Adaptive sensing for compressive imaging

Mark A. Davenport

Georgia Institute of Technology
School of Electrical and Computer Engineering
Compressive sensing

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

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

- 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*
How to recover $x$?

- Lots and lots of algorithms
  - $\ell_1$-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

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

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

satisfies

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

[Candès and Tao (2005)]
Room for improvement?

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

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

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

$a_i$ 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 $\hat{x}$, 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

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

Compressive sensing is already operating at the limit

[Candès and Davenport (2013)]
Suppose that $y = x + z$ with $z \sim \mathcal{N}(0, I)$ and that $k = 1$

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

$$\sqrt{\log n} \quad \| z \|_{\infty} \approx \sqrt{\log n}$$

$x + z$
Adaptive Sensing
Adaptive sensing

Think of sensing as a game of 20 questions

\[
\begin{align*}
\text{Adaptive sensing} & = \text{Game of 20 questions} \\
& + \text{Additional steps}
\end{align*}
\]
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.

\[ m/2k = \begin{array}{c}
\begin{array}{c}
\text{support} \\
\text{estimated}
\end{array}
\end{array} \]

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

\[ \mathbb{E} \|\hat{x} - x\|_2^2 = \frac{2k}{m} k\sigma^2 \ll \frac{n}{m} k\sigma^2 \log n \]
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
    \[
    \mathbb{E} \| \hat{x} - x \|_2^2 = \frac{n}{m} k \sigma^2
    \]
    \[
    \mathbb{E} \| \hat{x} - x \|_2^2 = \frac{k}{m} k \sigma^2
    \]
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 (2013)]
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 \( \hat{x} \), 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,

\[
\mathbb{E} \|\hat{x} - x\|_2^2 \geq \frac{\mu^2}{4} \mathbb{E} |\hat{S} \Delta S| \\
\geq k \cdot \frac{\mu^2}{4} \left(1 - \frac{\mu}{2} \sqrt{\frac{m}{n}}\right)
\]

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

\[
\mathbb{E} \|\hat{x} - x\|_2^2 \geq \frac{4}{27} \cdot \frac{kn}{m} \geq \frac{1}{7} \cdot \frac{kn}{m}
\]
Adaptivity in Practice
Adaptive imaging

[Duarte, Davenport, Takhar, Laska, Sun, Kelly, Baraniuk (2008)]
Incredibly simplified model

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

Our goal is to find $j^*$ and estimate $\mu$

We will assume a fixed budget of time available for sensing
- rather than forcing ourselves to use $m$ equally weighted rows we simply require that the total energy in the (adaptively chosen) sensing matrix is fixed

We will split our “energy budget” into two phases
1. Identify $j^*$ via compressive binary search
2. Estimate the value of $\mu$ by directly sampling it with the remaining sensing energy
Compressive 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

[Coiera and Tewfik (2011), Davenport and Arias-Castro (2012), Malloy and Nowak (2012)]
Experimental results

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

• Our lower bound shows that no method can find the location of the nonzero when \( \frac{\mu}{\sigma} \approx \sqrt{\frac{n}{m}} \)

• With careful allocation of the energy budget across the stages, compressive binary search will succeed with high probability provided \( \frac{\mu}{\sigma} > 4\sqrt{\frac{n}{m}} \)

• By randomly splitting the image into smaller sets and iteratively applying the compressive binary search idea, we can extend this approach to \( k \)-sparse signals

• Open questions
  - noise models for low-light imaging
  - alternative sparsity models
  - alternative measurement models
Thank You!