Compressive Sensing
Part III: Compressive Sensing in Practice

Mark A. Davenport
Stanford University
Department of Statistics
Important Practical Challenges

• Noise!
  - noisy measurements
  - noisy signals
  - interference

• Quantization
  - quantization error
  - saturation effects

• Good signal models
  - is sparsity sometimes not enough?
  - what dictionaries should we use in practice?
Measurement and Signal Noise
Sparse Signal Recovery

- Optimization / $\ell_1$-minimization
- Greedy algorithms
  - matching pursuit
  - orthogonal matching pursuit (OMP)
  - regularized OMP
  - CoSaMP, Subspace Pursuit, IHT, ...
Exact Recovery

If we can determine $\Lambda = \text{supp}(x)$, then the problem becomes over-determined.

In the absence of noise,

$$\Phi_{\Lambda}^\dagger y = (\Phi_{\Lambda}^T \Phi_{\Lambda})^{-1} \Phi_{\Lambda}^T y$$

$$= (\Phi_{\Lambda}^T \Phi_{\Lambda})^{-1} \Phi_{\Lambda}^T \Phi_{\Lambda} x$$

$$= x$$
Signal Recovery in Noise

Given \[ y = \Phi x + e \]

find \( x \)

- Optimization-based methods
  - basis pursuit, basis pursuit de-noising, Dantzig selector

\[
\hat{x} = \arg\min_{x \in \mathbb{R}^N} \|x\|_1 \\
\text{s.t. } \|y - \Phi x\|_2 \leq \epsilon
\]

- Greedy/Iterative algorithms
  - OMP, StOMP, ROMP, CoSaMP, Thresh, SP, IHT, ...
Stable Signal Recovery

Suppose that we observe \( y = \Phi x + e \) and that \( \Phi \) satisfies the RIP of order \( S \).

Typical (worst-case) guarantee

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

Even if \( \Lambda = \text{supp}(x) \) is provided by an oracle, the error can still be as large as

\[
\| \hat{x} - x \|_2 = \frac{\| e \|_2}{1 - \delta}
\]
Expected Performance

- Worst-case bounds can be pessimistic

- What about the average error?
  - assume $e$ is white noise with variance $\sigma^2$

$$\mathbb{E} \left( \|e\|_2^2 \right) = M \sigma^2$$

- for oracle-assisted estimator

$$\mathbb{E} \left( \|\hat{x} - x\|_2 \right) \leq \frac{S \sigma^2}{1 - \delta}$$

- if $e$ is Gaussian, then for $\ell_1$-minimization

$$\mathbb{E} \left( \|\hat{x} - x\|_2^2 \right) \leq CS \sigma^2 \log N$$
What if our signal $x$ is contaminated with noise?

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

Suppose $\Phi$ satisfies the RIP and has orthogonal and equal-norm rows. If $n$ is white noise with variance $\sigma^2$, then $\Phi n$ is white noise with variance $\sigma^2 \frac{N}{M}$.

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

$$\text{SNR} = 10 \log_{10} \left( \frac{\|x\|_2^2}{\|\hat{x} - x\|_2^2} \right)$$

3dB loss per octave of subsampling
Noise Folding

SNR (dB)

$\log_2 (N/M)$

[D, Laska, Treichler, and Baraniuk - 2011]
Can We Do Better?

- Better choice of $\Phi$?
- Better recovery algorithm?

If we knew the support of $x$ \textit{a priori}, then we could achieve

$$\|\hat{x} - x\|_2^2 \approx \frac{S}{M} S\sigma^2 \ll C' \frac{N}{M} S\sigma^2 \log N$$

Is there any way to match this performance without knowing the support of $x$ in advance?

$$R^*_{mm}(\Phi) = \inf_{\hat{x}} \sup_{\|x\|_0 \leq S} \mathbb{E} \left[ \|\hat{x}(y) - x\|_2^2 \right]$$
No!

Theorem:

If $y = \Phi x + e$ with $e \sim \mathcal{N}(0, \sigma^2 I)$, then

$$R_{mm}^*(\Phi) \geq C \frac{N}{\|\Phi\|_F^2} S\sigma^2 \log(N/S).$$

If $y = \Phi(x + n)$ with $n \sim \mathcal{N}(0, \sigma^2 I)$, then

$$R_{mm}^*(\Phi) \geq C \frac{N}{M} S\sigma^2 \log(N/S).$$

Ingredients in proof:

- Fano’s inequality
- Random construction of packing set of sparse points
- Matrix Bernstein inequality to bound empirical covariance matrix of packing set

[Candès and D - 2011]
Interference

$$y = \Phi x + e$$

- What if $e$ represents corruption or *structured noise*, rather than Gaussian noise or arbitrary perturbations?

- Structured signal noise:
  $$y = \Phi x_S + \Phi x_I$$

- Structured measurement noise:
  $$y = \Phi x + \Omega e$$
Suppose \( x = x_S + x_I \) where \( x_S \) is sparse with unknown support and \( x_I \) is sparse with known support \( \Lambda \).

**Goal:** Design an \( M \times M \) matrix \( P \) such that

\[
\| P(\Phi x_I) \|_2 \approx 0
\]

\[
\| P(\Phi x_S) \|_2 \approx \| \Phi x_S \|_2
\]

\[
P = I - \Phi_\Lambda \Phi_\Lambda^\dagger
\]

Projection onto \( \mathcal{R}(\Phi_\Lambda) \)

\[
P \Phi_\Lambda = 0
\]
Lemma:
If $\Phi$ satisfies the RIP of order $S$, then

$$\left(1 - \frac{\delta}{1 - \delta}\right) \|x\|_2^2 \leq \|P\Phi x\|_2^2 \leq (1 + \delta) \|x\|_2^2$$

provided that $\|x\|_0 \leq S - |\Lambda|$ and $\text{supp}(x) \cap \Lambda = \emptyset$. 

[D, Boufounos, Wakin, and Baraniuk - 2010]
Interference Cancellation in Action

[Graph showing the effect of $K_1/K_S$ on Recovered SNR for different methods: Oracle, Cancel-then-recover, Recover-then-cancel.]

[D, Boufounos, Wakin, and Baraniuk - 2010]
Lemma:
If $\Phi$ satisfies the RIP of order $S$, then
\[
\left(1 - \frac{\delta}{1 - \delta}\right) \|x\|_2^2 \leq \|P\Phi x\|_2^2 \leq (1 + \delta) \|x\|_2^2
\]
provided that $\|x\|_0 \leq S - |\Lambda|$ and $\text{supp}(x) \cap \Lambda = \emptyset$.

\[
|\langle Py, P\Phi_j \rangle - x_j| \leq \frac{\delta}{1 - \delta} \|x_{\Lambda^c}\|_2
\]

[D, Boufounos, Wakin, and Baraniuk - 2010]
Aside: Orthogonal Matching Pursuit

OMP selects one index at a time

Iteration 1:

\[ j^* = \arg \max_j |\langle y, \Phi_j \rangle| \]

If \( \Phi \) satisfies the RIP of order \( \|u \pm v\|_0 \), then

\[ |\langle \Phi u, \Phi v \rangle - \langle u, v \rangle| \leq \delta \|u\|_2 \|v\|_2 \]

Set \( u = x \) and \( v = e_j \)

\[ |\langle y, \Phi_j \rangle - x_j| \leq \delta \|x\|_2 \]
Aside: Orthogonal Matching Pursuit

Subsequent Iterations:

\[ j^* = \arg \max_j |\langle Py, P\Phi_j \rangle| \]

\[ P = I - \Phi_\Lambda \Phi_\Lambda^\dagger \]

\[ P\Phi_\Lambda = 0 \quad \rightarrow \quad P\Phi x = P\Phi x^c_\Lambda \]

\[ |\langle Py, P\Phi_j \rangle - x_j| \leq \frac{\delta}{1-\delta} \| x^c_\Lambda \|_2 \]
Aside: Orthogonal Matching Pursuit

**Theorem:**
Suppose $x$ is $S$-sparse and $y = \Phi x$.
If $\Phi$ satisfies the RIP of order $S + 1$ with constant $\delta < \frac{1}{3\sqrt{S}}$, then the $j^*$ identified at each iteration will be a nonzero entry of $x$.

Exact recovery after $S$ iterations.

Argument provides simplified proofs for other orthogonal greedy algorithms (e.g. ROMP) that are robust to noise.

[D and Wakin - 2010]
What about structured measurement noise?
What about structured measurement noise?
Theorem:
If $\Phi$ is a sub-Gaussian matrix with
$$M = O \left( (S + \kappa) \log \left( \frac{N + M}{S + \kappa} \right) \right)$$
then $[\Phi \ I]$ satisfies the RIP of order $(S + \kappa)$ with probability at least $1 - 3e^{-CM}$.
We can recover sparse signals *exactly* in the presence of *unbounded* sparse noise

$$\text{Fixed } \|e\|_2 = 0.1$$

[Justice Pursuit](#)

[Laska, D, and Baraniuk - 2009](#)
Conclusions

• CS systems are sensitive to noisy signals
  - if our input signal is very noisy, it isn’t really very sparse
  - when noise is large, *measurements matter*
  - exploit sparsity in a different manner - e.g., adaptivity

• CS can be highly robust to *interference*
  - structured signal noise
  - structured measurement noise

• What about quantization noise?
Quantization Noise
Signal Recovery with Quantization

- Finite-range quantization leads to *saturation* and *unbounded errors*

- Quantization noise noise changes as we change the sampling rate
Saturation Strategies

- **Rejection:** Ignore saturated measurements

- **Consistency:** Retain saturated measurements. Use them only as inequality constraints on the recovered signal

- If the rejection approach works, the consistency approach should automatically do better
Rejection and Democracy

- The RIP is *not sufficient* for the rejection approach

- Example: $\Phi = I$
  - perfect isometry
  - *every* measurement must be kept

- We would like to be able to say that *any* submatrix of $\Phi$ with sufficiently many rows will still satisfy the RIP

- Strong, *adversarial* form of “democracy”
Sketch of Proof

- Step 1: Concatenate the identity to $\Phi$

Theorem:
If $\Phi$ is a sub-Gaussian matrix with

$$M = O \left( S \log \left( \frac{N + M}{S} \right) \right)$$

then $[\Phi \; I]$ satisfies the RIP of order $S$ with probability at least $1 - 3e^{-CM}$.

[D, Laska, Boufounos, and Baraniuk - 2009]
Sketch of Proof

• Step 2: Combine with the “interference cancellation” lemma

\[ \Lambda \]

\[ P\Phi = \tilde{\Phi} \]

• The fact that \([\Phi \ I]\) satisfies the RIP implies that if we take \(D\) extra measurements, then we can delete \(O(D)\) arbitrary rows of \(\Phi\) and retain the RIP.

• This is a strong adversarial notion of democracy

[D, Laska, Boufounos, and Baraniuk - 2009]
Rejection In Practice

$$\text{SNR} = 10 \log_{10} \left( \frac{\|x\|^2}{\|\hat{x} - x\|^2} \right)$$
Rejection In Practice

\[ \text{SNR} = 10 \log_{10} \left( \frac{\|x\|_2^2}{\|\hat{x} - x\|_2^2} \right) \]
Rejection In Practice

\[ \text{SNR} = 10 \log_{10} \left( \frac{\|x\|_2^2}{\|\hat{x} - x\|_2^2} \right) \]
Benefits of Saturation

\[ \text{saturation rate} \approx 0 \]

SNR (dB)

Saturation Rate

\[ T \]

[Laska, Boufounos, D, and Baraniuk - 2011]
Benefits of Saturation

\[ \approx 5 \text{ dB gain} \]

SNR (dB) vs. Saturation Rate

[Laska, Boufounos, D, and Baraniuk - 2011]
Potential for SNR Improvement?

By sampling at a lower rate, we can quantize to a higher bit-depth, allowing for potential gains.

[Le et al. - 2005]
Empirical SNR Improvement

SNR (dB)

log₂(N/M)

8 bits

4 bits

Oracle CS

CoSaMP CS

[D, Laska, Treichler, and Baraniuk - 2011]
Conclusions

• CS is robust to quantization noise in a non-traditional sense

• Democracy is a major advantage of CS measurements

• CS offers the potential to significantly boost dynamic range
  - can offset drawbacks associated with noise

• When is CS most useful?
  - performance is limited by quantization (high bandwidth apps)
  - when your signal is sparse (not too noisy)
Real-World Signal Models
## Candidate Analog Signal Models

<table>
<thead>
<tr>
<th></th>
<th>Model for $x(t)$</th>
<th>Basis for $x$</th>
<th>Sparsity level for $x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>multitone</td>
<td>sum of $S$ “on-grid” tones</td>
<td>$\Psi = \text{DFT}$</td>
<td>$S$-sparse</td>
</tr>
</tbody>
</table>

### Mathematical Expression

$$X(F) = \sum_{n=-\frac{B_{\text{nyq}}}{2}}^{\frac{B_{\text{nyq}}}{2}} X(n) e^{-j2\pi fn T}$$
### Candidate Analog Signal Models

<table>
<thead>
<tr>
<th>Model for $x(t)$</th>
<th>Basis for $x$</th>
<th>Sparsity level for $x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>multitone sum of $S$ “on-grid” tones</td>
<td>$\Psi = \text{DFT}$</td>
<td>$S$-sparse</td>
</tr>
<tr>
<td>multiband $K$ occupied bands of bandwidth $B$</td>
<td>$\Psi = ?$</td>
<td>?</td>
</tr>
</tbody>
</table>

![Diagram](chart.png)

- Landau
- Bresler, Feng, Venkataramani
- Eldar, Mishali

\[ X(F) = \frac{B_{\text{nyq}}}{2} \quad 0 \quad \frac{B_{\text{nyq}}}{2} \]
The Problem with the DFT

\[ x(t) = \int_{-\frac{B}{2}}^{\frac{B}{2}} X(F) e^{j2\pi Ft} \, dF \]

sampling

\[ x[n] = \int_{-W}^{W} X(f) e^{j2\pi fn} \, df, \quad \forall n \]

\[ W = \frac{B}{2B_{nyq}} \]
The Problem with the DFT

\[ x[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} X(f) e^{j2\pi f n} \, df, \quad \forall n \]

\[ x = \sum_{k=0}^{N-1} X_k e^{j \frac{2\pi k}{N}}, \quad e_f := \begin{bmatrix} e^{j2\pi f_0} \\ e^{j2\pi f} \\ \vdots \\ e^{j2\pi f(N-1)} \end{bmatrix} \]

time-limiting

NOT SPARSE
\[ x[n] = \int_{-W}^{W} X(f) e^{j2\pi fn} \, df, \quad \forall n \]

Diagram:

\[ \mathcal{T}_N(x[n]) = \int_{-W}^{W} X(f) \mathcal{T}_N(e^{j2\pi fn}) \, df, \quad \forall n \]
Time-limited complex exponentials form a “basis” for bandlimited signals

\[ x = \int_{-W}^{W} X(f) e_f \, df \]

\[ e_f := \begin{bmatrix} e^{j2\pi f_0} \\ e^{j2\pi f} \\ \vdots \\ e^{j2\pi f(N-1)} \end{bmatrix} \]

The problem: we need infinitely many of them.
Suppose that we wish to minimize
\[
\int_{-W}^{W} \| e_f - P_Q e_f \|_2^2 \, df
\]
over all subspaces $Q$ of dimension $k$.

Optimal subspace is spanned by the first $k$ “DPSS vectors”.
Discrete Prolate Spheroidal Sequences (DPSS’s)

**Slepian [1978]:** Given an integer $N$ and $\frac{W}{N} \leq \frac{1}{2}$, the DPSS’s are a collection of $N$ vectors

$$s_0, s_1, \ldots, s_{N-1} \in \mathbb{R}^N$$

that satisfy

$$\mathcal{T}_N(\mathcal{B}_W(s_\ell))) = \lambda_\ell s_\ell.$$  

The DPSS’s are perfectly time-limited, but when $\lambda_\ell \approx 1$ they are highly concentrated in frequency.
The first $\approx 2NW$ eigenvalues $\approx 1$.
The remaining eigenvalues $\approx 0$.  

$N = 1024$  
$W = \frac{1}{4}$  
$2NW = 512$
DPSS Examples

\[ N = 1024 \quad W = \frac{1}{4} \]

\[ \ell = 0 \quad \ell = 127 \quad \ell = 511 \]
Recall: Best Subspace Fit

Suppose that we wish to minimize

$$\int_{-W}^{W} \| e_f - P_Q e_f \|_2^2 \, df$$

over all subspaces $Q$ of dimension $k$.

Optimal subspace is spanned by the first $k$ “DPSS vectors”.

$$\int_{-W}^{W} \| e_f - P_Q e_f \|_2^2 \, df = \sum_{\ell=k}^{N-1} \lambda_\ell$$
Approximation of Bandlimited Signals

\[ \text{SNR} = 20 \log_{10} \left( \frac{\|e_f\|}{\|e_f - P_Q e_f\|} \right) \text{ dB} \]
Approximation of Bandlimited Signals

$$\text{SNR} = 20 \log_{10} \left( \frac{\|e_f\|}{\|e_f - P_Q e_f\|} \right) \text{ dB}$$
DPSS’s for Bandpass Signals
Modulate $k$ DPSS vectors to center of each band:

$$\Psi = [\Psi_1, \Psi_2, \ldots, \Psi_J]$$

approximately square if $k \approx 2NW$

Most multiband signals, when sampled and time-limited, are well-approximated by a sparse representation in $\Psi$. 
Theorem:
Suppose that $\Phi$ is sub-Gaussian and that the $\Psi_i$ are constructed with $k = (1 - \epsilon)2NW$. If

$$M \geq CS \log(N/S)$$

then with high probability $\Phi \Psi$ will satisfy the RIP of order $S$.

$K$ occupied bands $\quad\Rightarrow\quad S \approx KNB/B_{nyq}$

$$\frac{M}{N} \geq C' \frac{KB}{B_{nyq}} \log \left( \frac{B_{nyq}}{KB} \right)$$

[D and Wakin - 2011]
Block-Sparse Recovery

Nonzero coefficients of $\alpha$ should be clustered in blocks according to the occupied frequency bands

$$x = [\Psi_1, \Psi_2, \ldots, \Psi_J] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix}$$

This can be leveraged to reduce the required number of measurements and improve performance through “model-based CS”

- Baraniuk et al. [2008, 2009, 2010]
- Blumensath and Davies [2009, 2011]
Empirical Results: Noise

\[ N = 4096 \]
\[ M = 512 \]
\[ K = 5 \]
\[ \frac{B}{B_{\text{nyq}}} = \frac{1}{256} \]

[D and Wakin - 2011]
Empirical Results: Measurements

\[ N = 4096 \]

\[ \frac{B}{B_{\text{nyq}}} = \frac{1}{256} \]

[D and Wakin - 2011]
Empirical Results: Measurements

\[ N = 4096 \]

\[ \frac{B}{B_{nyq}} = \frac{1}{256} \]

Graph showing the relationship between recovery SNR (dB) and oversampling factor for different values of $K$. The graph includes data points and lines for $K = 5$, $K = 10$, and $K = 15$. The formula $\frac{M}{2NWK}$ is also shown in the legend box.
Empirical Results: Real-World Sensors

\[ N = 4096 \]
\[ \frac{B}{B_{\text{nyq}}} = \frac{1}{256} \]
\[ K = 5 \]
Empirical Results: DFT Comparison

\[ N = 4096 \]

\[ \frac{B}{B_{\text{nyq}}} = \frac{1}{256} \]

\[ K = 5 \]

[Empirical Results: DFT Comparison - D and Wakin - 2011]
Empirical Results: DFT Comparison

\[ N = 4096 \]

\[ \frac{B}{B_{nyq}} = \frac{1}{256} \]

\[ K = 5 \]

[D and Wakin - 2011]
Interference Cancellation

DPSS’s can be used to cancel bandlimited interferers without reconstruction.

\[ P = I - \Phi \Psi_i (\Phi \Psi_i)^\dagger \]

Extremely useful in compressive signal processing applications.
Conclusions

• DPSS’s can be used to efficiently represent *most* sampled multiband signals
  - far superior to DFT

• Two types of error: *approximation* + *reconstruction*
  - approximation: small for most signals
  - reconstruction: zero for DPSS-sparse vectors
  - delicate balance in practice, but there is a sweet spot

• This approach combines careful design of $\Psi$ with more sophisticated sparse models
  - relevant in many contexts beyond ADCs