Spectral Compressive Sensing

Marco F. Duarte

Portions are joint work with:

Richard G. Baraniuk (Rice University)
Hamid Dadkhahi (UMass Amherst)
Karsten Fyhn (Aalborg University)
Spectral Compressive Sensing

- Compressive sensing applied to *frequency-sparse signals*

\[
Y = \Phi X
\]

\(M \times 1\) linear measurements

\(M \times N\)

\(N \times 1\) frequency-sparse signal

\(K\) Fourier components

[E. Candès, J. Romberg, T. Tao; D. Donoho]
Spectral Compressive Sensing

- Compressive sensing applied to frequency-sparse signals

\[ y = \Phi \Psi \theta \]

- \( y \): \( M \times 1 \) linear measurements
- \( \Phi \): \( M \times N \) DFT Basis for frequency-sparse signals
- \( \Psi \): \( N \times 1 \) frequency-sparse signal
- \( \theta \): \( K \) nonzero DFT coefficients

DFT Basis for frequency-sparse signals
Frequency-Sparse Signals and the DFT Basis

\[ x = \sum_{k=1}^{K} a_k e(f_k) \quad X(\omega) = \sum_{k=1}^{K} a_k \delta(\omega - \omega_k) \]

\[ e(f) = \frac{1}{\sqrt{N}} \begin{bmatrix} e^{j2\pi f/N} & e^{j2\pi 2f/N} & \ldots & e^{j2\pi (N-1)f/N} \end{bmatrix} \]

\[ \theta = \Psi^{-1} x \]

\[ x[n] = \sin \left( \frac{2\pi n}{N} \times 10 \right) \]

\[ x[n] = \sin \left( \frac{2\pi n}{N} \times 10.5 \right) \]

\[ N = 1024 \]

\[ \|\theta\|_0 = 2, \|\theta - \theta_2\|_2 = 0 \]

\[ \|\theta\|_0 = 1024, \quad \|\theta - \theta_2\|_2 = 0.76\|\theta\|_2 \]
Frequency-Sparse Signals and the DFT Basis

Signal is sum of 10 sinusoids
Compressive Sensing for Frequency-Sparse Signals

$N = 1024$
$K = 20$
Compressive Sensing for Frequency-Sparse Signals

Number of measurements $M$

Average SNR, dB

$N = 1024$

$K = 20$
The Redundant DFT Frame

\[ x[n] = \sin \left( \frac{2\pi n}{N} \times 10.5 \right) \]

\[ N = 1024 \]
The Redundant DFT Frame

\[ x[n] = \sin \left( \frac{2\pi n}{N} \times 10.5 \right) \]

\[ N = 1024 \]
\[ \Psi(c), \ c = 10 \]

\[ \Psi(c) = \begin{bmatrix} e \left( \frac{1}{c} \right) & e \left( \frac{2}{c} \right) & \ldots & e \left( \frac{N - 1/c}{c} \right) \end{bmatrix} \]
The Redundant DFT Frame

\[ x[n] = \sin \left( \frac{2\pi n}{N} \times 10.5 \right) \]

\[ N = 1024 \]

\[ \Psi(c), \ c = 10 \]

\[ x = \Psi(c)\theta \]

\[ \|\theta\|_0 = 2, \|\theta - \theta_2\|_2 = 0 \]
The Redundant DFT Frame

\[ x[n] = \sin \left( \frac{2\pi n}{N} \times 10.5 \right) \]

\[ N = 1024 \]
\[ \Psi(c), \ c = 10 \]

Recovery algorithms operate similarly to “matched filtering”:

\[ p = \Psi(c)^T x \]
The Redundant DFT Frame

\[ x[n] = \sin \left( \frac{2\pi n}{N} \times 10.5 \right) \]

\[ N = 1024 \]

\[ \Psi(c), \ c = 10 \]

\[ x = \Psi(c)\theta \]

\[ \|\theta\|_0 = 2, \|\theta - \theta_2\|_2 = 0 \]

\[ \mu(\Psi(c)) \approx 0.98 \]

Sparse approximation algorithms fail

\[ \theta' = \Psi(c)^T x \]

\[ \|\theta'\|_0 = 9218 = (c - 1)N + 2, \]

\[ \|\theta' - \theta'_2\|_2 = 0.95\|\theta'\|_2 \]

[Candès, Needell, Eldar, Randall 2011]
Sparse Approximation of Frequency-Sparse Signals

Signal is sum of 10 sinusoids at arbitrary frequencies

Approximation sparsity $K$

Normalized approx. error

- Standard sparse approx. via DFT Basis
- Standard sparse approx. via DFT Frame
Structured Sparse Signals

- A $K$-sparse signal lives on the collection of $K$-dim subspaces aligned with coordinate axes.
- A $K$-structured sparse signal lives on a particular (reduced) collection of $K$-dimensional canonical subspaces.

[Baraniuk, Cevher, Duarte, Hegde 2010]
Structured Restricted Isometry Property (SRIP)

- Preserve the structure **only** between sparse signals that follow the structure model.
- Random (iid Gaussian, Rademacher) matrix has the SRIP with high probability if

\[ M = O(K + \log m_K) \]

\[ m_K \text{ K-dim planes} \]

[Blumensath, Davies; Lu, Do]
Many state-of-the-art sparse recovery algorithms (greedy and optimization solvers) rely on **thresholding** $x' = \mathcal{T}(x, K)$ [Daubechies, Defrise, and DeMol; Nowak, Figueiredo, and Wright; Tropp and Needell; Blumensath and Davies...]

$$x'(n) = \begin{cases} 
  x(n) & \text{if } |x(n)| \text{ is among } K \text{ largest,} \\
  0 & \text{otherwise.}
\end{cases}$$

Thresholding provides the best approximation of $x$ within $\Sigma_K$

$$x' = \arg \min_{\overline{x} \in \Sigma_K} \|x - \overline{x}\|_2$$
Structured Recovery Algorithms

- Modify existing approaches (optimization or greedy-based) to obtain **structure-aware recovery algorithms**: replace the thresholding step in IHT, CoSaMP, SP, ... with a **best structured sparse approximation step** that finds the closest point within union of subspaces

$\mathbb{R}^N \overset{x}{\rightarrow} \Omega_K \overset{x'}{\rightarrow} \Omega_K$

$x' = M(x, K) = \arg \min_{\bar{x} \in \Omega_K} \|x - \bar{x}\|_2$

Greedy structure-aware recovery algorithms **inherit guarantees** of generic counterparts

*(even though feasible set may be nonconvex)*
Structured Frequency-Sparse Signals

- A \textit{K-structured frequency-sparse} signal \( x \) consists of \( K \) sinusoids that are mutually incoherent:

\[ x = \sum_{k=1}^{K} a_k e(f_k) \in T_{K,c,\mu} \quad \text{if} \quad cf_K \in \mathbb{Z}, \quad |\langle e(f_k), e(f_{k'}) \rangle| \leq \mu \quad \forall \ k \neq k' \]

- If \( x \) is \( K \)-structured frequency-sparse, then there exists a \( K \)-sparse vector \( \theta \) such that \( x = \Psi(c)\theta \) and the nonzeros in \( \theta \) are spaced apart from each other (\textit{band exclusion}).
Structured Frequency-Sparse Signals

- If $x$ is $K$-structured frequency-sparse, then there exists a $K$-sparse vector $\theta$ such that $x = \Psi(c)\theta$ and the nonzeros in $\theta$ are spaced apart from each other.

- Preserve the structure only between sparse signals that follow the structured sparsity model

- Random (iid Gaussian, Bernoulli) matrix has the structured RIP with high probability if

\[
M = \mathcal{O} \left( K \log \left( \frac{c(N - K D_N^{-1}(\mu N))}{K} \right) \right)
\]
Structured Sparse Approximation

Algorithm 1: $\mathbb{T}(x, K, c, \mu)$

Integer Program

Inputs:
- Signal vector $x$
- Target sparsity $K$
- Redundancy factor $c$
- Maximum coherence $\mu$

Output:
- Approximation vector $\hat{x}$

- Compute coefficients: $\theta = \Phi(c)^T x$
  $w_\theta[i] = \theta[i]^2, \ i = 0, \ldots, cN - 1$
- Solve support:
  $s = \arg \max_{s \in \{0,1\}^{cN}} w_\theta^T s$
  s.t. $D_\mu s \leq 1, s^T 1 \leq K$
- Mask coefficients:
  $\hat{\theta}[i] \leftarrow \theta[i] s[i], \ i = 0, \ldots, cN - 1$
- Return $\hat{x} = \Phi(c) \hat{\theta}$
**Theorem:**
Assume we obtain noisy CS measurements of a signal $y = \Phi x + n$. If $\Phi$ has the structured RIP with $\delta < 0.1$, then the output of the structured IHT algorithm obeys

\[
\|x - \hat{x}\|_2 \leq C_1 \|x - \mathbb{M}(x, K)\|_2 + \frac{C_2}{\sqrt{K}} \|x - \mathbb{M}(x, K)\|_1 + C_3 \|n\|_2
\]

**CS recovery error**  
**signal $K$-term structured sparse approximation error**  
**noise**

In words, *instance optimality* based on **structured sparse approximation**

[Baraniuk, Cevher, Duarte, Hegde 2010]
Structured Sparse Approximation

**Algorithm 2:** $T_h(x, K, c, \mu)$

**Inhibition Heuristic**

**Inputs:**
- Signal vector $x$
- Target sparsity $K$
- Redundancy factor $c$
- Maximum coherence $\mu$

**Output:**
- Approximation vector $\hat{x}$

- Compute coefficients: $\theta = \Phi(c)^T x$
- Initialize: $\hat{\theta}[d] = 0$, $d = 0, \ldots, cN - 1$
- While $\theta$ is nonzero and $\|\hat{\theta}\|_0 \leq K$,
  - Find max abs entry $|\theta[n_{\max}]|$ of $\theta$
  - Copy entry $\hat{\theta}[n_{\max}] = \theta[n_{\max}]$
  - Inhibit “coherent” entries $\theta[n'] = 0$
- Return $\hat{x} = \Phi(c)\hat{\theta}$
Structured Sparse Approximation

\[ \tilde{\theta} = \mathcal{T}(\Psi(c)^T x, K) \]

DFT Frame + Thresholding equivalent to Maximum Likelihood Estimate of amplitudes and frequencies for frequency-sparse signal via Periodogram

\[ x = \sum_{k=1}^{K} a_k e(f_k) + n \]

Wideley-studied problem: Line spectral estimation
Structured Sparse Approximation

**Algorithm 3:** $\mathbb{T}_l(x, K)$

**Line Spectral Estimation**

**Inputs:**
- Signal vector $x$
- Target sparsity $K$

**Output:**
- Parameter estimates $\hat{a}_1, \ldots, \hat{a}_K$
- Signal estimate $\hat{x}$
  
  $\hat{f}_1, \ldots, \hat{f}_K$

\[
\mathbb{T}_l(x, K) \rightarrow \hat{x} = \sum_{k=1}^{K} \hat{a}_k e(\hat{f}_k)
\]
Sparse Approximation of Frequency-Sparse Signals

Signal is sum of 10 sinusoids at arbitrary frequencies

Approximation sparsity $K$

Normalized approx. error

- Standard sparse approx. via DFT Basis
- Standard sparse approx. via DFT Frame

Signal is sum of 10 sinusoids at arbitrary frequencies
Sparse Approximation of Frequency-Sparse Signals

Signal is sum of 10 sinusoids at arbitrary frequencies

- Standard sparse approx. via DFT Basis
- Standard sparse approx. via DFT Frame
- Structured sparse approx. via Alg. 1
- Structured sparse approx. via Alg. 2
- Structured sparse approx. via Alg. 3

Normalized approx. error vs. Approximation sparsity K

Signal is sum of 10 sinusoids at arbitrary frequencies
Structured CS: Performance

Number of measurements $M$

Average SNR, dB

- Root MUSIC on $M$ signal samples
- Standard IHT via DFT (Average)
- Standard IHT via DFT (Best/Worst)

$N = 1024$

$K = 20$
Structured CS: Performance

\[ N = 1024 \]
\[ K = 20 \]
From Recovery of Sparse Signals To Line Spectral Estimation

- Can “read” indices of nonzero DFTF coefficients to obtain *frequencies* of frequency-sparse signal components
- Equivalence: accurate recovery = accurate estimation?
- **Algorithms**: Alg. 3 essentially combines legacy line spectral estimation with CS recovery algorithms

![Graph](image)

- How to change *signal model* to further improve performance?
Interpolating the Projections (Dirichlet Kernel)

- Main lobe of Dirichlet kernel can be well approximated by a quadratic polynomial (parabola)
- **Three samples** around peak are required for interpolation
From Discrete to Continuous Models

• Both the DFT basis and the DFT frame can be conceived as **samplings** from an **infinite set** of signals $e(f)$ for a discrete set of values for the frequency $f \in [0, N)$

$$e(f) = \frac{1}{\sqrt{N}} \left[ e^{j2\pi f/N} \ e^{j2\pi 2f/N} \ldots \ e^{j2\pi (N-1)f/N} \right]$$

• Since the signal vector $e(f)$ varies smoothly in each entry as a function of $f$, we can represent the signal set as a one-dimensional **nonlinear manifold**:

$$\mathcal{M} = \{ e(f) : f \in [0, N) \}$$
From Discrete to Continuous Models

• For computational reasons, we wish to design methods that allow us to **interpolate** the manifold from the samples obtained in the DFT basis/frame to increase the resolution of the frequency estimates.

• An **interpolation-based** compressive line spectral estimation algorithm obtains projection values for sets of manifold samples and interpolates around peak on the rest of the manifold to get frequency estimate

\[ \mathcal{M} = \{e(f) : f \in [0, N)\} \]
Interpolating the Manifold: Polar Interpolation

- All points in manifold have equal norm; distance b/w samples is uniform
- Manifold must be contained within unit Euclidean ball (hypersphere)
- Project signal estimates into hypersphere
- Find closest point in manifold by interpolating from closest samples with polar coordinates
- Integrate band exclusion to get Band-Excluding Interpolating SP (BISP)
In BISP, find closest point in manifold by interpolating from closest samples with polar coordinates:

\[ e(f_0 - 1/c) \leftrightarrow \angle = \theta_0 - \Delta \]
\[ e(f_0) \leftrightarrow \angle = \theta_0 \]
\[ e(f_0 + 1/c) \leftrightarrow \angle = \theta_0 + \Delta \]
\[ \hat{x} \leftrightarrow \angle = ? \]

Map back from manifold to frequency estimates (parameter space)

Akin to Continuous Basis Pursuit (CBP) [Ekanadham, Tranchina, and Simoncelli 2011]
Compressive Line Spectral Estimation: Performance Evaluation

\( N = 100, \; K = 4, \; c = 5, \Delta f = 0.2 \text{ Hz} \)

\( \ell_1 \)-analysis
SIHT
SDP
BOMP
CBP
BISP

Table 1. We observed that most algorithms exhibit computation time roughly independent of the number of measurements. Moreover, the relaxation in BISP that accounts for the presence of noise reduces its computation time, counts for the presence of noise reduces its computation time, and varies according to the measurement scheme. Additionally, the relaxation in BISP that accounts for the presence of noise reduces its computational complexity, as well as its lack of flexibility on the measurement scheme. Moreover, the relaxation in BISP that accounts for the presence of noise reduces its computation time, counts for the presence of noise reduces its computation time, and varies according to the measurement scheme.

BOMP [Fannjiang and Liao 2012]
SDP: Atomic Norm Minimization
[Tang, Rhaskar, Shah, Recht 2012]
Compressive Line Spectral Estimation: Performance Evaluation (Noise)

$N = 100$, $K = 4$, $M = 50$, $c = 5$, $\Delta f = 0.2$ Hz
Compressive Line Spectral Estimation: Computational Expense

<table>
<thead>
<tr>
<th>Time (seconds)</th>
<th>Noiseless</th>
<th>Noisy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_1$-analysis</td>
<td>9.5245</td>
<td>8.8222</td>
</tr>
<tr>
<td>SIHT</td>
<td>0.2628</td>
<td>0.1499</td>
</tr>
<tr>
<td>SDP</td>
<td>8.2355</td>
<td>9.9796</td>
</tr>
<tr>
<td>BOMP</td>
<td>0.0141</td>
<td>0.0101</td>
</tr>
<tr>
<td>CBP</td>
<td>46.9645</td>
<td>40.3477</td>
</tr>
<tr>
<td>BISP</td>
<td>5.4265</td>
<td>1.4060</td>
</tr>
</tbody>
</table>
Conclusions

• Spectral CS provides significant improvements on frequency-sparse signal recovery
  – address coherent dictionaries via structured sparsity
  – simple-to-implement modifications to recovery algs
  – can leverage decades of work on spectral estimation
  – robust to model mismatch, presence of noise

• Compressive line spectral estimation:
  – recovery via parametric dictionaries provides compressive parameter estimation
  – dictionary elements as samples from manifold models
  – from dictionaries to manifolds via interpolation techniques
  – from recovery to parameter estimation from compressive measurements
  – localization, bearing estimation, radar imaging, ...

http://www.ecs.umass.edu/~mduarte    mduarte@ecs.umass.edu