Sparse Parameter Estimation: Compressed Sensing meets Matrix Pencil

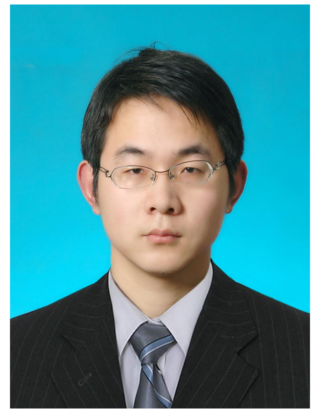
Yuejie Chi
Departments of ECE and BMI
The Ohio State University

Colorado School of Mines
December 9, 2014

THE OHIO STATE UNIVERSITY

Acknowledgement

- Joint work with Yuxin Chen (Stanford) and Yuanxin Li (Ohio State).



  - Y. Chen and Y. Chi, *Robust Spectral Compressed Sensing via Structured Matrix Completion*, IEEE Trans. Information Theory, `http://arxiv.org/abs/1304.8126`
  - Y. Li and Y. Chi, *Off-the-Grid Line Spectrum Denoising and Estimation with Multiple Measurement Vectors*, submitted, `http://arxiv.org/abs/1408.2242`

Sparse Fourier Analysis

In many sensing applications, one is interested in identification of a parametric signal:

$$x(\boldsymbol{t}) = \sum_{i=1}^{r} d_i e^{j2\pi\langle \boldsymbol{t}, \boldsymbol{f}_i \rangle}, \quad \boldsymbol{t} \in [\![n_1]\!] \times \ldots \times [\![n_K]\!]$$

($\boldsymbol{f}_i \in [0,1)^K$: frequencies, d_i : amplitudes, r : model order)

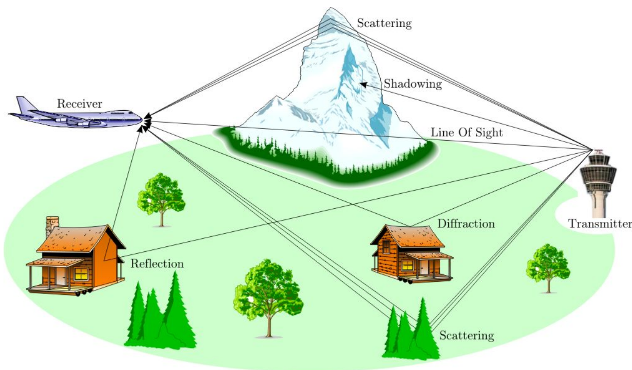
- Occam's razor: the number of modes r is small.

- Sensing with a minimal cost: how to identify the parametric signal model from a small subset of entries of $x(\boldsymbol{t})$?

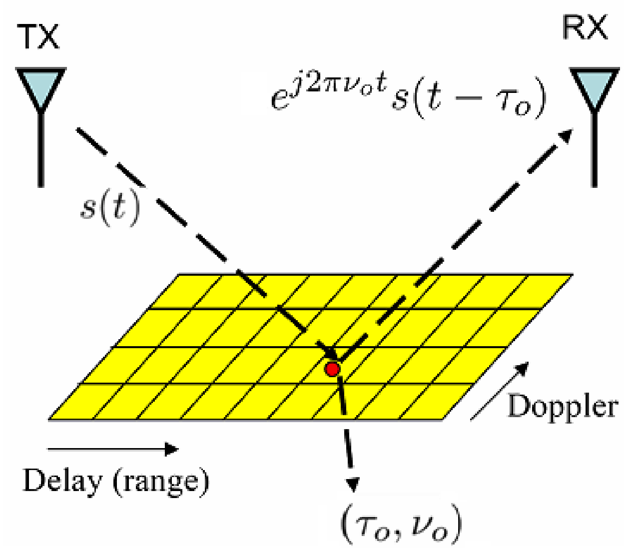
- This problem has many (classical) applications in communications, remote sensing, and array signal processing.

Applications in Communications and Sensing

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



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

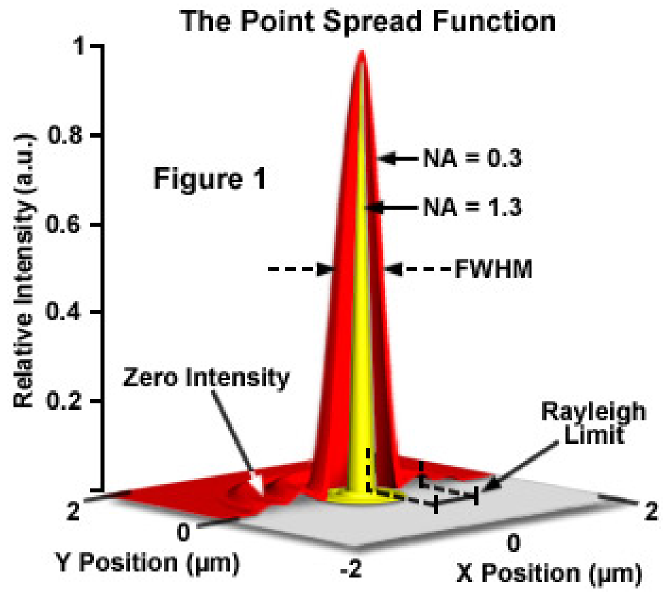


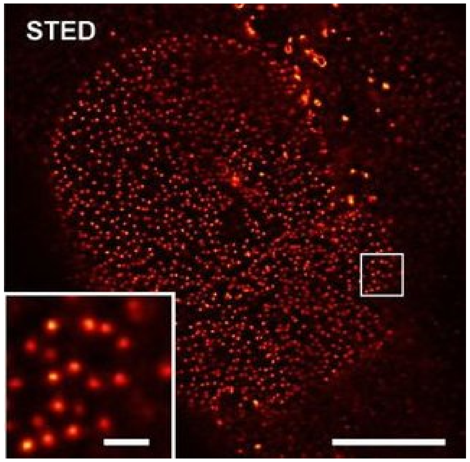
Applications in Imaging

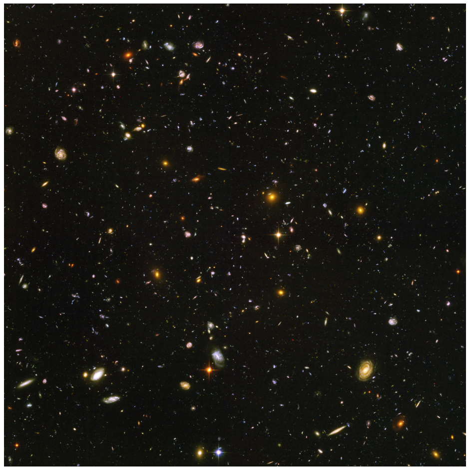
- Swap time and frequency:

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

- Applications in microscopy imaging and astrophysics.
- Resolution is limited by the point spread function of the imaging system







Something Old: Parametric Estimation

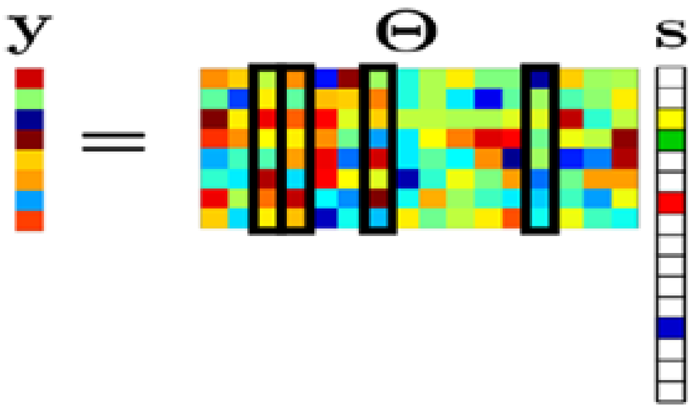
Exploring Physically-meaningful Constraints: *shift invariance* of the harmonic structure

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

- Prony's method: root-finding.
- SVD based approaches: ESPRIT [RoyKailath'1989], MUSIC [Schmidt'1986], matrix pencil [HuaSarkar'1990, Hua'1992].
- spectrum blind sampling [Bresler' 1996], finite rate of innovation [Vetterli' 2001], Xampling [Eldar' 2011].
- Pros: perfect recovery from (equi-spaced) $O(r)$ samples
- Cons: sensitive to noise and outliers, usually require prior knowledge on the model order.

Something New: Compressed Sensing

Exploring Sparsity: Compressed Sensing [Candes and Tao'2006, Donoho'2006] capture the attributes (sparsity) of signals from a small number of samples.



- Discretize the frequency and assume a sparse representation over the discretized basis

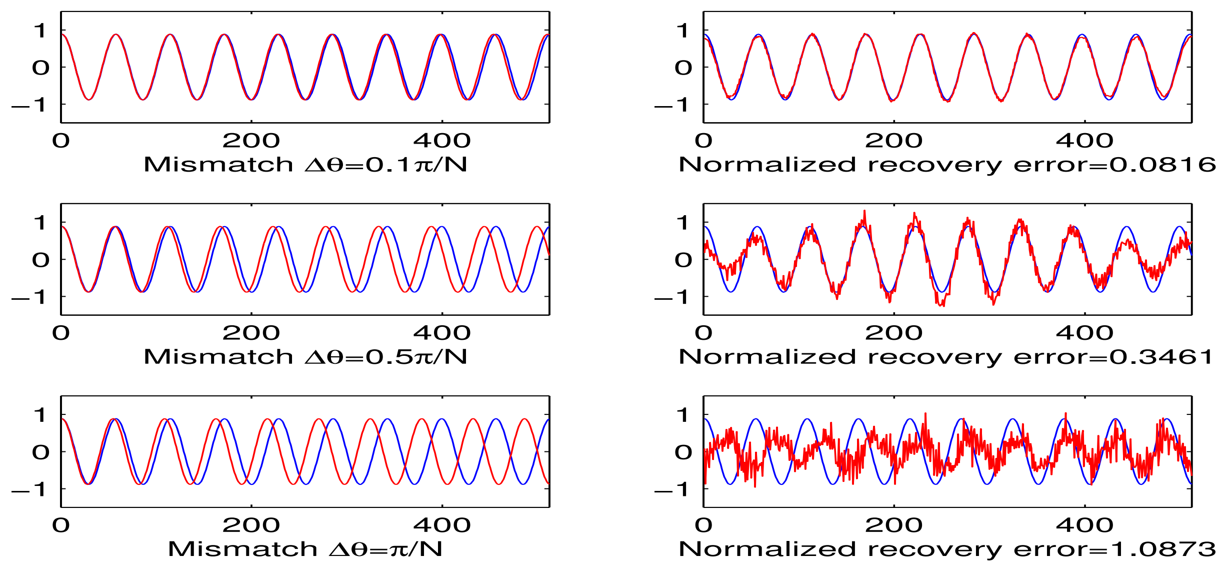
$$f_i \in \mathcal{F} = \left\{ \frac{0}{n_1}, \ldots, \frac{n_1 - 1}{n_1} \right\} \times \left\{ \frac{0}{n_2}, \ldots, \frac{n_2 - 1}{n_2} \right\} \times \ldots$$

- Pros: perfect recovery from $O(r \log n)$ *random* samples, robust against noise and outliers

- Cons: sensitive to gridding error

Sensitivity to Basis Mismatch

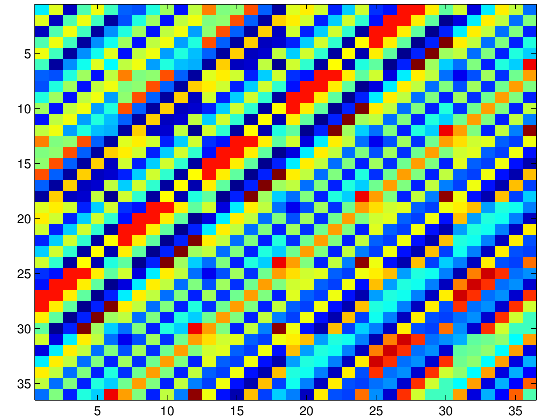
- A toy example: $x(t) = e^{j2\pi f_0 t}$:
  - The success of CS relies on sparsity in the DFT basis.
  - Basis mismatch: Physics places f off grid by a frequency offset.
    - *Basis mismatch translates a sparse signal into an incompressible signal.*



  - Finer grid does not help, and one never estimates the true continuous-valued frequencies! [Chi, Scharf, Pezeshki, Calderbank 2011]

Our Approach

- Conventional approaches enforce physically-meaningful constraints, but not sparsity;
- Compressed sensing enforces sparsity, but not physically-meaningful constraints;
- Approach: We combine sparsity with physically-meaningful constraints, so that we can stably estimate the continuous-valued frequencies from a minimal number of time-domain samples.
  - revisit matrix pencil proposed for array signal processing
  - revitalize matrix pencil by combining it with convex optimization



Two-Dimensional Frequency Model

- Stack the signal $x(\boldsymbol{t}) = \sum_{i=1}^{r} d_i e^{j2\pi\langle \boldsymbol{t}, \boldsymbol{f}_i \rangle}$ into a matrix $\boldsymbol{X} \in \mathbb{C}^{n_1 \times n_2}$.
- The matrix $\boldsymbol{X}$ has the following ***Vandermonde decomposition***:

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

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

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

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

- Goal: We observe a *random subset of entries* of $\boldsymbol{X}$, and wish to recover the missing entries.

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

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

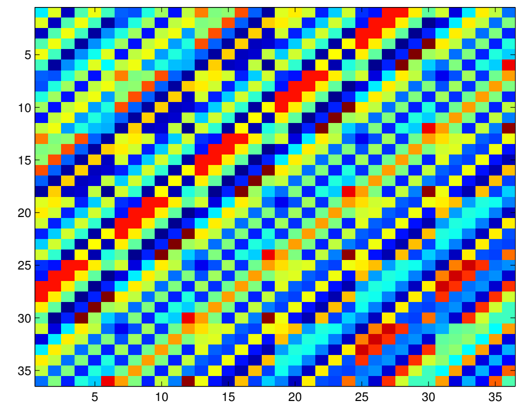
- Quick observation: $\boldsymbol{X}$ can be a low-rank matrix with $\operatorname{rank}(\boldsymbol{X}) = r$.
- Question: can we apply *Matrix Completion* algorithms directly on $\boldsymbol{X}$?

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

- Yes, but it yields sub-optimal performance.
  - It requires at least $r \max\{n_1, n_2\}$ samples.
- No, $\boldsymbol{X}$ is no longer low-rank if $r > \min(n_1, n_2)$
  - Note that r can be as large as $n_1 n_2$

Revisiting Matrix Pencil: Matrix Enhancement

Given a data matrix $\boldsymbol{X}$, Hua proposed the following matrix enhancement for two-dimensional frequency models [Hua 1992]:



- Choose two pencil parameters k_1 and k_2;
- An **enhanced form** $\boldsymbol{X}_{\rm e}$ is an $k_1 \times (n_1 - k_1 + 1)$ *block Hankel matrix* :

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

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

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

Low Rankness of the Enhanced Matrix

- Choose pencil parameters $k_1 = \Theta(n_1)$ and $k_2 = \Theta(n_2)$, the dimensionality of X_{e} is proportional to $n_1 n_2 \times n_1 n_2$.

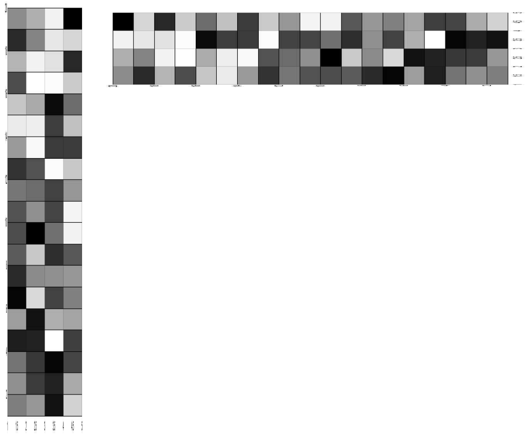
- The enhanced matrix can be decomposed as follows [Hua 1992]:

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

  - $\boldsymbol{Z}_{\mathsf{L}}$ 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.
  - $\mathrm{rank}\left(\boldsymbol{X}_{\mathsf{e}}\right) \leq r$
  - Spectral Sparsity $\Rightarrow$ Low Rankness

- holds even with damping modes.



Enhanced Matrix Completion (EMaC)

- The natural algorithm is to find the enhanced matrix with the minimal rank satisfying the measurements:

$$\begin{aligned} \underset{\boldsymbol{M} \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad & \operatorname{rank}(\boldsymbol{M}_{\mathsf{e}}) \\ \text{subject to} \quad & \boldsymbol{M}_{i,j} = \boldsymbol{X}_{i,j}, \forall (i,j) \in \Omega \end{aligned}$$

  where Ω denotes the sampling set.

- Motivated by Matrix Completion, we will solve its convex relaxation,

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

  where $\|\cdot\|_*$ denotes the nuclear norm.

- The algorithm is referred to as *Enhanced Matrix Completion (EMaC)*.

Enhanced Matrix Completion (EMaC)

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

- existing MC result won't apply – requires at least $\mathcal{O}(nr)$ samples
- **Question**: How many samples do we need?

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

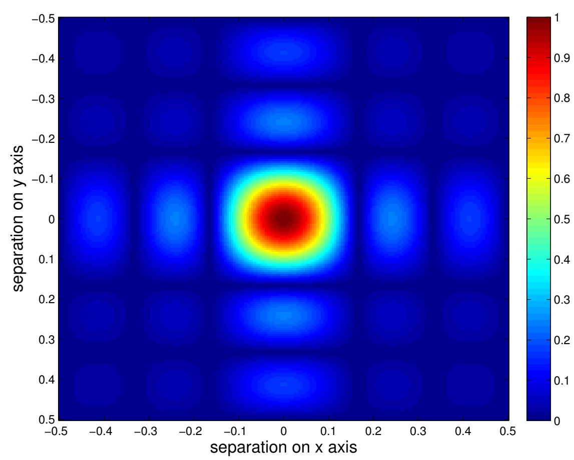
Introduce Coherence Measure

- Define the 2-D Dirichlet kernel:

$$\mathcal{D}(k_1, k_2, f_1, f_2) := \frac{1}{k_1 k_2} \left(\frac{1 - e^{-j2\pi k_1 f_1}}{1 - e^{-j2\pi f_1}} \right) \left(\frac{1 - e^{-j2\pi k_2 f_2}}{1 - e^{-j2\pi f_2}} \right),$$

- Define $\boldsymbol{G}_{\mathsf{L}}$ and $\boldsymbol{G}_{\mathsf{R}}$ as $r \times r$ Gram matrices such that

$$(\boldsymbol{G}_{\mathsf{L}})_{i,l} = \mathcal{D}(k_1, k_2, f_{1i} - f_{1l}, f_{2i} - f_{2l}),$$
$$(\boldsymbol{G}_{\mathsf{R}})_{i,l} = \mathcal{D}(n_1 - k_1 + 1, n_2 - k_2 + 1, f_{1i} - f_{1l}, f_{2i} - f_{2l}).$$



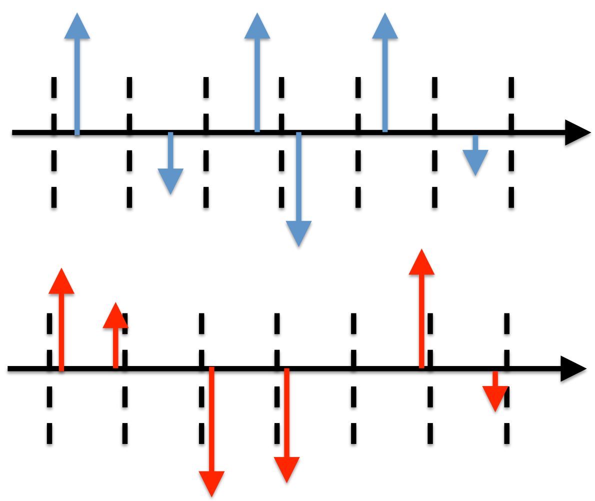
Dirichlet Kernel

Introduce Incoherence Measure

- **Incoherence condition** holds w.r.t. μ if

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

- Examples: $\mu = \Theta(1)$ under many scenarios:
  - Randomly generated frequencies;
  - (Mild) perturbation of grid points;
  - In 1D, let $k_1 \approx \frac{n_1}{2}$: well-separated frequencies (Liao and Fannjiang, 2014): $\Delta = \min_{i \neq j} |f_i - f_j| \gtrsim \frac{2}{n_1}$, which is about 2 times Rayleigh limits.



Theoretical Guarantees for Noiseless Case

- **Theorem [Chen and Chi, 2013] (Noiseless Samples)** Let $n = n_1 n_2$. If all measurements are noiseless, then EMaC recovers $\boldsymbol{X}$ perfectly with high probability if

$$m > C \mu r \log^4 n.$$

  where C is some universal constant.

- **Implications**
  - deterministic signal model, random observation;
  - coherence condition μ only depends on the frequencies but not amplitudes.
  - near-optimal within logarithmic factors: $\Theta(r \text{polylog} n)$.
  - general theoretical guarantees for Hankel (Toeplitz) matrix completion.
    — *see applications in control, MRI, natural language processing, etc*

Phase Transition

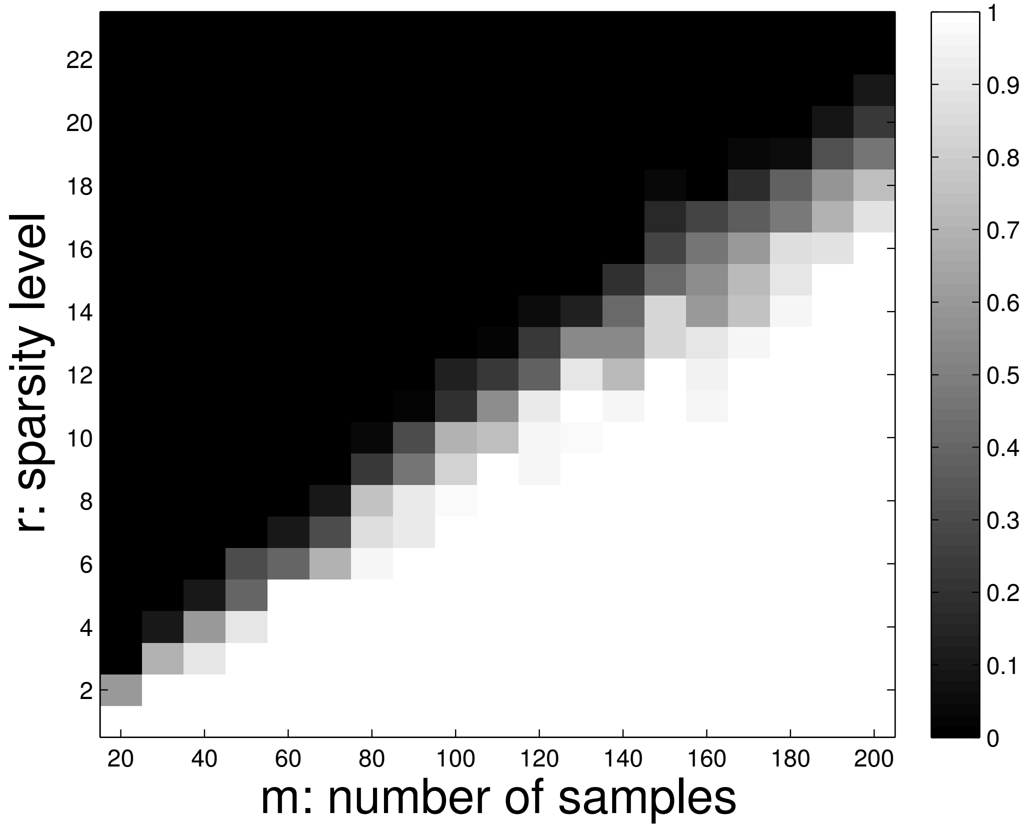


Figure 1: Phase transition diagrams where spike locations are randomly generated. The results are shown for the case where $n_1 = n_2 = 15$.

Robustness to Bounded Noise

Assume the samples are noisy $X = X^o + N$, where N is bounded noise:

$$\text{(EMaC-Noisy)}: \quad \underset{\boldsymbol{M}\in\mathbb{C}^{n_1\times n_2}}{\text{minimize}} \quad \|\boldsymbol{M}_{\mathsf{e}}\|_*$$
$$\text{subject to} \quad \|\mathcal{P}_\Omega(\boldsymbol{M}-\boldsymbol{X})\|_{\mathsf{F}} \le \delta,$$

- **Theorem [Chen and Chi, 2013] (Noisy Samples)** Suppose $\boldsymbol{X}^{\mathrm{o}}$ is a noisy copy of $\boldsymbol{X}$ that satisfies

$$\|\mathcal{P}_\Omega(\boldsymbol{X}-\boldsymbol{X}^{\mathrm{o}})\|_F \le \delta.$$

Under the conditions of Theorem 1, the solution to **EMaC-Noisy** satisfies

$$\|\hat{\boldsymbol{X}}_{\mathrm{e}} - \boldsymbol{X}_{\mathrm{e}}\|_F \le \left\{2\sqrt{n} + 8n + \frac{8\sqrt{2}n^2}{m}\right\}\delta$$

with probability exceeding $1-n^{-2}$.

- **Implications:** The average entry inaccuracy is bounded above by $\mathcal{O}(\frac{n}{m}\delta)$. In practice, EMaC-Noisy usually yields better estimate.

Singular Value Thresholding (Noisy Case)

- Several optimized solvers for Hankel matrix completion exist, for example [Fazel et. al. 2013, Liu and Vandenberghe 2009]

Algorithm 1 Singular Value Thresholding for EMaC

1: **initialize** Set $\boldsymbol{M}_0 = \boldsymbol{X}_{\mathrm{e}}$ and $t = 0$.
2: **repeat**
3: 1) $\boldsymbol{Q}_t \leftarrow \mathcal{D}_{\tau_t}(\boldsymbol{M}_t)$ (*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

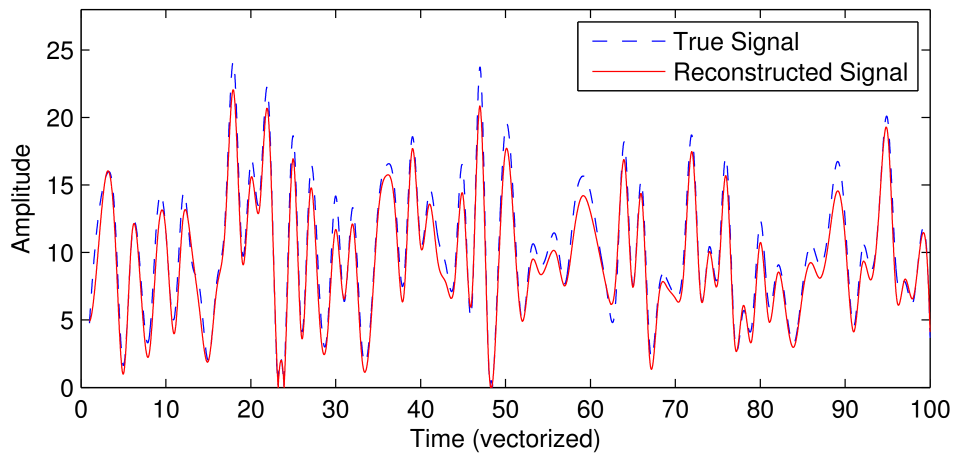
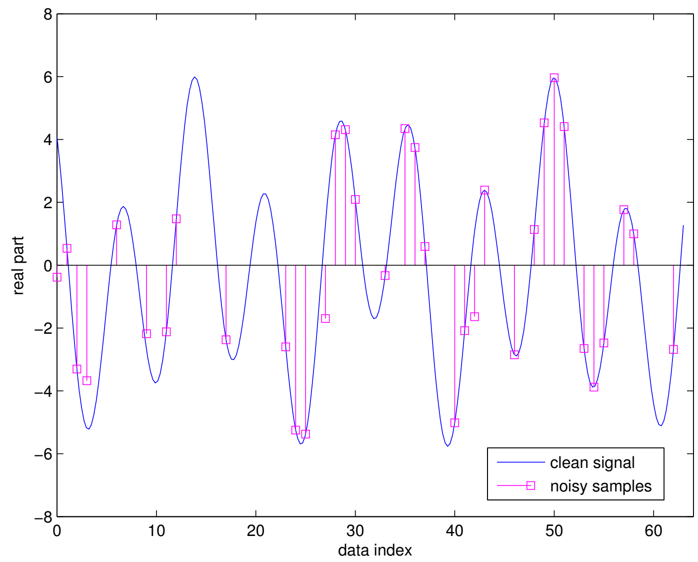


Figure 2: dimension: 101×101, $r = 30$, $\frac{m}{n_1 n_2} = 5.8\%$, SNR $= 10$dB.

Robustness to Sparse Outliers

- What if a constant portion of measurements are arbitrarily corrupted?



  - Robust PCA approach [Candes et. al. 2011, Chandrasekaran et. al. 2011]
  - Solve the following algorithm:

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

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

Theoretical Guarantees for Robust Recovery

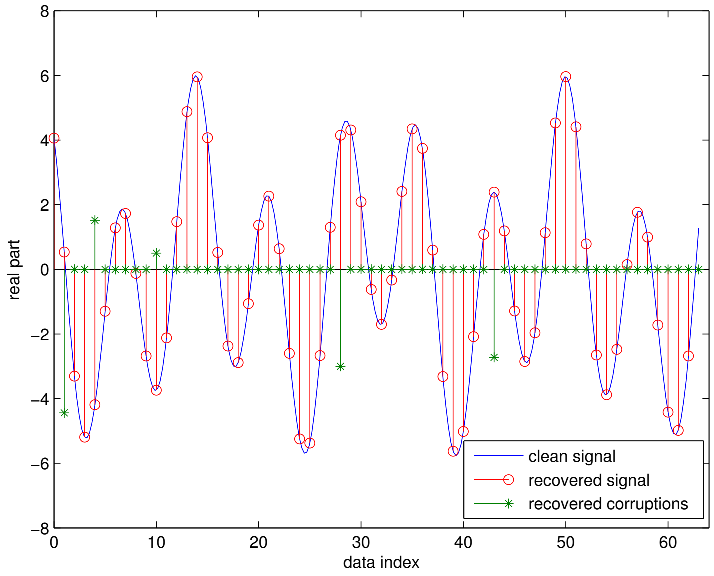
- **Theorem [Chen and Chi, 2013] (Sparse Outliers)** Set $n = n_1 n_2$ and $\lambda = \frac{1}{\sqrt{m \log n}}$. Let the percent of corrupted entries $s \leq 20\%$ selected uniformly at random, then RobustEMaC recovers $\boldsymbol{X}$ with high probability if

$$m > C\mu r^2 \log^3 n,$$

where C is some universal constant.

- **Implications:**
  - slightly more samples $m \sim \Theta(r^2 \log^3 n)$;
  - robust to a constant portion of outliers: $s \sim \Theta(m)$;

- In summary, EMaC achieves robust recovery with respect to dense and sparse errors from a near-optimal number of samples.



Phase Transition for Line Spectrum Estimation

Fix the amount of corruption as 10% of the total number of samples:

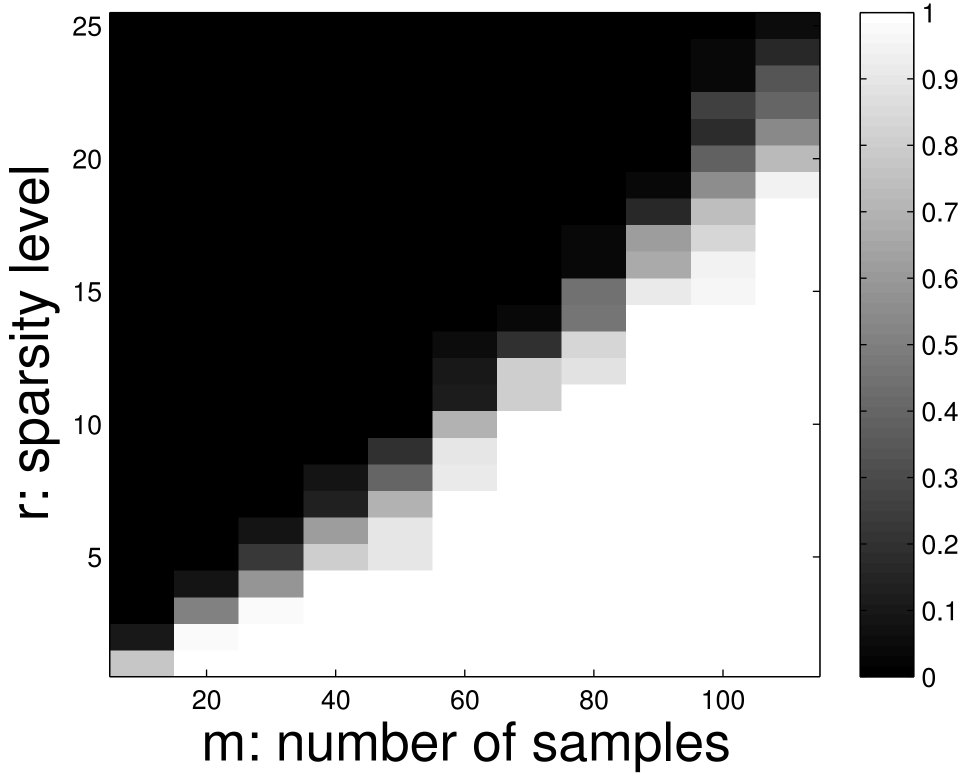


Figure 3: Phase transition diagrams where spike locations are randomly generated. The results are shown for the case where $n = 125$.

Related Work

- Cadzow's denoising with full observation: non-convex heuristic to denoise line spectrum data based on the Hankel form.
- Atomic norm minimization with random observation: recently proposed by [Tang et. al., 2013] for compressive line spectrum estimation off the grid.

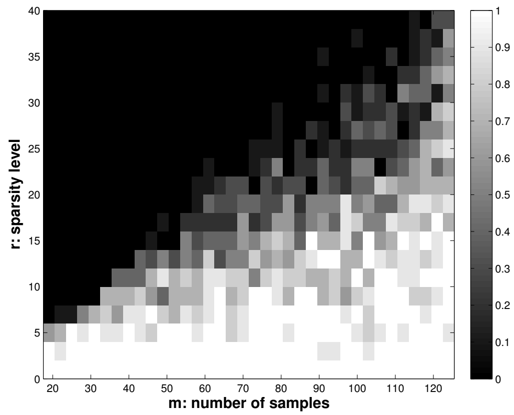
$$\min_{s} \quad \|\boldsymbol{s}\|_{\mathcal{A}} \quad \text{subject to} \quad \mathcal{P}_{\Omega}(\boldsymbol{s}) = \mathcal{P}_{\Omega}(\boldsymbol{x}),$$

where the atomic norm is defined as $\|\boldsymbol{x}\|_{\mathcal{A}} = \inf \left\{ \sum_i |d_i| \middle| \boldsymbol{x}(t) = \sum_i d_i e^{j2\pi f_i t} \right\}$.

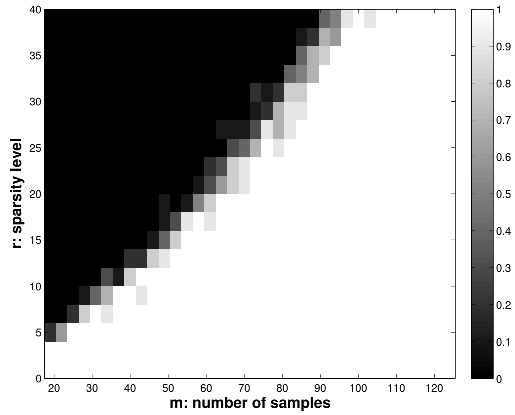
  - Random signal model: if the frequencies are **separated** by 4 times the Rayleigh limit and the phases are random, then perfect recovery with $O(r \log r \log n)$ samples;
  - no stability result with random observations;
  - extendable to multi-dimensional frequencies [Chi and Chen, 2013], but the SDP characterization is more complicated [Xu et. al. 2013].

(Numerical) Comparison with Atomic Norm Minimization

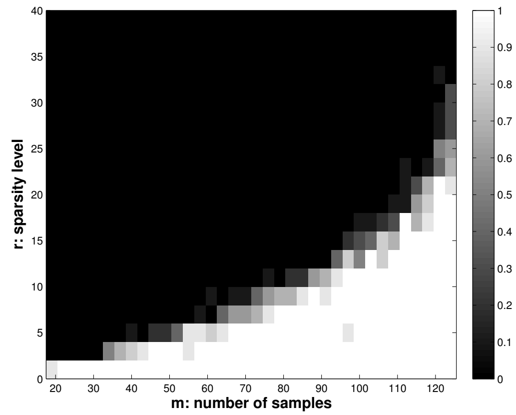
Phase transition for 1D spectrum estimation: the phase transition for atomic norm minimization is very sensitive to the separation condition. The EMaC, in contrast, is insensitive to the separation.



(a)



(b)

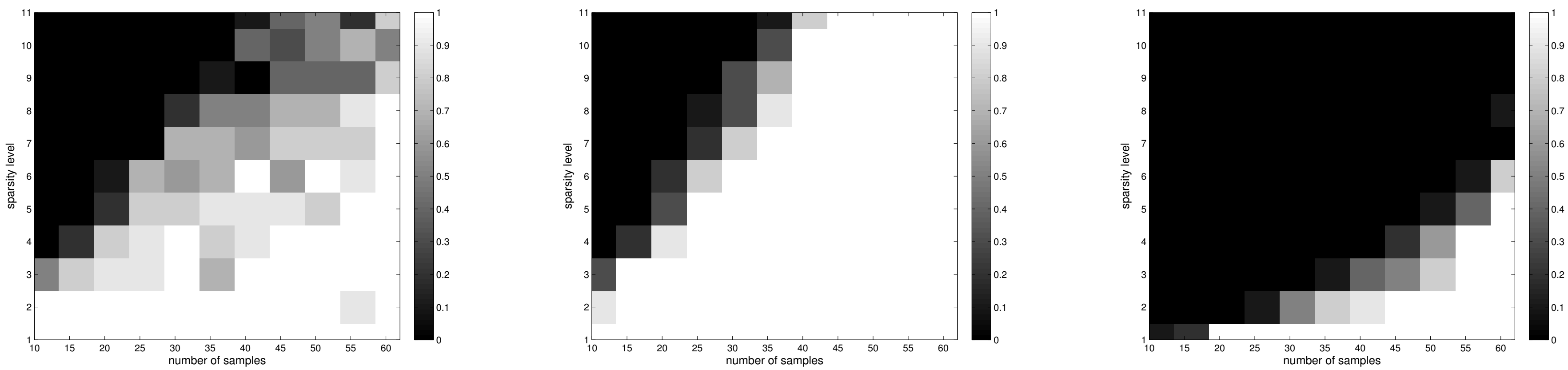


(c)

Figure: phase transition for atomic norm minimization without separation (a), with separation (b); and EMaC without separation (c). The inclusion of separation doesn't change the phase transition of EMaC.

(Numerical) Comparison with Atomic Norm Minimization

Phase transition for 2D spectrum estimation: the phase transition for atomic norm minimization is very sensitive to the separation condition. The EMaC, in contrast, is insensitive to the separation. Here the problem dimension $n_1 = n_2 = 8$ is relatively small and the atomic norm minimization approach seems in favor despite of separation.



(a) (b) (c)

Figure: phase transition for atomic norm minimization without separation (a), with separation (b); and EMaC without separation (c). The inclusion of separation doesn't change the phase transition of EMaC.

Extension to Multiple Measurement Vectors Model

- When multiple snapshots available, it is possible to exploit the covariance structure to reduce the number of sensors. Without loss of generality, consider 1D:

$$x_\ell(t) = \sum_{i=1}^{r} d_{i,\ell} e^{j2\pi t f_i}, \quad t \in \{0, 1, \dots, n-1\}$$

  where $\boldsymbol{x}_\ell = [x_\ell(0), x_\ell(1), \dots, x_\ell(n-1)]^T$, $\ell = 1, \dots, L$.

- We assume the coefficients $d_{i,\ell} \sim \mathcal{CN}(0, \sigma_i^2)$, then the covariance matrix

$$\boldsymbol{\Sigma} = \mathbb{E}\left[\boldsymbol{x}_\ell \boldsymbol{x}_\ell^H\right] = \mathrm{toep}(\boldsymbol{u})$$

  is a **PSD block Toeplitz matrix with rank**$(\boldsymbol{\Sigma}) = r$.

- The frequencies can be estimated without separation from $\boldsymbol{u}$ using ℓ_1 minimization with nonnegative constraints [Donoho and Tanner' 2005].

Observation with the Sparse Ruler

- **Observation pattern:** Instead of random observations, we assume deterministic observation pattern Ω over a (minimum) sparse ruler for all snapshots:

$$\boldsymbol{y}_\ell = \mathbf{x}_{\Omega,\ell} = \{x_\ell(t), \quad t \in \Omega\}, \quad \ell = 1, \ldots, L.$$

- **Sparse ruler in 1D:** for $\Omega \in \{0, \ldots, n-1\}$
  - Define the difference set:

$$\Delta = \{|i-j|, \quad \forall i, j \in \Omega\}$$

  - Ω is called a length-n sparse ruler if $\Delta = \{0, \ldots, n-1\}$.
  - Examples:
    - when $n = 21$, $\Omega = \{0, 1, 2, 6, 7, 8, 17, 20\}$.
    - nested arrays, co-prime arrays [Pal and Vaidyanathan' 2010, 2011]
    - minimum sparse rulers [Redei and Renyi, 1949]
- roughly $|\Omega| = O(\sqrt{n})$.

Covariance Estimation on Sparse Ruler Entries

- Consider the observation on $\Omega = \{0, 1, 2, 5, 8\}$,

$$\mathbb{E}\left[\boldsymbol{y}_\ell \boldsymbol{y}_\ell^H\right] = \mathbb{E}\begin{bmatrix} x_\ell(0) \\ x_\ell(1) \\ x_\ell(2) \\ x_\ell(5) \\ x_\ell(8) \end{bmatrix}\left[x_\ell^H(0)\ x_\ell^H(1)\ x_\ell^H(2)\ x_\ell^H(5)\ x_\ell^H(8)\right]$$

$$= \begin{bmatrix} u_0 & u_1^H & u_2^H & u_5^H & u_8^H \\ u_1 & u_0 & u_1^H & u_4^H & u_7^H \\ u_2 & u_1 & u_0 & u_3^H & u_6^H \\ u_5 & u_4 & u_3 & u_0 & u_3^H \\ u_8 & u_7 & u_6 & u_3 & u_0 \end{bmatrix}$$

- which gives the exact full covariance matrix $\boldsymbol{\Sigma} = \mathrm{toep}(\boldsymbol{u})$ in the absence of noise and an infinite number of snapshots.

Two-step Structured Covariance Estimation

- In practice, measurements will be noisy with a finite number of snapshots:

$$\boldsymbol{y}_\ell = \boldsymbol{x}_{\Omega,\ell} + \boldsymbol{w}_\ell, \quad \ell = 1, \ldots, L,$$

  where $\boldsymbol{w}_\ell \sim \mathcal{CN}(\sigma^2, \boldsymbol{I})$.

- **Two-step covariance estimation:**
  - formulate the sample covariance matrix of $\boldsymbol{y}_\ell$:

$$\boldsymbol{\Sigma}_{\Omega,L} = \frac{1}{L}\sum_{\ell=1}^{L} \boldsymbol{y}_\ell \boldsymbol{y}_\ell^H;$$

  - determine the Teoplitz covariance matrix with SDP:

$$\hat{\boldsymbol{u}} = \underset{\boldsymbol{u}:\mathrm{toep}(\boldsymbol{u})\succeq 0}{\operatorname{argmin}} \ \frac{1}{2}\left\|\mathcal{P}_\Omega(\mathrm{toep}(\boldsymbol{u})) - \boldsymbol{\Sigma}_{\Omega,L}\right\|_F^2 + \lambda \mathrm{Tr}(\mathrm{toep}(\boldsymbol{u})),$$

    where λ is some regularization parameter.

Theoretical Guarantee

- **Theorem [Li and Chi]** Let $\boldsymbol{u}^\star$ be the ground truth. Set

$$\lambda \geq C \max\left\{\sqrt{\frac{r\log(Ln)}{L}}, \frac{r\log(Ln)}{L}\right\} \|\boldsymbol{\Sigma}_\Omega^\star\|$$

with $\boldsymbol{\Sigma}_\Omega^\star = \mathbb{E}\left[\boldsymbol{y}_m \boldsymbol{y}_m^H\right]$ for some constant C, then with probability at least $1 - L^{-1}$, the solution satisfies

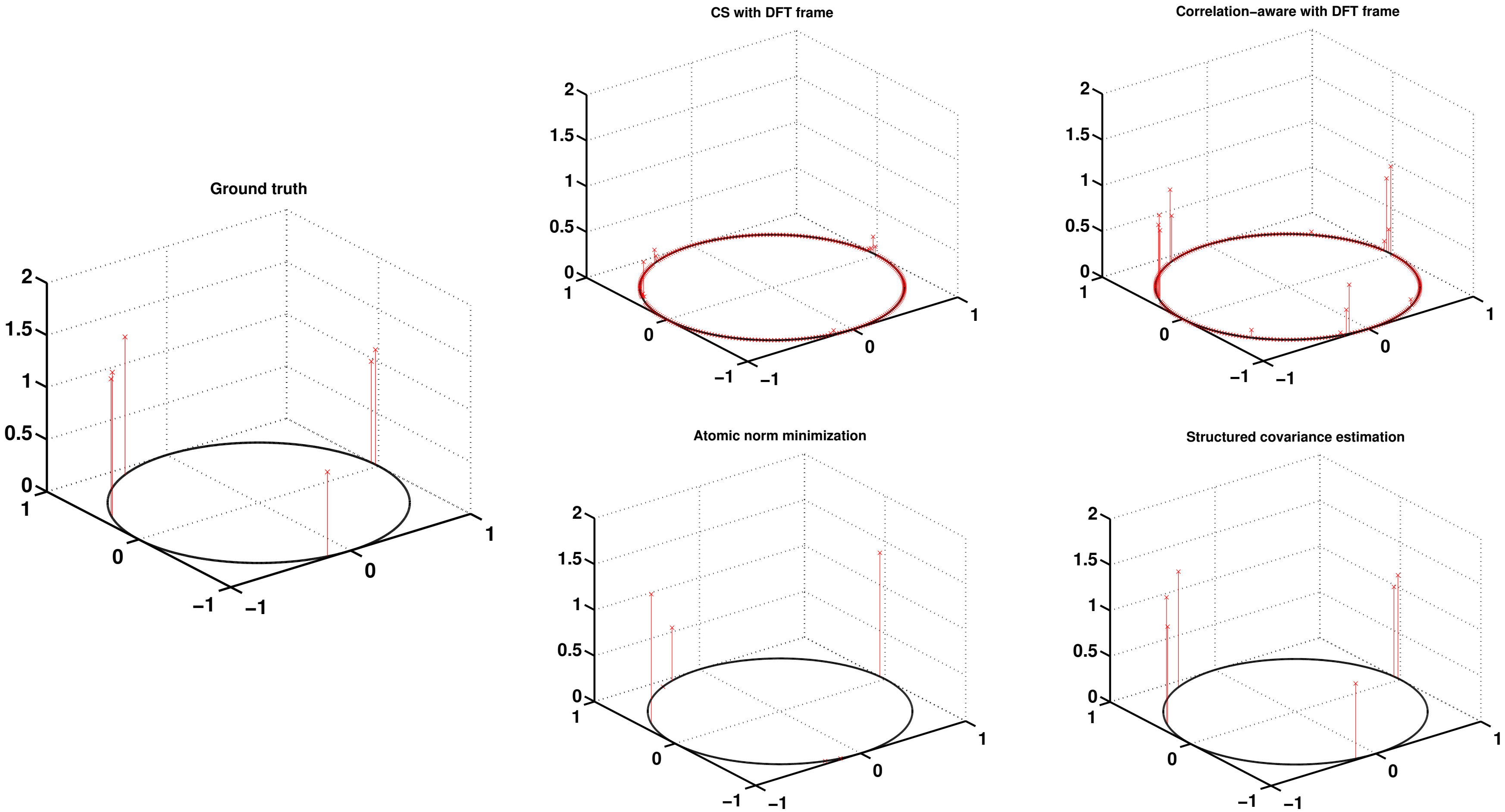
$$\frac{1}{\sqrt{n}}\|\hat{\boldsymbol{u}} - \boldsymbol{u}^\star\|_F \leq 16\lambda\sqrt{r}$$

if Ω is a complete sparse ruler.

- **Remark:**
  - $\frac{1}{\sqrt{n}}\|\hat{\boldsymbol{u}} - \boldsymbol{u}^\star\|_F$ is small as soon as $L \gtrsim O(r^2 \log n)$;
  - As the rank r (as large as n) can be larger than $|\Omega|$ (as small as $\sqrt{n}$), this allows frequency estimation even the snapshots cannot be recovered.

Numerical Simulations

The algorithm also applies to other observation patterns, e.g. random. Setting: $n = 64$, $L = 400$, $|\Omega| = 5$, and $r = 6$.



Final Remarks

- Sparse parameter estimation is possible leveraging shift-invariance structures embedded in matrix pencil with recent matrix completion techniques;

- Fundamental performance is determined by the proximity of the frequencies measured by the conditioning number of the Gram matrix formed by the sampling the Dirichlet kernel;

- Recovering more lines than the number of sensors is made possible by exploiting the second-order statistics;

- **Future work:** how to compare conventional algorithms with those of convex optimization?

Q&A

Publications available on arXiv:

- Robust Spectral Compressed Sensing via Structured Matrix Completion, IEEE Trans. Information Theory, http://arxiv.org/abs/1304.8126
- Off-the-Grid Line Spectrum Denoising and Estimation with Multiple Measurement Vectors, submitted, http://arxiv.org/abs/1408.2242

Thank You! Questions?