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 $x$?

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

Suppose that $A$ has unit-norm rows.

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

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

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

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

\[
\mathbb{E} (\hat{x}_i - x_i)^2 = \frac{2k}{m} \sigma^2
\]

\[
\mathbb{E} \| \hat{x} - x \|_2^2 = \frac{2k}{m} k\sigma^2 \ll \frac{n}{m} k\sigma^2 \log n
\]
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

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

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

*Which is it?*
Which Is It?

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

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

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 $\hat{x}$, define $\hat{S} := \{ j : |\hat{x}_j| \geq \mu/2 \}$

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

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

$$\mathbb{E} \|\hat{x} - x\|_2^2 \geq \frac{\mu^2}{4} \mathbb{E} |\hat{S} \Delta S|$$
Proof of Main Result

Lemma
Under the Bernoulli prior, \( \text{any estimate } \hat{S} \) satisfies

\[
\mathbb{E} |\hat{S} \Delta S| \geq k \left( 1 - \frac{\mu}{2} \sqrt{\frac{m}{n}} \right).
\]

Thus,

\[
\mathbb{E} \| \hat{x} - x \|^2 \geq \frac{\mu^2}{4} \mathbb{E} |\hat{S} \Delta S| \\
\geq k \cdot \frac{\mu^2}{4} \left( 1 - \frac{\mu}{2} \sqrt{\frac{m}{n}} \right)
\]

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

\[
\mathbb{E} \| \hat{x} - x \|^2 \geq \frac{4}{27} \cdot \frac{kn}{m} \geq \frac{1}{7} \cdot \frac{kn}{m}
\]
Key Ideas in Proof of Lemma

\[ P_{0,j}(y_1, \ldots, y_m) = P(y_1, \ldots y_m | x_j = 0) \]
\[ P_{1,j}(y_1, \ldots, y_m) = P(y_1, \ldots, y_m | x_j = \mu) \]

\[ \mathbb{E}|\hat{S}\Delta S| \geq \frac{k}{n} \sum_j \left( 1 - \|P_{1,j} - P_{0,j}\|_{TV} \right) \]
\[ \geq k - \frac{k}{\sqrt{n}} \sqrt{\sum_j \|P_{1,j} - P_{0,j}\|_{TV}^2} \]

\[ \sum_j \|P_{1,j} - P_{0,j}\|_{TV}^2 \leq \frac{\mu^2}{4} m \quad \Rightarrow \quad \mathbb{E}|\hat{S}\Delta S| \geq k \left( 1 - \frac{\mu}{2} \sqrt{\frac{m}{n}} \right) \]
Key Ideas in Proof of Lemma

Pinsker’s Inequality

\[ \|P - Q\|_{TV} \leq \sqrt{K(P, Q)/2} \]

\[ \|P_{1,j} - P_{0,j}\|_{TV}^2 \leq \frac{\pi_0}{2} K(P_{0,j}, P_{1,j}) + \frac{\pi_1}{2} K(P_{1,j}, P_{0,j}) \]

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

\[ \sum_j \|P_{1,j} - P_{0,j}\|_{TV}^2 \leq \frac{\mu^2}{4} \sum_{i,j} \mathbb{E} \ a_{i,j}^2 = \frac{\mu^2}{4} m \]
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
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
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

\[ \| \hat{x} - x \|_2 \]

\[ 0 \quad 5 \quad 10 \quad 15 \quad 20 \]

\[ n = 512 \quad m = 256 \]

[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.