Compressive Parameter Estimation with Earth Mover’s Distance via K-means Clustering

Dian Mo and Marco F. Duarte
Compressive Sensing (CS)

- Integrates linear acquisition with dimensionality reduction

$$M \times 1 \text{ linear measurements} \quad = \quad \Phi \quad N \times 1 \text{ discrete-time signal}$$

$M \times N$ nonzero components in suitable transform

$$x \in \Sigma_K$$

[Candès, Romberg, Tao; Donoho]
Parametric Dictionaries for Sparsity

- Integrates sparsity/CS with **parameter estimation**
- **Parametric dictionaries** (PDs) collect observations for a set of values of parameter of interest (one per column)
  \[ \Theta = \{\theta_1, \ldots, \theta_N\} \]
- Simple signals (e.g., few localization targets) can be expressed via PDs using sparse coefficient vectors

\[ x = \Psi \begin{bmatrix} 1 & \ldots & n & \ldots & N \end{bmatrix} \alpha \]

[Goordntisky and Rao 1997] [Malioutov, Cetin, Willsky 2005]
[Cevher, Duarte, Baraniuk 2008] [Cevher, Gurbuz, McClellan, Chellapa 2008][...]
Issues with Parametric Dictionaries

- As parameter resolution (e.g., number of grid points) increases, PD becomes increasingly coherent, hampering sparse approximation algorithms.
- PD’s high coherence is a manifestation of resolution issues in underlying estimation problem.

- **Structured sparsity models** can mitigate this issues by preventing PD elements with coherence above target maximum $\nu$ from appearing simultaneously in recovered signal.

![Near-Field Localization](image1.png) ![Far-Field Localization](image2.png)

[Duarte and Baraniuk 2010] [Duarte, 2012]
**Standard Sparse Signal Recovery**

**Iterative Hard Thresholding**

**Inputs:**
- Measurement vector $y$
- Measurement matrix $\Phi \Psi$
- Sparsity $K$

**Initialize:** $\hat{\alpha}_0 = 0, r = y, i = 0$

**While halting criterion false,**
- $i \leftarrow i + 1$
- $b \leftarrow \hat{\alpha}_{i-1} + \Psi^T \Phi^T r$ \hspace{1cm} (*estimate signal*)
- $\hat{\alpha}_i \leftarrow \mathcal{T}(b, K)$ \hspace{1cm} (*obtain best sparse approx.*)
- $r \leftarrow y - \Phi \Psi \hat{\alpha}_i$ \hspace{1cm} (*calculate residual*)

**Return estimate** $\hat{\alpha} = \hat{\alpha}_i$

[Blumensath and Davies 2009]
Structured Sparse Signal Recovery

Structured Iterative Hard Thresholding (SIHT)

**Inputs:**
- Measurement vector \( y \)
- Measurement matrix \( \Phi \Psi \)
- Structured sparse approx. algorithm \( \mathbb{M}(x, K) \)

**Initialize:** \( \hat{\alpha}_0 = 0, r = y, i = 0 \)

**While** halting criterion false,
- \( i \leftarrow i + 1 \)
- \( b \leftarrow \hat{\alpha}_{i-1} + \Psi^T \Phi^T r \) (estimate signal)
- \( \hat{\alpha}_i \leftarrow \mathbb{M}(b, K) \) (obtain best **structured** sparse approx.)
- \( r \leftarrow y - \Phi \Psi \hat{\alpha}_i \) (calculate residual)

**Return estimate** \( \hat{\alpha} = \hat{\alpha}_i \)

Can be applied to a variety of greedy algorithms (CoSaMP, OMP, Subspace Pursuit, etc.)

[Baraniuk, Cevher, Duarte, Hegde 2009]
Issues with Parametric Dictionaries

- Structured sparsity models need **careful** control of maximal coherence parameter $\nu$

- **Correlation function** provides measure of coherence between PD elements

- Correlation function connects parameter resolution to maximal coherence value $\nu$

- In most cases, coherence control equals **band exclusion**

[Duarte and Baraniuk, 2010] [Duarte, 2012] [Fannjiang and Liao, 2012]
Issues with Parametric Dictionaries

- Structured sparsity models need careful control of maximal coherence parameter $\nu$

- *Correlation function* provides measure of coherence between PD elements

- Correlation function *connects parameter resolution* to maximal coherence value $\nu$

- In most cases, coherence control equals *band exclusion*

---

[Duarte and Baraniuk, 2010] [Duarte, 2012] [Fannjiang and Liao, 2012]
Issues with PDs/Structured Sparsity: Sensitivity to Maximal Coherence Value

- Example: Compressive Time Delay Estimation (TDE) with PD and CS
- Performance depends on measurement ratio $\kappa = \frac{M}{N}$
- Structured sparsity used to enable high-resolution TDE

- Parameter $\nu$ set to **optimal value** for chirp of length 1 $\mu$s
- As length of chirp wave increases, performance of compressive TDE **varies widely**
- **Shape** of correlation function dependent on chirp length
Issues with PDs:
Euclidean Norm Guarantees

- Most recovery methods provide guarantees to keep Euclidean norm error $\|\alpha - \hat{\alpha}\|_2$ small.
- This metric, however, is not connected to quality of parameter estimates.
- Example: both estimates have same Euclidean error $\|\alpha - \hat{\alpha}_1\|_2 = \|\alpha - \hat{\alpha}_2\|_2$ but provide very different location estimates.
- We search for a performance metric better suited to the use of PD coefficient vectors (i.e., $\|\theta - \hat{\theta}\|$).
Improved Performance Metric: Earth Mover’s Distance

- Earth Mover’s Distance (EMD) is based on concept of “mass” flowing between the entries of first vector in order to match the second.
- EMD value is minimum “work” needed (measured as mass \(\times\) transport distance) for first vector to match second:

\[
\text{EMD}(\alpha, \hat{\alpha}) := \min_f \sum_{i,j=1}^{L} f_{ij} |i - j| \\
s.t. \sum_j f_{ij} = |\alpha_i| \quad \forall i = 1, \ldots, N, \\
\sum_i f_{ij} = |\hat{\alpha}_j| \quad \forall j = 1, \ldots, N.
\]
Improved Performance Metric: Earth Mover’s Distance

- When PDs are used, EMD captures *parameter estimation error* by measuring distance traveled by “mass”
- Parameter values must be *proportional to indices* in PD coefficient vector
- How to introduce EMD metric into CS recovery process?

\[
\text{EMD}(\alpha, \hat{\alpha}) := \min_f \sum_{i,j=1}^{L} f_{ij} |i - j| \\
\text{s.t.} \sum_{j} f_{ij} = |\alpha_i| \ \forall \ i = 1, \ldots, N, \\
\sum_{i} f_{ij} = |\hat{\alpha}_j| \ \forall \ j = 1, \ldots, N.
\]
Sparse Approximation with Earth Mover’s Distance

- To integrate into greedy algorithms, we will need to solve the **EMD-optimal K-sparse approximation problem**

\[
\hat{x}_K = \arg \min_{\bar{x} \in \Sigma_K} \text{EMD}(x, \bar{x})
\]

- It can be shown that approximation can be obtained by performing **K-median clustering** on set of points at locations \(\{1, \ldots, N\}\) with respective weights \(\{|x[1]|, \ldots, |x[N]|\}\)

- Cluster centroids provide **support** of \(\hat{x}_K\), values can be easily computed to minimize EMD/estimation error

[Indyk and Price 2009]
Greedy algorithms generally compute a signal “proxy”/residual of the form

\[ \tilde{x} = \Psi^T \Phi^T \psi y = \Psi^T \Phi^T \phi \psi \alpha \]

- For sufficiently large number of measurements, signal proxy will resemble correlation function convolved with original sparse coefficient vector.
- EMD sparse approximation (K-median clustering) converts each correlation function into single cluster with centroid at correlation peak.
- Small estimation biases may appear if translated correlation function is asymmetric or noise is present.
Structured Sparse Signal Recovery

**Band-Excluding IHT**

**Inputs:**
- Measurement vector $y$
- Measurement matrix $\Phi\Psi$
- Structured sparse approx. algorithm $\mathbb{M}(x, K)$

**Initialize:** $\hat{\alpha}_0 = 0$, $r = y$, $i = 0$

**While** halting criterion false,
- $i \leftarrow i + 1$
- $b \leftarrow \hat{\alpha}_{i-1} + \Psi^T\Phi^Tr$ (estimate signal)
- $\hat{\alpha}_i \leftarrow \mathbb{M}(b, K)$ (obtain **band-excluding** sparse approx.)
- $r \leftarrow y - \Phi\Psi\hat{\alpha}_i$ (calculate residual)

**Return** estimate $\hat{\alpha} = \hat{\alpha}_i$

Can be applied to a variety of greedy algorithms (CoSaMP, OMP, Subspace Pursuit, etc.)

[Baraniuk, Cevher, Duarte, Hegde 2009] [Fannjiang and Liao, 2012]
EMD + Sparse Signal Recovery

**Clustered IHT**

**Inputs:**
- Measurement vector $y$
- Measurement matrix $\Phi \Psi$
- Sparsity $K$

**Output:**
- PD coefficient estimate $\hat{\alpha}$

**Initialize:** $\hat{\alpha}_0 = 0, r = y, i = 0$

**While halting criterion false,**
- $i \leftarrow i + 1$
- $b \leftarrow \hat{\alpha}_{i-1} + \Psi^T \Phi^T r$
- $\hat{\alpha}_i \leftarrow \arg \min_{\bar{b} \in \Sigma_K} \text{EMD}(b, \bar{b})$ \hspace{1cm} (*estimate signal*)
- $r \leftarrow y - \Phi \Psi \hat{\alpha}_i$ \hspace{1cm} (*calculate residual*)

Return estimate $\hat{\alpha} = \hat{\alpha}_i$

Can be applied to a variety of greedy algorithms (CoSaMP, OMP, Subspace Pursuit, etc.)
Numerical Results

Example: Compressive TDE with PD & random projections

Performance depends on \textit{measurement ratio} $\kappa = \frac{M}{N}$

TDE performance \textit{varies widely} as chirp length increases

Consistent behavior for EMD-based signal recovery, but \textit{consistent bias observed}

Bias partially due to \textit{parameter space discretization}
Another PD Issue: Discretization

- Every PD can be conceived as a **sampling** from an **infinite set** of parametrizable signals $s(\theta)$ for a discrete set of parameter values $\Theta \in [\theta_{\text{min}}, \theta_{\text{max}}]$
- When signal vector $s(\theta)$ varies smoothly as a function of $\theta$, signal set can be represented by **nonlinear manifold**
- If manifold is well behaved, resolution can be improved by **interpolating** between PD samples

\[ \mathcal{M} = \{ s(\theta) : \theta \in [\theta_{\text{min}}, \theta_{\text{max}}] \} \]

[Fyhn, Dadkhahi, Duarte 2013] [Fyhn, Jensen, Duarte 2013]
Numerical Results

- Example: Compressive TDE with PD and CS
- Performance depends on measurement ratio $\kappa = M/N$
- When integrated with *polar interpolation*, performance of compressive TDE improves significantly
- Sensitivity of Band-Excluding SP becomes more *severe*, while Clustered SP *remains robust*
Conclusions

• Retrofitting sparsity via parametric dictionaries is not enough!
  – PDs enable use of CS, but often are coherent
  – band exclusion can help, but must be highly precise
  – issues remain with guarantees (Euclidean is not useful)
  – PDs also discretize parameter space, limiting resolution

• Earth Mover’s Distance is a suitable alternative
  – easily implementable by leveraging $K$-median clustering
  – EMD is suitable for dictionaries with well-behaved (compact) correlation functions
  – from PDs to manifolds via interpolation techniques
  – ongoing work: theoretical guarantees, bias issues, sensitivity to noise...
  – localization, bearing estimation, radar imaging, ...

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