Compressive Sensing in Noise and the Role of Adaptivity

Mark A. Davenport

Georgia Institute of Technology
School of Electrical and Computer Engineering
Compressive Sensing in Noise

When (and how well) can we estimate $x$ from the measurements $y$?
Nonadaptive Compressive Sensing
Stable Signal Recovery

Given \( y = \Phi x + e \), find \( x \)

Typical (worst-case) guarantee: If \( \Phi \) satisfies the RIP

\[
\| \hat{x} - x \|_2^2 \leq C \| e \|_2^2
\]

Even if \( \Lambda = \text{supp}(x) \) is provided by an oracle, the error can still be as large as \( \| \hat{x} - x \|_2^2 = \| e \|_2^2 / (1 - \delta) \).
Suppose now that $\Phi$ satisfies

$$A(1 - \delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq A(1 + \delta)\|x\|_2^2 \quad \|x\|_0 \leq 2S$$

In this case our guarantee becomes

$$\|\hat{x} - x\|_2^2 \leq \frac{C}{A} \|e\|_2^2$$

Unit-norm rows

$$\|\hat{x} - x\|_2^2 \leq C\left(\frac{N}{M}\right)\|e\|_2^2$$
Expected Performance

- Worst-case bounds can be pessimistic

- What about the average error?
  - Assume $e$ is white noise with variance $\sigma^2$
    \[
    \mathbb{E} \left( \|e\|^2 \right) = M \sigma^2
    \]
  - For oracle-assisted estimator
    \[
    \mathbb{E} \left( \|\hat{x} - x\|^2 \right) \leq \frac{S \sigma^2}{A(1 - \delta)}
    \]
  - If $e$ is Gaussian, then for $\ell_1$-minimization
    \[
    \mathbb{E} \left( \|\hat{x} - x\|^2 \right) \leq \frac{C'}{A} S \sigma^2 \log N
    \]
White Signal Noise

What if our signal $x$ is contaminated with noise?

$$y = \Phi (x + n) = \Phi x + \Phi n$$

Suppose $\Phi$ has orthogonal rows with norm equal to $\sqrt{B}$. If $n$ is white noise with variance $\sigma^2$, then $\Phi n$ is white noise with variance $B\sigma^2$.

$$\mathbb{E} \left[ \| \hat{x} - x \|_2^2 \right] \leq C' \frac{B}{A} S\sigma^2 \log N$$
White Signal Noise

What if our signal $x$ is contaminated with noise?

\[ y = \Phi(x + n) = \Phi x + \Phi n \]

Suppose \( \Phi \) has orthogonal rows with norm equal to \( \sqrt{B} \). If \( n \) is white noise with variance \( \sigma^2 \), then \( \Phi n \) is white noise with variance \( B\sigma^2 \).

\[
\mathbb{E} \left[ \| \hat{x} - x \|_2^2 \right] \leq C' \frac{N}{M} S\sigma^2 \log N
\]

\[ \text{SNR} = 10 \log_{10} \left( \frac{\| x \|^2_2}{\| \hat{x} - x \|^2_2} \right) \quad \text{3dB loss per octave of subsampling} \]
Noise Folding

\[ \log_2 \left( \frac{N}{M} \right) \]

**SNR (dB)**

- 3dB per octave
- Oracle CS
- CoSaMP CS

[Davenport, Laska, Treichler, Baraniuk - 2011]
There exists matrices $\Phi$ (with unit-norm rows) such that for any (sparse) $x$ we have

$$\mathbb{E} \| \hat{x} - x \|_2^2 \leq C \frac{N}{M} S\sigma^2 \log N.$$ 

$$y_i = \langle \phi_i, x \rangle + e_i$$

$\phi_i$ 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?

Via a better choice of $\Phi$? Via a better recovery algorithm?

If $y = \Phi x + e$ with $e \sim \mathcal{N}(0, \sigma^2 I)$, then there exists an $x$ such that for any $\hat{x}$ and any $\Phi$

$$
\mathbb{E} \left[ \| \hat{x}(\Phi x + e) - x \|_2^2 \right] \geq C \frac{N}{\| \Phi \|_F^2} S \sigma^2 \log(N/S).
$$

If $y = \Phi(x + n)$ with $n \sim \mathcal{N}(0, \sigma^2 I)$, then there exists an $x$ such that for any $\hat{x}$ and any $\Phi$

$$
\mathbb{E} \left[ \| \hat{x}(\Phi(x + n)) - x \|_2^2 \right] \geq C \frac{N}{M} S \sigma^2 \log(N/S).
$$

$\Phi = U\Sigma V^*$, $y' = \Sigma^{-1} U^* y = V^* x + V^* n$, $\| V^* \|_F^2 = M$

[Candès and Davenport - 2011]
Suppose that $y = x + n$ with $n \sim \mathcal{N}(0, I)$ and that $S = 1$

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

$\sqrt{\log N}$

$\| n \|_\infty \approx \sqrt{\log N}$

$x + n$
Proof Recipe

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

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

- Lemma 2: Define $R_{mm}^*(\Phi) = \inf_{\hat{x}} \sup_{\|x\|_0 \leq S} \mathbb{E} \left[ \|\hat{x}(\Phi x + e) - x\|_2^2 \right]$.

  Suppose $\mathcal{X}$ is a set of $S$-sparse vectors such that
  $\|x_i - x_j\|_2^2 \geq 8N R_{mm}^*(\Phi)$ for all $x_i, x_j \in \mathcal{X}$.

  Then $\frac{1}{2} \log |\mathcal{X}| - 1 \leq \frac{1}{2|\mathcal{X}|^2} \sum_{i,j} \|\Phi x_i - \Phi x_j\|_2^2$.

Instructions

Combine ingredients and add a dash of linear algebra.
\[ \mu = \frac{1}{|x|} \sum_i x_i \quad Q = \frac{1}{|x|} \sum_i x_i x_i^* \]

\[
\frac{S}{4} \log(N/S) - 2 \leq \frac{1}{|x|^2} \sum_{i,j} \| \Phi x_i - \Phi x_j \|_2^2 \\
= \text{Tr} \left( \Phi^* \Phi \left( \frac{1}{|x|^2} \sum_{i,j} (x_i - x_j)(x_i - x_j)^* \right) \right) \\
= \text{Tr} \left( \Phi^* \Phi \left( 2(Q - \mu \mu^*) \right) \right) \\
\leq 2 \text{Tr} (\Phi^* \Phi Q) \\
\leq 2 \text{Tr} (\Phi^* \Phi) \| Q \| \\
\leq 2 \| \Phi \|_F^2 \cdot 16 R^*_\text{mm}(\Phi)(1 + \beta)
\]

\[ R^*_\text{mm}(\Phi) \geq \frac{S \log(N/S)}{128(1 + \beta)\| \Phi \|_F^2} \]
Lemma 1

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

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

**Strategy**

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

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

Repeat for $|\mathcal{X}| = (N/S)^{S/4}$ 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.

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

\[
\mathbb{E} \|\hat{x} - x\|_2^2 = \frac{2S}{M} S \sigma^2 \ll \frac{N}{M} S \sigma^2 \log N
\]
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 \( O(S \log(N/S)) \) to \( O(S \log \log(N/S)) \) [Indyk et al. - 2011]

- **Noisy setting**
  - distilled sensing [Haupt et al. - 2007, 2010]
  - adaptivity can reduce the estimation error to
    \[
    \mathbb{E} \| \hat{x} - x \|_2^2 = \frac{N}{M} S \sigma^2
    \]
    \[
    \mathbb{E} \| \hat{x} - x \|_2^2 = \frac{S}{M} S \sigma^2
    \]
Suppose we have a budget of $M$ measurements of the form $y_i = \langle \phi_i, x \rangle + e_i$ where $\|\phi_i\|_2 = 1$ and $e_i \sim \mathcal{N}(0, \sigma^2)$.

The vector $\phi_i$ can have an arbitrary dependence on the measurement history, i.e., $(\phi_1, y_1), \ldots, (\phi_{i-1}, y_{i-1})$.

**Theorem**

There exist $x$ with $\|x\|_0 \leq S$ such that for any adaptive measurement strategy and any recovery procedure $\hat{x}$,

$$\mathbb{E} \|\hat{x}(y) - x\|_2^2 \geq C \frac{N}{M} S \sigma^2.$$

Thus, in general, adaptivity does not seem to 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 \( S \)-sparse prior:

\[
x_j = \begin{cases} 
0 & \text{with probability } 1 - S/N \\
\mu > 0 & \text{with probability } S/N
\end{cases}
\]
Proof of Main Result

Let \( T = \{ j : x_j \neq 0 \} \) and set \( \sigma^2 = 1 \).

For any estimator \( \hat{x} \), define \( \hat{T} := \{ j : |\hat{x}_j| \geq \mu/2 \} \).

Whenever \( j \in T \setminus \hat{T} \) or \( j \in \hat{T} \setminus T \), \( |\hat{x}_j - x_j| \geq \mu/2 \).

\[
\|\hat{x} - x\|_2^2 \geq \frac{\mu^2}{4} |T \setminus \hat{T}| + \frac{\mu^2}{4} |\hat{T} \setminus T| = \frac{\mu^2}{4} |\hat{T} \Delta T|
\]

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

Lemma
Under the Bernoulli prior, any estimate $\hat{T}$ satisfies

$$\mathbb{E} |\hat{T}\Delta T| \geq S \left(1 - \frac{\mu}{2} \sqrt{\frac{M}{N}}\right).$$

Thus,

$$\mathbb{E} \|\hat{x} - x\|_2^2 \geq \frac{\mu^2}{4} \mathbb{E} |\hat{T}\Delta T|$$

$$\geq S \cdot \frac{\mu^2}{4} \left(1 - \frac{\mu}{2} \sqrt{\frac{M}{N}}\right)$$

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

$$\mathbb{E} \|\hat{x} - x\|_2^2 \geq \frac{4}{27} \cdot \frac{SN}{M} \geq \frac{1}{7} \cdot \frac{SN}{M}$$
Key Ideas in Proof of Lemma

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

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

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

**Pinsker’s Inequality**

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

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

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

\[ \sum_j \|\mathbb{P}_{1,j} - \mathbb{P}_{0,j}\|_{TV}^2 \leq \frac{\mu^2}{4} \sum_{i,j} \mathbb{E} \phi_{i,j}^2 = \frac{\mu^2}{4} M \]
Adaptivity In Practice

Suppose that $S = 1$ 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 $S = 1$ 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 $S = 1$ 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 $S = 1$ 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

[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 \( \frac{\mu}{\sigma} \geq C \sqrt{\frac{N}{M}} \)  
  [Davenport and Arias-Castro; Malloy and Nowak - 2012]

- Practical algorithms that work well for all values of \( \mu \)

- Optimal algorithms for \( S > 1 \)

- 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