# Convex Optimization Techniques for Super-resolution Parameter Estimation

Yuejie Chi[o] and Gongguo Tang[c]
[o]ECE and BMI, The Ohio State University
[c]ECE, Colorado School of Mines

IEEE International Conference on Acoustics, Speech and Signal Processing

Shanghai, China

March 2016

# Acknowledgements

# Parameter Estimation or Image Inversion I

- **Image:** Observable image $\mathbf{y} \sim p(\mathbf{y}; \boldsymbol{\theta})$, whose distribution is parameterized by unknown parameters $\boldsymbol{\theta}$.
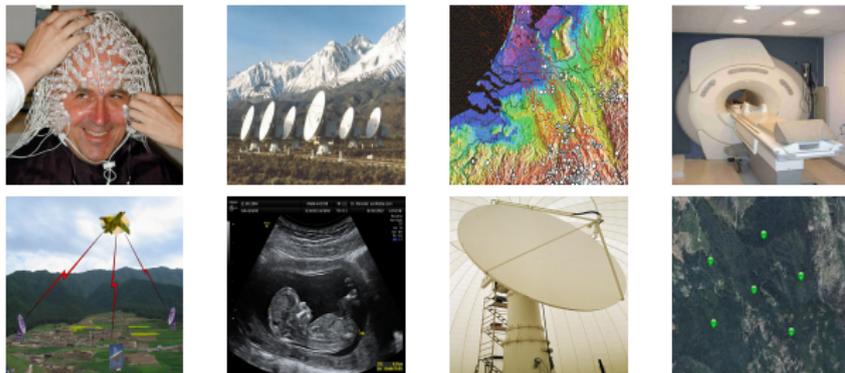


- **Inversion:** Estimate $\boldsymbol{\theta}$, given a set of samples of $\mathbf{y}$.
  - Source location estimation in MRI and EEG
  - DOA estimation in sensor array processing
  - Frequency and amplitude estimation in spectrum analysis
  - Range, Doppler, and azimuth estimation in radar/sonar

# Parameter Estimation or Image Inversion II

- **Canonical Model:** Supperposition of modes:

$$y(t) = \sum_{i=1}^{r} \psi(t; \nu_i)\alpha_i + n(t)$$

  - $p = 2r$ unknown parameters: $\boldsymbol{\theta} = [\nu_1, \ldots, \nu_r, \alpha_1, \ldots, \alpha_r]^T$
  - Parameterized modal function: $\psi(t; \nu)$
  - Additive noise: $n(t)$

- **After Sampling:**

$$\begin{bmatrix} y(t_0) \\ y(t_1) \\ \vdots \\ y(t_{n-1}) \end{bmatrix} = \sum_{i=1}^{r} \begin{bmatrix} \psi(t_0; \nu_i) \\ \psi(t_1; \nu_i) \\ \vdots \\ \psi(t_{n-1}; \nu_i) \end{bmatrix} \alpha_i + \begin{bmatrix} n(t_0) \\ n(t_1) \\ \vdots \\ n(t_{n-1}) \end{bmatrix}$$

or

$$\mathbf{y} = \boldsymbol{\Psi}(\boldsymbol{\nu})\boldsymbol{\alpha} + \mathbf{n} = \sum_{i=1}^{r} \boldsymbol{\psi}(\nu_i)\alpha_i + \mathbf{n}$$

- Typically, $t_i$'s are uniformly spaced and almost always $n > p$.

# Parameter Estimation or Image Inversion III

**Canonical Model:**

$$\mathbf{y} = \mathbf{\Psi}(\boldsymbol{\nu})\boldsymbol{\alpha} + \mathbf{n} = \sum_{i=1}^{r} \boldsymbol{\psi}(\nu_i)\alpha_i + \mathbf{n}$$

▶ DOA estimation and spectrum analysis:

$$\boldsymbol{\psi}(\nu) = [e^{jt_0\nu}, e^{jt_1\nu}, \ldots, e^{jt_{m-1}\nu}]^T$$

where $\nu$ is the DOA (electrical angle) of a radiating point source.

▶ Radar and sonar:

$$\boldsymbol{\psi}(\nu) = [w(t_0 - \tau)e^{j\omega t_0}, w(t_1 - \tau)e^{j\omega t_1}, \ldots, w(t_{m-1} - \tau)e^{j\omega t_{m-1}}]^T$$

where $w(t)$ is the transmit waveform and $\nu = (\tau, \omega)$ are delay and Doppler coordinates of a point scatterer.

# New Challenges for Parameter Estimation

- **Limited Sampling Rate:**
  ultra-wideband signals, large antenna arrays, etc.

- **Noise, corruptions and missing data:**
  sensor failures, attacks, outliers, etc.

- **Multi-modal data:** the received signal exhibits
  superpositions of *multiple* modal functions:
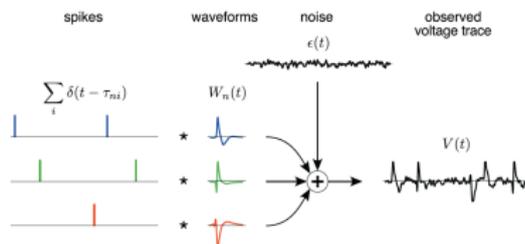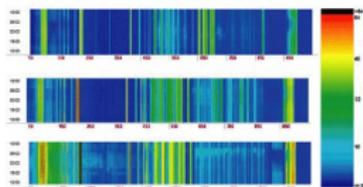


  which occurs frequently in multi-user/multi-channel environments.

- **Calibration and Blind Super-resolution:** the modal
  function needs to be calibrated or estimated
  before performing parameter estimation.

# Motivating applications: Super-resolution Imaging

▶ Single-molecule based superresolution techniques (STORM/PALM) achieve nanometer spatial resolution by integrating the temporal information of the switching dynamics of fluorophores (emitters).
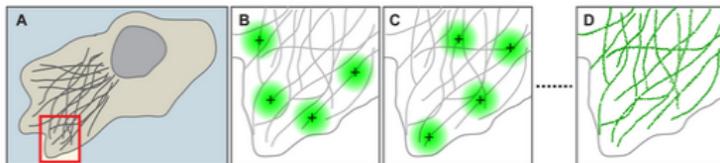


▶ In each frame, our goal is to localize a point source model via observing its convolution with a point spread function (PSF) $g(t)$:

$$z\left(t\right) = \left(\sum_{i=1}^{r} d_i \delta(t - t_i)\right) * g(t) = \sum_{i=1}^{r} d_i g(t - t_i)$$

▶ The final image is obtained by superimposing the reconstruction of each frame.

▶ The reconstruction requires estimating locations of point sources.

# Three-Dimensional Super-resolution Imaging

- This principle can be extended to reconstruct 3-D objects from 2-D images, by modulating the shape, e.g. ellipticity, of the PSFs along the $z$-dimension.
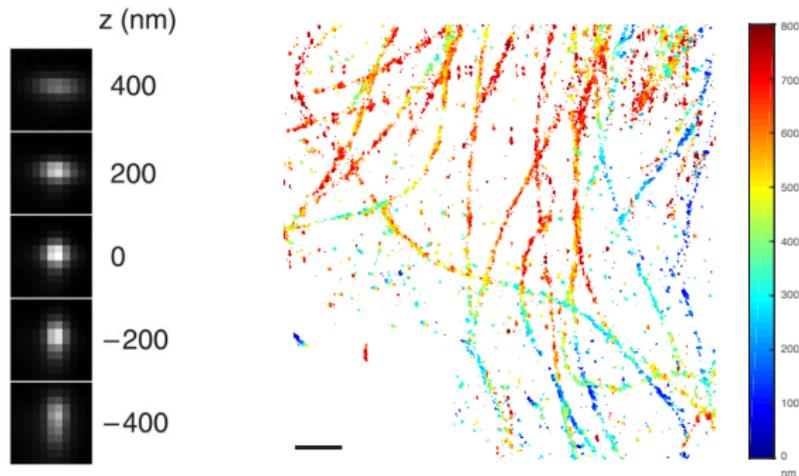


- The reconstruction requires separation of point sources modulated by different PSFs.

M. J. Rust, M. Bates, X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)", Nature Methods 3, 793-795 (2006).

J. Huang, M. Sun, K. Gumpper, Y. Chi and J. Ma, "3D Multifocus Astigmatism and Compressed Sensing (3D MACS) Based Superresolution Reconstruction", Biomedical Optics Express, 2015.

# Motivating Applications: Neural Spike Sorting

▶ The electrode measures firing activities of neighboring neurons with *unknown* characteristic functions (or PSF).
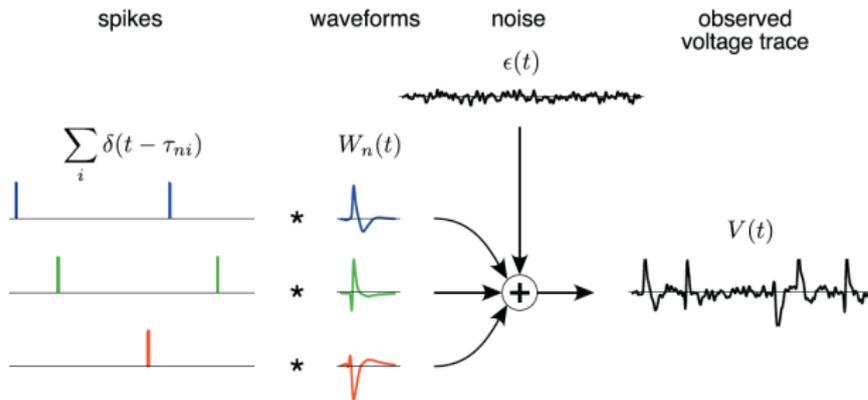


▶ The goal is to identify and separate the firing times of each neuron from the observed voltage trace at the electrode.

▶ The reconstruction requires simultaneous estimation of the activation time and the PSF.

# Motivating Applications: Blind multi-path channel identification

- In multi-user communication systems, each user transmits a waveform $g(t)$ modulated by unknown data symbols, which arrives at the receiver asynchronously

$$y(t) = \sum_{i=1}^{r} \alpha_i g_i(t - t_i)$$
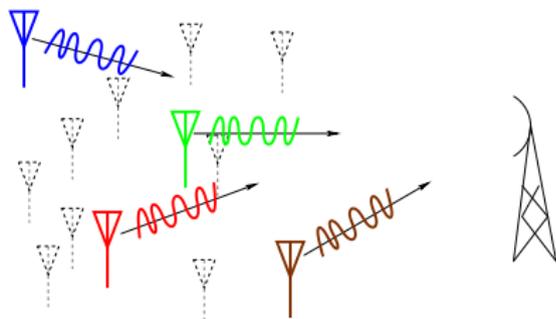


- The goal is to simultaneously decode and estimate the multi-path delay.

# Tutorial Outline

- Review conventional parameter estimation methods, with a focus on spectrum estimation.
- Super-resolution Parameter Estimation via $\ell_1$-minimization: consequences of basis mismatch
- Super-resolution Parameter Estimation via atomic norm minimization
- Super-resolution Parameter Estimation via structured matrix completion
- Final remarks

# Classical Parameter Estimation: Matched Filtering I

- **Matched filtering**

  - Sequence of rank-one subspaces, or 1D test images, is matched to the measured image by filtering, correlating, or phasing.

  - Test images are generated by scanning a prototype image (e.g., a waveform or a steering vector) through frequency, wavenumber, doppler, and/or delay at some desired resolution $\Delta\nu$.

  $$P(\ell) = \|\boldsymbol{\psi}(\ell\Delta\nu)^H \mathbf{y}\|_2^2$$
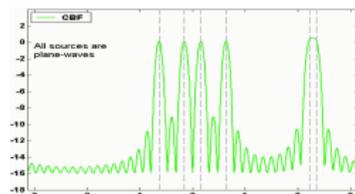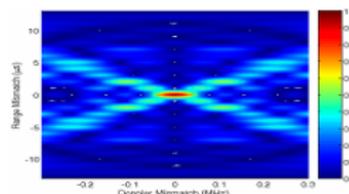
  

  Bearing Response      Cross-Ambiguity

  - Peak locations are taken as estimates of $\nu_i$ and peak values are taken as estimates of source powers $|\alpha_i|^2$.
  - Resolution: Rayleigh Limit (RL), inversely proportional to the number of measurements

# Classical Parameter Estimation: Matched Filtering II

▶ **Matched filtering (Cont.)**

  ▶ Extends to subspace matching for those cases in which the model for the image is comprised of several dominant modes.

  ▶ Extends to whitened matched filter, or minimum variance unbiased (MVUB) filter, or generalized sidelobe canceller.

H. L. Van Trees, "Detection, Estimation, and Modulation Theory: Part I",

D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE,* vol. 70, pp. 10551096, Sep. 1982.

T.-C.Lui and B. D. Van Veen, "Multiple window based minimum variance spectrum estimation for multidimensional random fields," *IEEE Trans. Signal Process.,* vol. 40, no. 3, pp. 578–589, Mar. 1992.

L. L. Scharf and B. Friedlander, "Matched subspace detectors," IEEE Trans. Signal Process., vol. 42, no. 8, pp. 21462157, Aug. 1994.

A. Pezeshki, B. D. Van Veen, L. L. Scharf, H. Cox, and M. Lundberg, "Eigenvalue beamforming using a multi-rank MVDR beamformer and subspace selection," *IEEE Trans. Signal Processing,* vol. 56, no. 5, pp. 1954–1967, May 2008.

# Classical Parameter Estimation: ML Estimation I

- **ML Estimation in Separable Nonlinear Models**

    - **Low-order** separable modal representation for the image:

    $$\mathbf{y} = \mathbf{\Psi}(\boldsymbol{\nu})\boldsymbol{\alpha} + \mathbf{n} = \sum_{i=1}^{r} \boldsymbol{\psi}(\nu_i)\alpha_i + \mathbf{n}$$

    Parameters $\boldsymbol{\nu}$ in $\mathbf{\Psi}$ are nonlinear parameters (like frequency, delay, and Doppler) and $\boldsymbol{\alpha}$ are linear parameters (comples amplitudes).

    - Estimates of linear parameters (complex amplitudes of modes) and nonlinear mode parameters (frequency, wavenumber, delay, and/or doppler) are extracted, usually based on maximum likelihood (ML), or some variation on linear prediction, using $\ell_2$ minimization.

# Classical Parameter Estimation: ML Estimation II

- **Estimation of Complex Exponential Modes**
  - Physical model:

$$y(t) = \sum_{i=1}^{r} \alpha_i \nu_i^t + n(t); \quad \psi(t;\nu_i) = \nu_i^t$$

where $\nu_i = e^{d_i + j\omega_i}$ is a complex exponential mode, with damping $d_i$ and frequency $\omega_i$.

  - Uniformly sampled measurement model:

$$\mathbf{y} = \mathbf{\Psi}(\boldsymbol{\nu})\boldsymbol{\alpha}$$

$$\mathbf{\Psi}(\boldsymbol{\nu}) = \begin{bmatrix} \nu_1^0 & \nu_2^0 & \cdots & \nu_r^0 \\ \nu_1^1 & \nu_2^1 & \cdots & \nu_r^1 \\ \nu_1^2 & \nu_2^2 & \cdots & \nu_r^2 \\ \vdots & \vdots & \ddots & \vdots \\ \nu_1^{n-1} & \nu_2^{n-1} & \cdots & \nu_r^{n-1} \end{bmatrix}.$$

Here, without loss of generality, we have taken the samples at $t = \ell t_0$, for $\ell = 0, 1, \ldots, n-1$, with $t_0 = 1$.

# Classical Parameter Estimation: ML Estimation III

- **ML Estimation of Complex Exponential Modes**

$$\min_{\boldsymbol{\nu},\boldsymbol{\alpha}} \|\mathbf{y} - \boldsymbol{\Psi}(\boldsymbol{\nu})\boldsymbol{\alpha}\|_2^2$$

$$\hat{\boldsymbol{\alpha}}_{ML} = \boldsymbol{\Psi}(\boldsymbol{\nu})^\dagger \mathbf{y}$$

$$\hat{\boldsymbol{\nu}}_{ML} = \arg\min \mathbf{y}^H \mathbf{P}_{\mathbf{A}(\boldsymbol{\nu})} \mathbf{y}; \quad \mathbf{A}^H \boldsymbol{\Psi} = \mathbf{0}$$

Prony's method (1795), modified least squares, linear prediction, and Iterative Quadratic Maximum Likelihood (IQML) are used to solve exact ML or its modifications.

- Rank-reduction is used to combat noise.

- Requires to estimate the modal order.

D. W. Tufts and R. Kumaresan, "Singular value decomposition and improved frequency estimation using linear prediction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, no. 4, pp. 671675, Aug. 1982.

D. W. Tufts and R. Kumaresan, "Estimation of frequencies of mul- tiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE.*, vol. 70, pp. 975989, 1982.

L. L. Scharf "Statistical Signal Processing," Prentice Hall, 1991.

▶ **Example**: Exact recovery via Linear Prediction
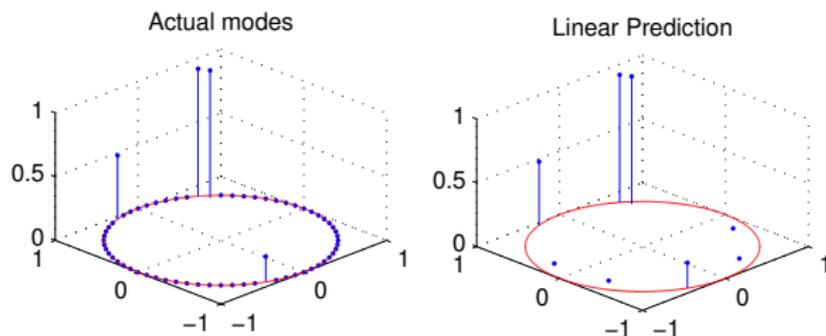


Two damped and two undamped modes

# Classical Parameter Estimation: Fundamental Limits

- **Estimation-theoretic fundamental limits and performance bounds:**

  - Fisher Information
  - Kullback-Leibler divergence
  - Cramér-Rao bounds
  - Ziv-Zakai bound
  - SNR Thresholds



Fisher     Edgeworth     Kullback

Leibler     Cramér     Rao

- **Key fact:** Any subsampling of the measured image (e.g. compressed sensing) has consequences for resolution (or bias) and for variability (or variance) in parameter estimation.

---

L. L. Scharf "Statistical Signal Processing," Prentice Hall, 1991.

# CS and Fundamental Estimation Bounds I

- Canonical model before compression:

$$\mathbf{y} = \mathbf{\Psi}(\boldsymbol{\nu})\boldsymbol{\alpha} + \mathbf{n} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{n}$$

  where $\boldsymbol{\theta}^T = [\boldsymbol{\nu}^T, \boldsymbol{\alpha}^T] \in \mathbb{C}^p$ and $\mathbf{s}(\boldsymbol{\theta}) = \mathbf{\Psi}(\nu)\boldsymbol{\alpha} \in \mathbb{C}^n$.

- Canonical model after compression (of noisy data):

$$\mathbf{\Phi}\mathbf{y} = \mathbf{\Phi}(\mathbf{\Psi}(\boldsymbol{\nu})\boldsymbol{\alpha} + \mathbf{n}) = \mathbf{\Phi}(\mathbf{s}(\boldsymbol{\theta}) + \mathbf{n})$$

  where $\mathbf{\Phi} \in \mathbb{C}^{m \times n}$, $m \ll n$, is a compressive sensing matrix.

- Observation: $\mathbf{y} \sim p(\mathbf{y}; \boldsymbol{\theta})$ (or $\mathbf{\Phi}\mathbf{y} \sim p(\mathbf{\Phi}\mathbf{y}; \boldsymbol{\theta})$ after compression)

# CS and Fundamental Estimation Bounds II

**Estimation-theoretic measures:**

- Fisher information matrix: Covariance of Fisher score

$$\{\mathbf{J}(\boldsymbol{\theta})\}_{i,j} = E\left[\left(\frac{\partial}{\partial\theta_i}\log p(\mathbf{y};\boldsymbol{\nu})\right)\left(\frac{\partial}{\partial\theta_j}\log p(\mathbf{y};\boldsymbol{\theta})\right)|\boldsymbol{\theta}\right]$$

$$= -E\left[\frac{\partial^2}{\partial^2\theta_i\theta_j}\log p(\mathbf{y};\boldsymbol{\theta})|\boldsymbol{\theta}\right]$$

- Cramér-Rao lower bound (CRB): Lower bounds the error covariance of any unbiased estimator $T(\mathbf{y})$ of the parameter vector $\boldsymbol{\theta}$ from measurement $\mathbf{y}$:

$$\mathrm{tr}[\mathrm{cov}_{\boldsymbol{\theta}}(T(\mathbf{y}))] \geq \mathrm{tr}[\mathbf{J}^{-1}(\boldsymbol{\theta})]$$

In particular, the $i$th diagonal element of $\mathbf{J}^{-1}(\boldsymbol{\theta})$ lower bounds the MSE of any unbiased estimator $T_i(\mathbf{y})$ of the $i$th parameter $\theta_i$ from $\mathbf{y}$.

# CS, Fisher Information, and CRB

**Question:** What is the impact of compression (e.g. CS) on the Fisher information matrix and the CRB for estimating parameters?

> ## Theorem (Pakrooh, Pezeshki, Scharf, Chi '13)
>
> *(a) For any compression matrix, we have*
>
> $$(\mathbf{J}^{-1}(\boldsymbol{\theta}))_{ii} \leq (\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} \leq 1/\lambda_{min}(\mathbf{G}^T(\boldsymbol{\theta})\mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G}(\boldsymbol{\theta}))$$
>
> *(b) For a random compression matrix, we have*
>
> $$(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} \leq \frac{\lambda_{max}(\mathbf{J}^{-1}(\boldsymbol{\theta}))}{C(1-\epsilon)}$$
>
> *with probability at least $1 - \delta - \delta'$.*

**Remarks:**

- $(\hat{\mathbf{J}}^{-1})_{ii}$ is the CRB in estimating the $i$th parameter $\theta_i$.
- CRB always gets worse after compressive sampling.
- Theorem gives a confidence interval and a confidence level for the increase in CRB after random compression.

# CS, Fisher Information, and CRB

$$(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} \leq \frac{\lambda_{max}(\mathbf{J}^{-1}(\boldsymbol{\theta}))}{C(1-\epsilon)}$$

- $\delta$ satisfies

$$Pr\left(\forall \mathbf{q} \in \langle \mathbf{G}(\boldsymbol{\theta}) \rangle : (1-\epsilon)\|\mathbf{q}\|_2^2 \leq \|\boldsymbol{\Phi}\mathbf{q}\|_2^2 \leq (1+\epsilon)\|\mathbf{q}\|_2^2\right) \geq 1 - \delta.$$

- $1 - \delta'$ is the probability that $\lambda_{min}((\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1})$ is larger than $C$.

- If entries of $\boldsymbol{\Phi}_{m \times n}$ are i.i.d. $\mathcal{N}(0, 1/m)$, then
    - $\delta \leq \lceil (2\sqrt{p}/\epsilon')^p \rceil e^{-m(\epsilon^2/4 - \epsilon^3/6)}$, where

    $$(\frac{3\epsilon'}{1-\epsilon'})^2 + 2(\frac{3\epsilon'}{1-\epsilon'}) = \epsilon.$$

    - $\delta'$ is determined from the distribution of the largest eigenvalue of a Wishart matrix, and the value of $C$, from a hypergeometric function.

P. Pakrooh, L. L. Scharf, A. Pezeshki and Y. Chi, "Analysis of Fisher information and the Cramer-Rao bound for nonlinear parameter estimation after compressed sensing", *in Proc. 2013 IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Vancouver May 26-31, 2013.

# CRB after Compression

Example: Estimating the DOA of a point source at boresight $\theta_1 = 0$ in the presence of a point interferer at electrical angle $\theta_2$.
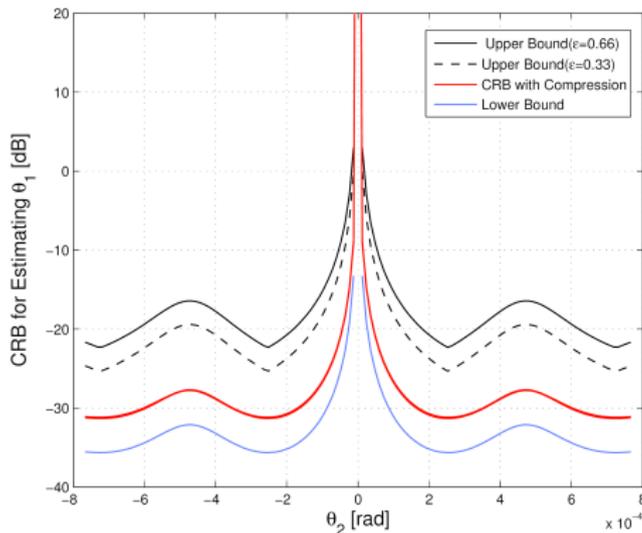


- The figure shows the after compassion CRB (red) for estimating $\theta_1 = 0$ as $\theta_2$ is varied inside the $(-2\pi/n, 2\pi/n]$ interval. Gaussian compression is done from dimension $n = 8192$ to $m = 3000$. Bounds on the after compression CRB are shown in blue and black. The upper bounds in black hold with probability at least $1 - \delta - \delta'$, where $\delta' = 0.05$.

# Applying $\ell_1$ minimization to Parameter Estimation

- Convert the nonlinear modal representation into a linear system via discretization of the parameter space at desired resolution:

$$\mathbf{s}(\boldsymbol{\theta}) = \sum_{i=1}^{r} \boldsymbol{\psi}(\boldsymbol{\nu}_i)\alpha_i$$

$$= \boldsymbol{\Psi}_{ph}\boldsymbol{\alpha}$$

$$\mathbf{s} \approx [\boldsymbol{\psi}(\omega_1), \cdots, \boldsymbol{\psi}(\omega_n)] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \boldsymbol{\Psi}_{cs}\mathbf{x}$$

Over-determined & nonlinear

Under-determined linear & sparse
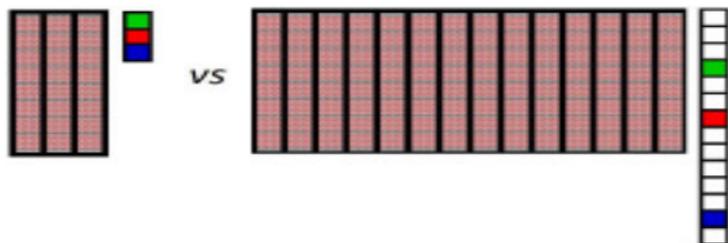


- The set of candidate $\nu_i \in \Omega$ is quantized to $\tilde{\Omega} = \{\omega_1, \cdots, \omega_n\}$, $n > m$; $\boldsymbol{\Psi}_{ph}$ unknown and $\boldsymbol{\Psi}_{cs}$ assumed known.

# Basis Mismatch: A Tale of Two Models

**Mathematical (CS) model:**

$$\mathbf{s} = \mathbf{\Psi}_{cs}\mathbf{x}$$

The basis $\mathbf{\Psi}_{cs}$ is assumed, typically a gridded imaging matrix (e.g., $n$ point DFT matrix or identity matrix), and $\mathbf{x}$ is presumed to be $k$-sparse.
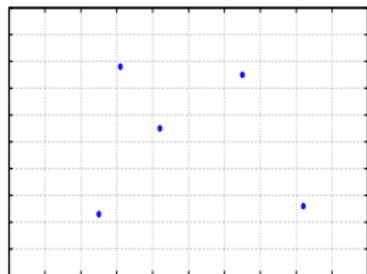


**Physical (true) model:**

$$\mathbf{s} = \mathbf{\Psi}_{ph}\boldsymbol{\alpha}$$

The basis $\mathbf{\Psi}_{ph}$ is unknown, and is determined by a point spread function, a Green's function, or an impulse response, and $\boldsymbol{\alpha}$ is $k$-sparse and unknown.

**Key transformation:**

$$\mathbf{x} = \mathbf{\Psi}_{mis}\boldsymbol{\alpha} = \mathbf{\Psi}_{cs}^{-1}\mathbf{\Psi}_{ph}\boldsymbol{\alpha}$$

$\mathbf{x}$ is sparse in the unknown $\mathbf{\Psi}_{mis}$ basis, not in the identity basis.

# Basis Mismatch: From Sparse to Incompressible

DFT Grid Mismatch:

$$\Psi_{mis} = \Psi_{cs}^{-1}\Psi_{ph} = \begin{bmatrix} L(\Delta\theta_0 - 0) & L(\Delta\theta_1 - \frac{2\pi(n-1)}{n}) & \cdots & L(\Delta\theta_{n-1} - \frac{2\pi}{n}) \\ L(\Delta\theta_0 - \frac{2\pi}{n}) & L(\Delta\theta_1 - 0) & \cdots & L(\Delta\theta_{n-1} - \frac{2\pi\cdot 2}{n}) \\ \vdots & \vdots & \ddots & \vdots \\ L(\Delta\theta_0 - \frac{2\pi(n-1)}{n}) & L(\Delta\theta_1 - \frac{2\pi(n-2)}{n}) & \cdots & L(\Delta\theta_{n-1} - 0) \end{bmatrix}$$

where $L(\theta)$ is the Dirichlet kernel:

$$L(\theta) = \frac{1}{n}\sum_{\ell=0}^{n-1} e^{j\ell\theta} = \frac{1}{n}e^{j\frac{\theta(n-1)}{2}}\frac{\sin(\theta n/2)}{\sin(\theta/2)}.$$
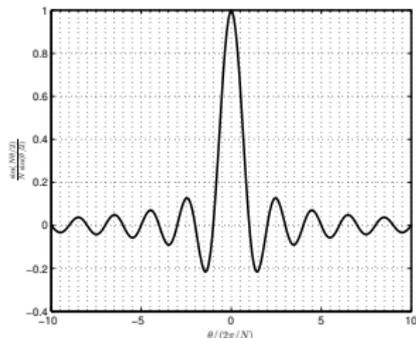


Slow decay of the Dirichlet kernel means that the presumably sparse vector $\mathbf{x} = \Psi_{mis}\boldsymbol{\alpha}$ is in fact incompressible.

# Basis Mismatch: Fundamental Question

Question: What is the consequence of assuming that $\mathbf{x}$ is $k$-sparse in $\boldsymbol{I}$, when in fact it is only $k$-sparse in an *unknown* basis $\boldsymbol{\Psi}_{mis}$, which is determined by the mismatch between $\boldsymbol{\Psi}_{cs}$ and $\boldsymbol{\Psi}_{ph}$?



**Physical Model**

$\mathbf{s} = \boldsymbol{\Psi}_{ph}\boldsymbol{\alpha}$

**CS Sampler**

$\mathbf{y} = \boldsymbol{\Phi}\mathbf{s}$

**CS Inverter**

$\min \|\mathbf{x}\|_1$

s.t. $\mathbf{y} = \boldsymbol{\Phi}\boldsymbol{\Psi}_{cs}\mathbf{x}$
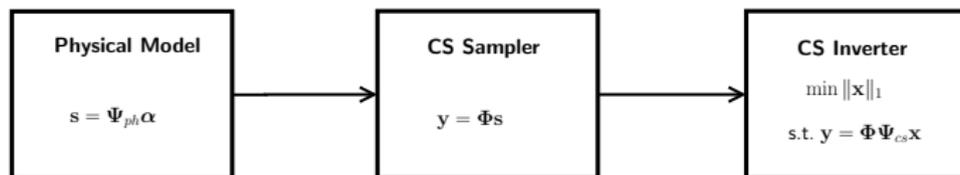
# Sensitivity to Basis Mismatch

▶ CS Inverter: Basis pursuit solution satisfies

$$\text{Noise-free:} \quad \|\mathbf{x}^* - \mathbf{x}\|_1 \leq C_0 \|\mathbf{x} - \mathbf{x}_k\|_1$$
$$\text{Noisy:} \quad \|\mathbf{x}^* - \mathbf{x}\|_2 \leq C_0 k^{-1/2} \|\mathbf{x} - \mathbf{x}_k\|_1 + C_1 \epsilon$$

where $\mathbf{x}_k$ is the best $k$-term approximation to $\mathbf{x}$.

▶ Similar bounds CoSaMP and ROMP.

▶ Where does mismatch enter? $k$-term approximation error.

$$\mathbf{x} = \mathbf{\Psi}_{mis} \boldsymbol{\alpha} = \mathbf{\Psi}_{cs}^{-1} \mathbf{\Psi}_{ph} \boldsymbol{\alpha}$$

▶ **Key:** Analyze the sensitivity of $\|\mathbf{x} - \mathbf{x}_k\|_1$ to basis mismatch.

# Degeneration of Best $k-$Term Approximation

## Theorem (Chi, Scharf, Pezeshki, Calderbank, 2011)

Let $\boldsymbol{\Psi}_{mis} = \boldsymbol{\Psi}_{cs}^{-1}\boldsymbol{\Psi}_{ph} = \mathbf{I} + \mathbf{E}$, where $\mathbf{x} = \boldsymbol{\Psi}_{mis}\boldsymbol{\alpha}$. Let $1 \leq p, q \leq \infty$ and $1/p + 1/q = 1$.

- If the rows $\mathbf{e}_\ell^T \in \mathbb{C}^{1 \times n}$ of $\mathbf{E}$ are bounded as $\|\mathbf{e}_\ell\|_p \leq \beta$, then

$$\|\mathbf{x} - \mathbf{x}_k\|_1 \leq \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\|_1 + (n-k)\beta\|\boldsymbol{\alpha}\|_q.$$

- The bound is achieved when the entries of $\mathbf{E}$ satisfy

$$e_{mn} = \pm\beta \cdot e^{j(\arg(\alpha_m) - \arg(\alpha_n))} \cdot (|\alpha_n|/\|\boldsymbol{\alpha}\|_q)^{q/p}.$$

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

# Bounds on Image Inversion Error

## Theorem (inversion error)

Let $\mathbf{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}_{mis}$ satisfy $\delta_{2k}^{\mathbf{A}} < \sqrt{2} - 1$ and $1/p + 1/q = 1$. If the rows of $\mathbf{E}$ satisfy $\|\mathbf{e}_m\|_p \leq \beta$, then

$$\|\mathbf{x} - \mathbf{x}^*\|_1 \leq C_0(n-k)\beta\|\boldsymbol{\alpha}\|_q. \quad \text{(noise-free)}$$

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq C_0(n-k)k^{-1/2}\beta\|\boldsymbol{\alpha}\|_q + C_1\epsilon. \quad \text{(noisy)}$$

▶ **Message:** In the presence of basis mismatch, *exact or near-exact sparse recovery cannot be guaranteed*. Recovery *may* suffer large errors.

---

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

# Mismatch of DFT Basis in Modal Analysis I

▶ Frequency mismatch
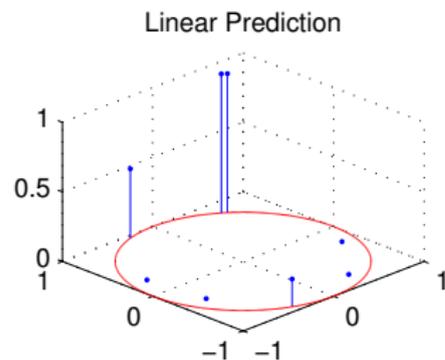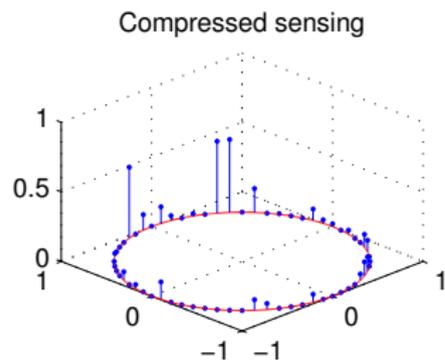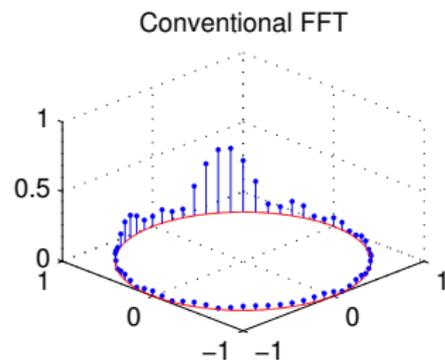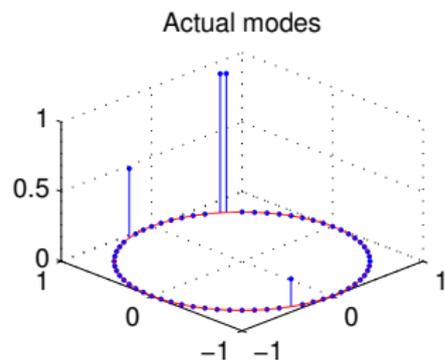
# Mismatch of DFT Basis in Modal Analysis II

- Damping mismatch
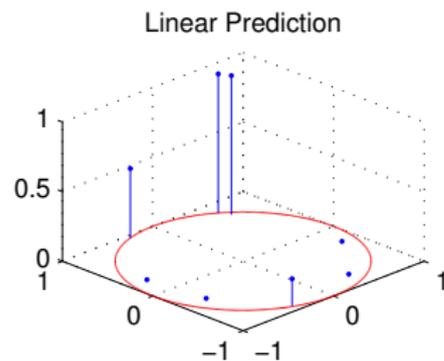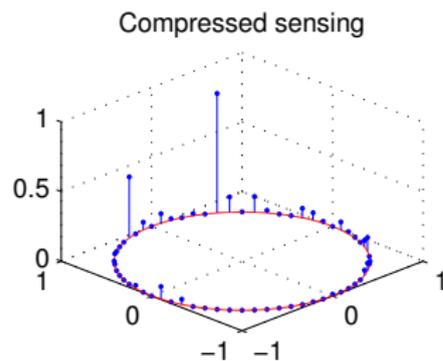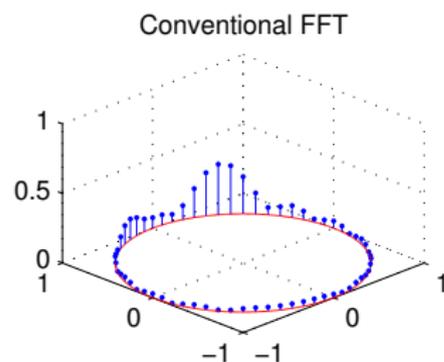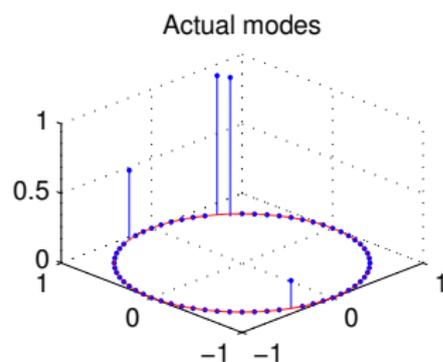
# Mismatch of DFT Basis in Modal Analysis III

▶ Frequency mismatch–noisy measurements



Actual modes

Conventional FFT

Compressed sensing

Linear Prediction with Rank Reduction
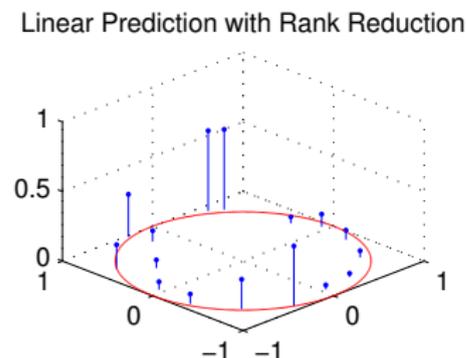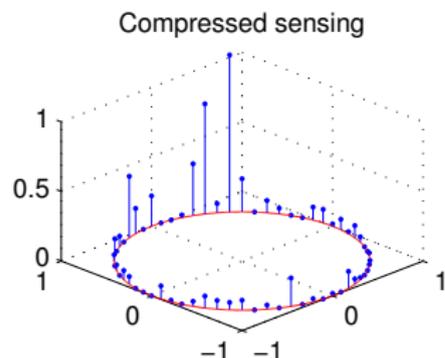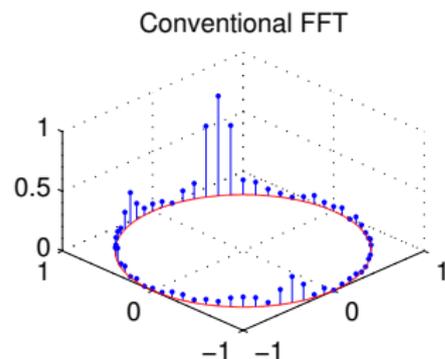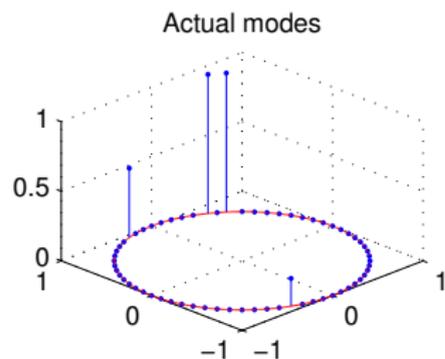
# Mismatch in DFT Frame for Modal Analysis I

**But what if we make the grid finer and finer?**

- Over-resolution experiment:
    - $m = 25$ samples
    - Equal amplitude complex tones at $f_1 = 0.5$ Hz and $f_2 = 0.52$ Hz (half the Rayleigh limit apart), mismatched to mathematical basis.
    - Mathematical model is $\mathbf{s} = \mathbf{\Psi}_{cs}\mathbf{x}$, where $\mathbf{\Psi}_{cs}$ is the $m \times mL$, "DFT" frame that is over-resolved to $\Delta f = 1/mL$.

$$\mathbf{\Psi}_{cs} = \frac{1}{\sqrt{m}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{j\frac{2\pi}{mL}} & \cdots & e^{j\frac{2\pi(mL-1)}{mL}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{j\frac{2\pi(m-1)}{mL}} & \cdots & e^{j\frac{2\pi(m-1)(mL-1)}{mL}} \end{bmatrix}.$$

# Mismatch in DFT Frame for Modal Analysis II

- ▶ MSE of inversion is noise-defeated, noise-limited, or null-space limited — depending on SNR.
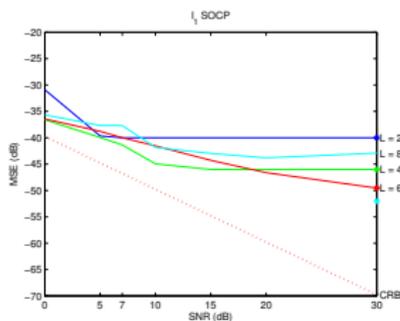


$\ell_1$ inversions
for $L = 2, 4, 6, 8$
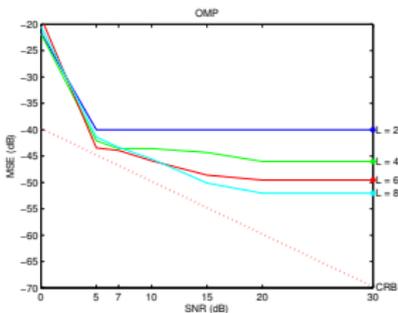
OMP
for $L = 2, 4, 6, 8$
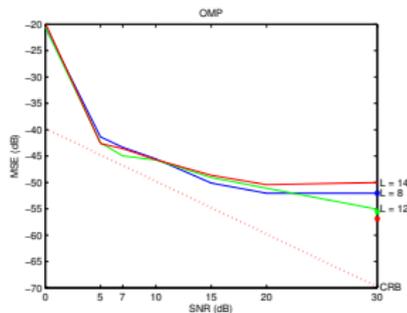
OMP
for $L = 8, 12, 14$

- ▶ The results are worse for a weak mode in the presence of a strong interfering mode.

L. L. Scharf, E. K. P. Chong, A. Pezeshki, and J. R. Luo, "Sensitivity considerations in compressed sensing," *in Conf. Rec. Asilomar'11*, Pacific Grove, CA,, Nov. 2011, pp. 744–748.

# Intermediate Recap: Sensitivity of CS to Basis Mismatch

- Basis mismatch is inevitable when exploiting $\ell_1$ minimization and sensitivities of CS to basis mismatch need to be fully understood. No matter how finely we grid the parameter space, the actual modes almost never lie on the grid.

- The consequence of over-resolution (very fine gridding) is that performance follows the Cramer-Rao bound more closely at low SNR, but at high SNR it departs more dramatically from the Cramer-Rao bound.

- This matches intuition that has been gained from more conventional modal analysis where there is a qualitatively similar trade-off between bias and variance. That is, bias may be reduced with frame expansion (over-resolution), but there is a penalty to be paid in variance.

# References on Model Mismatch in CS

- Y. Chi, A. Pezeshki, L. L. Scharf, and R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *in Proc. ICASSP'10,* Dallas, TX, Mar. 2010, pp. 3930 –3933.
- Y. Chi, L.L. Scharf, A. Pezeshki, and A.R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing,* vol. 59, no. 5, pp. 2182–2195, May 2011.
- L. L. Scharf, E. K. P. Chong, A. Pezeshki, and J. R. Luo, "Compressive sensing and sparse inversion in signal processing: Cautionary notes," *in Proc. DASP'11,* Coolum, Queensland, Australia, Jul. 10-14, 2011.
- L. L. Scharf, E. K. P. Chong, A. Pezeshki, and J. R. Luo, "Sensitivity considerations in compressed sensing," *in Conf. Rec. Asilomar'11,* Pacific Grove, CA,, Nov. 2011, pp. 744–748.
- M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE J. Selected Topics in Signal Processing,* vol. 4, no. 2, pp. 342349, Apr. 2010.
- D. H Chae, P. Sadeghi, and R. A. Kennedy, "Effects of basis-mismatch in compressive sampling of continuous sinusoidal signals," *Proc. Int. Conf. on Future Computer and Commun.,* Wuhan, China, May 2010.

# Some Remedies to Basis Mismatch : A Partial List

These approaches still assume a grid.

- H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Transactions on Signal Processing,* vol. 59, May 2011.

- M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis,* Vol. 35, No. 1, pp. 111-129, 2013.

- A. Fannjiang and W. Liao, "Coherence-Pattern Guided Compressive Sensing with Unresolved Grids," *SIAM Journal of Imaging Sciences*, Vol. 5, No. 1, pp. 179-202, 2012.

# Inspirations for Atomic Minimization I

▶ Prior information to exploit: there are only a few active parameters (sparse!), the exact number of which is unknown.

▶ In compressive sensing, a sparse signal is simple – it is a parsimonious sum of the canonical basis vectors $\{\mathbf{e}_k\}$.

▶ These basis vectors are building blocks for sparse signals.

▶ The $\ell_1$ norm enforces sparsity w.r.t. the canonical basis vectors.

▶ The unit $\ell_1$ norm ball is $\mathrm{conv}\{\pm\mathbf{e}_k\}$, the convex hull of the basis vectors.

▶ A hyperplane will most likely touch the $\ell_1$ norm ball at spiky points, which correspond to sparse solutions.
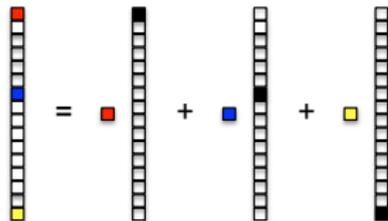


▶ This is the geometrical reason that $\mathrm{minimize}\ \|\mathbf{x}\|_1$ subject to $\mathbf{y} = A\mathbf{x}$ will produce a sparse solution.

# Inspirations for Atomic Minimization II

- Given a finite dictionary $D = \begin{bmatrix} \mathbf{d}_1 & \cdots & \mathbf{d}_p \end{bmatrix}$, we can consider simple signals that have sparse decompositions w.r.t. building blocks $\{\mathbf{d}_k\}$.

- We promote sparsity w.r.t. $D$ by using the norm:

$$\|\mathbf{x}\|_D = \min\{\|\boldsymbol{\alpha}\|_1 : \mathbf{x} = D\boldsymbol{\alpha}\}$$

- The unit norm ball is precisely the convex hull $\mathrm{conv}\{\pm\mathbf{d}_k\}$.

- Minimizing $\|\cdot\|_D$ subject to linear constraint is likely to recover solutions that are sparse w.r.t. $D$.
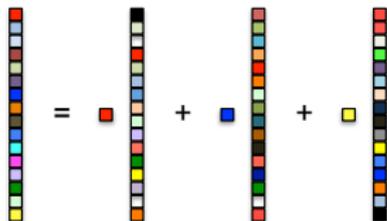
# Inspirations for Atomic Minimization III

- A low rank matrix has a sparse representation in terms of unit-norm, rank-one matrices.
- The dictionary $D = \{\mathbf{u}\mathbf{v}^T : \|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1\}$ is continuously parameterized and has infinite number of primitive signals.
- We enforce low-rankness using the nuclear norm:

$$\|X\|_* = \min\{\|\boldsymbol{\sigma}\|_1 : X = \sum_i \sigma_i \mathbf{u}_i \mathbf{v}_i^T\}$$

- The nuclear norm ball is the convex hull of unit-norm, rank-one matrices.
- A hyperplane touches the nuclear norm ball at low-rank solutions.
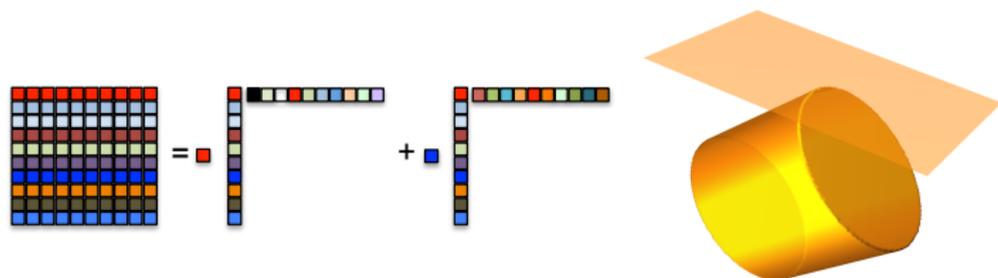
# Atomic Norms I

**Convex geometry.**

- Consider a dictionary or set of atoms $\mathcal{A} = \{\psi(\nu) : \nu \in N\} \subset \mathbb{R}^n$ or $\mathbb{C}^n$.
- The parameter space $N$ can be finite, countably infinite, or continuous.
- The atoms $\{\psi(\nu)\}$ are building blocks for signal representation.
- Examples: canonical basis vectors, a finite dictionary, rank-one matrices.
- **Line spectral atoms:**

$$\mathbf{a}(\nu) = [1, e^{j2\pi\nu}, \ldots, e^{j2\pi(n-1)\nu}]^T : \nu \in [0,1]$$

- **2D line spectral atoms:**

$$\mathbf{a}(\nu_1, \nu_2) = \mathbf{a}(\nu_1) \otimes \mathbf{a}(\nu_2), \nu_1, \nu_2 \in [0,1]$$

- Tensor atoms: $\mathcal{A} = \{\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} \in \mathbb{R}^{m \times n \times p} : \|\mathbf{u}\| = \|\mathbf{v}\| = \|\mathbf{w}\| = 1\}$, unit-norm, rank-one tensors.
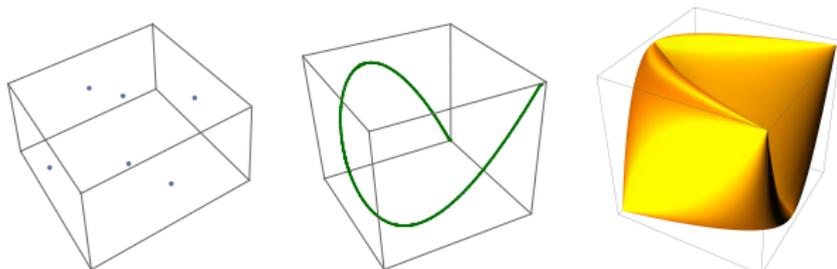
# Atomic Norms II

▶ Prior information: the signal is simple w.r.t. $\mathcal{A}$— it has a parsimonious decomposition using atoms in $\mathcal{A}$

$$\mathbf{x} = \sum_{k=1}^{r} \alpha_k \boldsymbol{\psi}(\nu_k)$$

▶ The atomic norm of any $\mathbf{x}$ is defined as (Chandrasekaran, Recht, Parrilo, & Willsky, 2010)

$$\|\mathbf{x}\|_{\mathcal{A}} = \inf\{\|\boldsymbol{\alpha}\|_1 : \mathbf{x} = \sum_k \alpha_k \boldsymbol{\psi}(\nu_k)\} = \inf\{t > 0 : \mathbf{x} \in t \operatorname{conv}(\mathcal{A})\}$$

▶ The unit ball of the atomic norm is the convex hull of the atomic set $\mathcal{A}$.
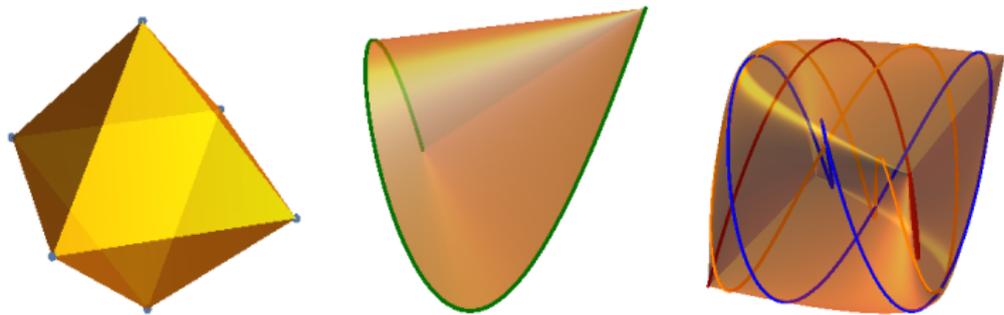
# Atomic Norms III

**Finite optimization.**

- ▶ Given linear measurements of a signal $\mathbf{x}^\star$, possibly with missing data and corrupted by noise and outliers, we want to recover the signal.

- ▶ Suppose we have some prior information that the signal is simple – it has a sparse representation with respect to an atomic set $\mathcal{A}$.

- ▶ We can recover the signal by solving convex optimizations:

$$\underline{\text{Basis Pursuit:}} \text{ minimize } \|\mathbf{x}\|_{\mathcal{A}} \text{ subject to } \mathbf{y} = A\mathbf{x}$$

$$\underline{\text{LASSO:}} \text{ minimize } \frac{1}{2}\|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_{\mathcal{A}}$$

$$\underline{\text{Demixing:}} \text{ minimize } \|\mathbf{x}\|_{\mathcal{A}} + \lambda\|\mathbf{z}\|_1 \text{ subject to } \mathbf{y} = \mathbf{x} + \mathbf{z}.$$
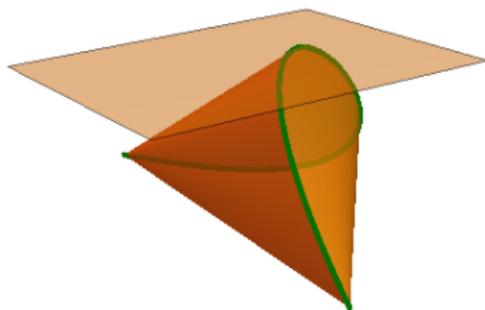
## Atomic Norms IV

- The dual atomic norm is defined as

$$\|\mathbf{q}\|_{\mathcal{A}}^* := \sup_{\mathbf{x}:\|\mathbf{x}\|_{\mathcal{A}} \le 1} |\langle \mathbf{x}, \mathbf{q} \rangle| = \sup_{\mathbf{a} \in \mathcal{A}} |\langle \mathbf{a}, \mathbf{q} \rangle|$$

- For **line spectral atoms**, the dual atomic norm is the maximal magnitude of a complex trigonometric polynomial.

$$\|\mathbf{q}\|_{\mathcal{A}}^* = \sup_{\mathbf{a} \in \mathcal{A}} |\langle \mathbf{a}, \mathbf{q} \rangle| = \sup_{\nu \in [0,1]} \left| \sum_{k=0}^{n-1} q_k e^{j2\pi k \nu} \right|$$

| Atoms | Atomic Norm | Dual Atomic Norm |
|-------|-------------|------------------|
| canonical basis vectors | $\ell_1$ norm | $\ell_\infty$ norm |
| finite atoms | $\|\cdot\|_D$ | $\|D^T \mathbf{q}\|_\infty$ |
| unit-norm, rank-one matrices | nuclear norm | spectral norm |
| unit-norm, rank-one tensors | tensor nuclear norm | tensor spectral norm |
| line spectral atoms | $\|\cdot\|_{\mathcal{A}}$ | $\|\cdot\|_{\mathcal{A}}^*$ |

# Atomic Norms V

**Measure optimization.**

- Rewrite the decomposition $\mathbf{x} = \sum_{k=1}^{r} \alpha_k \boldsymbol{\psi}(\nu_k)$ as

$$\mathbf{x} = \int_N \boldsymbol{\psi}(\nu)\mu(d\nu)$$

where $\mu = \sum_{k=1}^{r} \alpha_k \delta(\nu - \nu_k)$ is a discrete signed measure defined on the parameter space $N$.

- The atomic norm $\|\mathbf{x}\|_{\mathcal{A}}$ equals the optimal value of an infinite dimensional $\ell_1$ minimization:

$$\underset{\mu \in \mathcal{M}(N)}{\text{minimize}} \|\mu\|_{TV} \text{ subject to } \mathbf{x} = \int_N \boldsymbol{\psi}(\nu)\mu(d\nu)$$

- Here $\mathcal{M}(N)$ is the set of all measures defined on $N$, and $\|\mu\|_{TV}$ is the total variation norm of a measure.

## Atomic Norms VI

- When $\mu = \sum_{k=1}^{r} \alpha_k \delta(\nu - \nu_k)$ is a discrete measure,
  $\|\mu\|_{TV} = \sum_{k=1}^{r} |\alpha_k|$.

- When $\mu$ has a density function $\rho(\nu)$, $\|\mu\|_{TV} = \int_N |\rho(\nu)| d\nu = \|\rho(\nu)\|_{L_1}$.

- The equivalent measure optimization definition allows us to apply optimization theory and convex analysis to study atomic norm problems.

- The dual problem is a semi-infinite program:

$$\text{maximize } \langle \mathbf{q}, \mathbf{x} \rangle \text{ subject to } \underbrace{|\langle \mathbf{q}, \boldsymbol{\psi}(\nu) \rangle| \leq 1, \forall \nu \in N}_{\|\mathbf{q}\|_{\mathcal{A}}^* \leq 1}$$

# Problems I

**Fundamentals:**

- ▶ Atomic decomposition: Given a signal, which decompositions achieve the atomic norm?

- ▶ Recovery from noise-free linear measurements: how many measurements do we need to recover a signal that has a sparse representation w.r.t. an atomic set?

- ▶ Denoising: how well can we denoise a signal by exploiting its simplicity structure?

- ▶ Support recovery: how well can we approximately recover the active parameters from noisy data?

- ▶ Resolution limit: what's the fundamental limit in resolving active parameters?

- ▶ Computational methods: how shall we solve atomic norm minimization problems?

# Problems II

**Special cases and applications:**

- Atomic norm of tensors: how to find atomic decompositions of tensors?
- Atomic norm of spectrally-sparse ensembles: how to define the atomic norm for multiple measurement vector (MMV) models?
- Super-resolution of mixture models: how to solve the problem when multiple forms of atoms exist?
- Blind super-resolution: how to solve the problem when the form of the atoms are not known precisely?
- Applications on single-molecule imaging.

## Atomic Decomposition I

- Consider a parameterized set of atoms $\mathcal{A} = \{\boldsymbol{\psi}(\nu), \nu \in N\}$ and a signal $\mathbf{x}$ with decomposition

$$\mathbf{x} = \sum_{k=1}^{r} \alpha_k^\star \boldsymbol{\psi}(\nu_k^\star),$$

  under what conditions on the parameters $\{\alpha_k^\star, \nu_k^\star\}$, we have

$$\|\mathbf{x}\|_{\mathcal{A}} = \|\boldsymbol{\alpha}^\star\|_1?$$

- For $\mathcal{A} = \{\pm\mathbf{e}_k\}$, this question is trivial.
- For $\mathcal{A} = \{\mathbf{u}\mathbf{v}^T : \|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1\}$, the composing atoms should be orthogonal (Singular Value Decomposition).
- For $\mathcal{A} = \{\pm\mathbf{d}_k\}$, a sufficient condition is that the dictionary matrix $D$ satisfies restricted isometry property.

# Atomic Decomposition II

**Optimality condition.**

- Define $\mu^\star = \sum_{k=1}^r \alpha_k^\star \delta(\nu - \nu_k^\star)$. We are asking when $\mu^\star$ is the optimal solution of

$$\underset{\mu \in \mathcal{M}(N)}{\text{minimize}} \ \|\mu\|_{TV} \ \text{subject to} \ \mathbf{x} = \int_N \boldsymbol{\psi}(\nu)\mu(d\nu)$$

- Atomic decomposition studies the parameter estimation ability of total variation minimization in the full-data, noise-free case.

- Recall the dual problem:

$$\text{maximize} \ \langle \mathbf{q}, \mathbf{x} \rangle \ \text{subject to} \ \underbrace{|\langle \mathbf{q}, \boldsymbol{\psi}(\nu) \rangle| \leq 1, \forall \nu \in N}_{\|\mathbf{q}\|_{\mathcal{A}}^* \leq 1}$$

- Optimality condition: $\mu^\star$ is optimal if and only if there exists a dual certificate $\mathbf{q}$ such that

$$|\langle \mathbf{q}, \boldsymbol{\psi}(\nu) \rangle| \leq 1, \forall \nu \in N$$
$$\langle \mathbf{q}, \mathbf{x} \rangle = \|\mu^\star\|_{TV}$$

# Atomic Decomposition III

▶ Define a function $q(\nu) = \langle \mathbf{q}, \boldsymbol{\psi}(\nu) \rangle$. The optimality condition becomes

$$\text{dual feasibility: } \|q(\nu)\|_{L_\infty} \leq 1$$
$$\text{complementary slackness: } q(\nu_k^\star) = \text{sign}(\alpha_k^\star), k \in [r]$$

▶ To ensure the uniqueness of the optimal solution $\mu^\star$, we strengthen the optimality condition to:

$$\text{strict boundeness: } |q(\nu)| < 1, \nu \in N/\{\nu_k^\star, k \in [r]\}$$
$$\text{interpolation: } q(\nu_k^\star) = \text{sign}(\alpha_k^\star), k \in [r]$$
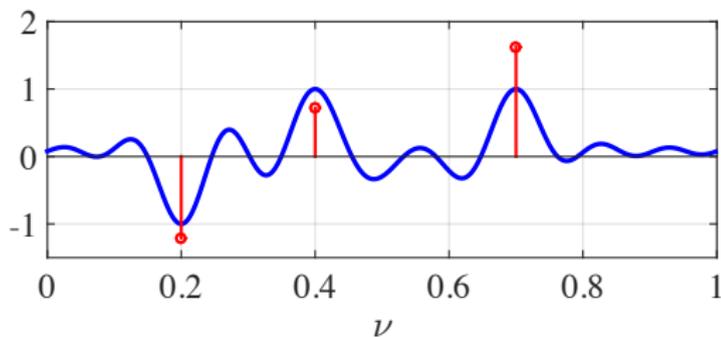
# Atomic Decomposition IV

**Subdifferential**

- The subdifferential of $\| \cdot \|_{\mathcal{A}}$ at $\mathbf{x}$ is

$$\partial \|\mathbf{x}\|_{\mathcal{A}} = \{\mathbf{q} : \|\mathbf{q}\|_{\mathcal{A}}^* \leq 1, \langle \mathbf{q}, \mathbf{x} \rangle = \|\mathbf{x}\|_{\mathcal{A}}\},$$

which coincides with the optimality condition.

- Therefore, the dual certificate is a subgradient of the atomic norm.

- Example: For the nuclear norm, if the reduced SVD of a matrix $X$ is $U\Sigma V^T$, then the subdifferential has the characterization

$$\partial \|X\|_* = \{Q : Q = UV^T + W, U^T W = 0, WV = 0, \|W\| \leq 1\}$$

- For general atomic norms, it seems hopeless to fully characterize the subdifferential.

- To find atomic decomposition conditions, a dual certificate is usually constructed, which merely finds *one* subgradient in the subdifferential.

# Atomic Decomposition V

**Minimal energy dual certificate.**

- The boundedness and interpolation conditions imply that the function $q(\nu)$ achieves maximum or minimum at $\nu = \nu_k^\star$.

- We require that a pre-certificate function to satisfy

$$\frac{\partial}{\partial \nu} q(\nu_k^\star) = 0, k \in [r]$$
$$q(\nu_k^\star) = \text{sign}(\alpha_k^\star), k \in [r]$$

- To ensure that $|q(\nu)|$ is small, we push it down by minimizing the (possibly weighted) energy of $\mathbf{q}$ to get a pre-certificate as the solution of

$$\text{minimize } \frac{1}{2}\mathbf{q}^T W^{-1}\mathbf{q}$$
$$\text{subject to } \langle \mathbf{q}, \frac{\partial}{\partial \nu}\boldsymbol{\psi}(\nu_k^\star)\rangle = 0, \langle \mathbf{q}, \boldsymbol{\psi}(\nu_k^\star)\rangle = \text{sign}(\alpha_k^\star), k \in [r]$$

# Atomic Decomposition VI

▶ This leads to the following kernel expansion of the pre-certificate function

$$q(\nu) = \sum_{k=1}^{r} c_k K(\nu, \nu_k^\star) + \sum_{k=1}^{r} d_k \partial K(\nu, \nu_k^\star)$$

where the kernel $K(\nu, \xi) = \boldsymbol{\psi}(\nu)^T W \boldsymbol{\psi}(\xi)$.

▶ For **line spectral atoms**, when $W = \mathrm{diag}(w)$ with $w$ being the autocorrelation sequence of the triangle function, the corresponding $K(\nu, \xi) = K(\nu - \xi)$ is the Jackson kernel (squared Fejér), which decays rapidly.
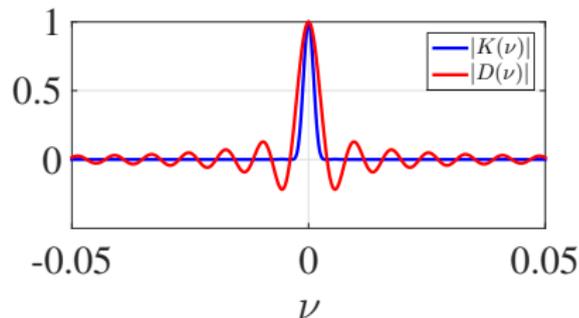
# Atomic Decomposition VII

**Line spectral decomposition.**

- Using these ideas, for line spectal atoms
  $\mathbf{a}(\nu) = \begin{bmatrix} 1 & e^{j2\pi\nu} & \cdots & e^{j2\pi n\nu} \end{bmatrix}^T$, Candès and Fernandez-Granda
  obtained the following theorem

---

## Theorem (Candès & Fernandez-Granda, 2012)

*If the true paramters $\{\nu_k^\star\}$ are separated by $\frac{4}{n}$, the atomic norm
$\|\mathbf{x}\|_{\mathcal{A}} = \sum_{k=1}^r |\alpha_k^\star|$.*

---

- The critical separation was improved to $\frac{2.52}{n}$ (Fernandez-Granda, 2015).
- The separation condition is in a flavor similar to the restricted isometry property for finite dictionaries, and the orthogonality condition for singular value decomposition.
- For atomic decomposition results (full-data, noise-free), the sparsity level is typically only restricted by the separattion constraint and can be large.

# Atomic Decomposition VIII

**Other decomposition results.**

- Finite dictionary: restricted isometry property [Candès, Romberg, Tao, 2004]
- 2D line spectral atoms: separation of parameters [Candès & Fernandez-Granda, 2012].
- Symmetric rank-1 tensors: soft-orthogonality of the factors [Tang & Shah 2015].
- Non-symmetric rank-1 tensors: incoherence, Gram isometry, etc. [Li, Prater, Shen & Tang, 2015]
- Translation invariant signals: separation of translations [Tang & Recht 2013; Bendory, Dekel & Feuer 2014]
- Spherical harmonics: separation of parameters [Bendory, Dekel & Feuer 2014]
- Radar signals: separation of time-frequency shifts [Heckel, Morgenshtern & Soltanolkotabi, 2015]

# Resolution Limits I

**Why there is a resolution limit?**

- To simultaneously interpolate $\text{sign}(\alpha_i^\star) = +1$ and $\text{sign}(\alpha_j^\star) = -1$ at $\nu_i^\star$ and $\nu_j^\star$ respectively while remain bounded imposes constraints on the derivative of $q(\nu)$:

$$\|\nabla q(\hat{\nu})\|_2 \geq \frac{|q(\nu_i^\star) - q(\nu_j^\star)|}{\Delta_{i,j}} = \frac{2}{\Delta_{i,j}}$$

- For $N \subset \mathbb{R}$, there exists $\hat{\nu} \in (\nu_i^\star, \nu_j^\star)$ such that

$$q'(\hat{\nu}) = 2/(\nu_j^\star - \nu_i^\star)$$
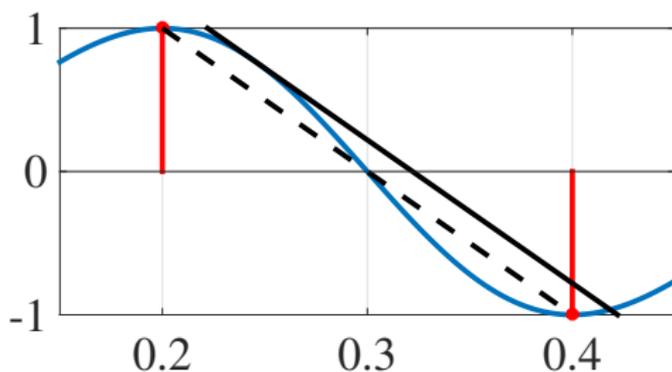
# Resolution Limits II

- For certain classes of functions $\mathcal{F}$, if the function values are uniformly bounded by $1$, this limits the maximal achievable derivative, i.e.,

$$\sup_{g \in \mathcal{F}} \frac{\|g'\|_\infty}{\|g\|_\infty} < \infty.$$

- For $\mathcal{F} = \{$trigonometric polynomials of degree at most $n\}$,

$$\|g'(\nu)\|_\infty \leq 2\pi n \|g(\nu)\|_\infty.$$

- This is the classical Markov-Bernstein's inequality.

- Resolution limit for line spectral signals: If $\min_{i \neq j} |\nu_i^\star - \nu_j^\star| < \frac{1}{\pi n}$, then there is a sign pattern for $\{\alpha_k^\star\}$ such that $\sum_k \alpha_k^\star \mathbf{a}(\nu_k^\star)$ is not an atomic decomposition.

# Resolution Limits III

- Using a theorem by Turán about the roots of trigonometric polynomials, Duval and Peyŕe obtained a better critical separation bound

$$\min_{i \neq j} |\nu_i^\star - \nu_j^\star| > \frac{1}{n}.$$

- Sign pattern of $\{\alpha_j^\star\}$ plays a big role. There is no resolution limit if, e.g., all $\alpha_j^\star$ are positive ([Schiebinger, Robeva & Recht, 2015]).

# Recovery from Gaussian Measurements I

- Given $\mathbf{y} = A\mathbf{x}^\star$ where the entries of $A$ are i.i.d. Gaussian, we recover $\mathbf{x}^\star$ by solving

$$\text{minimize} \ \|\mathbf{x}\|_{\mathcal{A}} \ \text{subject to} \ \mathbf{y} = A\mathbf{x}.$$

- Highlight the power of atomic regularization.

- When does this work? How many generic (Gaussian) measurements do we need to recover $\mathbf{x}^\star$ exactly?

- Summary of atomic minimization recovery bounds (Chandrasekaran, Recht, Parrilo, & Willsky, 2010):

| Underlying model | Convex heuristic | # Gaussian measurements |
|---|---|---|
| $s$-sparse vector in $\mathbb{R}^p$ | $\ell_1$ norm | $2s(\log(p/s - 1) + 1)$ |
| $m \times m$ rank-$r$ matrix | nuclear norm | $3r(2m - r)$ |
| sign-vector $\{-1, +1\}^p$ | $\ell_\infty$ norm | $p/2$ |
| $m \times m$ permutation matrix | norm induced by Birkhoff polytope | $9\,m \log(m)$ |
| $m \times m$ orthogonal matrix | spectral norm | $(3m^2 - m)/4$ |

# Recovery from Gaussian Measurements II

- Tangent cone: set of directions that decrease the norm at $\mathbf{x}^\star$

$$T_{\mathcal{A}}(\mathbf{x}^\star) = \{\mathbf{d} : \|\mathbf{x}^\star + \alpha\mathbf{d}\|_{\mathcal{A}} \leq \|\mathbf{x}^\star\|_{\mathcal{A}} \text{ for some } \alpha > 0\}$$
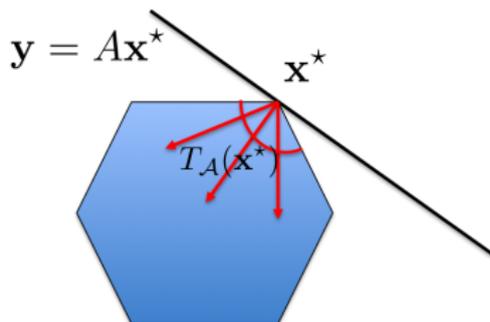


- $\mathbf{x}^\star$ is the unique minimizer iff $\mathrm{null}(A) \bigcap T_{\mathcal{A}}(\mathbf{x}^\star) = \{0\}$.
- When does the random subspace $\mathrm{null}(A)$ intersect the decent cone $T_{\mathcal{A}}(\mathbf{x}^\star)$ only at the origin?
- The size of the descent cone matters as measured by the mean width: we need

$$m \geq n w(T_{\mathcal{A}}(\mathbf{x}^\star) \bigcap \mathbb{S}^{n-1})^2$$

for the recovery of $\mathbf{x}^\star$.

# Recovery from Gaussian Measurements III

- Here the mean width

$$w(T_{\mathcal{A}}(\mathbf{x}^\star) \bigcap \mathbb{S}^{n-1}) := \frac{1}{2} \int_{\mathbb{S}^{n-1}} \sup_{\mathbf{x} \in T_{\mathcal{A}}(\mathbf{x}^\star), \|\mathbf{x}\|_2 = 1} \langle \mathbf{x}, \mathbf{u} \rangle d\mathbf{u}$$

$$\leq \frac{1}{2} \int_{\mathbb{S}^{n-1}} \inf_{\mathbf{z} \in N_{\mathcal{A}}(\mathbf{x}^\star)} \|\mathbf{z} - \mathbf{u}\|_2 d\mathbf{u}$$

- The normal cone $N_{\mathcal{A}}(\mathbf{x}^\star)$ is the polar cone of the descent cone, the cone induced by the subdifferential at $\mathbf{x}^\star$.

- Find a $\mathbf{z} \in N_{\mathcal{A}}(\mathbf{x}^\star)$ that is good enough (depending on $\mathbf{u}$), which requires some knowledge of the subdifferential.
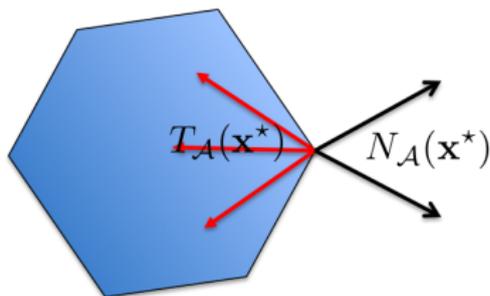
# Recovery with Missing Data I

- Suppose we observe only a (random) portion of the full signal $\mathbf{x}^\star$, $\mathbf{y} = \mathbf{x}_\Omega^\star$, and would like to complete the rest.
- E.g., matrix completion, recovery from partial Fourier transform in compressive sensing
- Optimization formulation:

$$\underset{\mathbf{x}}{\text{minimize}}\,\|\mathbf{x}\|_\mathcal{A} \text{ subject to } \mathbf{x}_\Omega = \mathbf{x}_\Omega^\star.$$

- Results for line spectral signals:

### Theorem (Tang, Bhaskar, Shah & Recht, 2012)

*If we observe $\mathbf{x}^\star = \sum_{k=1}^r \alpha_k^\star \mathbf{a}(\nu_k^\star)$ on a size-$O(r \log(r) \log(n))$ random subset of $\{0, 1, \ldots, n-1\}$ and the true parameters are separated by $\frac{4}{n}$, then atomic norm minimization successfully completes the signal.*

### Theorem (Chi and Chen, 2013)

*Similar results hold for multi-dimensional spectral signals.*

# Recovery with Missing Data II
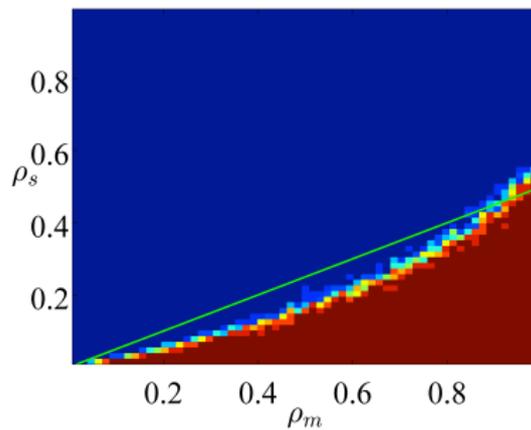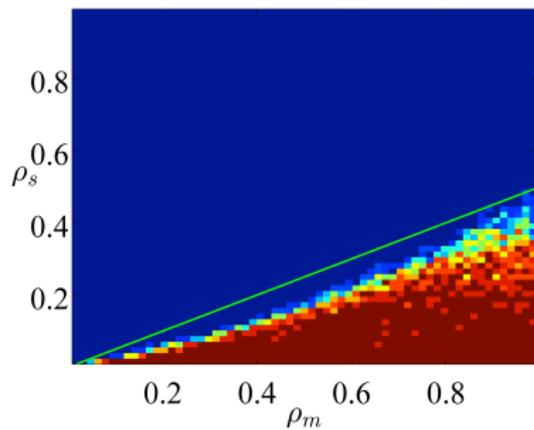
# Recovery with Missing Data III

▶ Dual certificate: $\mathbf{x}^\star$ is the unique minimizer iff there exists a dual certificate vector $\mathbf{q}$ such that the dual certificate function $q(\nu) = \langle \mathbf{q}, \mathbf{a}(\nu) \rangle$ satisfies

$$q(\nu_k^\star) = \mathrm{sign}(\alpha_k^\star), k \in [r]$$
$$|q(\nu)| < 1, \forall \nu \notin \{\nu_k^\star, k \in [r]\}$$
$$\mathbf{q}_i = 0, \forall i \notin \Omega.$$
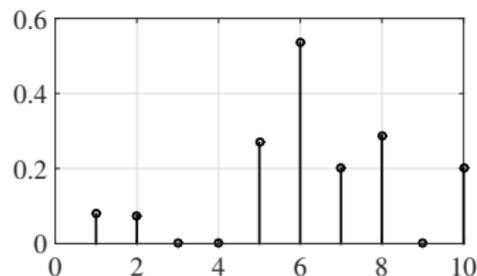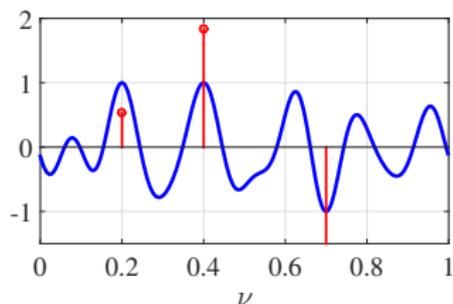
# Recovery with Missing Data IV

▶ The minimal energy construction yields

$$q(\nu) = \sum_{k=1}^{r} c_k K_r(\nu - \nu_k^\star) + \sum_{k=1}^{r} d_k \partial K_r(\nu - \nu_k^\star)$$

where the (random) kernel

$$K_r(\nu) = \mathbf{a}(0)^H W \mathbf{a}(\nu) = \sum_l w_l I_{l \in \Omega} e^{j2\pi\nu l}$$

▶ When the observation index set $\Omega$ is random, argue that $q(\nu)$ is close to the Candès-Fernandez-Granda decomposition dual certificate function using concentration of measure.
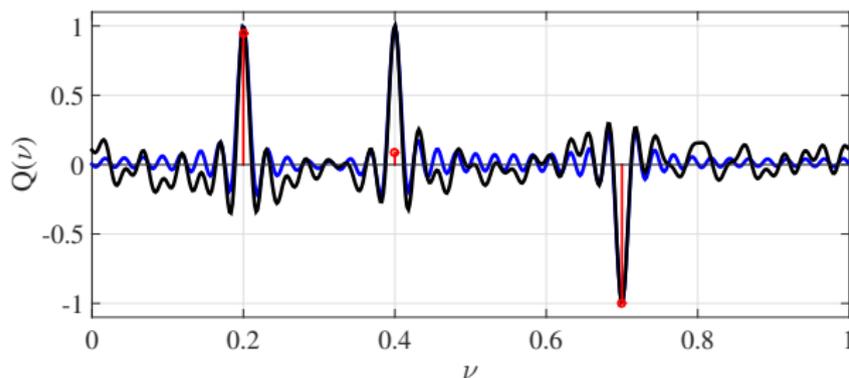
# Denoising I

**Slow rate for general atomic norms**

- Observe noisy measurements: $\mathbf{y} = \mathbf{x}^\star + \mathbf{w}$ with $\mathbf{w}$ a noise.
- Denoise $\mathbf{y}$ to obtain

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_{\mathcal{A}}.$$

- Choose $\lambda \geq \mathbb{E} \|\mathbf{w}\|_{\mathcal{A}}^*$.

### Theorem (Bhaskar, Tang & Recht, 2012)

*Error Rate:* $\frac{1}{n} \mathbb{E} \|\hat{\mathbf{x}} - \mathbf{x}^\star\|_2^2 \leq \frac{\lambda}{n} \|\mathbf{x}^\star\|_{\mathcal{A}}$.

- Specialize to line spectral signals: suppose the signal $\mathbf{x}^\star = \sum_{k=1}^{r} \alpha_k^\star \mathbf{a}(\nu_k^\star)$ and the noise $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 I_n)$.
- We can choose $\lambda = \sigma \sqrt{n \log n}$.

### Theorem (Bhaskar, Tang & Recht, 2012)

*Error Rate:* $\frac{1}{n} \mathbb{E} \|\hat{\mathbf{x}} - \mathbf{x}^\star\|_2^2 \leq \sigma \sqrt{\frac{\log(n)}{n}} \sum_{l=1}^{r} |\alpha_l^\star|$.

# Denoising II

**Fast rate with well-separated frequency parameters.**

## Theorem (Tang, Bhaskar & Recht, 2013)

*Fast Rate:* $\frac{1}{n}\|\hat{\mathbf{x}} - \mathbf{x}^\star\|_2^2 \leq \frac{C\sigma^2 r \log(n)}{n}$ *if the parameters are separated.*

The rate is minimax optimal:

**No algorithm** can do better than

$$\mathbb{E}\frac{1}{n}\|\hat{\mathbf{x}} - \mathbf{x}^\star\|_2^2 \geq \frac{C'\sigma^2 r \log(n/r)}{n}$$

even if the parameters are
well-separated.

**No algorithm** can do better than

$$\frac{1}{n}\|\hat{\mathbf{x}} - \mathbf{x}^\star\|_2^2 \geq \frac{C'\sigma^2 r}{n}$$

even if we know a priori the
well-separated parameters.

# Denoising III
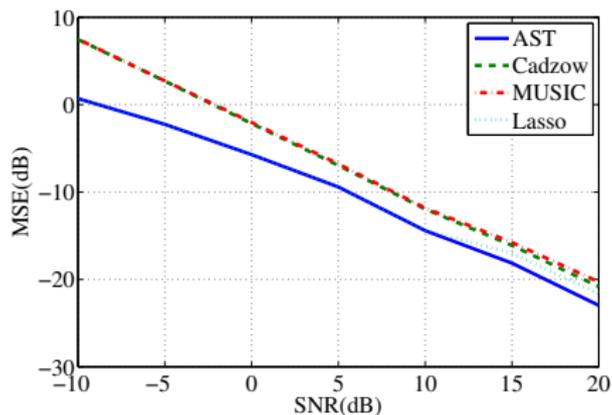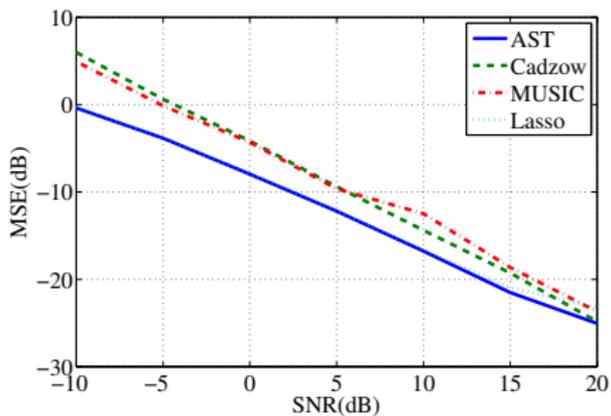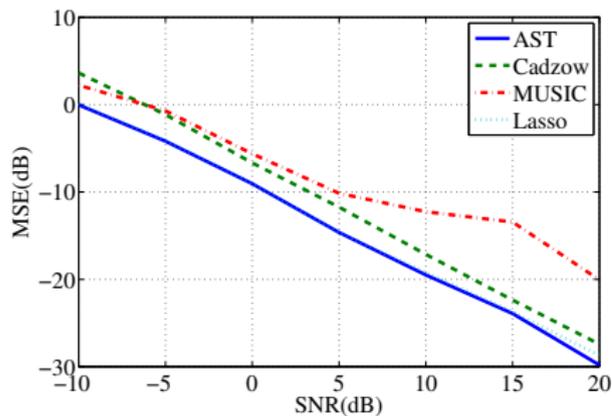
# Noisy Support Recovery/Parameter Estimation I

**Gaussian noise (Tang, Bhaskar & Recht, 2013)**

- When the noise $\mathbf{w}$ is Gaussian, we denoise the signal and recover the frequencies using:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_{\mathcal{A}}.$$

- Dual problem projects $\mathbf{y}$ onto the dual norm ball of radius $\lambda$.

$$\text{maximize} \ \frac{1}{2} \left( \|\mathbf{y}\|_2^2 - \|\mathbf{y} - \mathbf{z}\|_2^2 \right)$$
$$\text{subject to} \ \|\mathbf{z}\|_{\mathcal{A}}^* \leq \lambda.$$

- Optimality condition: The dual certificate for $\hat{\mathbf{x}}$, $\mathbf{q} = (\mathbf{y} - \hat{\mathbf{x}})/\lambda$, is a scaled version of the noise estimator.
- The places where $|\langle \hat{\mathbf{q}}, \psi(\nu) \rangle| = 1$ correspond to support.

# Noisy Support Recovery/Parameter Estimation II
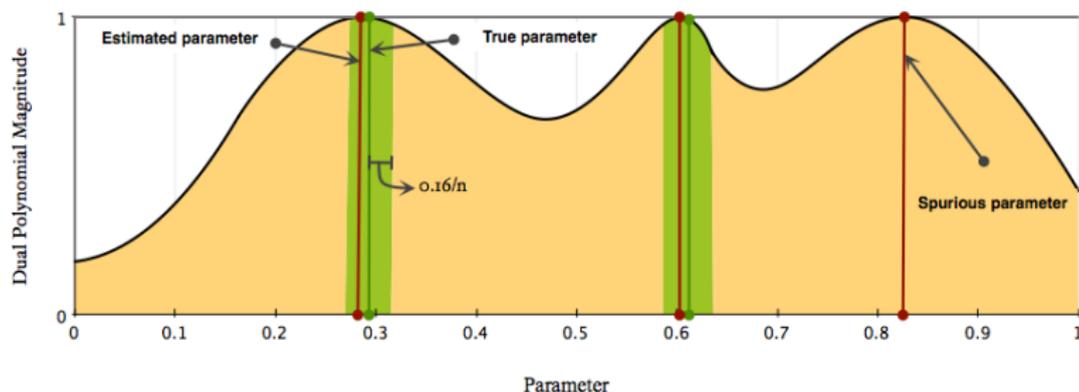


- Spurious amplitudes: $\sum_{l:\hat{\nu}_l \in F} |\hat{\alpha}_l| \leq C_1 \sigma \sqrt{\frac{r^2 \log(n)}{n}}$.
- Frequency deviation:
  $$\sum_{l:\hat{\nu}_l \in N_j} |\hat{\alpha}_l| \left\{ n \min_{\nu_j^\star} d(\nu_j^\star, \hat{\nu}_l) \right\}^2 \leq C_2 \sigma \sqrt{\frac{r^2 \log(n)}{n}}.$$
- Near-region approximation: $\left| \alpha_j^\star - \sum_{l:\hat{\nu}_l \in N_j} \hat{\alpha}_l \right| \leq C_3 \sigma \sqrt{\frac{r^2 \log(n)}{n}}$.

# Noisy Support Recovery/Parameter Estimation III

▶ For any $\nu_i^\star$ such that $\alpha_i^\star > C_3\sigma\sqrt{\frac{r^2\log(n)}{n}}$, there exists a recovered frequency $\hat{\nu}_i$ such that

$$|\nu_i^\star - \hat{\nu}_i| \leq \frac{\sqrt{C_2/C_3}}{n}\left(\frac{|\alpha_i^\star|}{C_3\sigma\sqrt{\frac{r^2\log(n)}{n}}} - 1\right)^{-\frac{1}{2}}$$

**Bounded noise (Fernandez-Granda, 2013)**

▶ When the noise $\mathbf{w}$ is bounded, $\|\mathbf{w}\|_2 \leq \delta$, we denoise the signal and recover the frequencies by solving:

$$\text{minimize } \|\mathbf{x}\|_{\mathcal{A}} \text{ subject to } \|\mathbf{y} - \mathbf{x}\|_2 \leq \delta.$$

▶ Spurious amplitudes: $\sum_{l:\hat{\nu}_l \in F} |\hat{\alpha}_l| \leq C_1\delta$.

▶ Frequency deviation: $\sum_{l:\hat{\nu}_l \in N_j} |\hat{\alpha}_l| \left\{n\min_{\nu_j^\star} d(\nu_j^\star, \hat{\nu}_l)\right\}^2 \leq C_2\delta$.

▶ Near-region approximation: $\left|\alpha_j - \sum_{l:\hat{\nu}_l \in N_j} \hat{\alpha}_l\right| \leq C_3\delta$.

# Noisy Support Recovery/Parameter Estimation IV

▶ For any $\nu_i^\star$ such that $\alpha_i^\star > C_3\delta$, there exists a recovered frequency $\hat{\nu}_i$ such that

$$|\nu_i^\star - \hat{\nu}_i| \leq \frac{1}{n}\sqrt{\frac{C_2\delta}{|\alpha_i^\star| - C_3\delta}}$$

**Small noise.**

> ### Theorem (Duval & Peyŕe, 2013)
> *Suppose the frequency parameters are well-separated and the coefficients $\{\alpha_i^\star\}$ are real, when both the noise $\mathbf{w}$ and the regularization parameter $\lambda$ are small, regularized atomic norm minimization will recover exactly $r$ parameters in a small neighborhood of the true parameters.*

# Computational Methods I

**Semidefinite Reformulations/Relaxations.**

- The dual problem involves a dual norm constraint of the form

$$\|\mathbf{z}\|_{\mathcal{A}}^* \leq 1 \Leftrightarrow |\langle \mathbf{z}, \boldsymbol{\psi}(\nu) \rangle| \leq 1 \ \forall \nu \in N$$

- Line spectral atoms:

$$\|\mathbf{z}\|_{\mathcal{A}}^* \leq 1 \Leftrightarrow |\sum_{k=0}^{n-1} \mathbf{z}_k e^{j2\pi\nu k}| \leq 1 \ \forall \nu \in [0,1]$$

- The latter states that the magnitude of a complex trigonometric polynomial is bounded by 1 everywhere.

- Bounded real lemma (Dumitrescu, 2007):

$$|\sum_{k=0}^{n-1} \mathbf{z}_k e^{j2\pi\nu k}| \leq 1 \ \forall \nu \in [0,1]$$

$$\Leftrightarrow \begin{bmatrix} Q & \mathbf{z} \\ \mathbf{z}^H & 1 \end{bmatrix} \succeq 0,$$

$$\text{trace}(Q, j) = \delta(j = 0), j = 0, \ldots, n-1.$$

# Computational Methods II

▶ This leads to an exact semidefinite representation of the line spectral atomic norm (Bhaskar, Tang & Recht, 2012):

$$\|\mathbf{x}\|_{\mathcal{A}} = \inf \left\{ \frac{1}{2}(t + u_0) : \begin{bmatrix} \text{Toep}(\mathbf{u}) & \mathbf{x} \\ \mathbf{x}^H & t \end{bmatrix} \succeq 0 \right\}$$

▶ Therefore, line spectral atomic norm regularized problems have exact semidefinite representations, e.g.,

$$\text{minimize} \|\mathbf{x}\|_{\mathcal{A}} \text{ subject to } \mathbf{x}_{\Omega} = \mathbf{x}_{\Omega}^{\star}$$

$$\Leftrightarrow$$

$$\text{minimize} \frac{1}{2}(t + u_0) \text{ subject to } \begin{bmatrix} \text{Toep}(\mathbf{u}) & \mathbf{x} \\ \mathbf{x}^H & t \end{bmatrix} \succeq 0, \mathbf{x} = \mathbf{x}_{\Omega}^{\star}$$

# Computational Methods III

**Discretization.**

▶ The dual atomic problem involves a semi-infinite constraint

$$\|\mathbf{z}\|_{\mathcal{A}}^* \le 1 \Leftrightarrow |\langle \mathbf{z}, \boldsymbol{\psi}(\nu)\rangle| \le 1 \ \forall \nu \in N$$

▶ When the dimension of $N$ is small, discretize the parameter space to get a finite number of grid points $N_m$.

▶ Enforce finite number of constraints:

$$|\langle \mathbf{z}, \boldsymbol{\psi}(\nu_j)\rangle| \le 1, \ \forall \nu_j \in N_m$$

▶ Equivalently, we replace the set of atoms with a discrete one

$$\|\mathbf{x}\|_{\mathcal{A}_m} = \inf\{\|\boldsymbol{\alpha}\|_1 : \mathbf{x} = \sum_j \alpha_j \boldsymbol{\psi}(\nu_j), \nu_j \in N_m\}$$

# Computational Methods IV

- What happens to the solutions when

$$\rho(N_m) = \max_{\nu \in N} \min_{\nu' \in N_m} d(\nu, \nu') \to 0$$

## Theorem (Tang, Bhaskar & Recht, 2014; Duval & Peyŕe, 2013)

- *The optimal values converge to the original optimal values.*
- *The dual solutions converge with speed $O(\rho_m)$.*
- *The primal optimal measures converge in distribution.*
- *When the SNR is large enough, the solution of the discretized problem is supported on pairs of parameters which are neighbors of the true parameters.*

# Computational Methods V
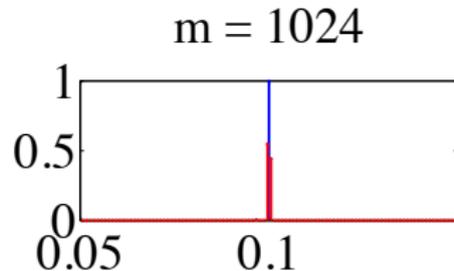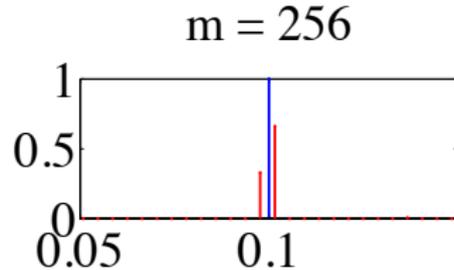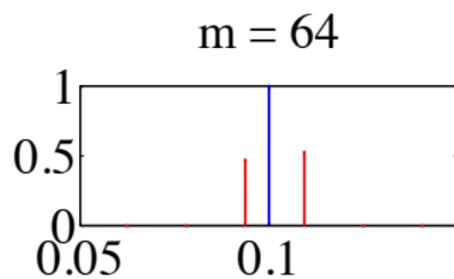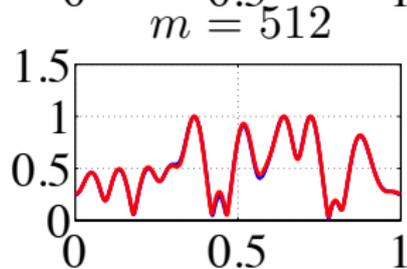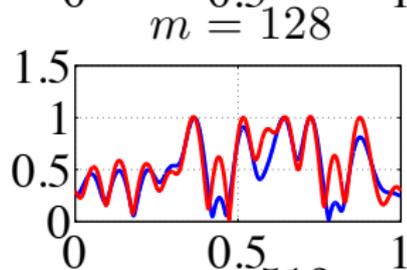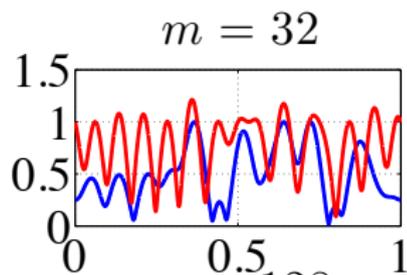
# Problems

**Special cases and applications:**

- Atomic norm of tensors: how to find the atomic decomposition of tensors?
- Atomic norm of spectrally-sparse ensembles: how to define the atomic norm for multiple measurement vector (MMV) models?
- Super-resolution of mixture models: how to solve the problem when multiple forms of atoms exist?
- Blind super-resolution: how to solve the problem when the form of the atoms are not known precisely?
- Applications on single-molecule imaging.

# Atomic Decomposition of Tensors I

**Tensor decomposition.**

- Given a tensor decomposition

$$T = \sum_{i=1}^{r} \alpha_i^\star \mathbf{u}_i^\star \otimes \mathbf{v}_i^\star \otimes \mathbf{w}_i^\star = \int_{\mathbb{K}} \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} d\mu^\star$$

  where the parameter space $\mathbb{K} = \mathbb{S}^{n-1} \times \mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$, the decomposition measure $\mu^\star = \sum_{i=1}^{r} \alpha_i^\star \delta(\mathbf{u} - \mathbf{u}_i^\star, \mathbf{v} - \mathbf{v}_i^\star, \mathbf{w} - \mathbf{w}_i^\star)$ is a nonnegative measure defined on $\mathbb{K}$.

- We propose recovering the decomposition measure $\mu^\star$ by solving,

$$\text{minimize} \, \mu(\mathbb{K}) \text{ subject to } T = \int_{\mathbb{K}} \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} d\mu.$$

- The optimal value of this optiimization defines the tensor nuclear norm.

- To certify the optimality of $\mu^\star$, we a construct a pre-certificate following the minimal energy principle to get

$$q(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \langle Q, \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} \rangle = \sum_{i=1}^{r} (\mathbf{a}_i \otimes \mathbf{v}_i^\star \otimes \mathbf{w}_i^\star + \mathbf{u}_i^\star \otimes \mathbf{b}_i \otimes \mathbf{w}_i^\star + \mathbf{u}_i^\star \otimes \mathbf{v}_i^\star \otimes \mathbf{c}_i)$$

# Atomic Decomposition of Tensors II
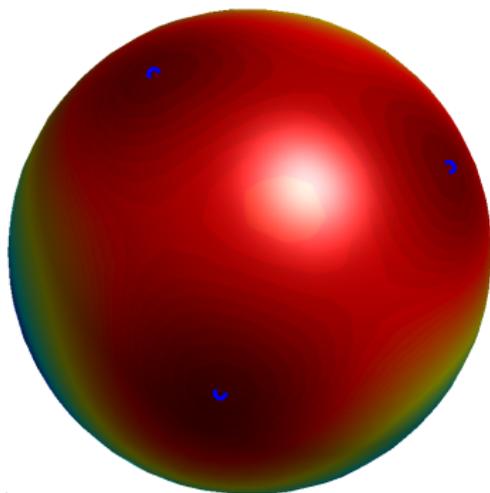


- This pre-certificate satisfies the tensor eigenvalue-eigenvector relationships such as

$$\sum_{j,k} Q_{:,j,k} \mathbf{v}_i^\star(j) \mathbf{w}_i^\star(k) = \mathbf{u}_i^\star, i \in [r]$$

# Atomic Decomposition of Tensors III

## Theorem (Li, Prater, Shen, Tang, 2015)

*Suppose*

- *Incoherence:* $\max_{p \neq q}\{|\langle \mathbf{u}_p^\star, \mathbf{u}_q^\star \rangle|, |\langle \mathbf{v}_p^\star, \mathbf{v}_q^\star \rangle|, |\langle \mathbf{w}_p^\star, \mathbf{w}_q^\star \rangle|\} \leq \frac{\mathrm{polylog}(n)}{\sqrt{n}}$

- *Bounded spectra:* $\max\{\|U^\star\|, \|V^\star\|, \|W^\star\|\} \leq 1 + c\sqrt{\frac{r}{n}}$

- *Gram isometry:* $\|(U^{\star\prime}U^\star) \odot (V^{\star\prime}V^\star) - I_r\| \leq \mathrm{polylog}(n)\frac{\sqrt{r}}{n}$ *and similar bounds for* $U^\star, W^\star,$ *and* $V^\star, W^\star$

- *Low-rank (but still overcomplete):* $r = O(n^{17/16}/\mathrm{polylog}(n))$

*Then* $\mu^\star$ *is the optimal solution of the total mass minimization problem as certified by the minimal energy dual certificate.*

## Corollary (Li, Prater, Shen, Tang, 2015)

*Suppose that the factors* $\{\mathbf{u}_p^\star\}, \{\mathbf{v}_p^\star\}$ *and* $\{\mathbf{w}_p^\star\}$ *follow uniform distributions on the unit sphere, then the first three assumptions are satisfied with high probability.*

# Atomic Decomposition of Tensors IV

**SOS Relaxations.**

- Symmetric tensor atoms:

$$\|Z\|_{\mathcal{A}}^* \leq 1 \Leftrightarrow \sum_{i,j,k} Z_{ijk} u_i u_j u_k \leq 1 \ \forall \|\mathbf{u}\|_2 = 1$$

- The latter states that a third order multivariate polynomial is bounded by 1, or $1 - \sum_{i,j,k} Z_{ijk} u_i u_j u_k$ is nonnegative on the unit sphere.

- The general framework of Sum-of-Squares (SOS) for non-negative polynomials over semi-algebraic sets leads to a hierarchy of increasingly tight semidefinite relaxations for the symmetric tensor spectral norm.

- Taking the dual yields a hierarchy of increasingly tight semidefinite approximations of the (symmetric) tensor nuclear norm.

# Atomic Decomposition of Tensors V

## Theorem (Tang & Shah, 2015)

*For a symmetric tensor $T = \sum_{k=1}^{r} \lambda_k \mathbf{x}_k \otimes \mathbf{x}_k \otimes \mathbf{x}_k$, if the tensor factors $X = [\mathbf{x}_1, \cdots, \mathbf{x}_r]$ satisfy $\|X'X - I_r\| \leq 0.0016$, then the (symmetric) tensor nuclear norm $\|T\|_*$ equals both $\sum_{k=1}^{r} \lambda_k$ and the optimal value of the smallest SOS approximation.*

# Atomic Decomposition of Tensors VI

**Low-rank Factorization.**

- Matrix atoms: $\{\mathbf{u} \otimes \mathbf{v} : \|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1\}$
- Tensor atoms: $\{\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} : \|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = \|\mathbf{w}\|_2 = 1\}$
- For a matrix $X$ with rank $r$, when $\tilde{r} \geq r$, the matrix nuclear norm equals the optimal value of

$$\underset{\{(\mathbf{u}_p, \mathbf{v}_p)\}_{p=1}^{\tilde{r}}}{\text{minimize}} \frac{1}{2} \left( \sum_{p=1}^{\tilde{r}} [\|\mathbf{u}_p\|_2^2 + \|\mathbf{v}_p\|_2^2] \right) \text{ subject to } X = \sum_{p=1}^{\tilde{r}} \mathbf{u}_p \mathbf{v}_p^T$$

- For a tensor $T$ with rank $r$, when $\tilde{r} \geq r$, the tensor nuclear norm equals the optimal value of

$$\underset{\{(\mathbf{u}_p, \mathbf{v}_p, \mathbf{w}_p)\}_{p=1}^{\tilde{r}}}{\text{minimize}} \frac{1}{3} \left( \sum_{p=1}^{\tilde{r}} [\|\mathbf{u}_p\|_2^3 + \|\mathbf{v}_p\|_2^3 + \|\mathbf{w}_p\|_2^3] \right)$$

$$\text{subject to } T = \sum_{p=1}^{\tilde{r}} \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p$$

# Atomic Decomposition of Tensors VII

- ▶ Incorporate these nonlinear reformulations into atomic norm regularized problems.

## Theorem (Haeffele & Vidal, 2015)

- ▶ When $\tilde{r} > r$, any local minimizer such that one component is zero, e.g., $\mathbf{u}_{i_0} = \mathbf{v}_{i_0} = \mathbf{w}_{i_0} = 0$.
- ▶ There exists a non-increasing path an initial point $(\mathbf{u}^{(0)}, \mathbf{v}^{(0)}, \mathbf{w}^{(0)})$ to a global minimizer of the nonlinear formulation.
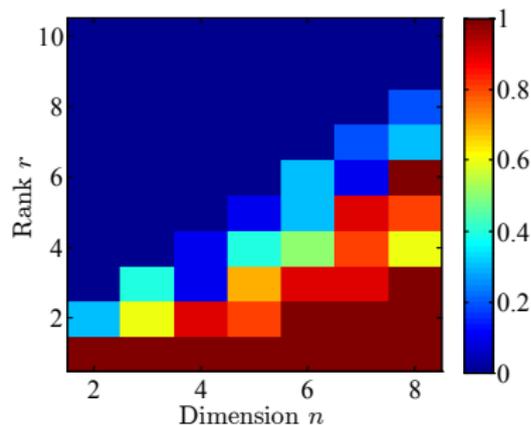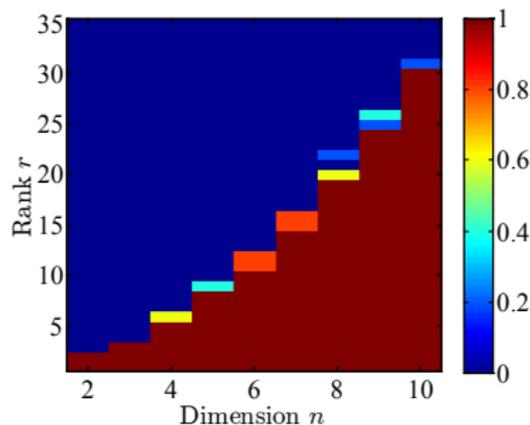
# Atomic Norm for Ensemble of Spectral Signals I

**Signal model.**

- In applications such as array signal processing, we receive multiple snapshots of observations impinging on the array.

- Recall the atoms for line spectrum is defined as

$$\mathbf{a}(\nu) = \left[1, e^{j2\pi\nu}, \ldots, e^{j2\pi(n-1)\nu}\right]^T, \quad \nu \in [0, 1).$$

- we consider $L$ signals, stacked in a matrix, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_L]$, where each $\mathbf{x}_l \in \mathbb{C}^n$ is composed of the same set of atoms

$$\mathbf{x}_l = \sum_{i=1}^r c_{i,l}\mathbf{a}(\nu_i), \quad l = 1, \ldots, L.$$

- Continuous-analog of group sparsity.

# Atomic Norm for Ensemble of Spectral Signals II

- We define the atomic set as

$$\mathcal{A} = \left\{ \mathbf{A}(\nu, \mathbf{b}) = \mathbf{a}(\nu)\mathbf{b}^H, \quad \|\mathbf{b}\|_2 = 1. \right\}$$

- The atomic norm $\|\mathbf{X}\|_{\mathcal{A}}$ is defined as

$$\|\mathbf{X}\|_{\mathcal{A}} = \inf \left\{ t > 0 : \ \mathbf{X} \in t \, \mathsf{conv}(\mathcal{A}) \right\}$$

- The atomic norm $\|\mathbf{X}\|_{\mathcal{A}}$ can be written equivalently as

$$\|\boldsymbol{X}\|_{\mathcal{A}} = \inf_{\boldsymbol{u} \in \mathbb{C}^n, \mathbf{W} \in \mathbb{C}^{L \times L}} \left\{ \frac{1}{2} u_0 + \frac{1}{2} \operatorname{Tr}(\mathbf{W}) \Big| \begin{bmatrix} \operatorname{toep}(\boldsymbol{u}) & \mathbf{X} \\ \mathbf{X}^H & \mathbf{W} \end{bmatrix} \succeq \mathbf{0} \right\}.$$

- The dual norm of $\|\mathbf{X}\|_{\mathcal{A}}$ can be defined as

$$\|\mathbf{Y}\|_{\mathcal{A}}^* = \sup_{f \in [0,1)} \|\mathbf{Y}^* \mathbf{a}(f)\|_2 \triangleq \sup_{f \in [0,1)} \|\mathbf{Q}(f)\|_2,$$

where $\mathbf{Q}(f) = \mathbf{Y}^H \mathbf{a}(f)$ is a length-$L$ vector with each entry a polynomial in $f$.

# Atomic Norm for Ensemble of Spectral Signals III

- Recovery of missing data:

$$\min \|\mathbf{X}\|_{\mathcal{A}} \quad \text{subject to} \quad \mathbf{Y}_{\Omega} = \mathbf{X}_{\Omega}.$$

- For noncoherently generated snapshots, increasing the number of measurement vectors will increase the localization resolution.
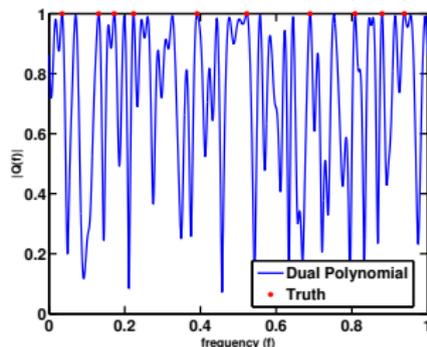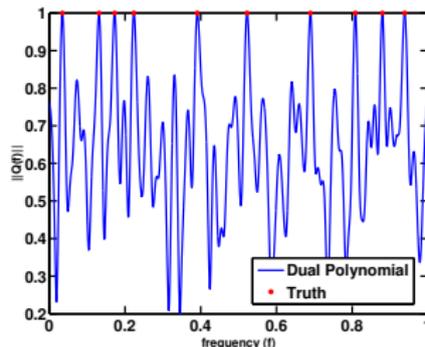


(a) $L = 1$        (b) $L = 3$

Figure : The reconstructed dual polynomial for randomly generated spectral signals with $r = 10$, $n = 64$, and $m = 32$: (a) $L = 1$, (b) $L = 3$.

## Atomic Norm for Ensemble of Spectral Signals IV

- Denoising: consider noisy data $\mathbf{Z} = \mathbf{X} + \mathbf{N}$, where each entry of $\mathbf{N}$ is $\mathcal{CN}(0, \sigma^2)$.

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{X} - \mathbf{Z}\|_F^2 + \tau \|\mathbf{X}\|_{\mathcal{A}}.$$

### Theorem (Li and Chi, 2014)

*Set* $\tau = \sigma \left(1 + \frac{1}{\log n}\right)^{\frac{1}{2}} \left(L + \log(\alpha L) + \sqrt{2L \log(\alpha L)} + \sqrt{\frac{\pi L}{2}} + 1\right)^{\frac{1}{2}}$, *where* $\alpha = 8\pi n \log n$, *then the expected error rate is bounded as*

$$\frac{1}{L} \mathbb{E} \left\| \hat{\boldsymbol{X}} - \boldsymbol{X}^\star \right\|_F^2 \leq \frac{2\tau}{L} \|\boldsymbol{X}^\star\|_{\mathcal{A}}.$$

- As $\tau$ is set on the order of $\sqrt{L}$, if $\|\boldsymbol{X}^\star\|_{\mathcal{A}} = o\left(\sqrt{L}\right)$, then the per-measurement vector MSE vanishes as $L$ increases.

# Super-resolution of Mixture Models I

**Mixture Model for Multi-modal Data**

- Formally, consider **inverting** the following mixture model:

$$y(t) = \sum_{i=1}^{I} x_i(t) * g_i(t) + w(t),$$

where $*$ is the convolution operator,

  - $I$ is the total number of mixtures, assumed known;
  - $x_i(t)$ is a parametrized point source signal with $K_i$ unknown:

$$x_i(t) = \sum_{j=1}^{K_i} a_{ij}\delta(t - t_{ij}), \quad t_{ij} \in [0, 1], \ a_{ij} \in \mathbb{C};$$

  - $g_i(t)$ is a point spread function with a finite cut-off frequency $2M$;
  - $w(t)$ is the additive noise;

- The goal is to invert the locations and amplitudes of the point sources for each mixture, $\{a_{ij}, \tau_{ij}\}_{j=1}^{K_i}$, $1 \leq i \leq I$.

## Super-resolution of Mixture Models II

- ▶ Set $I = 2$ for simplicity. Analysis generalizes to cases $I \geq 2$.
- ▶ In the frequency domain, we have the vector-formed signal

$$\mathbf{y} = \boldsymbol{g}_1 \odot \mathbf{x}_1^\star + \boldsymbol{g}_2 \odot \mathbf{x}_2^\star + \boldsymbol{w},$$

where $\odot$ denotes point-wise product, $\boldsymbol{g}_i$ is the DTFT of the PSF $g_i(t)$, and $\mathbf{x}_i$'s are spectrally-sparse signals:

$$\mathbf{x}_1^\star = \sum_{k=1}^{K_1} a_{1k} \mathbf{c}\left(\tau_{1k}\right), \quad \mathbf{x}_2^\star = \sum_{k=1}^{K_2} a_{2k} \mathbf{c}\left(\tau_{2k}\right),$$

where $\mathbf{c}(\tau) = \left[e^{-j2\pi\tau(-2M)}, \ldots, 1, \ldots, e^{-j2\pi\tau(2M)}\right]^T$.

- ▶ Conventional methods such as MUSIC and ESPRIT do not apply due to interference between different components.

# Super-resolution of Mixture Models III

- **Convex Demixing:** motivate the spectral sparsity of both components via minimizing the atomic norm:

$$\{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2\} = \operatorname*{argmin}_{\mathbf{x}_1, \mathbf{x}_2} \|\mathbf{x}_1\|_{\mathcal{A}} + \|\mathbf{x}_2\|_{\mathcal{A}}, \quad \text{s.t.} \quad \mathbf{y} = \boldsymbol{g}_1 \odot \mathbf{x}_1 + \boldsymbol{g}_2 \odot \mathbf{x}_2.$$

- **Incoherence condition:** Each entry of the sequences $\boldsymbol{g}_1, \boldsymbol{g}_2$ is generated i.i.d. from a uniform distribution on the complex unit circle.



  - The PSF functions should be incoherent across components

## Theorem (Li and Chi, 2015)

*Under the incoherence condition, assume the signals are generated with random signs from the unit circle satisfying the separation of $4/n$, then the recovery of convex demixing is unique with high probability if*

$$M/\log M \gtrsim (K_1 + K_2)\log(K_1 + K_2).$$

# Super-resolution of Mixture Models IV

**Phase Transition:** Set the separation condition $\Delta = 2/n$.



(a) $M = 8$    (b) $M = 16$
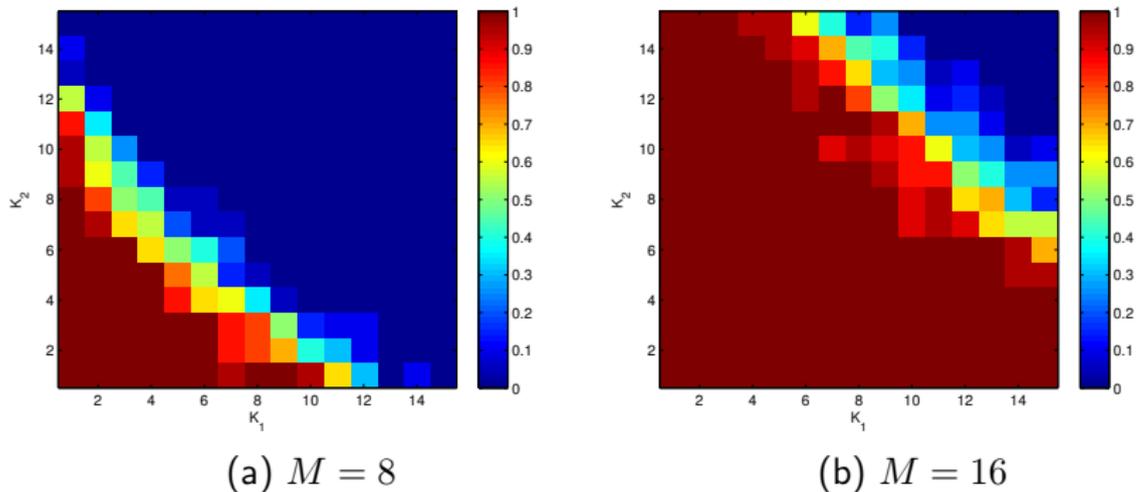
Figure : Successful rates of the convex demixing algorithm as a function of $(K_1, K_2)$ when (a) $M = 8$ and (b) $M = 16$.

# Super-resolution of Mixture Models V

**Comparison with CRB for Parameter Estimation:**

- We also compared with the Cramer-Rao Bound to benchmark the performance of parameter estimation in the noisy case when $K_1 = 1$, and $K_2 = 1$ for estimating source locations.
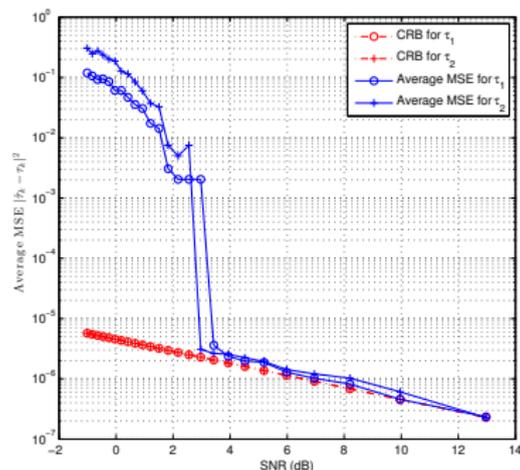


(a) $M = 10$
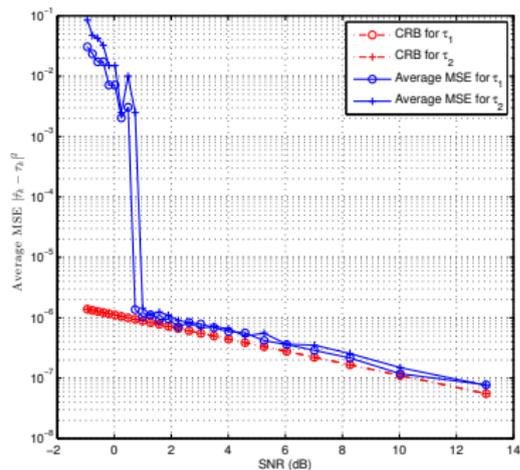


(b) $M = 16$

# Blind Super-resolution I

**Super-resolution with unknown point spread functions:**

- Model the observed signal as:

$$y(t) = \sum_{i=1}^{r} a_i g(t - \tau_i) = x(t) * g(t),$$

where $*$ is the convolution operator,

- $x(t)$ is a point source signal with *complex* amplitudes, where $K$ is unknown:

$$x(t) = \sum_{i=1}^{r} a_i \delta(t - \tau_i), \quad \tau_i \in [0, 1], \; a_i \in \mathbb{C};$$

- $g(t)$ is the *unknown* point spread function of the sensory system;

- In frequency domain, we have

$$\mathbf{y} = \boldsymbol{g} \odot \mathbf{x},$$

where $\mathbf{x} = \sum_{i=1}^{r} a_i \mathbf{c}(\tau_i)$.

# Blind Super-resolution II

- Extremely ill-posed without further constraints.
- **Subspace assumption:** We assume the PSF $g$ lies in some *known* low-dimensional subspace:

$$g = \mathbf{B}\mathbf{h} \in \mathbb{C}^{4M+1},$$

where $\mathbf{B} = [\mathbf{b}_{-2M}, \cdots, \mathbf{b}_{2M}]^T \in \mathbb{C}^{(4M+1)\times L}$, and $\mathbf{h} \in \mathbb{C}^L$.

  - **Self-calibration of unitary linear arrays:** the antenna gains $g$ may be well-approximated as lying in a low-dimensional (smooth) subspace.
  - **Blind channel estimation:** the transmitted data signal $g$ is coded by projection in a low-dimensional subspace (e.g. the generating matrix).

# Blind Super-resolution III

- Applying the **lifting trick:** and write the $i$-th entry of $\mathbf{y}$ as $y_i = x_i g_i$ as

$$y_i = x_i \cdot g_i = (\mathbf{e}_i^T \mathbf{x})(\mathbf{b}_i^T \mathbf{h}) = \mathbf{e}_i^T (\mathbf{x}\mathbf{h}^T)\mathbf{b}_i := \mathbf{e}_i^T \mathbf{Z}^\star \mathbf{b}_i,$$

  where $\boldsymbol{e}_i$ is the $i$th column of $\mathbf{I}_{4M+1}$, and $\boldsymbol{b}_i$ as the $i$th row of $\mathbf{B}$.

- Now $\mathbf{y}$ becomes *linear* measurements of $\mathbf{Z}^\star = \mathbf{x}\mathbf{h}^T \in \mathbb{C}^{(4M+1)\times L}$:

$$\mathbf{y} = \mathcal{X}(\mathbf{Z}^\star),$$

  with $(4M+1)$ equations and $(4M+1)L$ unknowns.

- $\mathbf{Z}^\star$ can be regarded as an ensemble of spectrally-sparse signals:

$$\mathbf{Z}^\star = \mathbf{x}\mathbf{h}^T = \left[\sum_{i=1}^{r} a_i \mathbf{c}(\tau_i)\right] \mathbf{h}^T.$$

# Blind Super-resolution IV

- Blind super-resolution via AtomicLift:

$$\min \ \|\boldsymbol{Z}\|_{\mathcal{A}} \quad \text{s.t.} \quad \mathbf{y} = \mathcal{X}(\mathbf{Z}).$$

- **Incoherence condition:** Each row of the subspace $\mathbf{B}$ is i.i.d. sampled from a population $F$, i.e. $\mathbf{b}_n \sim F$, that satisfies the following:
    - *Isometry property:* $\mathbb{E}\mathbf{b}\mathbf{b}^H = \mathbf{I}_L, \quad \mathbf{b} \sim F$.
    - *Incoherence property:* for $\mathbf{b} = [b_1, \ldots, b_L]^T \sim F$, define the coherence parameter $\mu$ of $F$ as the smallest number such that

$$\max_{1 \leq i \leq L} |b_i|^2 \leq \mu.$$

## Theorem (Chi, 2015)

*Assume $\mu = \Theta(1)$. For deterministic point source signals satisfying the separation condition of $1/M$, $M/\log M = O(r^2 L^2)$ is sufficient for successful recovery of $\mathbf{Z}$ with high probability.*

# Blind Super-resolution V
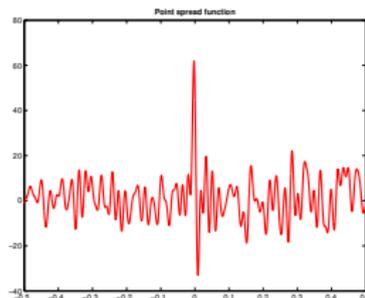


(a) PSF
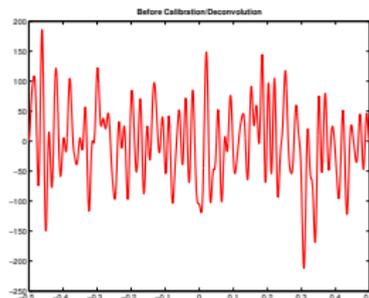
(b) Convolution with the PSF
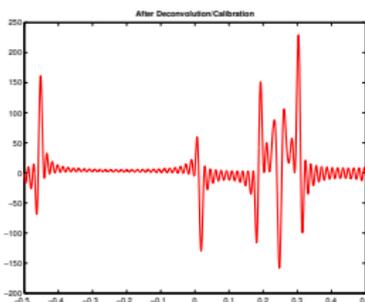
(c) Deconvolution
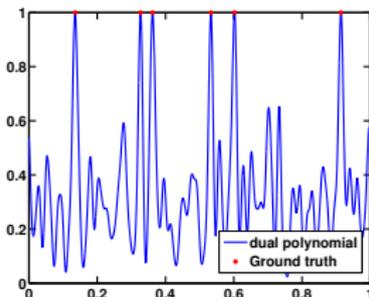
(d) Localization

Figure : Blind spikes deconvolution using AtomicLift: (a) PSF; (b) convolution between the PSF in (a) and a sparse spike signal; (c) deconvolution with the PSF using (b); (d) exact localization of the spikes via the dual polynomial.

# Blind Super-resolution VI

- Alternatively, consider different modulation for each point source:

$$y(t) = \sum_{i=1}^{r} \alpha_i g_i(t - \tau_i),$$

  motivated by asynchronous multi-user communications.

- The frequency domain model becomes

$$\mathbf{y} = \sum_{i=1}^{r} \alpha_i \mathbf{a}(\nu_i) \odot \mathbf{g}_i$$

- Assume all $\mathbf{g}_i$ lie in the same subspace $\mathbf{B}$ and apply the same lifting procedure, we obtain linear measurements of $\mathbf{Z} = \sum_{i=1}^{r} \alpha_i \mathbf{h}_i \mathbf{a}(\nu_i)^H$.

## Theorem (Yang, Tang, Wakin, 2015)

*For point sources with random signs satisfying the separation condition of $1/M$, $M = O(rL)$ is sufficient for successful recovery of $\mathbf{Z}$ with high probability.*

# Blind Super-resolution VII



Number of samples $N = 64$

$J$: Number of spikes

$K$: Dimension of subspace

Dimension of subspace $K = 4$

$N$: Number of samples

$J$: Number of spikes

Number of spikes $J = 4$

$N$: Number of samples

$K$: Dimension of subspace
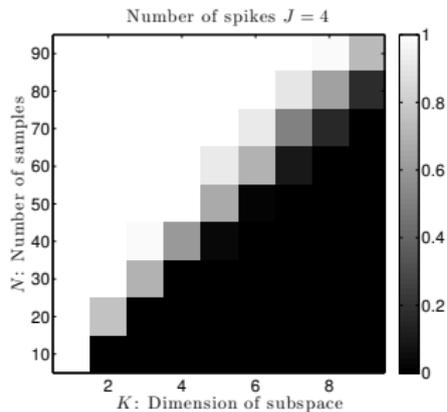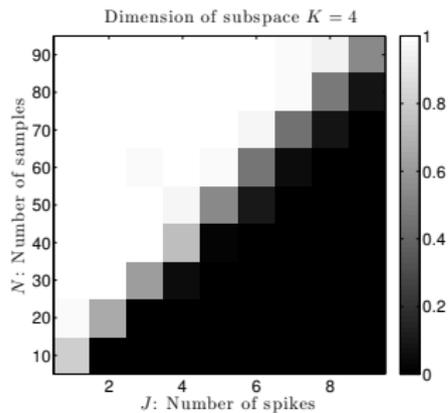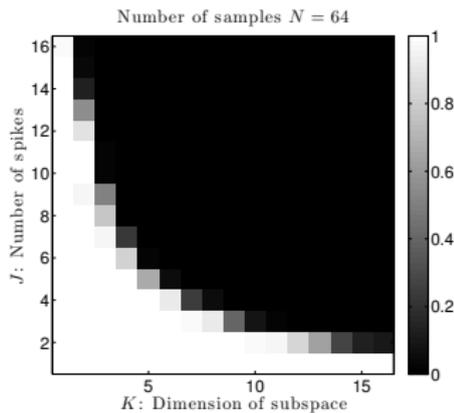
# Application to Single-molecule imaging I
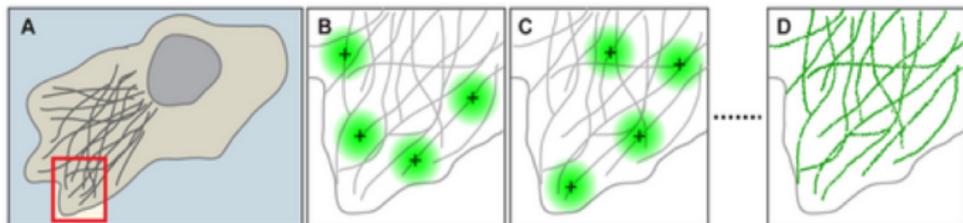


**Synthetic data: discretization-based reconstruction (CSSTORM)**

- Bundles of 8 tubes of 30 nm diameter
- Sparse density: 81049 molecules on 12000 frames
- Resolution: 64x64 pixels
- Pixel size: 100nmx100nm
- Field of view: 6400nmx6400nm
- Target resolution: 10nmx10nm
- Discretize the FOV into 640x640 pixels
- $I(x, y) = \sum_j c_j \text{PSF}(x - x_j, y - y_j),$
  $(x_j, y_j) \in [0, 6400]^2, (x, y) \in \{50, 150, \ldots, 6350\}^2$

# Application to Single-molecule imaging II
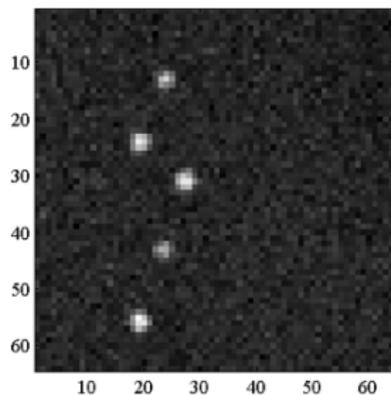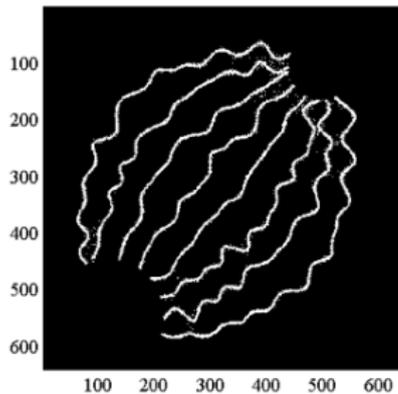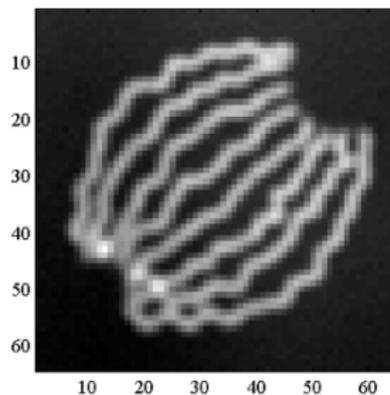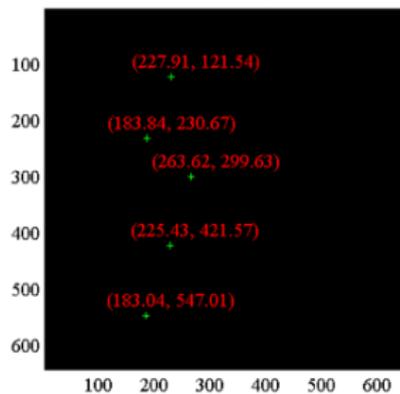
# Application to Single-molecule imaging III

**TVSTORM [Huang, Sun, Ma and Chi, 2016]:** atomic norm regularized Poisson MLE:

$$\hat{\chi} = \underset{\chi \in \mathcal{G}}{\text{argmin}}\ \ell(y|\chi) + \epsilon \|\chi\|_{\mathcal{A}}$$

Our algorithm avoids the realization of the dense dictionary introduced by discretization in CSSTORM.
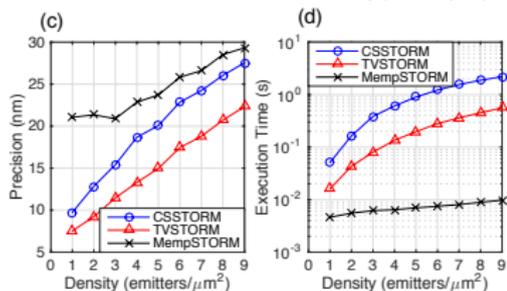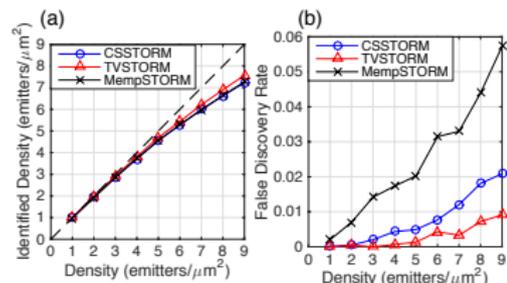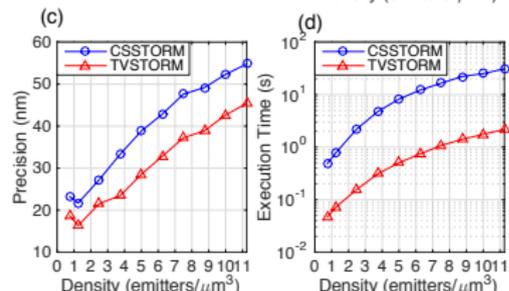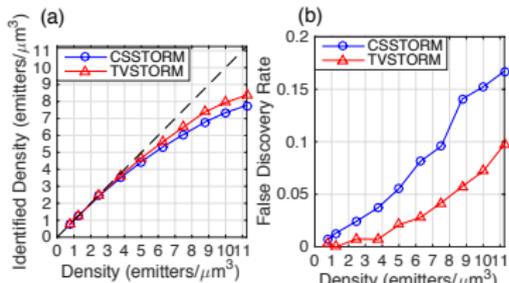


2D super-resolution          3D super-resolution

# Application to Single-molecule imaging IV

**Practical Super-resolution reconstruction on real data:**



(a)

(b)

(a)

(b)
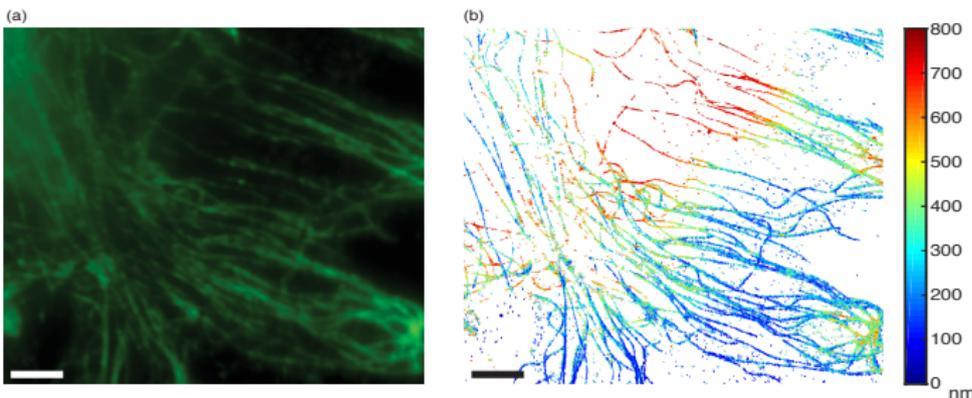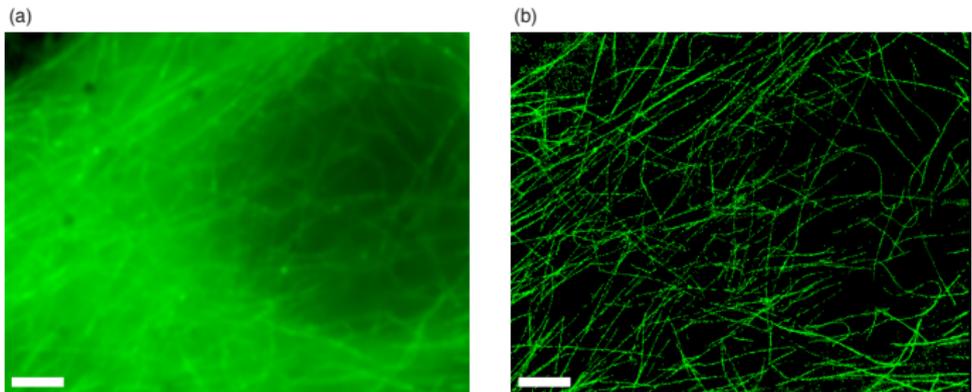
# Allowing Damping for Spectral Compressed Sensing

**Two-Dimensional Frequency Model**

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

- The matrix $\boldsymbol{X}$ has the following **Vandermonde decomposition**:

$$\boldsymbol{X} = \boldsymbol{Y} \cdot \boldsymbol{D} \cdot \boldsymbol{Z}^T.$$

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

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

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.

- Allow damping modes when $\boldsymbol{f}_i \in \mathbb{C}^2$.

# Revisiting Matrix Pencil: Matrix Enhancement

Given a data matrix $\mathbf{X}$, Hua proposed the following matrix enhancement for two-dimensional frequency models (MEMP):
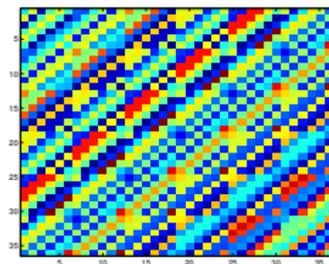


- Choose two pencil parameters $k_1$ and $k_2$;

- An **enhanced form** $\mathbf{X}_e$ is an $k_1 \times (n_1 - k_1 + 1)$ *block Hankel matrix* :

$$\mathbf{X}_e = \left[ \begin{array}{cccc} \mathbf{X}_0 & \mathbf{X}_1 & \cdots & \mathbf{X}_{n_1-k_1} \\ \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_{n_1-k_1+1} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_{k_1-1} & \mathbf{X}_{k_1} & \cdots & \mathbf{X}_{n_1-1} \end{array} \right],$$

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

$$\mathbf{X}_l = \left[ \begin{array}{cccc} x_{l,0} & x_{l,1} & \cdots & x_{l,n_2-k_2} \\ x_{l,1} & x_{l,2} & \cdots & x_{l,n_2-k_2+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{l,k_2-1} & x_{l,k_2} & \cdots & x_{l,n_2-1} \end{array} \right].$$

# 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:

$$\mathbf{X}_e = \begin{bmatrix} \mathbf{Z}_\mathsf{L} \\ \mathbf{Z}_\mathsf{L} \mathbf{Y}_\mathsf{d} \\ \vdots \\ \mathbf{Z}_\mathsf{L} \mathbf{Y}_\mathsf{d}^{k_1 - 1} \end{bmatrix} \mathbf{D} \begin{bmatrix} \mathbf{Z}_\mathsf{R}, \mathbf{Y}_\mathsf{d} \mathbf{Z}_\mathsf{R}, \cdots, \mathbf{Y}_\mathsf{d}^{n_1 - k_1} \mathbf{Z}_\mathsf{R} \end{bmatrix},$$

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

- The enhanced form $\mathbf{X}_e$ is low-rank.
  - $\mathsf{rank}\,(\mathbf{X}_e) \leq r$
  - Spectral Sparsity $\Rightarrow$ Low Rankness

- holds even with damping modes.

Hua, Yingbo. "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil." Signal Processing, IEEE Transactions on 40, no. 9 (1992): 2267-2280.

# Enhanced Matrix Completion (EMaC) I

- Motivated by Matrix Completion, we seek the low-rank solution via nuclear norm minimization:

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

- Define $\mathbf{G}_L$ and $\mathbf{G}_R$ as $r \times r$ Gram matrices such that

$$(\mathbf{G}_L)_{i,l} = \mathcal{K}(k_1, k_2, f_{1i} - f_{1l}, f_{2i} - f_{2l}),$$
$$(\mathbf{G}_R)_{i,l} = \mathcal{K}(n_1 - k_1 + 1, n_2 - k_2 + 1, f_{1i} - f_{1l}, f_{2i} - f_{2l}).$$

  where $\mathcal{K}(k_1, k_2, f_1, f_2)$ is the 2-D Dirichlet kernel.

- **Incoherence condition** holds w.r.t. $\mu$ if

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

  only depends on the locations of the frequency, not their amplitudes.

# Enhanced Matrix Completion (EMaC) II

▶ Performance Guarantee in the noise-free case:

## Theorem (Chen and Chi, 2013)

*Let $n = n_1 n_2$. If all measurements are noiseless, then EMaC recovers $\mathbf{X}$ perfectly with high probability if*

$$m > C\mu r \log^3 n.$$

*where $C$ is some universal constant.*

▶ $\mu = \Theta(1)$ holds (w.h.p.) under many scenarios:
  ▶ Randomly generated frequencies;
  ▶ Mild perturbation of grid points;
  ▶ In 1D, well-separated frequencies by 2RL [Liao and Fannjiang, 2014].
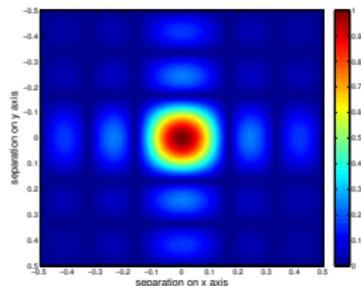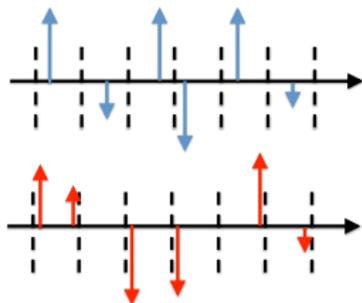
# Enhanced Matrix Completion (EMaC) III

**Robustness to Bounded Noise.**

- Assume the samples are noisy $\mathbf{X} = \mathbf{X}^o + \mathbf{N}$, where $\mathbf{N}$ is bounded noise:

(EMaC-Noisy) : $\displaystyle \operatorname*{minimize}_{\mathbf{M} \in \mathbb{C}^{n_1 \times n_2}} \quad \|\mathbf{M}_e\|_* \quad$ subject to $\|\mathcal{P}_\Omega (\mathbf{M} - \mathbf{X})\|_F \leq \delta$,

---

### Theorem (Chen and Chi, 2013)

*Suppose $\mathbf{X}^o$ satisfies $\|\mathcal{P}_\Omega(\mathbf{X} - \mathbf{X}^o)\|_F \leq \delta$. Under the conditions of Theorem 1, the solution to* **EMaC**-**Noisy** *satisfies*

$$\|\hat{\mathbf{X}}_e - \mathbf{X}_e\|_F \leq \left\{ 2\sqrt{n} + 8n + \frac{8\sqrt{2}n^2}{m} \right\} \delta$$

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

---

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

# Enhanced Matrix Completion (EMaC) IV

**Robustness to Sparse Outliers**

- Assume a constant portion of the measurements are arbitrarily corrupted as $\mathbf{X}_{i,l}^{\text{corrupted}} = \mathbf{X}_{i,l} + \mathbf{S}_{i,l}$, where $\mathbf{S}_{i,l}$ is of arbitrary amplitude.
- Reminiscent of the robust PCA approach [Candes et. al. 2011, Chandrasekaran et. al. 2011], solve the following algorithm:

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

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

## Theorem (Chen and Chi, 2013)

*Assume the percent of corrupted entries is $s$ is a small constant. Set $n = n_1 n_2$ and $\lambda = \frac{1}{\sqrt{m\log n}}$. Then RobustEMaC recovers $\mathbf{X}$ with high probability if*

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

*where $C$ is some universal constant.*

- Sample complexity: $m \sim \Theta(r^2 \log^3 n)$, slight loss than the previous case;
- Robust to a constant portion of outliers: $s \sim \Theta(1)$

# Comparisons between EMaC and ANM

|  | EMaC | Atomic Norm |
|---|---|---|
| **Signal model** | Deterministic | Random |
| **Observation model** | Random | Random |
| **Success Condition** | Coherence | Separation condition |
| **Sample Complexity** | $\Theta(r \log^3 n)$ | $\Theta(r \log r \log n)$ |
| **Bounded Noise** | Yes | Yes |
| **Sparse Corruptions** | Yes | Yes |
| **Damping Modes** | Yes | No |

# Comparisons of EMaC and ANM

Phase transition for line spectrum estimation: numerically, the EMaC approach seems less sensitive to the separation condition.

EMaC     Atomic Norm

• without separation
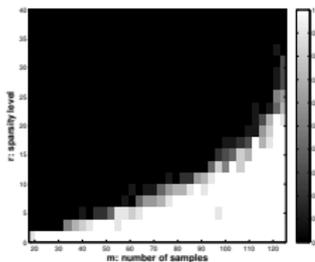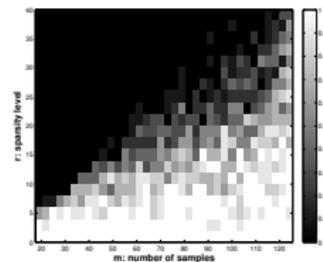


• with 1.5 RL separation
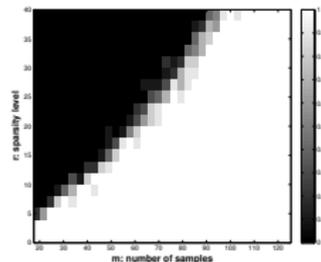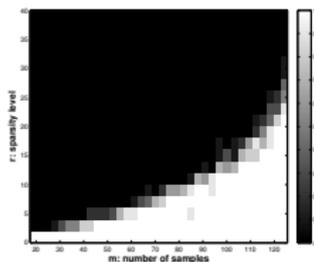
# References to Atomic Norm and Super Resolution I

- ▶ Chandrasekaran, Recht, Parrilo, Willsky (2010): general framework of atomic norm minimization.
- ▶ Tang, Bhaskar, Shah, Recht (2012): line spectrum estimation using atomic norm minimization with random sampling.
- ▶ Bhaskar, Tang, Recht (2012): line spectrum denoising using atomic norm minimization with consecutive samples.
- ▶ Candès and Fernandez-Granda (2012): Super-resolution using total variation minimization (equivalent to atomic norm) from low-pass samples.
- ▶ Chi (2013): line spectrum estimation using atomic norm minimization with multiple measurement vectors.
- ▶ Xu et. al. (2014): atomic norm minimization with prior information.
- ▶ Chen and Chi (2013): multi-dimensional frequency estimation via enhanced matrix completion.
- ▶ Xu et. al. (2013): exact SDP characterization of atomic norm minimization for high-dimensional frequencies.

# References to Atomic Norm and Super Resolution II

- Tang et. al. (2013): near minimax line spectrum denoising via atomic norm minimization.
- Chi and Chen (2013): higher dimensional spectrum estimation using atomic norm minimization with random sampling.
- Hua (1992): matrix pencil formulation for multi-dimensional frequencies.
- Liao and Fannjiang (2014): analysis of the MUSIC algorithm with separation conditions.
- Li and Chi (2014): atomic norm for multiple spectral signals.
- Chi (2015): Guaranteed blind super-resolution with AtomicLift.
- Yang, Tang and Wakin (2016): blind super-resolution with different modulations via AtomicLift.

# Concluding Remarks

- Compression, whether by linear maps (e.g, Gaussian) or by subsampling, has performance consequences for parameter estimation. Fisher information decreases, CRB increases, and the onset of breakdown threshold increases.

- Model mismatch can result in considerable performance degradation, and therefore sensitivities of CS to model mismatch need to be fully understood.

- Recent off-the-grid methods (atomic norm and structured matrix completion) provide a way forward for a class of problems, where modes to be estimated respect certain separation or coherence conditions. These methods are also useful for other problems where traditional methods cannot be applied.

- But sub-Rayleigh resolution still eludes us!