Robust Spectral Compressed Sensing via Structured Matrix Completion

Yuxin Chen
Electrical Engineering, Stanford University

Joint Work with Yuejie Chi
Fourier representation of a signal:

\[ x(t) = \sum_{i=1}^{r} d_i e^{j2\pi\langle t, f_i \rangle} \]

\((f_i : \text{frequencies}, \ d_i : \text{amplitudes}, \ r: \text{model order})\)

- **Sparsity**: nature is (approximately) sparse (small \(r\))
- **Goal**: identify the underlying frequencies from time-domain samples
Applications in Sensing

- **Multipath channels**: a (relatively) small number of strong paths.

- **Radar Target Identification**: a (relatively) small number of strong scatters.
Applications in Imaging

- Consider a time-sparse signal (a dual problem)

\[ z(t) = \sum_{i=1}^{r} d_i \delta(t - t_i) \]

- Resolution is limited by the point spread function of the imaging system

\[ \text{image} = z(t) \ast \text{PSF}(t) \]
Data Model

- **Signal Model:** a mixture of $K$-dimensional sinusoids at $r$ distinct frequencies

$$x(t) = \sum_{i=1}^{r} d_ie^{j2\pi\langle t,f_i \rangle}$$

where $f_i \in [0, 1]^K$ : frequencies; $d_i$ : amplitudes.

- **Observed Data:**

$$X = [x_{i_1, \ldots, i_K}] \in \mathbb{C}^{n_1 \times \ldots \times n_K}$$

  - **Continuous dictionary:** $f_i$ can assume ANY value in a unit disk
  - **Multi-dimensional model:** $f_i$ can be multi-dimensional
  - **Low-rate Data Acquisition:** obtain partial samples of $X$

- **Goal:** Identify the frequencies from partial measurements
Prior Art

- **Parametric Estimation:** *(shift-invariance of harmonic structures)*
  - Prony’s method, ESPRIT [RoyKailath’1989], Matrix Pencil [Hua’1992], Finite rate of innovation [DragottiVetterliBlu’2007][GedalyahuTurEldar’2011]...
  - perfect recovery from equi-spaced $O(r)$ samples
  - sensitive to noise and outliers
  - require prior knowledge on the model order.

- **Compressed Sensing:**
  - Discretize the frequency and assume a sparse representation
    
    $$ f_i \in \mathcal{F} = \left\{ \frac{0}{n_1}, \ldots, \frac{n_1 - 1}{n_1} \right\} \times \left\{ \frac{0}{n_2}, \ldots, \frac{n_2 - 1}{n_2} \right\} \times \ldots $$

  - perfect recovery from $O(r \log n)$ random samples
  - non-parametric approach
  - robust against noise and outliers
  - sensitive to gridding error
Basis Mismatch / Gridding Error

- A toy example: $x(t) = e^{j2\pi f_0 t}$:
  - The success of CS relies on sparsity in the DFT basis.
  - **Basis mismatch**: discrete v.s. continuous dictionary
    * Mismtach $\Rightarrow$ kernel leakage $\Rightarrow$ failure of CS (basis pursuit)

- Finer gridding does not help [ChiScharfPezeshkiCalderbank’2011]
Two Recent Landmarks in Off-the-grid Harmonic Retrieval (1-D)

- **Super-Resolution** (CandesFernandezGranda’2012)
  - Low-pass measurements
  - Total-variation norm minimization

- **Compressed Sensing Off the Grid** (TangBhaskarShahRecht’2012)
  - Random measurements
  - Atomic norm minimization
  - Require only $O(r \log r \log n)$ samples

**QUESTIONS:**
- How to deal with *multi-dimensional frequencies*?
- Robustness against *outliers*?
Our Objective

- **Goal:** seek an algorithm of the following properties
  - non-parametric
  - works for *multi-dimensional frequency model*
  - works for *off-the-grid frequencies*
  - requires a minimal number of measurements
  - robust against noise and sparse outliers
Concrete Example: 2-D Frequency Model

recall that \( x(t) = \sum_{i=1}^{r} d_i e^{j2\pi \langle t, f_i \rangle} \)

- For 2-D frequencies, we have the \textbf{Vandermonde decomposition}:

\[
X = Y \cdot D \cdot Z^T.
\]

diagonal matrix

Here, \( D := \text{diag} \{d_1, \cdots, d_r\} \) and

\[
Y := \begin{bmatrix}
1 & 1 & \cdots & 1 \\
1 & y_2 & \cdots & y_r \\
\vdots & \vdots & \ddots & \vdots \\
y_1 & y_2^{n_1-1} & \cdots & y_r^{n_1-1}
\end{bmatrix},
Z := \begin{bmatrix}
1 & 1 & \cdots & 1 \\
z_1 & z_2 & \cdots & z_r \\
\vdots & \vdots & \ddots & \vdots \\
z_1^{n_2-1} & z_2^{n_2-1} & \cdots & z_r^{n_2-1}
\end{bmatrix}
\]

\textbf{Vandemonde matrix}

where \( y_i = \exp(j2\pi f_{1i}) \), \( z_i = \exp(j2\pi f_{2i}) \).

- \textbf{Spectral Sparsity} \( \Rightarrow \) \( X \) may be low-rank for very small \( r \)
- \textbf{Reduced-rate Sampling} \( \Rightarrow \) observe partial entries of \( X \)
Matrix Completion?

\[ X = Y \cdot D \cdot Z^T, \]

where \( D := \text{diag}\{d_1, \ldots, d_r\} \), and

\[ Y := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_1 & y_2 & \cdots & y_r \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{n_1-1} & y_2^{n_1-1} & \cdots & y_r^{n_1-1} \end{bmatrix}, \quad Z := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_r \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{n_2-1} & z_2^{n_2-1} & \cdots & z_r^{n_2-1} \end{bmatrix} \]

- Question: can we apply Matrix Completion algorithms directly on \( X \)?

  - Yes, but it yields sub-optimal performance.
    - requires at least \( r \max\{n_1, n_2\} \) samples.
    - \( X \) is no longer low-rank if \( r > \min\{n_1, n_2\} \)
      * note that \( r \) can be as large as \( n_1 n_2 \)
  - Call for more effective forms.
Rethink Matrix Pencil: Matrix Enhancement

- An enhanced form \( X_e: \left( k_1 \times (n_1 - k_1 + 1) \right) \) block Hankel matrix [Hua’1992]

\[
X_e = \begin{bmatrix}
X_0 & X_1 & \cdots & X_{n_1-k_1} \\
X_1 & X_2 & \cdots & X_{n_1-k_1+1} \\
\vdots & \vdots & \ddots & \vdots \\
X_{k_1-1} & X_k & \cdots & X_{n_1-1}
\end{bmatrix},
\]

where each block is a \( k_2 \times (n_2 - k_2 + 1) \) Hankel matrix as follows

\[
X_l = \begin{bmatrix}
x_{l,0} & x_{l,1} & \cdots & x_{l,n_2-k_2} \\
x_{l,1} & x_{l,2} & \cdots & x_{l,n_2-k_2+1} \\
\vdots & \vdots & \ddots & \vdots \\
x_{l,k_2-1} & x_{l,k_2} & \cdots & x_{l,n_2-1}
\end{bmatrix}.
\]

- Incentive:
  - Lift the matrix to promote Harmonic Structure
  - Convert Sparsity to Low Rank
Low-Rank Structure of the Enhanced Matrix

- The enhanced matrix can be decomposed as follows.

\[
X_e = \begin{bmatrix}
Z_L \\
Z_L Y_d \\
\vdots \\
Z_L Y_d^{k_1 - 1}
\end{bmatrix} D \begin{bmatrix}
Z_R, Y_d Z_R, \cdots, Y_d^{n_1 - k_1} Z_R
\end{bmatrix},
\]

- \( Z_L \) and \( Z_R \) are Vandermonde matrices specified by \( z_1, \ldots, z_r \),
- \( Y_d = \text{diag}[y_1, y_2, \cdots, y_r] \).

- The enhanced form \( X_e \) is low-rank.
  - \( \text{rank} (X_e) \leq r \)
  - Spectral Sparsity \( \Rightarrow \) Low Rank
• Our recovery algorithm: Enhanced Matrix Completion (EMaC)

\[
\text{(EMaC)}: \quad \min_{M \in \mathbb{C}^{n_1 \times n_2}} \|M_e\|_* \\
\text{subject to} \quad M_{i,j} = X_{i,j}, \forall (i,j) \in \Omega
\]

where $\Omega$ denotes the sampling set, and $\| \cdot \|$ denotes the nuclear norm.

○ nuclear norm minimization (convex)

• existing MC result won’t apply – requires at least $O(nr)$ samples

• **Question**: How many samples do we need?
Coherence Measures

- **Notations:** $G_L$ is an $r \times r$ Gram matrices such that

$$\left( G_L \right)_{il} := \left\langle y^{(i)}, y^{(l)} \right\rangle \left\langle z^{(i)}, z^{(l)} \right\rangle$$

where $y^{(i)} := (1, y_i, y_i^2, \cdots, y_i^{k_1-1})$ and $y_i := e^{j2\pi f_i}$.

$z^{(i)}$ and $G_R$ are similarly defined with different dimensions...

- **Incoherence property** arises w.r.t. $\mu_1$ if

$$\sigma_{\min}(G_L) \geq \frac{1}{\mu_1}, \quad \sigma_{\min}(G_R) \geq \frac{1}{\mu_1}.$$  

- **Examples:**
  - Randomly generated frequencies
  - (Mild) perturbation of grid points
Theoretical Guarantees for Noiseless Case

- **Theorem 1 (Noiseless Samples)** Let $n = n_1 n_2$. If all measurements are noiseless, then EMaC recovers $X$ with high probability if:

  $$m \sim \Theta(\mu_1 r \log^3 n);$$

- **Implications**
  - minimum sample complexity: $O(r \log^3 n)$.
  - general theoretical guarantees for Hankel (Toeplitz) matrix completion. — see applications in control, MRI, natural language processing, etc
Proof Sketch: Inexact Dual + Golfing Scheme

Construct a relaxed dual certificate

- **Lemma (Relaxed Duality):** Let $T$ be the tangent space w.r.t. $X_e$. Suppose
  - $\Omega$ restricted to $T \cap \text{Hankel}$ is injective.

If there exists a matrix $W \in \text{Hankel}^\perp \cup \Omega^\perp$ that satisfies

$$\|P_T(W)\|_F \leq \frac{1}{2n^2}, \quad \text{and} \quad \|P_{T^\perp}(W)\| \leq \frac{1}{2},$$

then $X_e$ is the unique optimizer of EMaC.

- **Construction of dual certificate**
  - *the clever “golfing scheme”* introduced by D. Gross [Gross'2011].
Figure 1: Phase transition diagrams where spike locations are randomly generated. The results are shown for the case where $n_1 = n_2 = 15$. 

Phase Transition
Algorithm 1 Singular Value Thresholding for EMaC

1: initialize Set $M_0 = X_e$ and $t = 0$.
2: repeat
3: 1) $Q_t \leftarrow D_{\tau t}(M_t)$ (singular-value thresholding)
4: 2) $M_t \leftarrow \text{Hankel}_{X_0}(Q_t)$ (projection onto a Hankel matrix consistent with observation)
5: 3) $t \leftarrow t + 1$
6: until convergence

Figure 2: dimension: $101 \times 101$, $r = 30$, $\frac{m}{n_1 n_2} = 5.8\%$, signal-to-amplitude-ratio = 10.
Robustness to Sparse Outliers *(a brief discussion)*

- What if a constant portion of measurements are arbitrarily corrupted?
  - *Robust PCA approach* [CandesLiMaWright’2011]
  - Solve instead the following algorithm:

  \[
  \text{(RobustEMaC)} : \quad \text{minimize} \quad \|M_e\|_* + \lambda \|S_e\|_1 \\
  \text{subject to} \quad (M + S)_{i,l} = X_{i,l}^{\text{corrupted}}, \quad \forall (i, l) \in \Omega
  \]

- **Theorem 2 (Sparse Outliers)** Set \( \lambda = 1/\sqrt{m \log n} \), and outlier rate \( \leq 20\% \). Then RobustEMaC recovers \( X \) with high probability if

  \[
  m \sim \Theta(\mu_1^2 r^2 \log^3 n)
  \]

- **Robust to a constant portion of outliers!**
Super Resolution (2-D)

• Obtain low pass components ⇒ Extrapolate to high frequencies [CandesFernandezGranda’2012]

  (a) spatial illustration  (b) frequency extrapolation

• Might attempt 2-D super-resolution using EMaC...

  (a) Ground Truth  (b) Low Resolution Image  (c) Super-Resolution via EMaC
• Connect spectral compressed sensing with matrix completion

\[
\begin{bmatrix}
\sqrt{\mathbf{a}} \\
\sqrt{\mathbf{b}} \\
\sqrt{\mathbf{c}} \\
\sqrt{\mathbf{d}} \\
\sqrt{\mathbf{e}} \\
\sqrt{\mathbf{f}} \\
\sqrt{\mathbf{g}} \\
\sqrt{\mathbf{h}} \\
\sqrt{\mathbf{i}} \\
\end{bmatrix}
\]

\[
\mathbf{Σ}_{\text{decompose}} =
\begin{bmatrix}
\sqrt{\mathbf{a}} \\
\sqrt{\mathbf{b}} \\
\sqrt{\mathbf{c}} \\
\sqrt{\mathbf{d}} \\
\sqrt{\mathbf{e}} \\
\sqrt{\mathbf{f}} \\
\sqrt{\mathbf{g}} \\
\sqrt{\mathbf{h}} \\
\sqrt{\mathbf{i}} \\
\end{bmatrix}
\]

\[
\mathbf{Σ}_{\text{0}} + \mathbf{H} \mathbf{Σ}_{\text{0}} + \mathbf{H} \mathbf{K} \mathbf{H}
\]

\[
\vec{W} \mathbf{H} \mathbf{K} \mathbf{H}
\]

• Connect traditional approach (parametric harmonic retrieval) with recent advance (MC)

• Future work: performance guarantees for 2-D super resolution?
Q&A

Preprints available at arXiv:

Robust Spectral Compressed Sensing via Structured Matrix Completion

http://arxiv.org/abs/1304.8126

Thank You! Questions?
References (a partial list)


