Compressive Line Spectrum Estimation with Clustering and Interpolation

Dian Mo, Marco F. Duarte

Department of Electrical and Computer Engineering
University of Massachusetts Amherst
Spectrum Estimation

- **Observation:**
  - Noisy $y = x + w, w \sim \mathcal{N}(0, \sigma^2 I)$
  - Subsampled $y = x_T, T = \{t_1, t_2, ..., t_M\}$
  - Compressed $y = Ax, A \in \mathbb{R}^{M \times N}, M \ll N$

- **Goal:** estimate frequencies $f$ from signal observation $y$

- **Requirement:** simple spectrum structure – *sparsity*
Line Spectrum Estimation

- **Signal:**
  \[
  x[n] = \sum_{k=1}^{K} c_k e^{j2\pi f_k n}, n \in D = \{0, 1, 2, \ldots, N - 1\}
  \]

- **Spectrum:**
  \[
  X(f) = \sum_{k=1}^{K} c_k \delta(f - f_k), f_k \in \Theta = [0, 1]
  \]

- **Sparsity:** \( K \ll N \)

- **Observation:** \( y = x_T \)

[Candès and Fernández-Granda 2013][Tang, Bhaskar, Shah, and Recht 2013][...]
Atomic Norm

- Atoms: $a(f, \phi) = [e^{j(2\pi fn + \phi)}], n \in D$
- Atom set: $\mathcal{A} = \{a(f, \phi): f \in \Theta, \phi \in [0, 2\pi]\}$
- Atomic norm:

\[
\|x\|_\mathcal{A} = \inf\{t: x \in t\text{conv}(\mathcal{A})\} = \inf \left\{ \sum_{k=1}^{K} |c_k|: x = \sum_{k=1}^{K} |c_k| a(f_k, \phi_k) \right\}
\]

[Tang, Bhaskar, Shah, and Recht 2013]
Atomic Norm Minimization

- **Primal:**
  \[
  \min_x \|x\|_\mathcal{A}
  \text{ s.t. } x_T = y
  \]

- **Dual:**
  \[
  \min_{x, q} \langle q_T, y \rangle_{\mathbb{R}}
  \text{ s.t. } \|q\|_{\mathcal{A}}^* \leq 1, q_T c = 0
  \]

- **Recover** \(x\)

- **Estimate** \(f_k:\)
  \[
  |\langle \hat{q}, a(f, 0) \rangle| = 1 \iff f = f_k
  \]

[Tang, Bhaskar, Shah, and Recht 2013]


**Recovery Guarantee**

- Minimum separation: \( \zeta = \min_{i \neq j} |f_i - f_j| \)

- **Theorem:** Assume we select \( M \) samples from \( x \) uniformly at random. Under certain conditions, if \( \zeta \geq 4/N \), then one can guarantee with probability at least \( 1 - \delta \) that \( x \) can be uniquely recovered when
  \[
  M \geq C \max \left\{ \log^2 \frac{N}{\delta}, K \log \frac{K}{\delta} \log \frac{N}{\delta} \right\}
  \]

- Perfect recovery requires **sufficient separation**

  [Tang, Bhaskar, Shah, and Recht 2013]
Sparsity-Based Parameter Estimation

- Discrete parameters: \( \Omega = [\omega_1, \omega_2, ..., \omega_L] \subset \Theta \)
- Parametric Dictionary: \( \Psi = [\psi(\omega_1), \psi(\omega_2), ..., \psi(\omega_L)] \)
- Sparse signal: \( x = \Psi c \)

Parameter Estimation

Sparse Recovery

[Malisutov, Cetin, Willsky 2005] [Cevher, Duarte, Baraniuk 2008][...]

\( \Psi \)

\( \Theta \)

\( c \)
Parametric Dictionary Issues

- Discretization mismatch for arbitrary signal

- *Increasing dictionary coherence* with high resolution

- Sparse recovery relies on sparse approximation operation (hard thresholding)

- Euclidean norm guarantee on is *not connected to* parameter estimation error
Earth Mover’s Distance

- Earth mover’s distance measures *minimum “flow work” needed* for two vectors to be matched

\[
EMD(c, \hat{c}) = \min_g g_{ij} |i - j|
\]

s.t. \[
\sum_j g_{ij} = |c_i|, j = 1, 2, ..., L
\]

\[
\sum_i g_{ij} = |\hat{c}_j|, i = 1, 2, ..., L
\]

- EMD between coefficient vectors is proportion to estimation error

[Mo and Duarte 2015]
To integrate into greedy algorithms, we will need to solve the \textit{EMD-optimal K-sparse approximation} problem

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

Sparse approximation can be obtained by performing \textit{K-median clustering} on set of points at locations \{1, 2, ..., L\} with respective weights \{|c_1|, |c_2|, ..., |c_L|\}

Resulting centroids indicate \textit{support} of \(\hat{c}_K\)

[Indyk and Price 2009]
Performance Guarantee

- **Important factors:**
  - Autocorrelation $\lambda(\omega) = \psi^*(f)\psi(f + \omega)$
  - Cumulative autocorrelation $\Lambda(\omega) = \sum_{\theta \leq \omega} |\lambda(\theta)|$
  - Minimum separation: $\zeta = \min_{i \neq j}|f_i - f_j|$
  - Dynamic range: $r = \max_{i \neq j} |c_i|/|c_j|$

- **Theorem:** Let $\Delta$ be sufficiently small, if
  \[ \zeta \geq 2\Lambda^{-1} \left( 2\Lambda(0) \left( 1 - \frac{\Lambda(\sigma)/\Lambda(0) - 1}{2(K - 1)r + 1} \right) \right) \]
  Then estimation error is smaller than $\sigma$

- Clustering also requires **well separated** frequencies

[Mo and Duarte 2015]
EMD Frequency Estimation

- Approximation be applied to a variety of greedy algorithms (CoSaMP, IHT, **Subspace Pursuit**, etc.)
- **Input:** measurement vector \( y \), measurement matrix \( A \), PD \( \Psi \), clustering sparse approximation \( \mathcal{C}(x, K) \)
- **Output:** estimated parameter \( \hat{f} \), estimated signal \( \hat{x} \)
- **Initialize:** \( \hat{x} = 0 \), \( \hat{f} = \emptyset \)
- **Repeat:**
  - \( v = (A\Psi)^*(y - A\hat{x}) \) \{Obtain Signal Residual\}
  - \( \hat{f} = \hat{f} \cup \mathcal{C}(v, K) \) \{Augment estimates\}
  - \( \Psi_s = \psi(f) \) \{Obtain PD subset\}
  - \( c = (A\Psi_s)^+y \) \{Obtain estimate Proxy\}
  - \( \hat{f} = \mathcal{C}(c, K) \) \{Refine estimates\}

[Mo and Duarte 2014]
Atomic Norm vs. K-Median Clustering

- **Similarities:**
  - Both exploit signal sparsity models in frequency domain
  - Both require sufficient separation between frequencies

- **Differences**
  - Atomic norm use optimization; Clustering uses greedy algorithm
  - Atomic norm performs off-grid estimation; Clustering performs on-grid estimation *(interpolation can help)*
  - Clustering works on both subsampled and compressed observations

\[
x_T = Ax \iff A = \begin{cases} 
1, & j = t_i \\
0, & j \neq t_i 
\end{cases} \quad i = 1, 2, \ldots, M, t_i \in T
\]
Numerical Result

- Noiseless case with signal Length $N = 100$ and sinusoids number $K = 4$
- Estimation error depends on subsampling ratio $\kappa = M/N$
- Both subspace pursuit algorithms *cannot improve over* limit of sampling step $\Delta$
- Polar interpolation greatly improves greedy algorithm [Fyhn, Duarte, and Jensen 2015]

**Numerical Result**

- CSP : Clustering Subspace Pursuit
- BSP : Band-Exclusion Subspace Pursuit
- CISP : CSP with Interpolation
- BISP : BSP with Interpolation
- SDP : Atomic Norm Minimization (Semidefinite Program)

![Graph showing numerical results](image.png)
Numerical Results

- Noiseless case with $\kappa = 0.4$
- Estimation time depends on signal length
- SDP requires rapidly increasing computation time
- Clustering does not incur significant penalty in computation time
Summary

- Parametric dictionaries converts parameter estimation to sparse recovery
- EMD provides improved metric for sparsity-based parameter estimation
- $K$-median clustering achieves EMD-optimal sparse approximation
- Clustering methods has compatible error as Atomic norm methods
- Clustering methods shows advantage in time consuming and flexibility for different problems

Email: mo@umass.edu    Website: people.umass.edu/mo/