July 2013

Robust Spectral Compressed Sensing via Structured Matrix Completion

Yuxin Chen
Electrical Engineering, Stanford University

Joint Work with Yuejie Chi

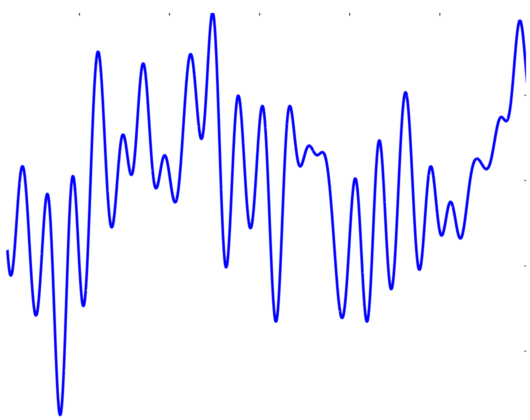
Sparse Fourier Representation/Approximation

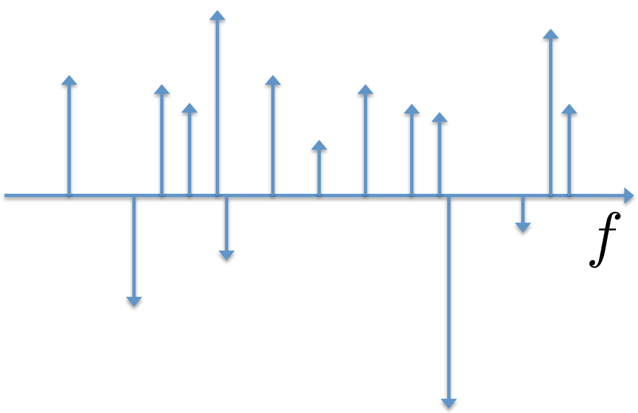
Fourier representation of a signal:

$$x(\boldsymbol{t}) = \sum_{i=1}^{r} d_i e^{j2\pi\langle \boldsymbol{t}, \boldsymbol{f}_i \rangle}$$

($\boldsymbol{f}_i$: frequencies, d_i : amplitudes, r: model order)

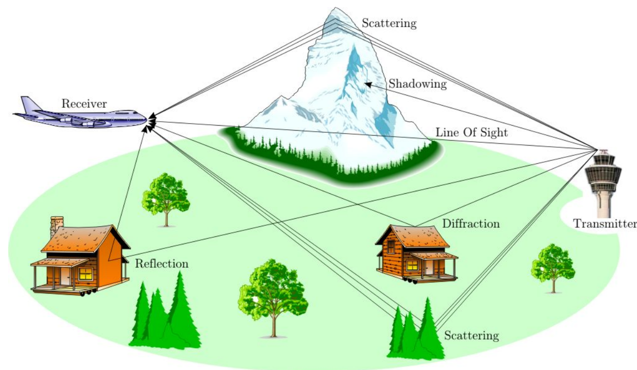
- Sparsity: nature is (approximately) sparse (small r)
- Goal: identify the underlying frequencies from time-domain samples



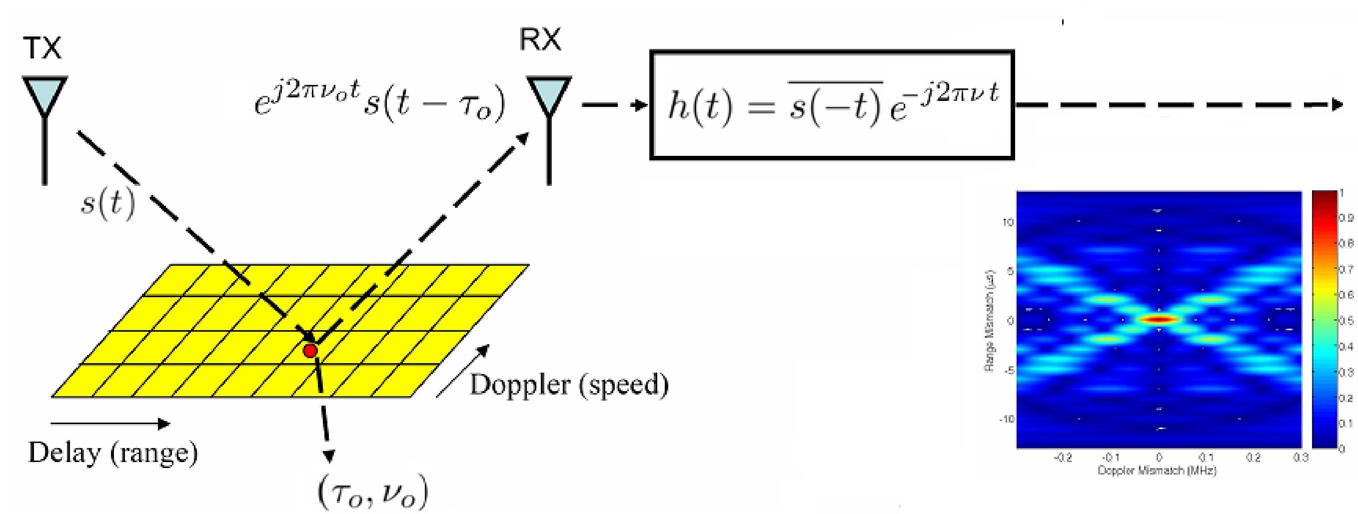


Applications in Sensing

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



- Radar Target Identification: a (relatively) small number of strong scatters.



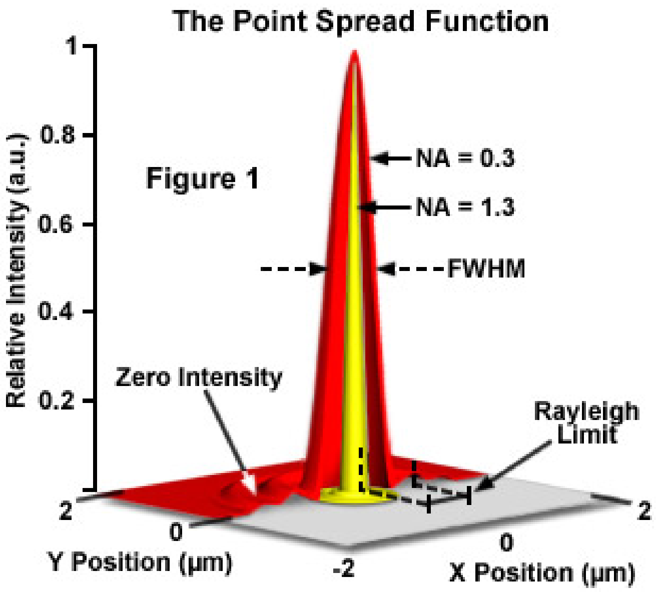
Applications in Imaging

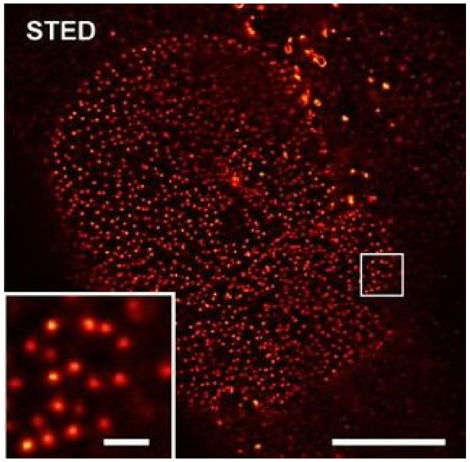
- Consider a time-sparse signal (a dual problem)

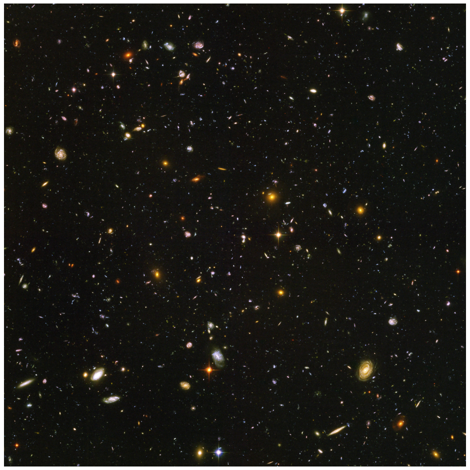
$$z(\boldsymbol{t}) = \sum_{i=1}^{r} d_i \delta(t - \boldsymbol{t}_i)$$

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

$$\text{image} = z(\boldsymbol{t}) * \text{PSF}(t)$$







Data Model

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

$$x(\boldsymbol{t}) = \sum_{i=1}^{r} d_i e^{j2\pi\langle \boldsymbol{t}, \boldsymbol{f}_i \rangle}$$

where $\boldsymbol{f}_i \in [0,1]^K$: frequencies; d_i : amplitudes.

- Observed Data:

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

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

- **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.

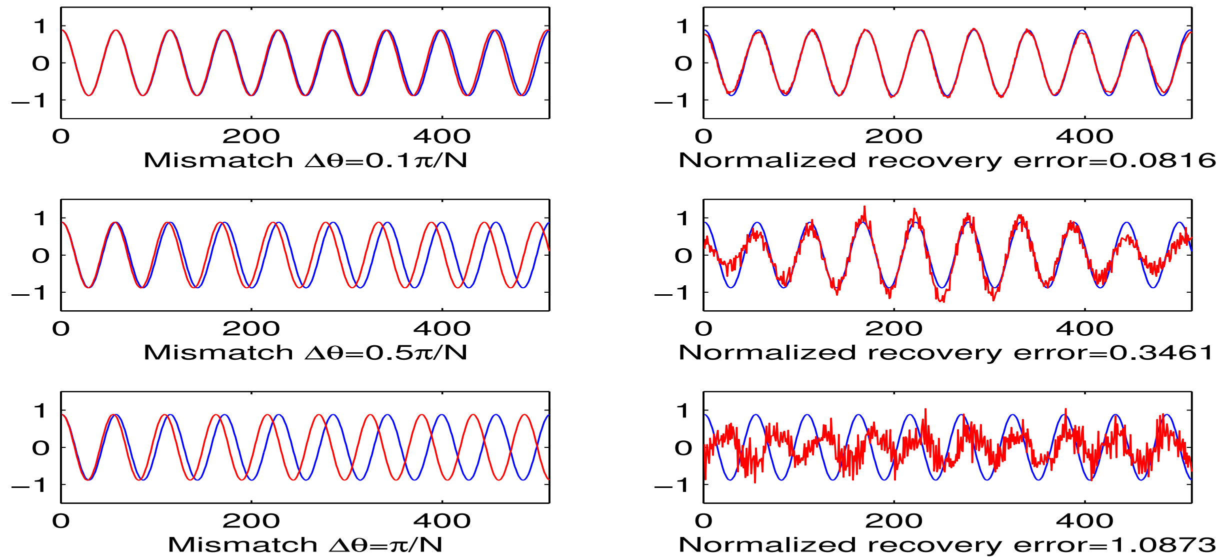
- **Compressed Sensing:**
  - Discretize the frequency and assume a sparse representation

$$f_i \in \mathcal{F} = \left\{ \frac{0}{n_1}, \dots, \frac{n_1 - 1}{n_1} \right\} \times \left\{ \frac{0}{n_2}, \dots, \frac{n_2 - 1}{n_2} \right\} \times \dots$$

  - 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: $\boldsymbol{x}(t) = e^{j2\pi f_0 t}$:
  - 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)*

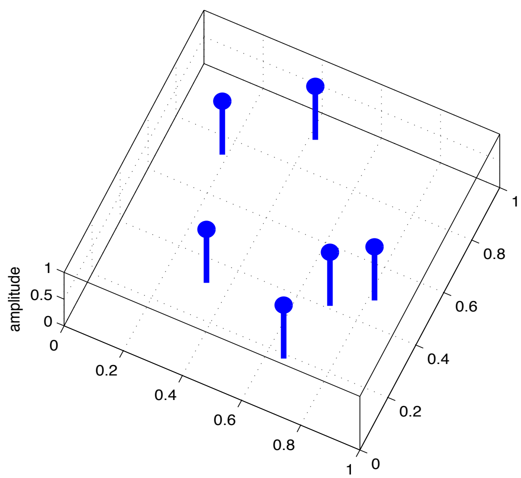


  - Finer gridding does not help [ChiScharfPezeshkiCalderbank'2011]

Two Recent Landmarks in Off-the-grid Harmonic Retrieval (1-D)

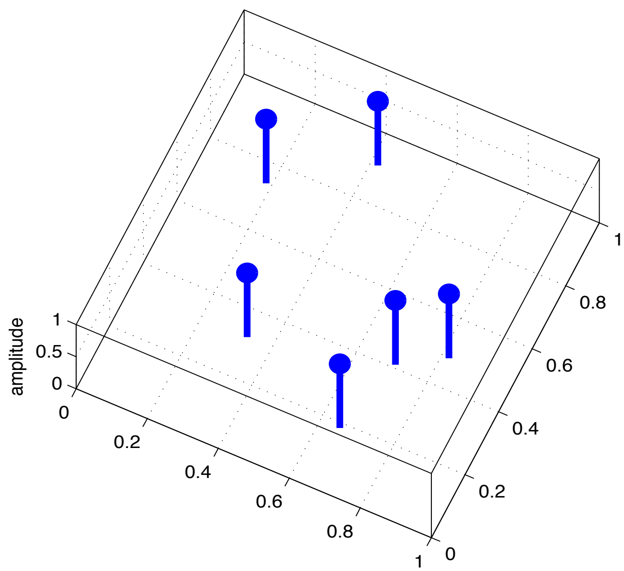
- **Super-Resolution** (CandesFernandezGranda'2012)
  - Low-pass measurements
  - Total-variation norm minimization

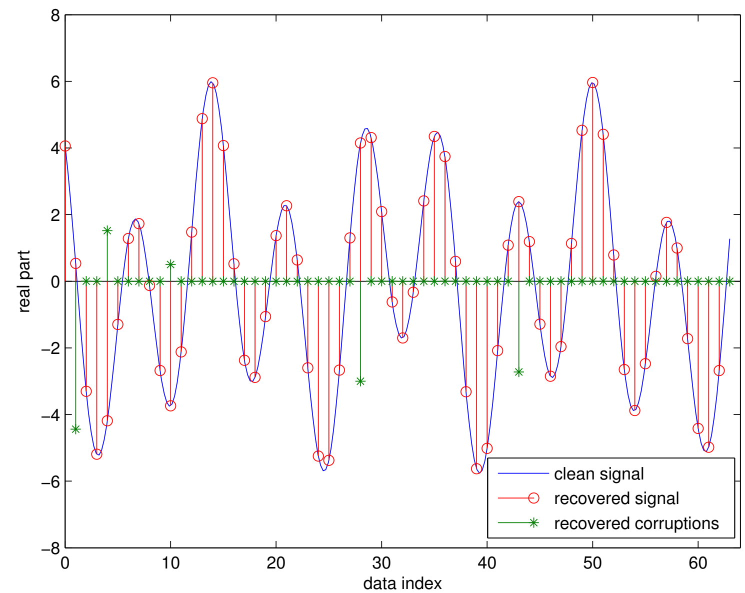
- **Compressed Sensing Off the Grid** (TangBhaskarShahRecht'2012)
  - Random measurements
  - Atomic norm minimization
  - Require only $\mathcal{O}(r \log r \log n)$ samples



- **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 $x(\boldsymbol{t}) = \sum_{i=1}^{r} d_i e^{j2\pi \langle \boldsymbol{t}, \boldsymbol{f}_i \rangle}$

- For 2-D frequencies, we have the ***Vandermonde decomposition***:

$$\boldsymbol{X} = \boldsymbol{Y} \cdot \underbrace{\boldsymbol{D}}_{\textbf{diagonal matrix}} \cdot \boldsymbol{Z}^T.$$

Here, $\boldsymbol{D} :=$ diag $\{d_1, \cdots, d_r\}$ and

$$\boldsymbol{Y} := \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_1 & y_2 & \cdots & y_r \\ \vdots & \vdots & \vdots & \vdots \\ y_1^{n_1-1} & y_2^{n_1-1} & \cdots & y_r^{n_1-1} \end{bmatrix}}_{\textbf{Vandemonde matrix}}, \boldsymbol{Z} := \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_r \\ \vdots & \vdots & \vdots & \vdots \\ z_1^{n_2-1} & z_2^{n_2-1} & \cdots & z_r^{n_2-1} \end{bmatrix}}_{\textbf{Vandemonde matrix}}$$

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

- **Spectral Sparsity** $\Rightarrow$ $\boldsymbol{X}$ may be *low-rank* for very small r
- **Reduced-rate Sampling** $\Rightarrow$ observe *partial entries* of $\boldsymbol{X}$

Matrix Completion?

$$\textit{recall that } \boldsymbol{X} = \underbrace{\boldsymbol{Y}}_{\textbf{Vandemonde}} \cdot \underbrace{\boldsymbol{D}}_{\textbf{diagonal}} \cdot \underbrace{\boldsymbol{Z}^T}_{\textbf{Vandemonde}} .$$

where $\boldsymbol{D} := \text{diag}\{d_1, \cdots, d_r\}$, and

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

- Question: can we apply *Matrix Completion* algorithms directly on $\boldsymbol{X}$?

$$\begin{bmatrix} \surd & ? & ? & \surd & \surd \\ ? & \surd & ? & \surd & \surd \\ ? & ? & \surd & \surd & ? \\ \surd & \surd & \surd & \surd & ? \\ \surd & \surd & ? & ? & \surd \end{bmatrix}$$

- Yes, but it yields sub-optimal performance.
  - requires at least $r \max\{n_1, n_2\}$ samples.
  - $\boldsymbol{X}$ 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.

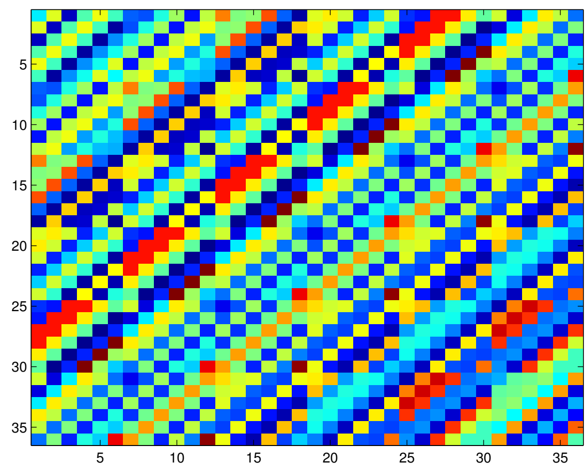
Rethink Matrix Pencil: Matrix Enhancement

- An **enhanced form** $\boldsymbol{X}_{\mathsf{e}}$: $(k_1 \times (n_1 - k_1 + 1)$ **block Hankel** *matrix [Hua'1992]*)

$$\boldsymbol{X}_{\mathsf{e}} = \begin{bmatrix} \boldsymbol{X}_0 & \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_{n_1-k_1} \\ \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_{n_1-k_1+1} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{X}_{k_1-1} & \boldsymbol{X}_{k_1} & \cdots & \boldsymbol{X}_{n_1-1} \end{bmatrix},$$

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

$$\boldsymbol{X}_l = \begin{bmatrix} x_{l,0} & x_{l,1} & \cdots & x_{l,n_2-k_2} \\ x_{l,1} & x_{l,2} & \cdots & x_{l,n_2-k_2+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{l,k_2-1} & x_{l,k_2} & \cdots & x_{l,n_2-1} \end{bmatrix}.$$



- **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.

$$\boldsymbol{X}_{\mathsf{e}} = \begin{bmatrix} \boldsymbol{Z}_{\mathsf{L}} \\ \boldsymbol{Z}_{\mathsf{L}}\boldsymbol{Y}_{\mathsf{d}} \\ \vdots \\ \boldsymbol{Z}_{\mathsf{L}}\boldsymbol{Y}_{\mathsf{d}}^{k_1-1} \end{bmatrix} \boldsymbol{D} \left[\boldsymbol{Z}_{\mathsf{R}}, \boldsymbol{Y}_{\mathsf{d}}\boldsymbol{Z}_{\mathsf{R}}, \cdots, \boldsymbol{Y}_{\mathsf{d}}^{n_1-k_1}\boldsymbol{Z}_{\mathsf{R}} \right],$$

  - $\boldsymbol{Z}_{\mathsf{L}}$ and $\boldsymbol{Z}_{\mathsf{R}}$ are Vandermonde matrices specified by $z_1, \ldots, z_r$,
  - $\boldsymbol{Y}_{\mathsf{d}} = \mathrm{diag}\left[y_1, y_2, \cdots, y_r\right]$.

- The enhanced form $\boldsymbol{X}_{\mathsf{e}}$ is low-rank.
  - rank $(\boldsymbol{X}_{\mathsf{e}}) \leq r$
  - Spectral Sparsity $\Rightarrow$ Low Rank

Enhancement Matrix Completion (EMaC)

- Our recovery algorithm: Enhanced Matrix Completion (EMaC)

$$
\text{(EMaC)}: \quad \underset{\boldsymbol{M}\in\mathbb{C}^{n_1\times n_2}}{\text{minimize}} \quad \|\boldsymbol{M}_{\mathsf{e}}\|_* \\
\text{subject to} \quad \boldsymbol{M}_{i,j} = \boldsymbol{X}_{i,j}, \forall (i,j)\in\Omega
$$

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

  - *nuclear norm minimization (convex)*

- existing MC result won't apply – requires at least $\mathcal{O}(nr)$ samples

- **Question**: How many samples do we need?

$$
\begin{bmatrix}
? & \surd & \surd & ? & \surd & ? & \surd & \surd & ? & ? & \surd & \surd \\
\surd & \surd & ? & ? & ? & \surd & \surd & ? & ? & \surd & \surd & ? \\
\surd & ? & ? & \surd & \surd & \surd & ? & \surd & \surd & \surd & ? & \surd \\
? & ? & \surd & ? & \surd & ? & \surd & ? & \surd & ? & \surd & \surd \\
\surd & ? & \surd & \surd & ? & ? & \surd & \surd & \surd & ? & \surd & \surd \\
? & \surd & \surd & ? & ? & \surd & \surd & ? & ? & \surd & \surd & ? \\
\surd & \surd & ? & \surd & \surd & \surd & ? & \surd & \surd & \surd & ? & \surd \\
\surd & ? & \surd & ? & \surd & ? & \surd & \surd & \surd & ? & \surd & ? \\
? & ? & \surd & \surd & \surd & ? & \surd & \surd & ? & ? & \surd & ? \\
? & \surd & \surd & ? & ? & \surd & \surd & ? & ? & \surd & ? & ? \\
\surd & \surd & ? & \surd & \surd & \surd & ? & \surd & \surd & ? & ? & \surd \\
\surd & ? & \surd & \surd & \surd & ? & \surd & ? & ? & ? & \surd & ?
\end{bmatrix}
$$

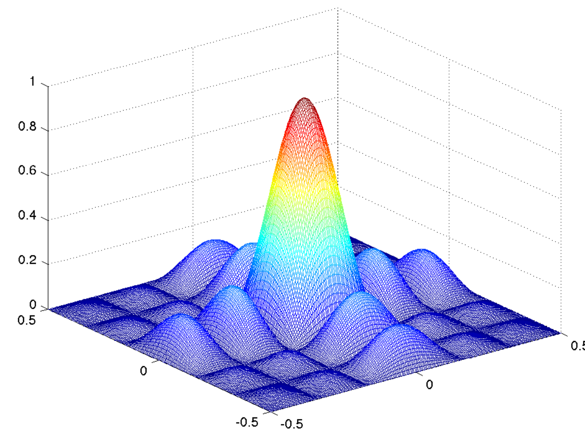
Coherence Measures

- Notations: $\boldsymbol{G}_{\mathsf{L}}$ is an $r \times r$ Gram matrices such that

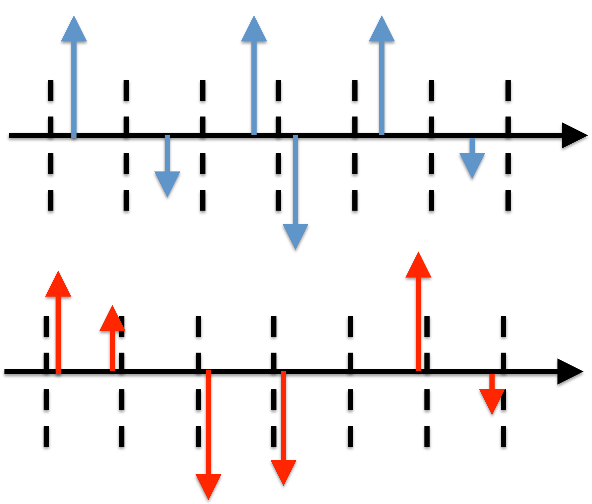
$$(\boldsymbol{G}_{\mathsf{L}})_{il} := \left\langle \boldsymbol{y}^{(i)}, \boldsymbol{y}^{(l)} \right\rangle \left\langle \boldsymbol{z}^{(i)}, \boldsymbol{z}^{(l)} \right\rangle$$

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

$\boldsymbol{z}^{(i)}$ and $\boldsymbol{G}_R$ are similarly defined with different dimensions...



Dirichlet Kernel



- **Incoherence property** arises w.r.t. μ_1 if

$$\sigma_{\min}(\boldsymbol{G}_L) \geq \frac{1}{\mu_1}, \quad \sigma_{\min}(\boldsymbol{G}_R) \geq \frac{1}{\mu_1}.$$

- 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 $\boldsymbol{X}$ with high probability if:

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

- **Implications**
  - minimum sample complexity: $\mathcal{O}(r \log^3 n)$.
  - general theoretical guarantees for Hankel (Toeplitz) matrix completion.
    — *see applications in control, MRI, natural languange 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. $\boldsymbol{X}_{\mathsf{e}}$. Suppose
  - Ω restricted to $T \cap$ Hankel is injective.

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

$$\|\mathcal{P}_T(\boldsymbol{W})\|_{\mathsf{F}} \leq \frac{1}{2n^2}, \quad \text{and} \quad \|\mathcal{P}_{T^{\perp}}(\boldsymbol{W})\| \leq \frac{1}{2},$$

  then $\boldsymbol{X}_{\mathsf{e}}$ is the unique optimizer of EMaC.

- **Construction of dual certificate**
  - *the clever "golfing scheme" introduced by D. Gross [Gross'2011].*

Phase Transition

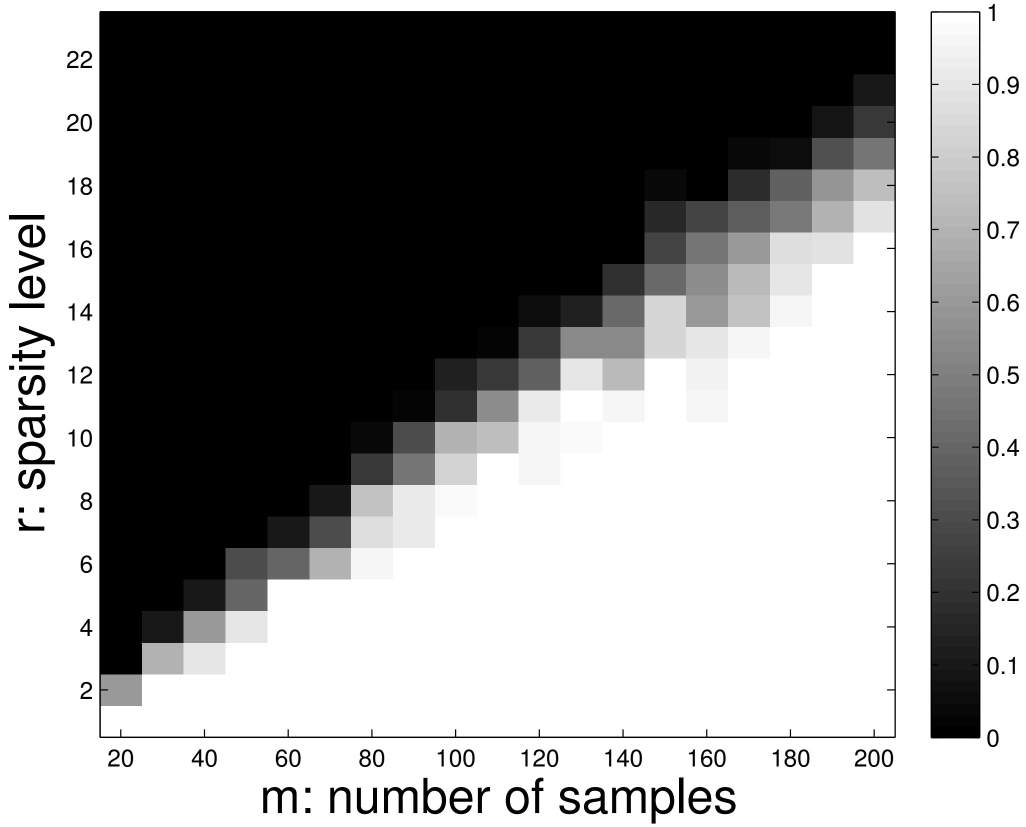


Figure 1: Phase transition diagrams where spike locations are randomly generated. The results are shown for the case where $n_1 = n_2 = 15$.

Singular Value Thresholding (Noisy Case)

Algorithm 1 Singular Value Thresholding for EMaC

1: **initialize** Set $\boldsymbol{M}_0 = \boldsymbol{X}_{\mathrm{e}}$ and $t = 0$.
2: **repeat**
3: 1) $\boldsymbol{Q}_t \leftarrow \mathcal{D}_{\tau_t}\left(\boldsymbol{M}_t\right)$ (*singular-value thresholding*)
4: 2) $\boldsymbol{M}_t \leftarrow \mathrm{Hankel}_{\boldsymbol{X}_0}(\boldsymbol{Q}_t)$ (*projection onto a Hankel matrix consistent with observation*)
5: 3) $t \leftarrow t + 1$
6: **until** convergence

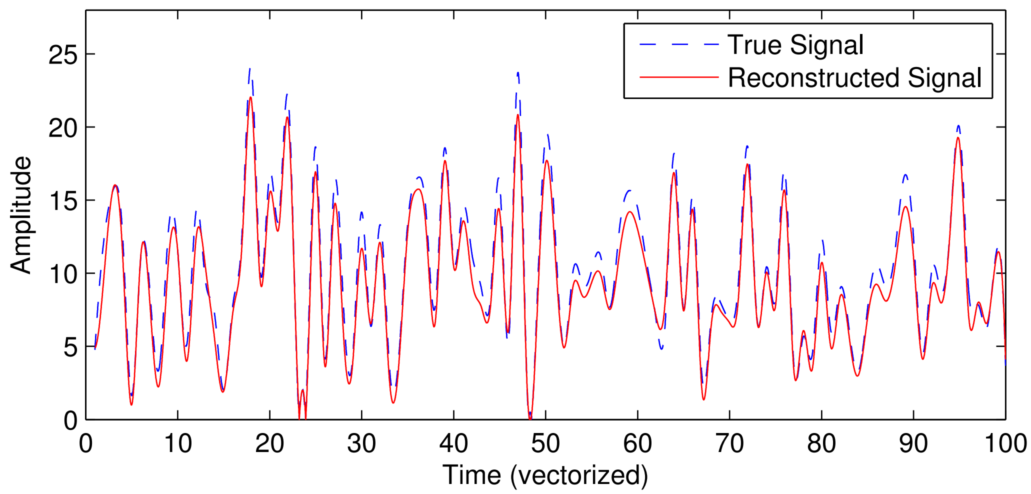


Figure 2: dimension: 101×101, $r = 30$, $\frac{m}{n_1 n_2} = 5.8\%$, signal-to-amplitude-ratio $= 10$.

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:

$$(\textbf{RobustEMaC}): \quad \underset{\boldsymbol{M},\boldsymbol{S}\in\mathbb{C}^{n_1\times n_2}}{\text{minimize}} \quad \|\boldsymbol{M}_{\mathsf{e}}\|_* + \lambda\|\boldsymbol{S}_{\mathsf{e}}\|_1$$

$$\text{subject to} \quad (\boldsymbol{M}+\boldsymbol{S})_{i,l} = \boldsymbol{X}^{\text{corrupted}}_{i,l}, \quad \forall(i,l)\in\Omega$$

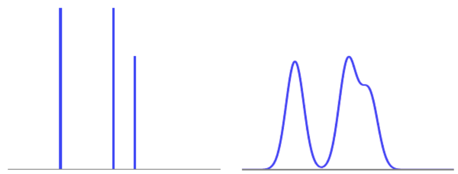
- **Theorem 2 (Sparse Outliers)** Set $\lambda = 1/\sqrt{m\log n}$, and outlier rate $\leq 20\%$. Then RobustEMaC recovers $\boldsymbol{X}$ with high probility if

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

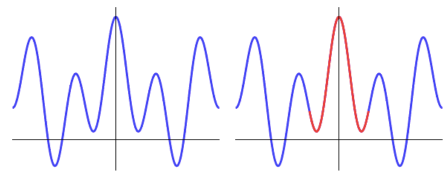
- **Robust to a constant portion of outliers!**

Super Resolution (2-D)

- Obtain low pass components $\Rightarrow$ Extrapolate to high frequencies [CandesFernandezGranda'2012]

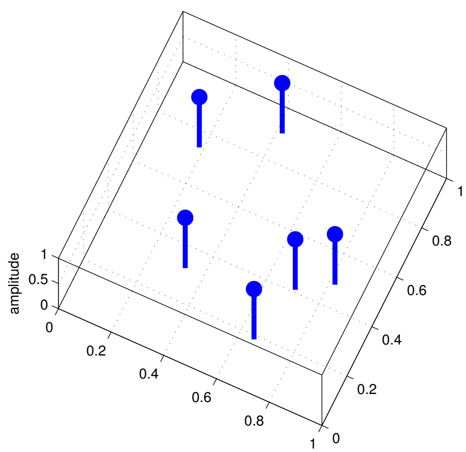


(a) spatial illustration

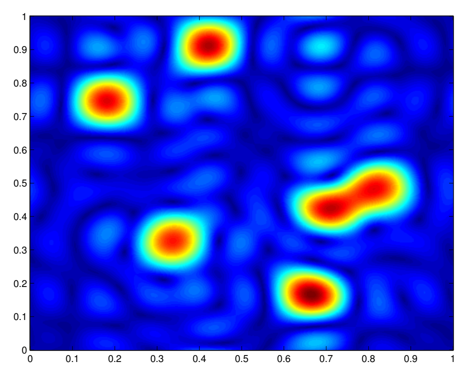


(b) frequency extrapolation

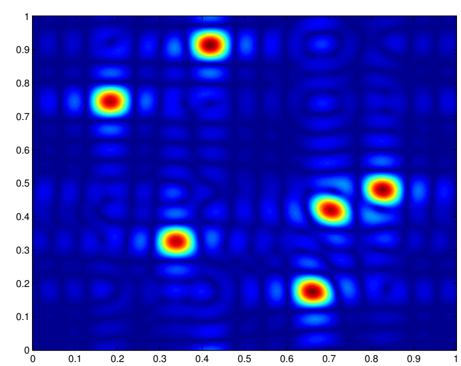
- Might attempt 2-D super-resolution using EMaC...



(a) Ground Truth



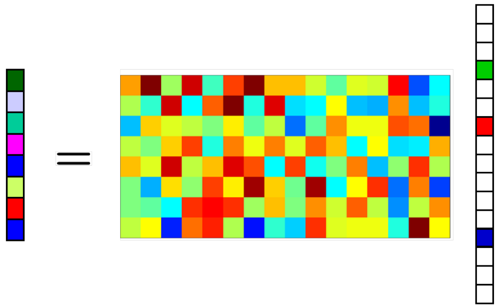
(b) Low Resolution Image



(c) Super-Resolution via EMaC

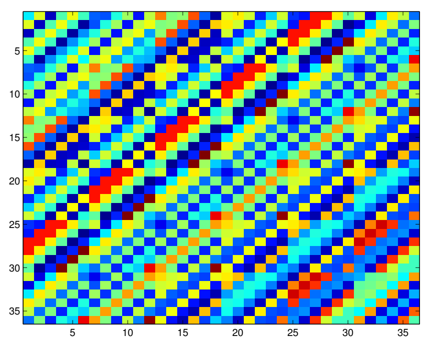
Final Remarks

- Connect spectral compressed sensing with matrix completion



$$\begin{bmatrix} \surd & ? & ? & \surd & \surd \\ ? & \surd & ? & \surd & \surd \\ ? & ? & \surd & \surd & ? \\ \surd & \surd & \surd & \surd & ? \\ \surd & \surd & ? & ? & \surd \end{bmatrix}$$

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



$$\begin{bmatrix} ? & \surd & \surd & ? & \surd & ? & \surd & \surd & ? & ? & \surd & \surd \\ \surd & \surd & ? & ? & ? & \surd & \surd & ? & ? & \surd & \surd & ? \\ \surd & ? & ? & \surd & \surd & \surd & ? & \surd & \surd & \surd & ? & \surd \\ ? & ? & \surd & ? & \surd & ? & \surd & ? & \surd & ? & \surd & \surd \\ \surd & ? & \surd & \surd & ? & ? & \surd & \surd & \surd & ? & \surd & \surd \\ ? & \surd & \surd & ? & ? & \surd & \surd & ? & ? & \surd & \surd & ? \\ \surd & \surd & ? & \surd & \surd & \surd & ? & \surd & \surd & \surd & ? & \surd \\ \surd & ? & \surd & ? & \surd & ? & \surd & \surd & \surd & ? & \surd & ? \\ ? & ? & \surd & \surd & \surd & ? & \surd & \surd & ? & ? & \surd & ? \\ ? & \surd & \surd & ? & ? & \surd & \surd & ? & ? & \surd & ? & ? \\ \surd & \surd & ? & \surd & \surd & \surd & ? & \surd & \surd & ? & ? & \surd \\ \surd & ? & \surd & \surd & \surd & ? & \surd & ? & ? & ? & \surd & ? \end{bmatrix}$$

- 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*)

[CaiCandesShen'2010] J. F. Cai, E. J. Candes, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[CandesFernandezGranda'2012] E. J. Candes and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Arxiv 1203.5871*, March 2012.

[CandesLiMaWright'2011] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of ACM*, 58(3):11:1–11:37, Jun 2011.

[ChiScharfPezeshkiCalderbank'2011] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[DragottiVetterliBlu'2007] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang-fix," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1741 –1757, May 2007.

[GedalyahuTurEldar'2011] K. Gedalyahu, R. Tur, and Y. C. Eldar, "Multichannel sampling of pulse streams at the rate of innovation," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1491–1504, 2011.

[Gross'2011] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, March 2011.

[Hua'1992] Y. Hua. Estimating two-dimensional frequencies by matrix enhancement and matrix pencil. *IEEE Transactions on Signal Processing*, 40(9):2267 –2280, Sep 1992.

[RoyKailath'1989] R. Roy and T. Kailath. Esprit-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(7):984 –995, Jul 1989.

[TangBhaskarShahRecht'2012] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *Arxiv 1207.6053*, July 2012.