Compressed Sensing

Compressive operator
\[ \mathbb{R}^n \leftrightarrow \mathbb{R}^m \]
\[ m \ll n \]

Emphases
1. reduced number of samples
2. non-adaptive sensing operator

\[ x \in \mathbb{R}^n \]
\[ y \in \mathbb{R}^m \]
Alternative

Goal: optimize sensing with a fixed energy/latency budget (rather than a sample budget)

Emphases

1. coping with noise
2. non-adaptive or adaptive?
Related Work and Motivation

Adaptive Compressed Sensing


Adaptive Multi-look Sensing

Detection/Estimation of Sparse Signals

How reliably can we infer sparse patterns?

fMRI  Genomics  Astrophysics
Now you see it, now you don’t!

\[
X = \text{sparse signal} + \text{noise}
\]

\(n \times 1\) vector with \(n^{1-\beta}\), \(0 < \beta < 1\), non-zero entries of magnitude \(\mu > 0\). Can the sparsity pattern be reliably perceived in presence of noise?

**Non-adaptive sensing:** Yes, iff \(\mu > \sqrt{2\beta \log n}\) (DDJ03).

**Adaptive sensing:** Yes, if \(\mu > a_n\), for any \(a_n \to \infty\).

Weak signals/patterns are perceptible only through adaptation!
Sparse Signal Model

Let \( \boldsymbol{\mu} = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n \) be a sparse vector

\[
\mu_i = \begin{cases} 
0 & i \in I_0 \\
\mu^* & i \in I_S
\end{cases}
\]

, where \( |I_S| \ll |I_0| \)

Example:

[Graph showing a sparse signal model]

In this talk we will assume \( \mu^* > 0 \).
Noisy Observation Model

\[ X_i = \mu_i + Z_i, \quad i \in \{0, \ldots, n\}, \]
where \( Z_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1) \)

Suppose we want to locate just one signal component:

\[ \hat{i} = \arg \max_i X_i \]

Because of noise, \( \max_i X_i \sim \sqrt{2 \log n} \) even if no signal is present

**Threshold of Perception:** Smallest \( \mu^*_i > 0 \) such that sparsity pattern can be reliable recovered from noisy observation.
Sparse Support Recovery

When testing a large number of hypotheses simultaneously we are bound to make errors.

Two approaches to statistically reliability:

- Non-adaptive threshold to control the probability of perfect localization (Bonferroni correction) – very conservative

- Adaptive threshold to control of the relative proportion of errors (Benjamini & Hochberg ’95)
False Discovery Rate Control

Recall the definition of the **signal support set**

\[ I_S = \{ i : \mu_i \neq 0 \} \]

**Goal:** Estimate the support as well as possible. Let \( \hat{I}_S(X) \) be the outcome of a support estimation procedure.

**False Discovery Proportion**

\[ \text{FDP} = \frac{I_S \setminus \hat{I}_S(X)}{|I_S(X)|} = \frac{\text{# falsely discovered components}}{\text{# discovered components}} \]

**Non Discovery Proportion**

\[ \text{NDP} = \frac{I_S \setminus \hat{I}_S(X)}{|I_S|} = \frac{\text{# missed components}}{\text{# true components}} \]

Desirable situation: \( \text{FDP}, \text{NDP} \approx 0 \)

Since \( n \) is typically very large it makes sense to study **asymptotic** performance, as \( n! \to 1 \).
Non-adaptive Sensing (Jin & Donoho ’03)

Assume the signal is very sparse:

\[ |I_s| = n^{1-\beta}, \text{ where } \beta \in (0, 1). \]

Number of signal components

Example: \(\beta = 3/4\)

- \(n = 10000\) \(\Rightarrow\) \(|I_s| = 10\)
- \(n = 1000000\) \(\Rightarrow\) \(|I_s| = 32\)

**Theorem:** If \(\mu^* > \sqrt{2\beta \log n}\) then BH thresholding applied to \(X\) drives both FDP and NDP to zero with probability tending to one as \(n \to \infty\). Conversely if \(\mu^* < \sqrt{2\beta \log n}\) no procedure can control simultaneously FDP and NDP.
Threshold of Perceptibility (Jin & Donoho ‘03)

\[ \mu^* = \sqrt{2r \log n} \]

These **asymptotic** results tell us how strong the signals need to be for sparsity pattern recovery.
A Generalization of the Sensing Model

Allow multiple observations $j = 0, 1, \ldots, k$

$$X_i^{(j)} = \phi_i^{(j)} \mu_i + Z_i^{(j)}, \quad i \in \{1, \ldots, n\}$$

where $Z_i^{(j)} \sim \mathcal{N}(0, 1)$

...subject to a sampling energy budget

$$\sum_{j=1}^{k} \sum_{i=1}^{n} (\phi_i^{(j)})^2 \leq n$$

$$\phi^{(j)} = (\phi_1^{(j)}, \ldots, \phi_n^{(j)})$$ are called the sensing vectors.

(Note: in the previous work a single observation was considered, where $\phi_i^{(0)} = 1, \quad i \in \{1, \ldots, n\}$)
Distilled Sensing

Proceeding in this fashion, gradually \textbf{focus} on the signal subspace
Distilled Sensing Example

- Original signal (~0.8% non-zero components)
- Noisy version of the image (\(k=0\))
- Noisy versions of the image (\(k=5\))
- Non-adaptive NDP
- Adaptive FDP
- Distilled sensing recovery (FDR = 0.01)
- Passive sensing recovery (FDR = 0.01)
Enhanced Sensitivity through Adaptivity

**Theorem 1 (J. Haupt, R. Castro and RN '08)**

Consider a \((k + 1)\)-step approach: at each step retain only the non-negative elements and sense at only those locations in next step (allocate equal fraction of sensing energy to each step). Then if

\[
\mu^* > \sqrt{2 \beta \left( \frac{k + 1}{2^k} \right)} \log n
\]

the BH thresholding procedure applied to \(X^{(k+1)}\) drives both the FDP and the NDP to zero with probability tending to one as \(n \to \infty\).

Furthermore if one does not allow adaptive sensing, then the previous results (equivalent to \(k=0\)) cannot be improved.
Proof Sketch (Theorem 1)

Main Idea: Quantify the effect of distillation $j$:

$\mathcal{m}(j) = |I_S \cap I(j)|$ - Retained true signal locations.

$\mathcal{\ell}(j) = |I_0 \cap I(j)|$ - Retained true non-signal locations.

Lemma:

$$\left(1 - \frac{1}{\log n}\right) \mathcal{m}(j) \leq \mathcal{m}(j+1) \leq \mathcal{m}(j)$$

$$\left(\frac{1}{2} - \frac{1}{\log n}\right) \mathcal{\ell}(j) \leq \mathcal{\ell}(j+1) \leq \left(\frac{1}{2} + \frac{1}{\log n}\right) \mathcal{\ell}(j)$$

with probability tending to one as $n \to \infty$

With high probability each distillation keeps almost all the non-zero components and rejects about half of the non-signal components.

Energy in signal subspace doubles at each step.
Thresholds of Perceptibility

\[ \mu^* = \sqrt{2r \log n} \]

These results suggest we might be able to estimate signal with amplitudes growing slower than \( \mu^* \sim \sqrt{\log n} \).
Universal Perceptibility

**Theorem 2** (J. Haupt, R. Castro and RN ‘09)

Consider a \((k + 1)\)-step approach (with geometrically decaying energy allocation), where \(k(n) = \log_2 \log n\). If

\[ \mu^* > a_n \]

where \(a_n \to \infty\), then a thresholding procedure applied to \(X^{(k(n)+1)}\) drives both the FDP and the NDP to zero with probability tending to one as \(n \to \infty\).

**Corollary**

If the sensing energy grows sublinearly like \(\frac{a_n}{\log n} n\), for any \(a_n \to \infty\), then so long as \(\mu^* > \sqrt{2 \beta \log n}\) the BH thresholding applied to \(X^{(k(n)+1)}\) drives the FDP and NDP to zero as \(n \to \infty\).
Now you see it, now you don’t!

\[ X = \text{sparse signal} + \text{noise} \]


**Non-adaptive sensing:** Yes, iff \( \mu > \sqrt{2\beta \log n} \).

**Adaptive sensing:** Yes, if \( \mu > a_n \), for any \( a_n \to \infty \).

Weak signals/patterns are imperceptible without adaptive sensing!