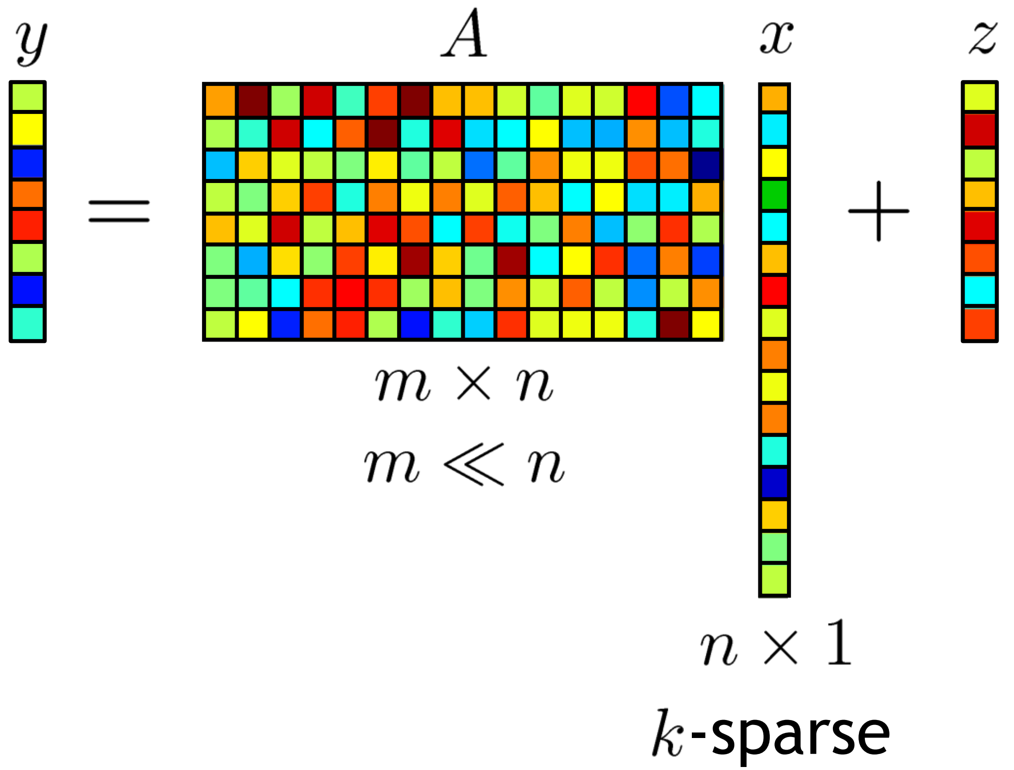
To Adapt or Not To Adapt

The Power and Limits of Adaptive Sensing

Mark A. Davenport

Stanford University
Department of Statistics

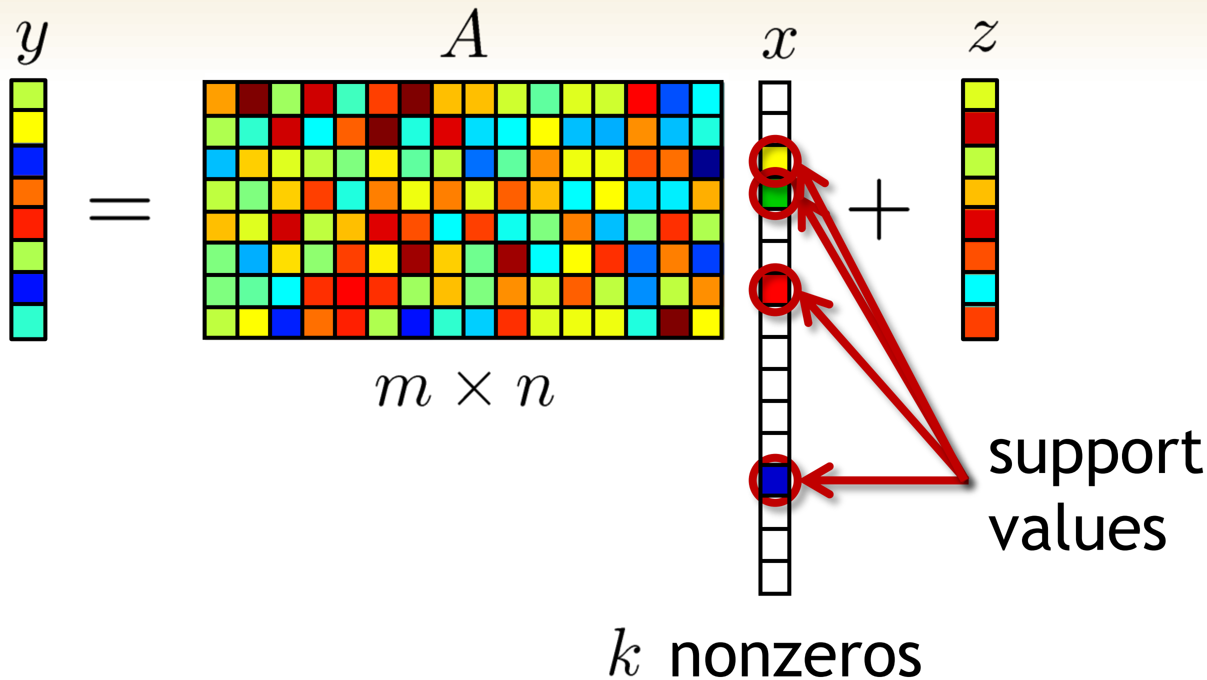
Compressive Sensing



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

Review of Nonadaptive Compressive Sensing

Compressive Sensing



- 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 unit-norm rows
- Pick A at *random!*
  - i.i.d. Gaussian entries (with variance $1/n$)
  - random rows from a unitary matrix
- As long as $m = O(k \log(n/k))$, with high probability a random A will satisfy the *restricted isometry property*
- Deep connections with *Johnson-Lindenstrauss Lemma*
  - see Baraniuk, Davenport, DeVore, and Wakin (2008)

How To Recover x?

- Lots and lots of algorithms
  - ℓ_1-minimization
  - greedy algorithms (matching pursuit, CoSaMP, IHT)

If A satisfies the RIP, $\|x\|_0 \leq k$, and $y = Ax + z$ with $z \sim \mathcal{N}(0, \sigma^2 I)$, then

$$\widehat{x} = \arg\min_{x' \in \mathbb{R}^n} \|x'\|_1$$

$$\text{s.t.} \quad \|A^*(y - Ax')\|_\infty \leq c\sqrt{\log n}\sigma$$

satisfies

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

[Candès and Tao - 2005]

Room For Improvement?

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

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

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

↑

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 (with unit-norm rows) 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

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

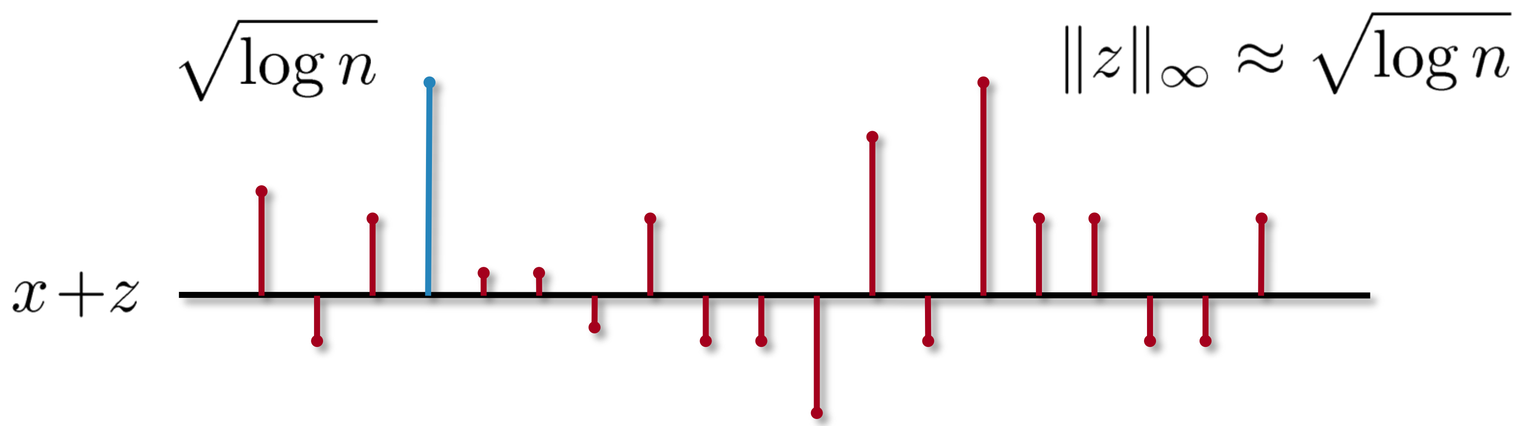
Compressive sensing is already operating at the limit

[Candès and Davenport - 2011]

Intuition

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

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



Proof Recipe

Ingredients (Makes $\sigma^2 = 1$ servings)

- Lemma 1: There exists a set $\mathcal{X}$ of k-sparse vectors such that
  - $|\mathcal{X}| = (n/k)^{k/4}$
  - $\|x_i - x_j\|_2 \geq \frac{1}{2}$ for all $x_i, x_j \in \mathcal{X}$
  - $\|\frac{1}{|\mathcal{X}|}\sum_i x_i x_i^* - \frac{1}{n}I\| \leq \frac{\beta}{n}$ for some $\beta > 0$

- Lemma 2: Define $R^*_{\mathrm{mm}}(A) = \inf_{\widehat{x}} \sup_{\|x\|_0 \leq k} \mathbb{E}\left[\|\widehat{x}(Ax+z) - x\|_2^2\right]$.

  Suppose $\mathcal{X}$ is a set of k-sparse vectors such that $\|x_i - x_j\|_2^2 \geq 8nR^*_{\mathrm{mm}}(A)$ for all $x_i, x_j \in \mathcal{X}$.
  Then $\frac{1}{2}\log|\mathcal{X}| - 1 \leq \frac{1}{2|\mathcal{X}|^2}\sum_{i,j}\|Ax_i - Ax_j\|_2^2$.

Instructions

Combine ingredients and add a dash of linear algebra.

The Details

$$\mu = \frac{1}{|\mathcal{X}|}\sum_i x_i \qquad Q = \frac{1}{|\mathcal{X}|}\sum_i x_i x_i^*$$

$$\begin{aligned}
\frac{k}{4}\log(n/k) - 2 &\leq \frac{1}{|\mathcal{X}|^2}\sum_{i,j} \|Ax_i - Ax_j\|_2^2 \\
&= \mathrm{Tr}\left(A^*A\left(\frac{1}{|\mathcal{X}|^2}\sum_{i,j}(x_i - x_j)(x_i - x_j)^*\right)\right) \\
&= \mathrm{Tr}\left(A^*A\left(2(Q - \mu\mu^*)\right)\right) \\
&\leq 2\mathrm{Tr}\left(A^*AQ\right) \\
&\leq 2\mathrm{Tr}\left(A^*A\right)\|Q\| \\
&\leq 2\|A\|_F^2 \cdot 16 R^*_{\mathrm{mm}}(A)(1+\beta)
\end{aligned}$$

$$\Rightarrow \quad R^*_{\mathrm{mm}}(A) \geq \frac{k\log(n/k)}{128(1+\beta)\|A\|_F^2}$$

Lemma 1

Lemma 1: There exists a set $\mathcal{X}$ of k-sparse points such that

- $|\mathcal{X}| = (n/k)^{k/4}$
- $\|x_i - x_j\|_2 \geq \frac{1}{2}$ for all $x_i, x_j \in \mathcal{X}$
- $\|\frac{1}{|\mathcal{X}|}\sum_i x_i x_i^* - \frac{1}{n}I\| \leq \frac{\beta}{n}$ for some $\beta > 0$

Strategy

Construct $\mathcal{X}$ by sampling (with replacement) from

$$\mathcal{U} = \left\{x \in \{0, \sqrt{1/k}, -\sqrt{1/k}\}^n : \|x\|_0 \leq k\right\}$$

Repeat for $|\mathcal{X}| = (n/k)^{k/4}$ iterations.

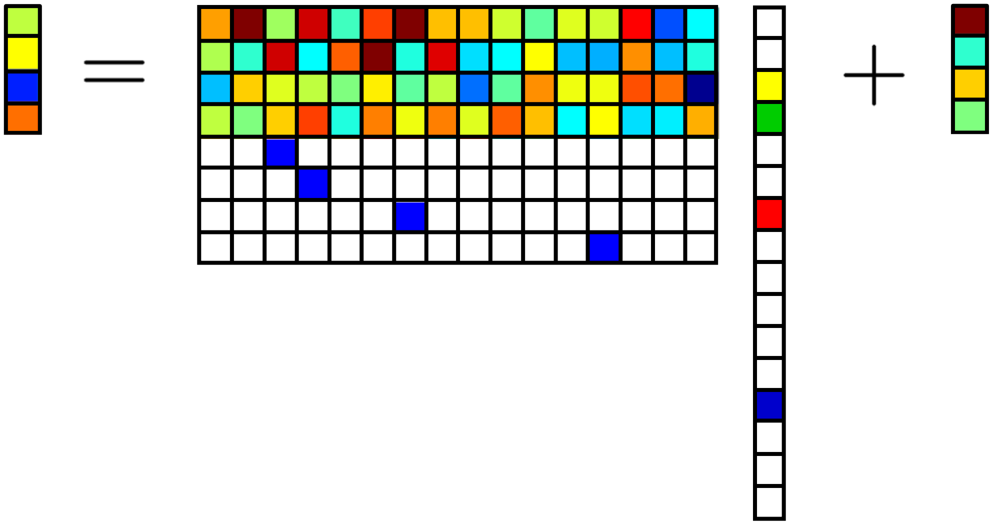
With probability > 0, the remaining properties are satisfied.

Key: *Matrix Bernstein Inequality* [Ahlswede and Winter, 2002]

Adaptive Sensing

Adaptive Sensing

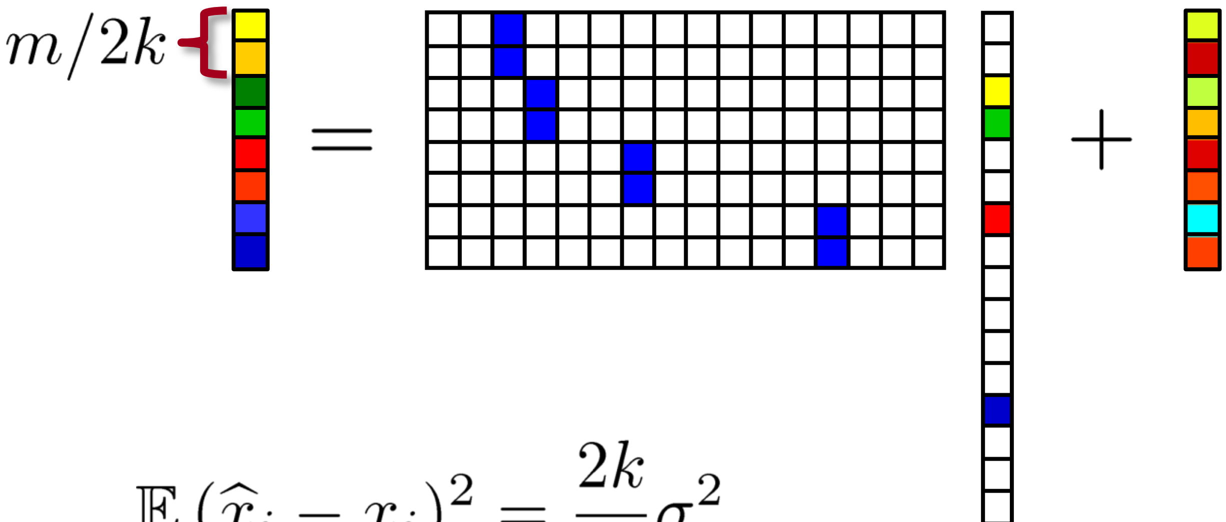
Think of sensing as a game of 20 questions



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

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

Which is it?

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

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 a prior on 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

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

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

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

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

Key Ideas in Proof of Lemma

Pinsker's Inequality

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

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

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

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

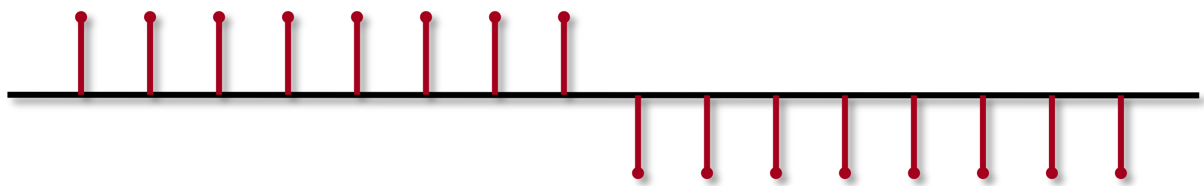
Adaptivity in Practice

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

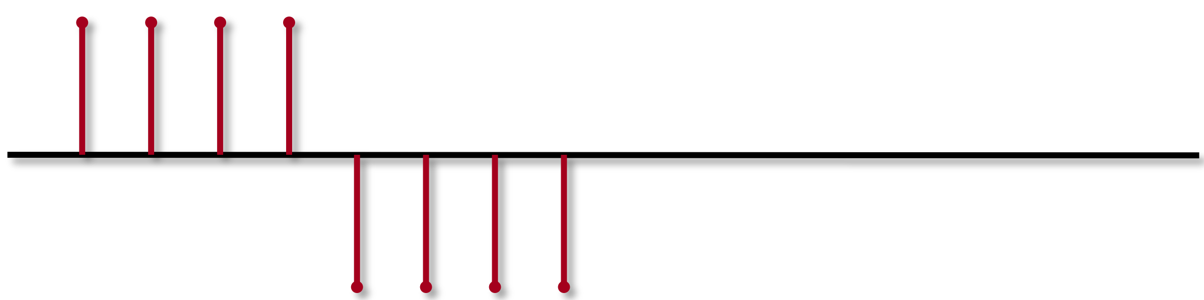


Adaptivity In Practice

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

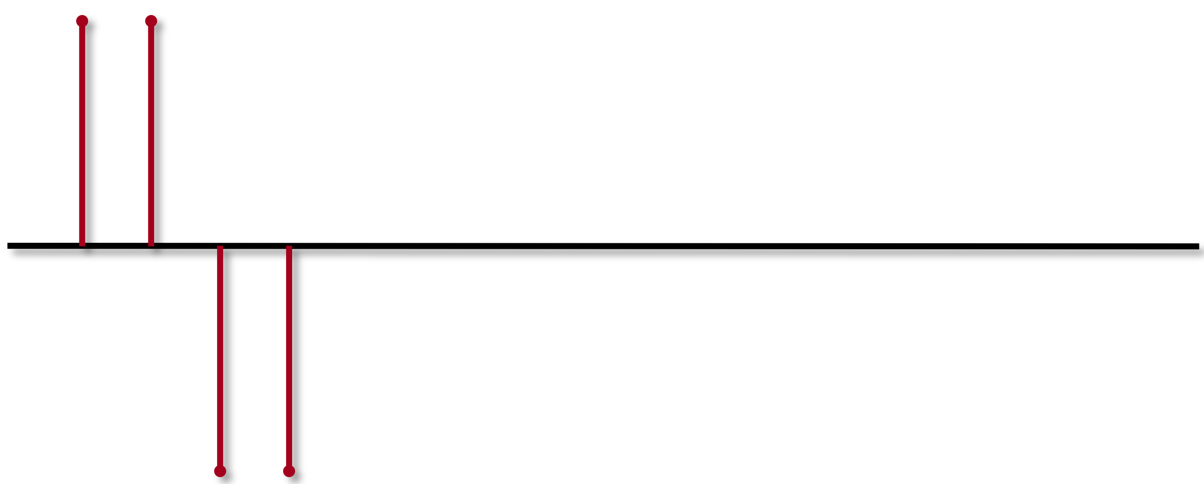


Adaptivity In Practice

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

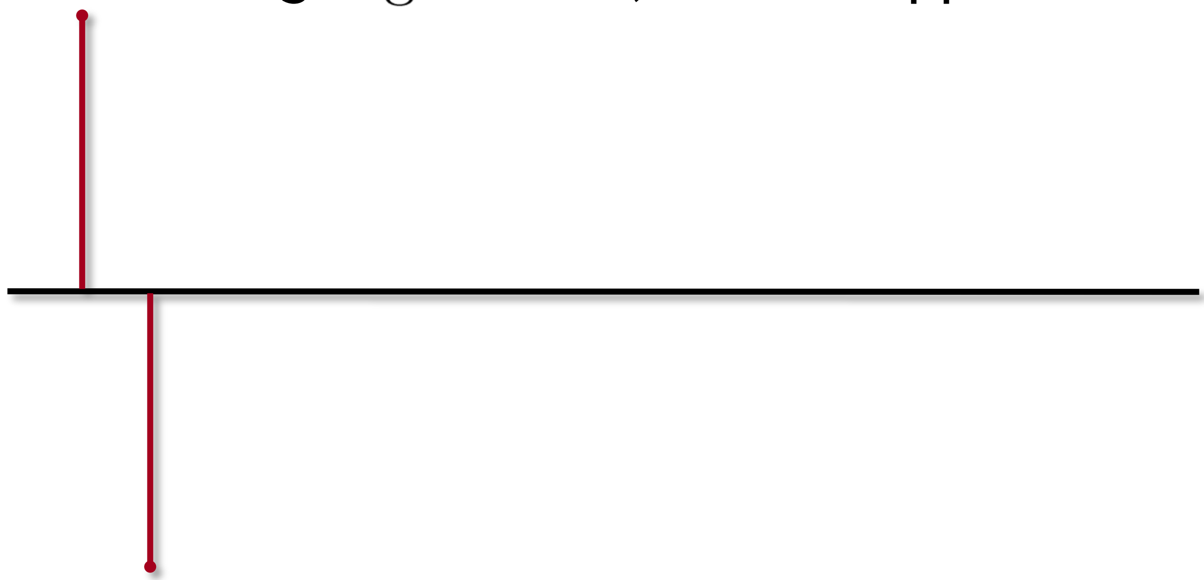


Adaptivity In Practice

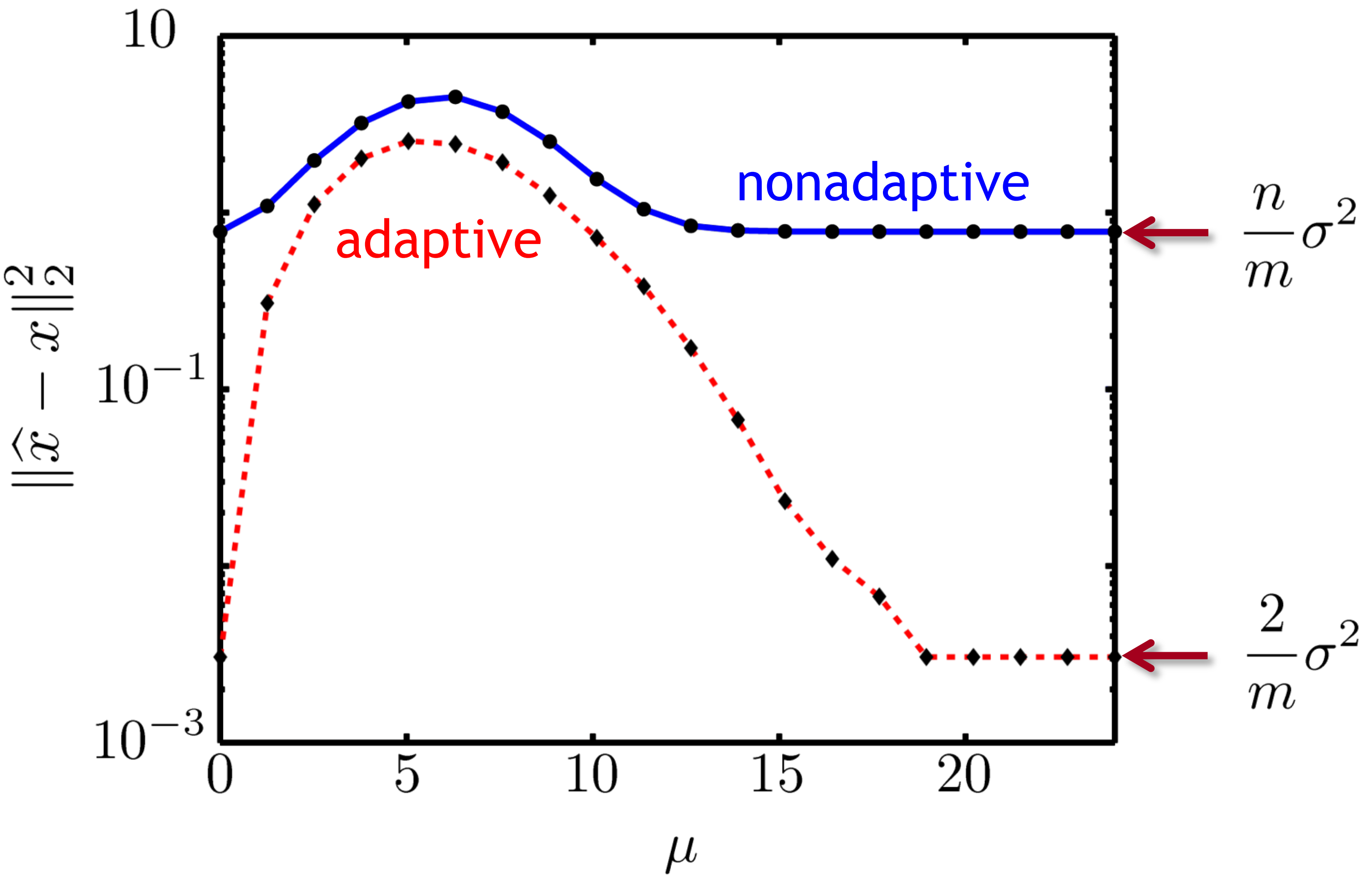
Suppose that $k = 1$ and that $x_{j^*} = \mu$

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]

Open Questions

- No method can succeed when $\frac{\mu}{\sigma} \approx \sqrt{\frac{n}{m}}$, but the binary search approach succeeds as long as $\frac{\mu}{\sigma} \geq C\sqrt{\frac{n}{m} \log \log n}$

  [Davenport and Arias-Castro - 2012]

- Practical algorithms that work well for all values of μ

- Practical algorithms for $k > 1$

- New theory for restricted adaptive measurements
  - single-pixel camera: 0/1 measurements
  - magnetic resonance imaging (MRI): Fourier measurements
  - analog-to-digital converters: linear filter measurements

- New sensors and architectures that can actually acquire adaptive measurements

More Information

http://stat.stanford.edu/~markad

markad@stanford.edu