Recovery of Compressible Signals in Unions of Subspaces

Marco F. Duarte

Joint work with Chinmay Hegde, Volkan Cevher, Richard Baraniuk
Sparsity / Compressibility

- Many signals are *sparse* or *compressible* in some representation/basis (Fourier, wavelets, ...)

\[ N \text{ pixels} \quad K \ll N \text{ large wavelet coefficients} \]

\[ N \text{ wideband signal samples} \quad K \ll N \text{ large Gabor coefficients} \]
Concise Signal Structure

- **Sparse** signal: only $K$ out of $N$ coordinates nonzero
  - model: **union of $K$-dimensional subspaces** aligned with coordinate axes

\[
\begin{align*}
|x_i| & \quad K \quad \text{sorted index} \\
\mathbb{R}^N & \quad x
\end{align*}
\]
Compressive Sensing

- **Sensing** with dimensionality reduction

\[ y = \Phi x \]

\[ M \times 1 \text{ measurements} \]

\[ M \approx K \ll N \]

\[ N \times 1 \text{ sparse signal} \]

\[ K \text{ information rate} \]
Restricted Isometry Property (RIP)

- Preserve the structure of sparse/compressible signals
- RIP of order $2K$ implies: for all $K$-sparse $x_1$ and $x_2$

\[
\left(1 - \delta_{2K}\right) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2^2}{\|x_1 - x_2\|_2^2} \leq (1 + \delta_{2K})
\]
Restricted Isometry Property (RIP)

- Preserve the structure of sparse/compressible signals
- Random (i.i.d. Gaussian, Bernoulli) matrix has the RIP with high probability if

\[ M = O(K \log(N/K)) \]
Beyond Sparse Models

- Sparse/compressible signal model captures **simplistic primary structure**

- Wavelets: natural images
- Gabor atoms: chirps/tones
- Pixels: background subtracted images
Beyond Sparse Models

- Sparse/compressible signal model captures simplistic primary structure
- Modern compression/processing algorithms capture richer secondary coefficient structure
Sparse Signals

- Defn: A $K$-sparse signal lives on the collection of $K$-dim subspaces aligned with coord. axes
Model-Sparse Signals

- Defn: A \textbf{\textit{\textbf{K-model sparse}}} signal lives on a particular (reduced) collection of \textit{K}-dim canonical subspaces \cite{Blumensath2008,Lu2009}.

\[
\begin{align*}
&\text{\textbf{R}}^N \\
&\text{\textit{m}_K \text{ \textit{K-dim planes}}} \\
&\text{\textbf{m}_k \ll \binom{N}{K}}
\end{align*}
\]
Model-Based RIP

- Preserve the structure only of sparse/compressible signals that follow the model.

- Random (i.i.d. Gaussian, Bernoulli) matrix has the RIP with high probability if

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

\[ m_K \] \( K \)-dim planes

[Blumensath and Davies]
Model-Sparse Signals

• Defn: A $K$-model sparse signal lives on a particular (reduced) collection of $K$-dim canonical subspaces

• Recovery: Adapt standard CS recovery algorithms to enforce signal model using *model-based sparse approximation*  

[Baraniuk, Cevher, Duarte, Hegde]
Tree-Sparse

- **Model:** $K$-sparse coefficients + nonzero coefficients lie on a **rooted subtree**

Typical of wavelet transforms of natural signals and images (piecewise smooth)
Ex: Tree-Sparse

- **Model:** $K$-sparse coefficients + nonzero coefficients lie on a rooted subtree

- Typical of wavelet transforms of natural signals and images (piecewise smooth)

- **Tree-sparse approx:** find best rooted subtree of coefficients
  - CSSA [Baraniuk], dynamic programming [Donoho]

- Number of measurements that a matrix $\Phi$ with i.i.d. Gaussian entries needs to have Tree-RIP:

  $$M = O(K) < O(K \log(N/K))$$
Simulation

- Recovery performance (MSE) vs. number of measurements

- Piecewise cubic signals + wavelets

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

![Graph showing average normalized error magnitude versus M/K for Model-based recovery and CoSaMP methods.](image)
Tree-Sparse Signal Recovery

\[ N = 1024 \quad M = 80 \]

- Target signal
- CoSaMP, (RMSE=1.12)
- \( \ell_1 \)-minimization, (RMSE=0.751)
- Tree-based CoSaMP, (RMSE=0.037)
Concise Signal Structure

- **Sparse** signal: only $K$ out of $N$ coordinates nonzero
  - model: *union of* $K$-dimensional subspaces

- **Compressible** signal: sorted coordinates decay rapidly to zero
  well-approximated by a $K$-sparse signal (simply by thresholding)
Concise Signal Structure

- **Sparse** signal: only $K$ out of $N$ coordinates nonzero
  - model: union of $K$-dimensional subspaces

- **Compressible** signal: sorted coordinates decay rapidly to zero
  well-approximated by a $K$-sparse signal (simply by thresholding)
  nested approximations

\[ |x_i| \]
Concise Signal Structure

- **Sparse** signal: only $K$ out of $N$ coordinates nonzero
  - model: union of $K$-dimensional subspaces

- **Compressible** signal: sorted coordinates decay rapidly to zero
  - model: weak $\ell_p$ ball: $|x_i| < S i^{-1/p}$
Concise Signal Structure

- **Sparse** signal: only $K$ out of $N$ coordinates nonzero
  - model: *union of $K$-dimensional subspaces*

- **Compressible** signal: sorted coordinates decay rapidly to zero
  - model: *weak $\ell_p$ ball: $|x_i| < Si^{-1/p}$*

\[
\sigma_K(x) := \|x - x_K\|_2 \leq (ps)^{-1/2} SK^{-s}
\]

- $s = \frac{1}{p} - \frac{1}{2}$
RIP and Recovery

• Using $\ell_1$ methods, CoSaMP, IHT

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

• **Compressible signals**
  - recovery as good as $K$-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 $K$-term approx error
- noise
**Model-Compressible Signals**

- **Model-compressible** $\iff$ well approximated by model-sparse
  - model-compressible signals lie close to a reduced union of subspaces
  - i.e.: model-approx error decays rapidly as $K \to \infty$
    
    $$\sigma_{\mathcal{M}_K}(x) = \| x - x_{\mathcal{M}_K} \|_2 \leq CK^{-s}$$

- **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** $\iff$ well approximated by model-sparse
  - model-compressible signals lie close to a reduced union of subspaces
  - i.e.: model-approx error decays rapidly as $K \to \infty$
    \[ \sigma_{\mathcal{M}_K}(x) = \| x - x_{\mathcal{M}_K} \|_2 \leq C K^{-s} \]

- **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** $<>$ well approximated by model-sparse
  - model-compressible signals lie close to a reduced union of subspaces
  - i.e.: model-approx error decays rapidly as $K \to \infty$
    \[
    \sigma_{\mathcal{M}_K}(x) = \|x - x_{\mathcal{M}_K}\|_2 \leq CK^{-s}
    \]

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

- **$K$-RIP:** controls amt of nonisometry of $\Phi$ on all $K$-dimensional subspaces
- Can control norm of $\|y - \Phi x_K\|_2$, account for contribution as *noise*
- Model-RIP is *not sufficient* for stable model-compressible recovery!

![Diagram](image_url)

- optimal $K$-term model recovery (error controlled by $\mathcal{M}_K$-RIP)
- optimal $2K$-term model recovery (error controlled by $\mathcal{M}_K$-RIP)
- residual subspace: *not* in model (error *not* controlled by $\mathcal{M}_K$-RIP)
Stable Model-Based Recovery

- Properties of **model-compressible signals:**
  - Structure on sparse approximation also yields *structure on residual subspaces* $\mathcal{R}_j, K$
    - $R_j$: Number of subspaces/supports that arise from growing a $jK$-model-sparse approx. to a $(j+1)K$-model-sparse approx.
  - Norm of sparse approximation residuals *also has power law decay*

![Diagram showing stability in model-based recovery](image)

- Optimal $K$-term model recovery (error controlled by $\mathcal{M}_K$-RIP)
- Optimal $2K$-term model recovery (error controlled by $\mathcal{M}_K$-RIP)
- Residual subspace: *not in model* (error *not* controlled by $\mathcal{M}_K$-RIP)
Stable Model-Based Recovery

• **RAmP:** Restricted Amplification Property controls amount of nonisometry of $\Phi$ for the residuals $x\mathcal{M}_{jK} - x\mathcal{M}_{(j+1)K}$
  - Still fewer subspaces than RIP, *fewer measurements*
  - Can *relax isometry* for subsequent residual subspaces
  - Goal: control norm of *projected approximation error*
    $$\|\Phi(x - x\mathcal{M}_K)\|_2$$

optimal $K$-term model recovery (error controlled by $\mathcal{M}_K$-RIP)

optimal $2K$-term model recovery (error controlled by $\mathcal{M}_K$-RIP)

residual subspace: *not* in model (error *controlled* by RAmP)
Restricted Amplification Property

A matrix $\Phi$ has the $(\epsilon_K, r)\text{--RAmP}$ for the residual subspaces $R_{j,K}$ of the signal model $M$ if

$$\| \Phi u \|_2^2 \leq (1 + \epsilon_K)j^{2r} \| u \|_2^2$$

for any $u \in R_{j,K}$ and for each $1 \leq j \leq \lceil N/K \rceil$.

optimal $K$-term model recovery (error controlled by $M_{K-RIP}$)

optimal $2K$-term model recovery (error controlled by $M_{K-RIP}$)

residual subspace: not in model (error controlled by RAmP)
Restricted Amplification Property

A matrix $\Phi$ has the $(\epsilon_K, r) -$RAmp for the residual subspaces $\mathcal{R}_{j,K}$ of the signal model $\mathcal{M}$ if

$$\|\Phi u\|_2^2 \leq (1 + \epsilon_K) j^{2r} \|u\|_2^2$$

for any $u \in \mathcal{R}_{j,K}$ and for each $1 \leq j \leq \lceil N/K \rceil$

**Theorem**: Let $x$ be an $s$-model compressible signal under a signal model $\mathcal{M}$ with the NAP. If $\Phi$ has the $(\epsilon_K, r)$-RAmp and $r = s - 1$, then we have

$$\|\Phi(x - x_{\mathcal{M}_K})\|_2 \leq \sqrt{1 + \epsilon_K CK^{-s} \ln \left[ \frac{N}{K} \right]}.$$

(see paper for details)
Restricted Amplification Property

A matrix $\Phi$ has the $(\epsilon_K, r)$–RAmP for the residual subspaces $\mathcal{R}_{j,K}$ of the signal model $\mathcal{M}$ if

$$\|\Phi u\|_2^2 \leq (1 + \epsilon_K)j^{2r} \|u\|_2^2$$

for any $u \in \mathcal{R}_{j,K}$ and for each $1 \leq j \leq \lceil N/K \rceil$

**Theorem**: Let $x$ be an $s$-model compressible signal under a signal model $\mathcal{M}$ with the NAP. If $\Phi$ has the $(\epsilon_K, r)$-RAmP and $r = s - 1$, then we have

$$\|x - \hat{x}\| \leq \frac{C_1S}{K^{-s}} + C_2 \left(\|n\|_2 + \sqrt{1 + \epsilon_KSK^{-s}} \ln \left\lceil \frac{N}{K} \right\rceil\right),$$

(see paper for details)
Restricted Amplification Property

A matrix $\Phi$ has the $(\epsilon_K, r)$–\textit{RAmP} for the residual subspaces $R_{j, K}$ of the signal model $\mathcal{M}$ if

$$\left\| \Phi u \right\|_2^2 \leq (1 + \epsilon_K)j^{2r} \left\| u \right\|_2^2$$

for any $u \in R_{j, K}$ and for each $1 \leq j \leq \lceil N/K \rceil$.

**Theorem:** Let $x$ be an $s$-model compressible signal under a signal model $\mathcal{M}$ with the NAP. If $\Phi$ has the $(\epsilon_K, r)$-RAmP and $r = s - 1$, then we have

$$\left\| x - \hat{x} \right\|_2 \leq C_1 \left\| x - x_{M_K} \right\|_2 + C_2 \frac{\left\| x - x_{M_K} \right\|_1}{K^{1/2}} + C_3 \epsilon$$

\begin{itemize}
  \item CS recovery error
  \item signal $K$-term approx error
  \item noise
\end{itemize}
Restricted Amplification Property

A matrix $\Phi$ has the $(\epsilon_K, r) -$ RAmP for the residual subspaces $\mathcal{R}_{j,K}$ of the signal model $\mathcal{M}$ if

$$\| \Phi u \|_2^2 \leq (1 + \epsilon_K) j^{2r} \| u \|_2^2$$

for any $u \in \mathcal{R}_{j,K}$ and for each $1 \leq j \leq \lfloor N/K \rfloor$

**Theorem**: A matrix $\Phi$ with i.i.d. subgaussian entries has the $(\epsilon_K, r)$-RAmP with probability $1 - e^{-t}$ if

$$M \geq \max_{1 \leq j \leq \lfloor N/K \rfloor} \frac{2K + 4 \ln \frac{R_j N}{K}}{(jr \sqrt{1 + \epsilon_K - 1})^2} + 2t$$

for each $1 \leq j \leq \lfloor N/K \rfloor$

(see paper for details)
Theorem: An $M \times N$ i.i.d. subgaussian random matrix has the Tree($K$)-RIP with constant $\delta_{TK}$ if

$$M \geq \begin{cases} \frac{2}{c\delta_{TK}^2} \left( K \ln \frac{48}{\delta_{TK}^2} + \ln \frac{512}{Ke^2} + t \right) & \text{if } K < \log_2 N \\ \frac{2}{c\delta_{TK}^2} \left( K \ln \frac{24e}{\delta_{TK}^2} + \ln \frac{2}{K+1} + t \right) & \text{if } K \geq \log_2 N \end{cases}$$

with probability $1 - e^{-t}$

Theorem: An $M \times N$ i.i.d. subgaussian random matrix has the Tree($K$)-RAmP with constant \_\_\_ if

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

with probability $1 - e^{-t}$
Simulation

- Number samples for guaranteed recovery
  \[ \| x - \hat{x} \|_2 \leq 2.5 \sigma_{T_K}(x) \]

- Piecewise cubic signals + wavelets

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

\[ \mathcal{O}(K \log N) \]
\[ \mathcal{O}(K) \]
Conclusions

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

• **Concise** models require *even fewer* measurements for recovery than simple sparsity models

• *Model-sparse and compressible signals* using correlations between coefficient values and locations
  – Can modify standard algorithms
  – Can obtain robustness, recovery guarantees
  – Further work: stochastic models, graphical models, optimization-based recovery

 dsp.rice.edu/cs