies

$$\mathbb{E}\,\|\widehat{x} - x\|_2^2 \leq C \frac{k\sigma^2}{\|A\|_F^2/n} \log n.$$

[Candès and Tao (2005)]

Room for improvement?

There exists matrices A such that for *any* (sparse) x we have

$$\mathbb{E}\,\|\widehat{x} - x\|_2^2 \leq C \frac{k\sigma^2}{\|A\|_F^2/n} \log n.$$

$$y_i = \langle a_i, x \rangle + z_i$$

a_i and x are almost orthogonal

- We are using most of our "sensing power" to sense entries that aren't even there!
- Tremendous loss in signal-to-noise ratio (SNR)
- It's hard to imagine any way to avoid this…

Can we do better?

Theorem

For *any* matrix A and *any* recovery procedure $\widehat{x}$, there exists an x with $\|x\|_0 \leq k$ such that if $y = Ax + z$ with $z \sim \mathcal{N}(0, \sigma^2 I)$, then

$$\mathbb{E}\, \|\widehat{x}(y) - x\|_2^2 \geq C' \frac{k\sigma^2}{\|A\|_F^2/n} \log(n/k).$$

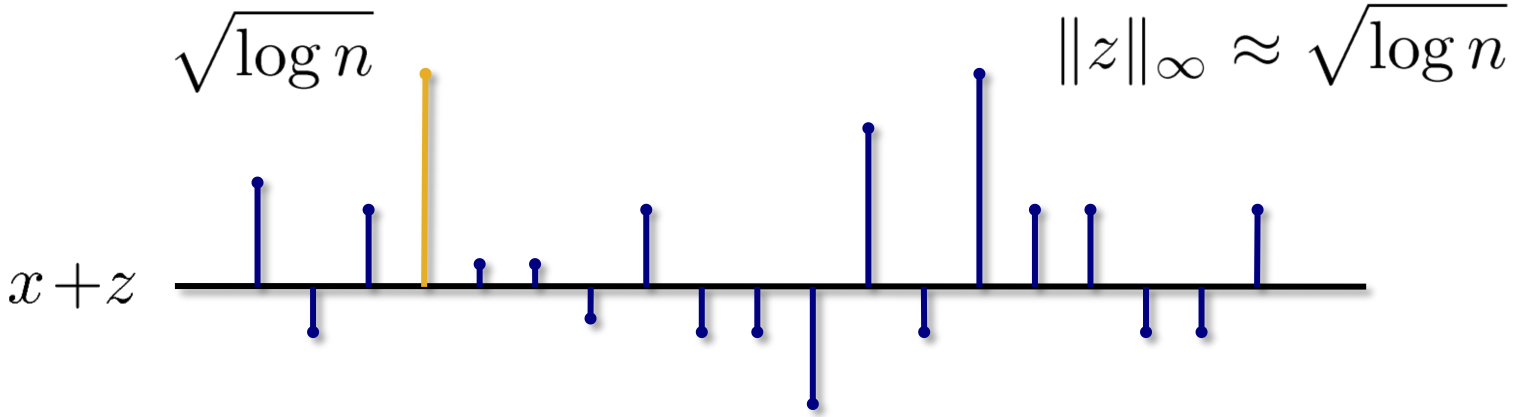
We are already operating at the limit

[Candès and Davenport (2013)]

Intuition

Suppose that $y = x + z$ with $z \sim \mathcal{N}(0, I)$ and that $k = 1$

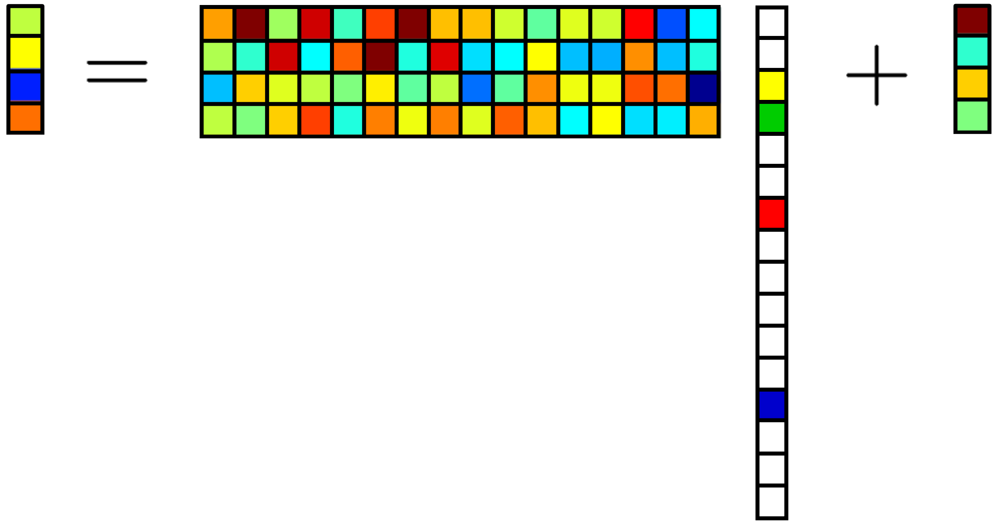
$$\mathbb{E}\, \|\widehat{x}(y) - x\|_2^2 \geq C' \log n$$



Adaptive Sensing

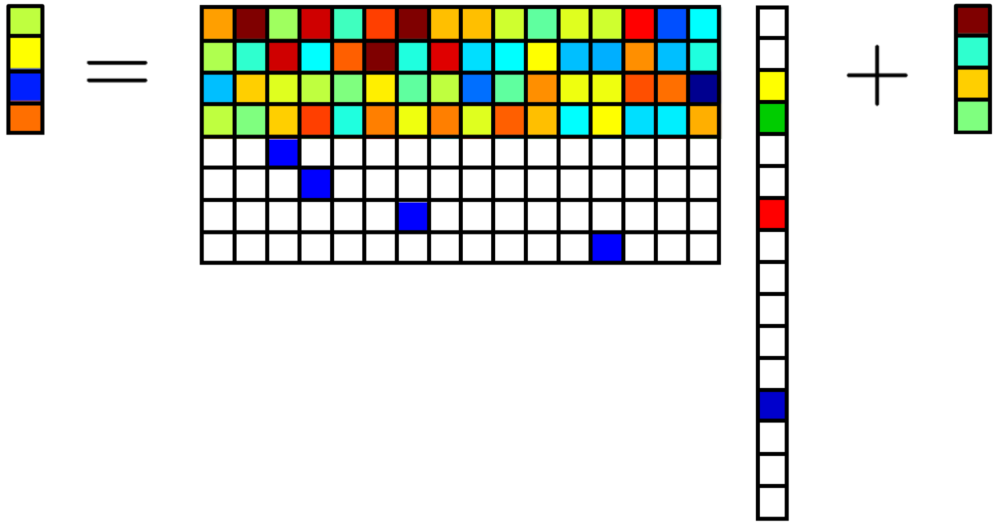
Adaptive sensing

Think of sensing as a game of 20 questions



Adaptive sensing

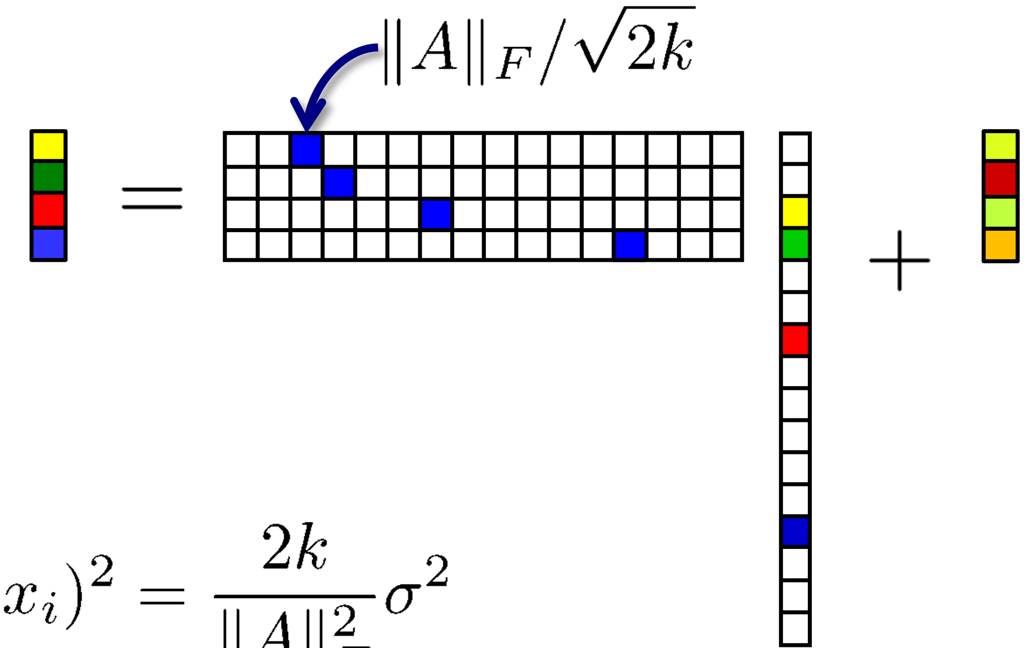
Think of sensing as a game of 20 questions



Simple strategy: Use half of our sensing energy to find the support, and the remainder to estimate the values.

Thought experiment

Suppose that after the first stage we have perfectly estimated the support



$$\mathbb{E}\,(\widehat{x}_i - x_i)^2 = \frac{2k}{\|A\|_F^2}\sigma^2$$

$$\mathbb{E}\,\|\widehat{x} - x\|_2^2 = \frac{k\sigma^2}{\|A\|_F^2/2k} \ll \frac{k\sigma^2}{\|A\|_F^2/n}\log n$$

Does adaptivity *really* help?

- Noise-free measurements, but non-sparse signal
  - adaptivity doesn't help if you want a uniform guarantee

    [Russians (1970s and 80s), Donoho (2004)]
  - probabilistic adaptive algorithms can reduce the required number of measurements to $O(k \log\log(n/k))$

    [Indyk, Price, and Woodruff (2011), Price and Woodruff (2013)]

- Noisy setting
  - distilled sensing [Haupt et al. (2007, 2010)]
  - adaptivity can reduce the estimation error to

$$\mathbb{E}\,\|\widehat{x} - x\|_2^2 \leq C\,\frac{k\sigma^2}{\|A\|_F^2/n}$$

Which is it?

$$\mathbb{E}\,\|\widehat{x} - x\|_2^2 \leq C'\,\frac{k\sigma^2}{\|A\|_F^2/k}$$

Which is it?

Suppose that we acquire measurements of the form $y_i = \langle a_i, x \rangle + z_i$ where $z_i \sim \mathcal{N}(0, \sigma^2)$ and the vector a_i can have an arbitrary dependence on the measurement history, i.e., $(a_1, y_1), \ldots, (a_{i-1}, y_{i-1})$

Theorem

There exist x with $\|x\|_0 \le k$ such that for *any* adaptive measurement strategy and *any* recovery procedure $\widehat{x}$,

$$\mathbb{E}\,\|\widehat{x}(y) - x\|_2^2 \ge C\,\frac{k\sigma^2}{\|A\|_F^2/n}.$$

Thus, in general, adaptivity does *not* significantly help!

[Arias-Castro, Candès, and Davenport (2013)]

Proof strategy

Step 1: Worst-case error is always bounded by average error over a class of possible x. Consider a prior on sparse signals with nonzeros of amplitude $\mu \approx \frac{\sigma}{\|A\|_F/\sqrt{n}}$

Step 2: Show that given our budget for $\|A\|_F$, it is impossible to detect the support very well

Step 3: Immediately translate this into a lower bound on MSE

To make things simpler, we will consider a Bernoulli prior $\pi(x)$ instead of a uniform k-sparse prior:

$$x_j = \begin{cases} 0 & \text{with probability } 1 - k/n \\ \mu > 0 & \text{with probability } k/n \end{cases}$$

Proof of main result

Let $S = \{j : x_j \neq 0\}$ and set $\sigma^2 = 1$

For any estimator $\widehat{x}$, define $\widehat{S} := \{j : |\widehat{x}_j| \geq \mu/2\}$

Whenever $j \in S \setminus \widehat{S}$ or $j \in \widehat{S} \setminus S$, $|\widehat{x}_j - x_j| \geq \mu/2$

$$\|\widehat{x} - x\|_2^2 \geq \frac{\mu^2}{4}|S \setminus \widehat{S}| + \frac{\mu^2}{4}|\widehat{S} \setminus S| = \frac{\mu^2}{4}|\widehat{S} \Delta S|$$

$$\Rightarrow \quad \mathbb{E}\,\|\widehat{x} - x\|_2^2 \geq \frac{\mu^2}{4}\mathbb{E}\,|\widehat{S} \Delta S|$$

Proof of main result

Lemma
Under the Bernoulli prior, *any* estimate $\widehat{S}$ satisfies

$$\mathbb{E}\,|\widehat{S}\Delta S| \geq k\left(1 - \frac{\mu}{2}\frac{\|A\|_F}{\sqrt{n}}\right).$$

Thus, $\mathbb{E}\,\|\widehat{x} - x\|_2^2 \geq \dfrac{\mu^2}{4}\mathbb{E}\,|\widehat{S}\Delta S|$

$$\geq k \cdot \frac{\mu^2}{4}\left(1 - \frac{\mu}{2}\frac{\|A\|_F}{\sqrt{n}}\right)$$

Plug in $\mu = \frac{8}{3}\sqrt{n}/\|A\|_F$ and this reduces to

$$\mathbb{E}\,\|\widehat{x} - x\|_2^2 \geq \frac{4}{27} \cdot \frac{kn}{\|A\|_F^2} \geq \frac{1}{7} \cdot \frac{k}{\|A\|_F^2/n}$$

Key ideas in proof of lemma

$$\mathbb{P}_{0,j}(y_1,\ldots,y_m) = \mathbb{P}(y_1,\ldots y_m | x_j = 0)$$
$$\mathbb{P}_{1,j}(y_1,\ldots,y_m) = \mathbb{P}(y_1,\ldots,y_m | x_j = \mu)$$

$$\mathbb{E}\,|\widehat{S}\Delta S| \geq \frac{k}{n}\sum_j (1 - \|\mathbb{P}_{1,j} - \mathbb{P}_{0,j}\|_{\mathrm{TV}})$$
$$\geq k - \frac{k}{\sqrt{n}}\sqrt{\sum_j \|\mathbb{P}_{1,j} - \mathbb{P}_{0,j}\|_{\mathrm{TV}}^2}$$

$$\sum_j \|\mathbb{P}_{1,j} - \mathbb{P}_{0,j}\|_{\mathrm{TV}}^2 \leq \frac{\mu^2}{4}\|A\|_F^2$$

$$\Longrightarrow \quad \mathbb{E}\,|\widehat{S}\Delta S| \geq k\left(1 - \frac{\mu}{2}\frac{\|A\|_F}{\sqrt{n}}\right)$$

Key ideas in proof of lemma

Pinsker's Inequality

$$\|\mathbb{P} - \mathbb{Q}\|_{\mathrm{TV}} \leq \sqrt{K(\mathbb{P}, \mathbb{Q})/2}$$

$$\|\mathbb{P}_{1,j} - \mathbb{P}_{0,j}\|_{\mathrm{TV}}^2 \leq \frac{\pi_0}{2} K(\mathbb{P}_{0,j}, \mathbb{P}_{1,j}) + \frac{\pi_1}{2} K(\mathbb{P}_{1,j}, \mathbb{P}_{0,j})$$

$$\leq \frac{\mu^2}{4} \sum_i \mathbb{E}\, a_{i,j}^2$$

$$\Rightarrow \sum_j \|\mathbb{P}_{1,j} - \mathbb{P}_{0,j}\|_{\mathrm{TV}}^2 \leq \frac{\mu^2}{4} \sum_{i,j} \mathbb{E}\, a_{i,j}^2 \leq \frac{\mu^2}{4} \|A\|_F^2$$

Adaptivity in Practice

Incredibly simplified model

Suppose that $k = 1$ and that $x_{j^*} = \mu$

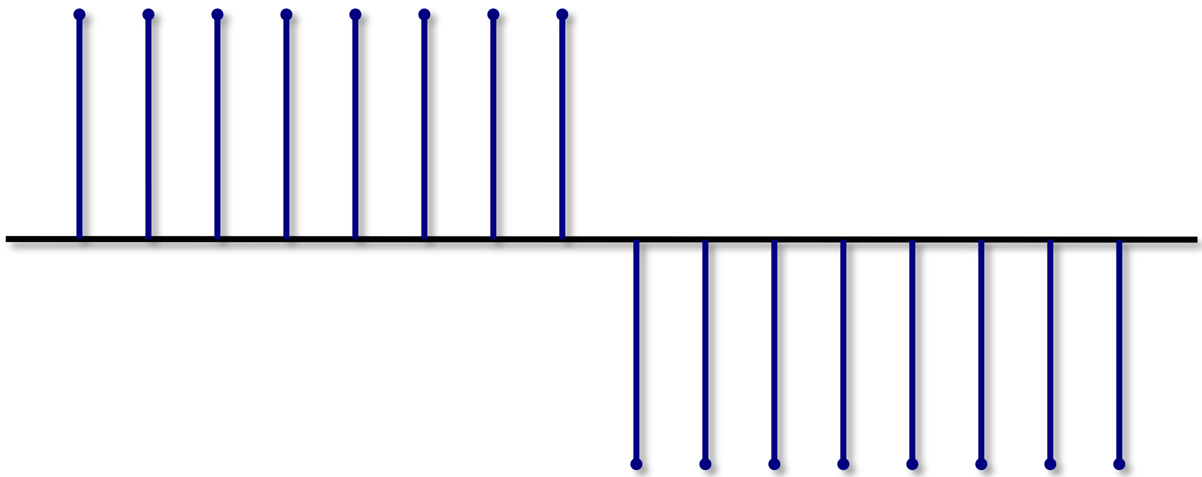
Our goal is to find j^* ***and*** estimate μ

We will split our budget for $\|A\|_F$ into two phases

1. Identify j^* via a simple binary search procedure
2. Estimate the value of μ by directly sampling it with the remaining sensing energy

Binary search

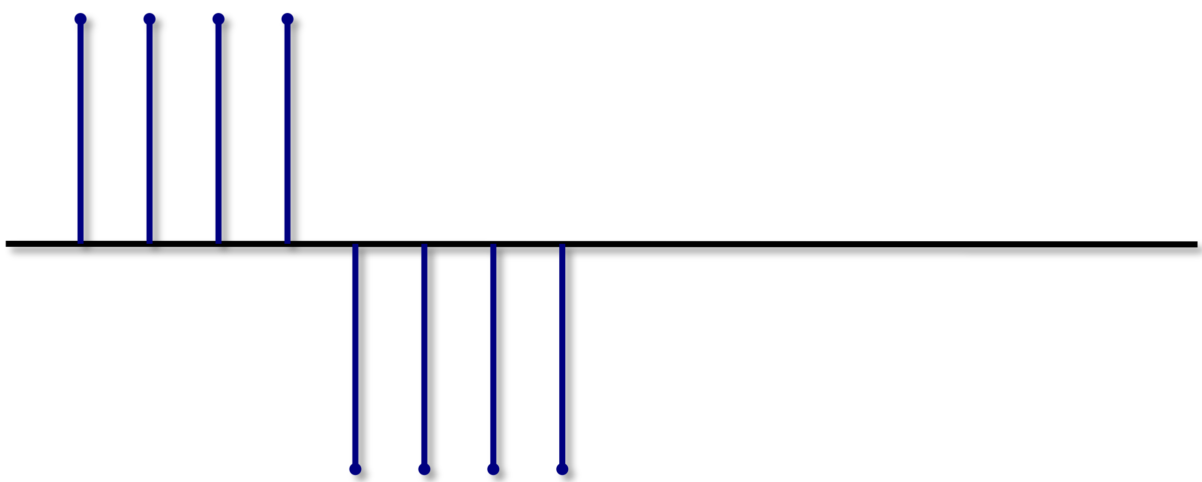
- Split measurements into $\log_2 n$ stages
- In each stage, use some of the "sensing energy" to determine if the nonzero is on the "left" or "right" of the active set



[Iwen and Tewfik (2011), Davenport and Arias-Castro (2012), Malloy and Nowak (2012)]

Binary search

- Split measurements into $\log_2 n$ stages
- In each stage, use some of the "sensing energy" to determine if the nonzero is on the "left" or "right" of the active set



[Iwen and Tewfik (2011), Davenport and Arias-Castro (2012), Malloy and Nowak (2012)]

Binary search

- Split measurements into $\log_2 n$ stages
- In each stage, use some of the "sensing energy" to determine if the nonzero is on the "left" or "right" of the active set

[Iwen and Tewfik (2011), Davenport and Arias-Castro (2012), Malloy and Nowak (2012)]

Binary search

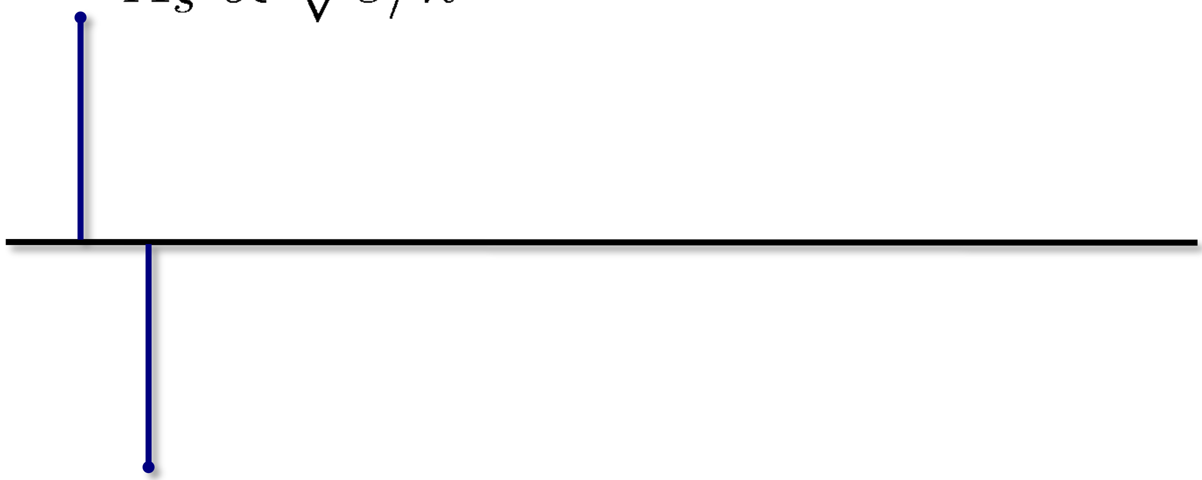
- Split measurements into $\log_2 n$ stages
- In each stage, use some of the "sensing energy" to determine if the nonzero is on the "left" or "right" of the active set

[Iwen and Tewfik (2011), Davenport and Arias-Castro (2012), Malloy and Nowak (2012)]

Binary search

- Split measurements into $\log_2 n$ stages
- In each stage, use some of the "sensing energy" to determine if the nonzero is on the "left" or "right" of the active set

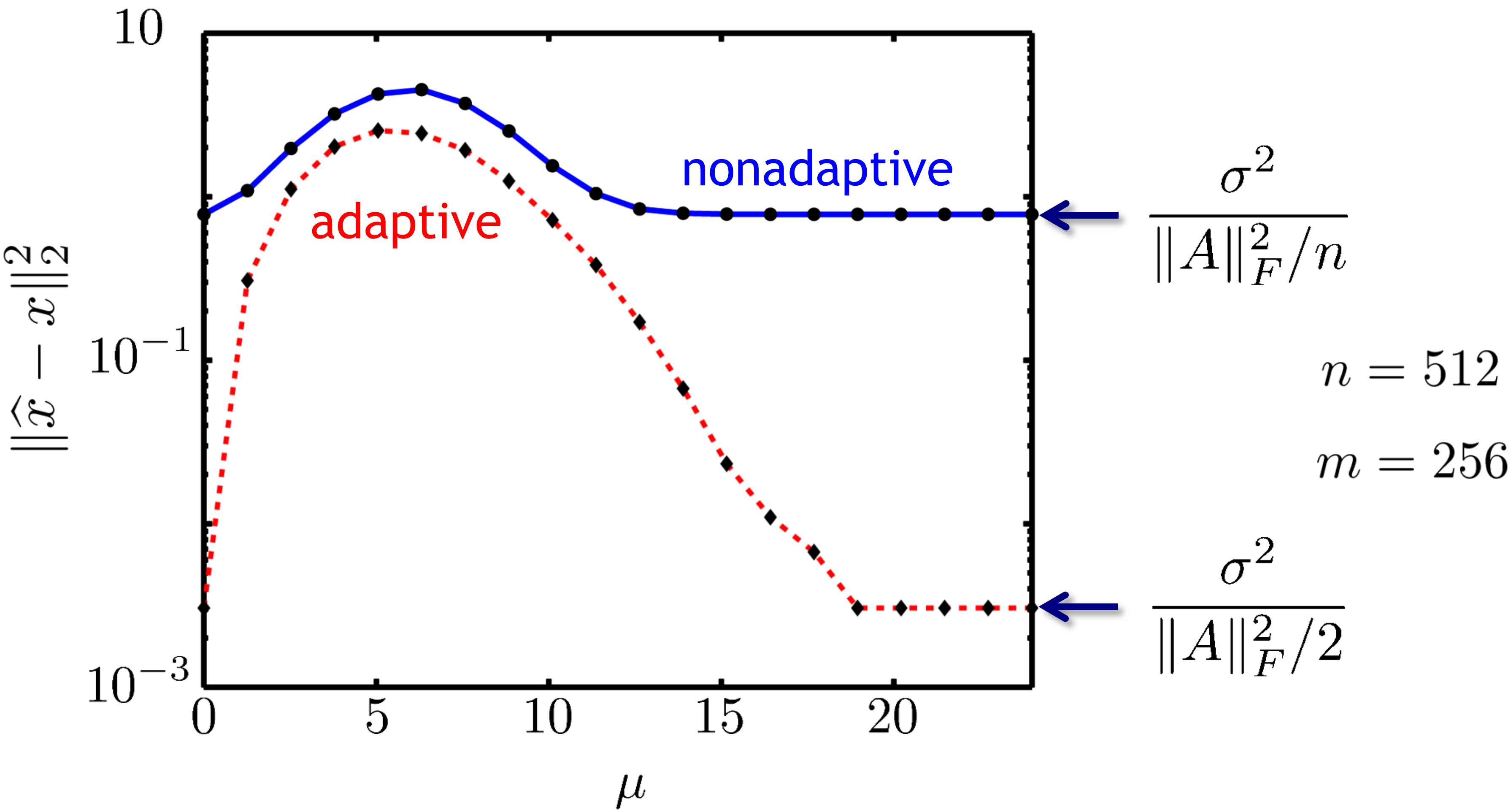
$$A_s \propto \sqrt{s/n}$$



- After subdividing $\log_2 n$ times, return estimated location

[Iwen and Tewfik (2011), Davenport and Arias-Castro (2012), Malloy and Nowak (2012)]

Experimental results



[Arias-Castro, Candès, and Davenport (2013)]

Discussion

- Our lower bound shows that no method can find the location of the nonzero when $\frac{\mu^2}{\sigma^2} \approx n/\|A\|_F^2$
- With careful allocation of the energy budget across the stages, compressive binary search will succeed with probability $1 - \delta$ provided $\frac{\mu^2}{\sigma^2} > 16n\log(\frac{1}{2\delta} + 1)/\|A\|_F^2$
- By randomly splitting the vector into smaller bins and iteratively applying the compressive binary search idea, we can extend this approach to k-sparse signals
- Open questions
  - alternative sparsity models
  - alternative measurement models

Thank You!