Structured Sparsity Models for Compressive Sensing

Chinmay Hegde
Rice University

Volkan Cevher
Richard Baraniuk
Marco Duarte
Concise Signal Model: Sparsity

• **Sparse** signal:

  - only $K$ out of $N$ coordinates nonzero
Concise Signal Model: Sparsity

• **Sparse** signal:
  - only $K$ out of $N$ coordinates nonzero

• Geometry: *union* of $K$-dimensional subspaces aligned w/ coordinate axes
Compressive Sensing

- **Sampling** via dimensionality reduction

\[
\begin{align*}
M \times 1 & \quad y & \Phi & \quad x \\
\text{Measurements} & & & \\
M \times N & & & \\
N \times 1 & \text{sparsity} & \quad K & \text{nonzero entries}
\end{align*}
\]
Restricted Isometry Property (RIP)

- Preserve the structure of sparse signals
• *Random* subgaussian matrix $\Phi$ has the RIP whp if

$$M = O(K \log(N/K))$$
Stable Recovery

- Efficient, stable algorithms that give back signal
Compressive Sensing

\[ M \times 1 \]
random measurements

\[ \begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \Phi \\ \end{bmatrix} \begin{bmatrix} x \end{bmatrix} \]

\[ N \times 1 \]
sparse signal

\[ K \]
nonzero entries

• \( \ell_1 \)-optimization
  \[[C, R, T]; [D]; [F, W, N]; [H, Y, Z]\]

• Greedy algorithms
  – OMP \[[G, T]\]
  – iterated thresholding \[[N, F]; [D, D, DeM]; [B, D]\]
  – CoSaMP \[[N, T]\]; Subspace Pursuit \[[D, M]\]
Iterated Thresholding

goal: given $y = \Phi x$, recover a sparse $x$
initialize: $\hat{x}_0 = 0$, $r = y$, $i = 0$
iteration:

- $i \leftarrow i + 1$

- $b \leftarrow \hat{x}_{i-1} + \Phi^T r$  \textbf{update signal estimate}

- $\hat{x}_i \leftarrow \text{thresh}(b, K)$  \textbf{prune signal estimate} (best $K$-term approx)

- $r \leftarrow y - \Phi \hat{x}_i$  \textbf{update residual}

return: $\hat{x} \leftarrow \hat{x}_i$
Performance

• **Sparse signals**
  - noise-free measurements: exact recovery
  - noisy measurements: stable recovery

• **Compressible signals**
  - recovery as good as best \( K \)-sparse approximation

\[
\|x - \hat{x}\|_{l_2} \leq C_1 \|x - x_K\|_{l_2} + C_2 \frac{\|x - x_K\|_{l_1}}{K^{1/2}} + C_3 \varepsilon
\]
From Sparsity to \textit{Structured} Sparsity
Beyond Sparse Models

- Sparsity captures \textit{simplistic primary structure}

5% sparse image
Beyond Sparse Models

- Most real-world apps exhibit *additional structure*
Beyond Sparse Models

wavelets: natural images

Gabor atoms: chirps/tones

pixels: background subtracted images
Sparse Signals

- Defn: **$K$-sparse signals** comprise all $K$-dimensional canonical subspaces
Model-Sparse Signals

- Defn: A **K-sparse signal model** comprises a particular (*reduced*) set of *K*-dim canonical subspaces $[B, D], [L, D]$
Model-Sparse Signals

Model-based CS
Sampling Bounds

- **RIP**: stable embedding

\[ M = O(K \log(N/K)) \]
Sampling Bounds

- **Model-RIP**: stable embedding
  \[
  [B, D]; [B,D,DeV,W]
  \]

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

goal: given \( y = \Phi x \), recover a sparse \( x \)
initialize: \( \hat{x}_0 = 0, \; r = y, \; i = 0 \)
iteration:

\begin{itemize}
  \item \( i \leftarrow i + 1 \)
  \item \( b \leftarrow \hat{x}_{i-1} + \Phi^T r \) \hspace{1cm} \text{update signal estimate}
  \item \( \hat{x}_i \leftarrow \text{thresh}(b, K) \) \hspace{1cm} \text{prune signal estimate}
  \hspace{1cm} \text{(best } K\text{-term approx)}
  \item \( r \leftarrow y - \Phi \hat{x}_i \) \hspace{1cm} \text{update residual}
\end{itemize}

return: \( \hat{x} \leftarrow \hat{x}_i \)
Iterated **Model** Thresholding

goal: given $y = \Phi x$, recover model-sparse $x$

initialize: $\hat{x}_0 = 0$, $r = y$, $i = 0$

iteration:

- $i \leftarrow i + 1$

- $b \leftarrow \hat{x}_{i-1} + \Phi^T r$ \hspace{1cm} **update signal estimate**

- $\hat{x}_i \leftarrow \mathcal{M}(b, K)$ \hspace{1cm} **prune signal estimate** (best $K$-term model approx)

- $r \leftarrow y - \Phi \hat{x}_i$ \hspace{1cm} **update residual**

return: $\hat{x} \leftarrow \hat{x}_i$
Recipe for CS recovery algorithm

Ingredients

- Model
- Sampling bound $M$
- Signal approximation algorithm $\mathcal{M}(\cdot, K)$
Application: 1D-signals

- Eg. Neuronal spike trains

[Lewicki, '98]
Application: spike trains

- Absolute refractoriness

![Graph showing the function P_0(s) versus s (ms)]
Application: spike trains

- Absolute refractoriness

![Graph showing $P_0(s)$ vs. s (ms)]
$x \in M(K, \Delta)$ if

- $\|x\|_0 = K$
- no pair of consecutive nonzeros occur within $\Delta$ locations of each other.
Ingredient 2: sampling bound

• # subspaces = # sparsity patterns

\[ m_K = \binom{N-(K-1)(\Delta-1)}{K} \]

• Number of measurements

\[ M \geq CK \log(N/K - \Delta) \]
Ingredient 3: approximation

• Given arbitrary $x \in \mathbb{R}^N$, find closest $x^* \in \mathcal{M}(K, \Delta)$

• Equivalent to finding optimal *binary* support pattern

$$s = (s_1, \ldots, s_N) \in \mathcal{M}(K, \Delta)$$

• Portion of signal lying within a given support pattern:

$$x|_s := (s_1x_1, s_2x_2, \ldots, s_Nx_N)$$

• Problem

$$\min_s \|x - x|_s\|_2, \ s \in \mathcal{M}$$
Ingredient 3: approximation

- Can be transformed into an integer program
  \[ s^* = \text{arg min } c^T s, \]
  \[ Ws \leq u. \]
- Can be relaxed into a linear program
Ingredient 3: approximation

- Can be transformed into a integer program
  \[ s^* = \arg \min c^T s, \]
  \[ Ws \leq u. \]
- Can be relaxed into a linear program
- \( W \) is totally unimodular, so polytope corners are integer, so solution is always integer
Ingredient 3: approximation

- Can be transformed into a integer program
  \[ s^* = \arg\min c^T s, \]
  \[ Ws \leq u. \]
- Can be relaxed into a linear program
- \( W \) is totally unimodular, so polytope corners are integer, so solution is always integer
If \( y = \Phi x + n, x \in \mathcal{M}(K, \Delta) \) and \( \Phi \) satisfies \((K, \Delta)\)-RIP, then the algorithm converges to an estimate \( \hat{x} \), such that

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

$N = 1024$, $K = 50$, $\Delta = 10$
Results

$M = 150$, Distortion $= 1.76$dB
Results

\[ M = 150, \text{ Distortion} = 25.53 \text{dB} \]
Related Work

- **Ex: wavelet-trees**
  
  [Duarte, Wakin, Baraniuk], [La, Do], [Baraniuk, Cevher, Duarte, H]

- **Ex: block sparsity / signal ensembles**
  
  [Tropp, Gilbert, Strauss], [Stojnic, Parvaresh, Hassibi],
  [Eldar, Mishali], [Baron, Duarte et al], [B, C, D, H]

- **Ex: clustered signals**
  
  [C, D, H, B], [C, Indyk, H, B]
Related Work

- **Ex: wavelet-trees**
  
  [Duarte, Wakin, Baraniuk], [La, Do], [Baraniuk, Cevher, Duarte, H]

- **Ex: block sparsity / signal ensembles**
  
  [Tropp, Gilbert, Strauss], [Stojnic, Parvaresh, Hassibi],
  [Eldar, Mishali], [Baron, Duarte et al], [B, C, D, H]

- **Ex: clustered signals**
  
  [C, D, H, B], [C, Indyk, H, B]

- **Model-compressible signals**
  
  - Restricted Amplification Property (RAmP)
  - Instance-optimal guarantees in some cases
    
    [Baraniuk, Cevher, Duarte, H]
Summary

• Why CS works: stable embedding for signals with concise geometric structure

• Sparse signals >> model-sparse signals

• Model-based recovery algorithms

  Advantages: provably fewer measurements, flexible framework for algorithm design, stable recovery

www.dsp.rice.edu/cs
• Blank slide
• Another blank slide
Compressible Signals

• Real-world signals are compressible, not sparse

• Recall: **compressible** <> approximable by sparse
  
  – compressible signals lie close to a union of subspaces
  – ie: approximation error decays rapidly as $K \to \infty$
  – nested in that $\text{supp}\{x_K\} \subset \text{supp}\{x'_K\}, \ K < K'$

• If $\Phi$ has RIP, then both sparse and compressible signals are stably recoverable via LP or greedy alg
Model-Compressible Signals

- **Model-compressible** $\Leftrightarrow$ approximable by model-sparse
  - model-compressible signals lie close to a reduced union of subspaces
  - i.e.: model-approx error decays rapidly as $K \to \infty$

- Nested approximation property (**NAP**): model-approximations nested in that
  $\text{supp}\{x_K\} \subset \text{supp}\{x'_K\}, \ K < K'$
Model-Compressible Signals

- **Model-compressible** $\iff$ approximable by model-sparse
  - model-compressible signals lie close to a reduced union of subspaces
  - i.e.: model-approx error decays rapidly as $K \to \infty$

- Nested approximation property (NAP): model-approximations nested in that
  \[
  \text{supp}\{x_K\} \subset \text{supp}\{x'_K\}, \quad K < K'
  \]
Model-Compressible Signals

- **Model-compressible** $\iff$ approximable by model-sparse
  - model-compressible signals lie close to a reduced union of subspaces
  - i.e.: model-approx error decays rapidly as $K \to \infty$

- Nested approximation property (**NAP**): model-approximations nested in that
  \[ \text{supp}\{x_K\} \subseteq \text{supp}\{x'_K\}, \quad K < K' \]
Model-Compressible Signals

- **Model-compressible** $\iff$ approximable by model-sparse
  - model-compressible signals lie close to a reduced union of subspaces
  - ie: model-approx error decays rapidly as $K \to \infty$

- Nested approximation property (**NAP**):
  model-approximations nested in that
  $\text{supp}\{x_K\} \subset \text{supp}\{x'_K\}, \ K < K'$
Model-Compressible Signals

- **Model-compressible** $\leftrightarrow$ approximable by model-sparse
  - model-compressible signals lie close to a reduced union of subspaces
  - ie: model-approx error decays rapidly as $K \to \infty$

- Nested approximation property (**NAP**): model-approximations nested in that
  $\text{supp}\{x_K\} \subset \text{supp}\{x'_K\}, \ K < K'$

- **New result**: while model-RIP enables stable model-sparse recovery,
  model-RIP is *not sufficient* for stable model-compressible recovery!
Model-Compressible Signals

- **Model-compressible** <-> approximable by model-sparse
  - model-compressible signals lie close to a reduced union of subspaces
  - i.e.: model-approx error decays rapidly as $K \to \infty$

- **New result:** while model-RIP enables stable model-sparse recovery, model-RIP is *not sufficient* for stable model-compressible recovery!

- Ex: If $\Phi$ has the Tree-RIP, then tree-sparse signals can be stably recovered with $M = O(K)$
- However, tree-compressible signals cannot be stably recovered
Stable Recovery

- **Result:** Stable model-compressible signal recovery requires that $\Phi$ have both:
  - RIP + **Restricted Amplification Property**

- **RAmP:** controls nonisometry of $\Phi$ in the approximation’s *residual subspaces*

optimal $K$-term model recovery (error controlled by $\Phi$ RIP)

optimal $2K$-term model recovery (error controlled by $\Phi$ RIP)

residual subspace (error *not* controlled by $\Phi$ RIP)
Tree-RIP, Tree-RAmP

**Theorem:** An $M \times N$ iid subgaussian random matrix has the Tree($K$)-RIP if

\[
M \geq \begin{cases} 
\frac{2}{c\delta^2_T K} \left( K \ln \frac{48}{\delta_T K} + \ln \frac{512}{K e^2} + t \right) & \text{if } K < \log_2 N \\
\frac{2}{c\delta^2_T K} \left( K \ln \frac{24e}{\delta_T K} + \ln \frac{2}{K+1} + t \right) & \text{if } K \geq \log_2 N 
\end{cases}
\]

**Theorem:** An $M \times N$ iid subgaussian random matrix has the Tree($K$)-RAmP if

\[
M \geq \begin{cases} 
\frac{2}{(\sqrt{1+\epsilon_K} - 1)^2} \left( 10K + 2 \ln \frac{N}{K(K+1)(2K+1)} + t \right) & \text{if } K \leq \log_2 N \\
\frac{2}{(\sqrt{1+\epsilon_K} - 1)^2} \left( 10K + 2 \ln \frac{601N}{K^3} + t \right) & \text{if } K > \log_2 N 
\end{cases}
\]
Performance

- Using model-based IHT, CoSaMP with RIP and RAmP

- **Model-sparse signals**
  - noise-free measurements: exact recovery
  - noisy measurements: stable recovery

- **Model-compressible signals**
  - recovery as good as $K$-model-sparse approximation

\[
\|x - \hat{x}\|_2 \leq C_1 \|x - x_K\|_2 + C_2 \frac{\|x - x_K\|_1}{K^{1/2}} + C_3 \epsilon
\]

- CS recovery error
- signal model $K$-term approx error
- signal model $K$-term approx error
- noise
Tree-Sparse Signal Recovery

$N=1024$
$M=80$

CoSaMP,
(MSE=1.12)

L1-minimization
(MSE=0.751)

Tree-sparse CoSaMP
(MSE=0.037)
Simulation

- Recovery performance (MSE) vs. number of measurements

- Piecewise cubic signals + wavelets

- Models/algorithms:
  - sparse (CoSaMP)
  - tree-sparse (tree-CoSaMP)
Simulation

- Number samples for correct recovery

- Piecewise cubic signals + wavelets

- Models/algorithms:
  - sparse (CoSaMP)
  - tree-sparse (tree-CoSaMP)

![Graph showing $O(K \log N)$ and $O(K)$]

\[
O(K \log N)
\]

\[
O(K)
\]
Block-Sparse Signal

\[ N = 4096 \]
\[ K = 6 \text{ active blocks} \]
\[ J = \text{block length} = 64 \]
\[ M = 2.5JK = 960 \text{ msnts} \]

CoSaMP (MSE = 0.723)

block-sparse model recovery (MSE=0.015)
Block-Compressible Signal

- Target
- CoSaMP ($\text{MSE}=0.711$)
- Best 5-block approximation ($\text{MSE}=0.116$)
- Block-sparse recovery ($\text{MSE}=0.195$)
Clustered Signals

- **Graphical model**

- Model *clustering of significant pixels* in space domain using Ising MRF

- Ising model approximation performed efficiently using *graph cuts*

![Images showing target, Ising-model recovery, CoSaMP recovery, and LP (FPC) recovery.](image-url)