# On The Fundamental Limits of Adaptive Sensing

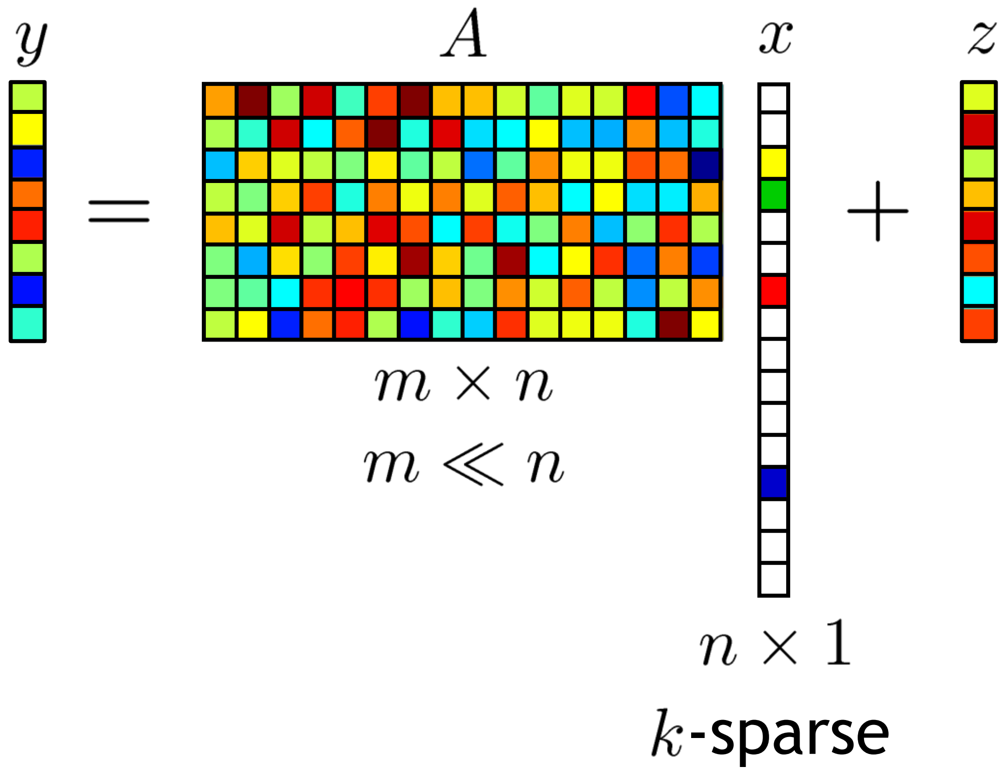
*Mark A. Davenport*

Stanford University

Department of Statistics

**Joint work with Ery Arias-Castro and Emmanuel Candès**

# Compressive Sensing




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

# How Well Can We Estimate $x$ ?

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

Suppose that $A$ has unit-norm rows.

There exist matrices $A$ such that for any $x$ with $\|x\|_0 \leq k$

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

For *any* choice of $A$ and *any* possible recovery algorithm,
there exists an $x$ with $\|x\|_0 \leq k$ such that

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

# Room For Improvement?

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

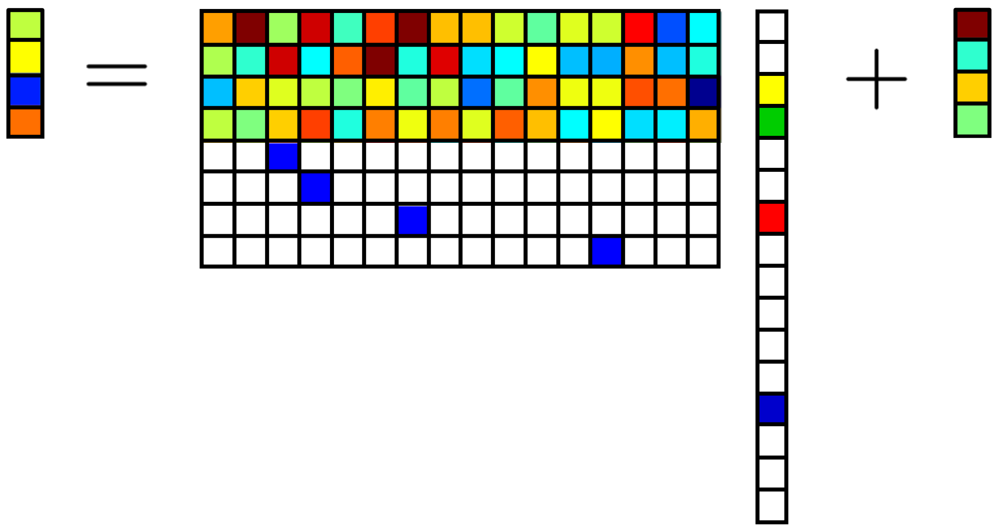
↑

$a_i$ 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…

# Adaptive Sensing

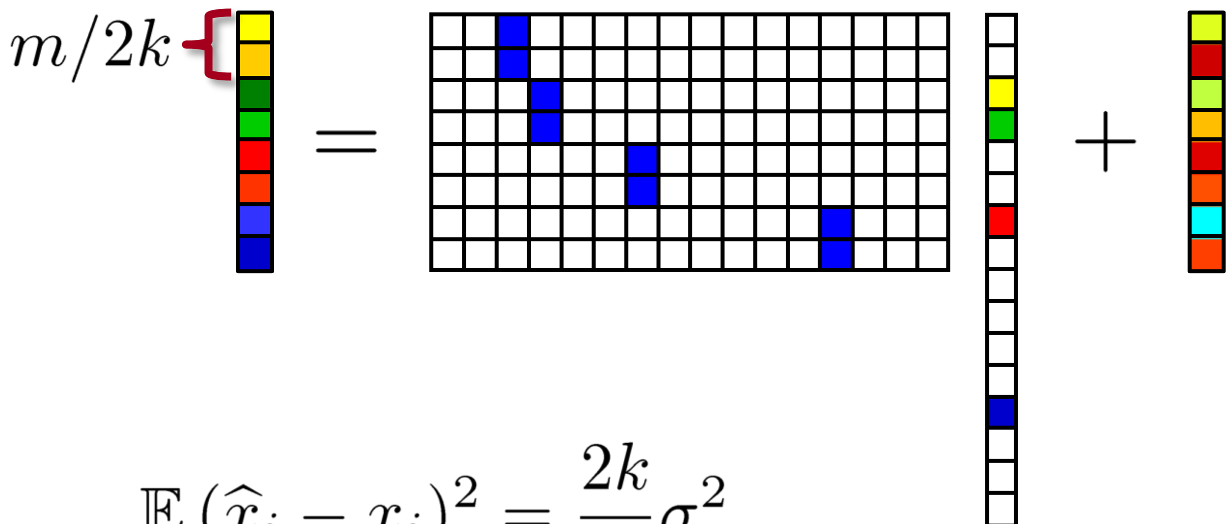
Think of sensing as a game of 20 questions



Simple strategy: Use $m/2$ measurements to find the support, and the remainder to estimate the values.

# Thought Experiment

Suppose that after $m/2$ measurements we have perfectly estimated the support.




$$\mathbb{E}\left(\widehat{x}_i - x_i\right)^2 = \frac{2k}{m}\sigma^2$$

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

# Does Adaptivity *Really* Help?

Sometimes…

- Noise-free measurements, but non-sparse signal
  - adaptivity doesn't help if you want a uniform guarantee
  - probabilistic adaptive algorithms can reduce the required number of measurements from $O(k \log(n/k))$ to $O(k \log \log(n/k))$ [Indyk et al. – 2011]

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

$$\mathbb{E}\,\|\widehat{x} - x\|_2^2 = \frac{n}{m} k\sigma^2$$

$$\mathbb{E}\,\|\widehat{x} - x\|_2^2 = \frac{k}{m} k\sigma^2$$

*Which is it?*

# Which Is It?

Suppose we have a budget of $m$ measurements of the form $y_i = \langle a_i, x \rangle + z_i$ where $\|a_i\|_2 = 1$ and $z_i \sim \mathcal{N}(0, \sigma^2)$

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

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

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

[Arias-Castro, Candès, and Davenport - 2011]

# Proof Strategy

Step 1: Consider sparse signals with nonzeros of amplitude $\mu \approx \sigma\sqrt{n/m}$

Step 2: Show that if given a budget of $m$ measurements, you cannot detect the support very well

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

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

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

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

Plug in $\mu = \frac{8}{3}\sqrt{\frac{n}{m}}$ and this reduces to

$$\mathbb{E}\,\|\widehat{x} - x\|_2^2 \geq \frac{4}{27} \cdot \frac{kn}{m} \geq \frac{1}{7} \cdot \frac{kn}{m}$$

# Key Ideas in Proof of Lemma

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

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

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

$$\sum_j \|\mathbb{P}_{1,j} - \mathbb{P}_{0,j}\|_{\mathrm{TV}}^2 \leq \frac{\mu^2}{4} m \quad \Longrightarrow \quad \mathbb{E}\,|\widehat{S} \Delta S| \geq k \left(1 - \frac{\mu}{2} \sqrt{\frac{m}{n}}\right)$$

# Key Ideas in Proof of Lemma

**Pinsker's Inequality**

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

$$\begin{aligned}\|\mathbb{P}_{1,j}-\mathbb{P}_{0,j}\|_{\mathrm{TV}}^2 &\le \frac{\pi_0}{2}K(\mathbb{P}_{0,j},\mathbb{P}_{1,j}) + \frac{\pi_1}{2}K(\mathbb{P}_{1,j},\mathbb{P}_{0,j}) \\ &\le \frac{\mu^2}{4}\sum_i \mathbb{E}\,a_{i,j}^2\end{aligned}$$

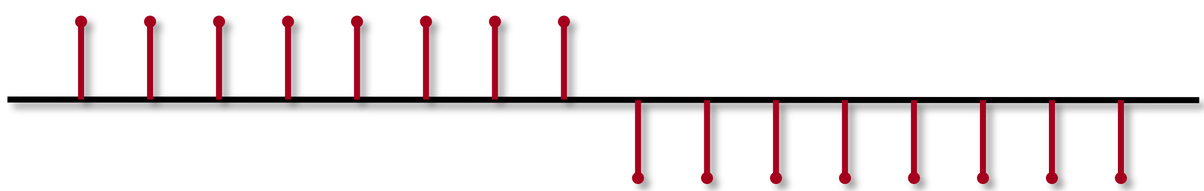
$$\Rightarrow \sum_j \|\mathbb{P}_{1,j}-\mathbb{P}_{0,j}\|_{\mathrm{TV}}^2 \le \frac{\mu^2}{4}\sum_{i,j}\mathbb{E}\,a_{i,j}^2 = \frac{\mu^2}{4}m$$

# Adaptivity In Practice

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

Binary Search [Iwen and Tewfik – 2011, Davenport and Arias-Castro – 2012]

- split measurements into $\log n$ stages
- in each stage, use measurements to decide if the nonzero is in the left or right half of the “active set”
- after subdividing $\log n$ times, return support

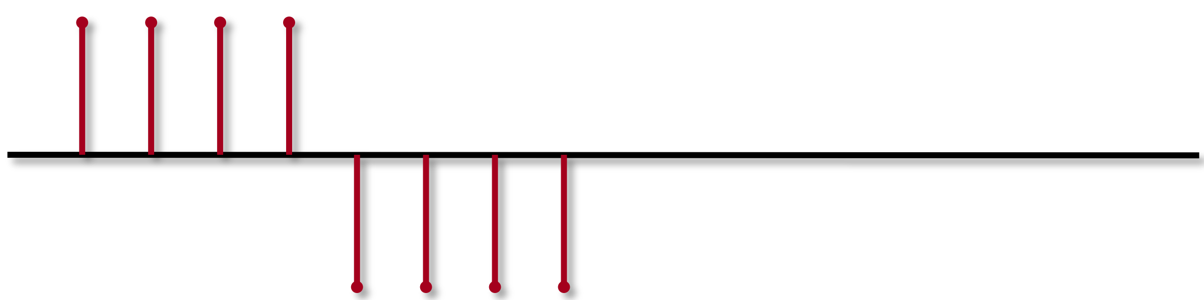


# Adaptivity In Practice

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

Binary Search [Iwen and Tewfik – 2011, Davenport and Arias-Castro – 2012]

- split measurements into $\log n$ stages
- in each stage, use measurements to decide if the nonzero is in the left or right half of the "active set"
- after subdividing $\log n$ times, return support



# Adaptivity In Practice

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

Binary Search [Iwen and Tewfik – 2011, Davenport and Arias-Castro – 2012]

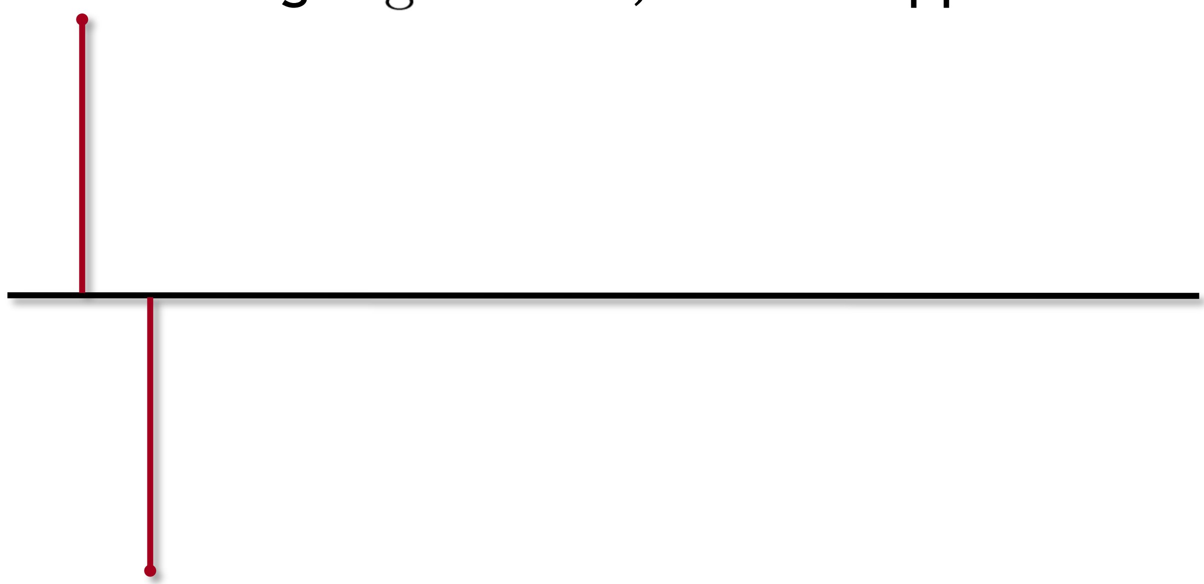
- split measurements into $\log n$ stages
- in each stage, use measurements to decide if the nonzero is in the left or right half of the "active set"
- after subdividing $\log n$ times, return support

# Adaptivity In Practice

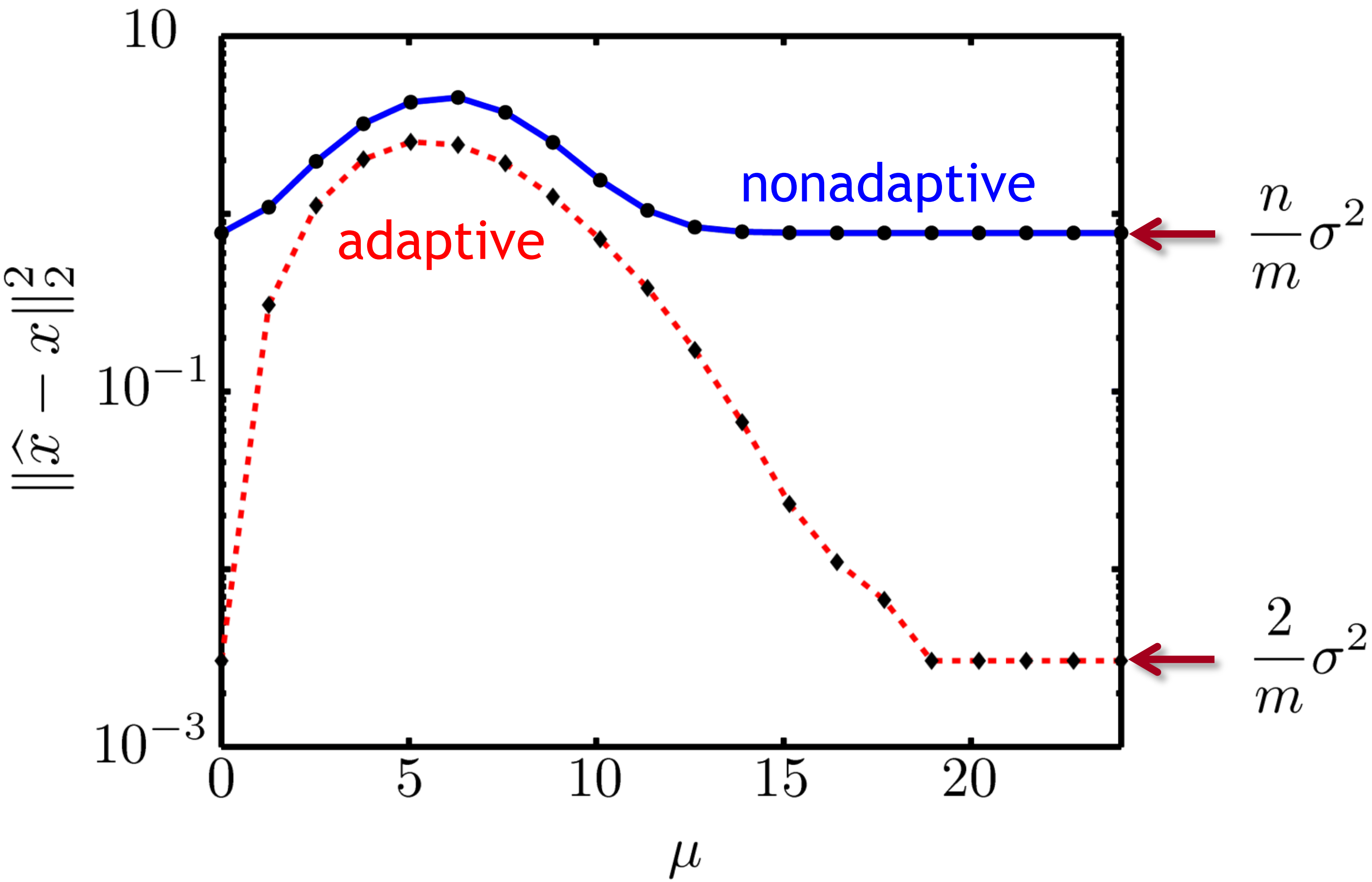
Suppose that $k = 1$ and that $x_{j^*} = \mu$

Binary Search [Iwen and Tewfik – 2011, Davenport and Arias-Castro – 2012]

- split measurements into $\log n$ stages
- in each stage, use measurements to decide if the nonzero is in the left or right half of the “active set”
- after subdividing $\log n$ times, return support



# Experimental Results




$n = 512$

$m = 256$

[Arias-Castro, Candès, and Davenport - 2011]

# Conclusions

- Surprisingly, adaptive algorithms, no matter how complex, cannot in general significantly improve over seemingly naively simple nonadaptive strategies

- Adaptivity might still be very useful in practice
  - how large does $\mu$ need to be to transition from the regime where adaptivity doesn't help to where it does?

$$\frac{\mu}{\sigma} \geq C\sqrt{(n/m)\log\log n}$$

  - improved practical algorithms that work well simultaneously for both large and small values of $\mu$
  - practical architectures and algorithms for implementing adaptive measurements in real-world settings