# Exploiting Geometry for High-Resolution Source Localization

Yuejie Chi

Department of Electrical and Computer Engineering Department of Biomedical Informatics



THE OHIO STATE UNIVERSITY

May 2017

# Motivations – New Imaging/Sensing Modalities

New imaging/sensing modalities allow us to probe the nature in unprecedented manners.



seismic imaging

The resulting large amount of data brings exciting opportunities that call for new signal processing tools.

### Single-Molecule Fluorescence Microscopy



The Nobel Prize in Chemistry 2014 "for the development of super-resolved fluorescence microscopy".



Photo credit: https://www.nobelprize.org/nobel\_prizes/chemistry/laureates/2014/.

# Single-Molecule Fluorescence Microscopy

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



#### High density implies better time resolution.

Figure credit: "The Nobel Prize in Chemistry 2014 - Popular Information".

#### Three-dimensional Single-Molecule Imaging

This imaging principle is extended to reconstruct 3-D objects from 2-D images, by modulating the shape of the PSFs along the depth.



#### Reconstruction is challenging due to the mutually interfering PSF profile.

B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-dimensional superresolution imaging by stochastic optical reconstruction microscopy," Science 2008.

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.

### Neural Signal Processing

The human brain is the most complicated biological structure in the known universe.

— Francis S. Collins



**Spike sorting:** to identify and separate the firing times of each neuron from the observed voltage trace at the electrode.



Simultaneous excitation of multiple neurons makes it challenging.

Figure taken from: C. Ekanadham, D. Tranchina, and E. P. Simoncelli. "A unified framework and method for automatic neural spike identification." Journal of neuroscience methods 222 (2014): 47-55.

#### Multi-user detection and IoT

Internet of things and 5G: By

2020, industry analysts predict 50 billion devices will be connected to mobile networks worldwide.



**Channel estimation** in a multi-user, multi-path environment is the "elephant in the room".



Blind channel estimation is desirable for reducing overheads.

### High-Resolution Source Localization

• **Observations:** Superposition of returns from sources:

$$y(t) = \sum_{k=1}^{K} \alpha_k g(t - \nu_k) + n(t),$$

where g(t) is the **point spread function**,  $\{\alpha_k\}_{k=1}^K$  and  $\{\nu_k\}_{k=1}^K$  are the source parameters, and n(t) is additive noise.



- Inversion: Estimate  $\{\alpha_k, \nu_k\}_{k=1}^K$ , given a set of samples of y(t).
  - DOA estimation in sensor array processing
  - Frequency and amplitude estimation in spectrum analysis
  - Range, Doppler, and azimuth estimation in radar/sonar
  - Source location estimation in MRI, EEG, NMR spectroscopy

### Something Old: Parameter Estimation

**Exploring Physically-meaningful Constraints:** shift-invariance in the frequency domain:

$$\boldsymbol{y} = \boldsymbol{g} \odot \boldsymbol{x} = \boldsymbol{g} \odot \left( \sum_{k=1}^{K} \alpha_k \begin{bmatrix} e^{j2\pi 2M\tau_k} \\ \vdots \\ 1 \\ \vdots \\ e^{-j2\pi 2M\tau_k} \end{bmatrix} \right)$$



where 2M is the cut-off frequency of the  $\mathsf{PSF}.$ 

- Prony's method [1795]: root-finding.
- SVD based approaches: ESPRIT [RoyKailath'1989], MUSIC [Schmidt'1986], matrix pencil [HuaSarkar'1990, Hua'1992].
- Finite rate of innovation [Vetterli' 2001].

**Performance:** Perfect recovery from (equi-spaced) O(K) samples. Performance is well understood via estimation-theoretic bounds under Gaussian noise.

#### Challenges

However, traditional methods no longer apply under more complicated sensing modalities discussed earlier, due to sensitivity to noise, missing data, interference and outliers.

• Subsampling or missing data:

ultra-wideband signals, channel estimation using fewer pilots.

• Noise and corruptions:

sensor failures, attacks, outliers, etc.

- Calibration: the PSF functions may be unknown and needs to be calibrated.
- **Multi-modality or interference:** the received signal exhibits superpositions of *multiple* PSF functions.







# Something New: Compressed Sensing

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





- Discretize the parameters and assume a sparse representation over the discretized basis: τ<sub>k</sub> ∈ T<sub>n</sub> = {0/n,..., n-1/n};
- Run  $\ell_1$  minimization in the discretized parameter space.

**Performance:** recovery from  $O(K \log n)$  samples, and robust against irregular sampling, noise and outliers enabled by convex optimization.

#### Sensitivity to Basis Mismatch

#### ALERT: Nature does not place the source on the grid!



Basis mismatch translates a sparse signal into an incompressible signal.

Y. Chi, L. L. Scharf, A. Pezeshki, and R. Calderbank, "Sensitivity of Basis Mismatch to Compressed Sensing," IEEE Signal Processing Society Young Author Best Paper Award.

### Our Approach

- Traditional approaches enforce physically-meaningful constraints, *but not* as much sparsity;
- Compressed sensing enforces sparsity, *but not* as much physically-meaningful constraints;

**Geometric Representations:** embed both sparsity and physically-meaningful constraints.

**Convex Relaxations:** enable provable and robust source localization in the presence of non-idealities.

**Applications:** single-molecule fluorescence microscopy.



# **Geometric Representations**

### Two-Dimensional Spectral Sparsity Model

 Sampling the two-dimensional complex harmonic on a uniform grid, which gives a data matrix  $X \in \mathbb{C}^{n_1 \times n_2}$ :

$$X_{\ell_1,\ell_2} = \sum_{k=1}^{K} a_k z_{1,k}^{\ell_1} z_{2,k}^{\ell_2}, \quad 0 \le \ell_1 < n_1, \ 0 \le \ell_2 < n_2$$

where  $z_k = (z_{1,k}, z_{2,k})$  corresponds to the kth source location.

- single-molecule imaging: NMR spectroscopy:  $z = e^{-j2\pi\tau}, \tau \in [0,1];$   $z = \rho e^{-j2\pi\tau}, \rho > 0, \tau \in [0,1];$







source locations

time domain

freg. domain

#### Parsimonious Representation via Matrix Enhancement

Given a data matrix X, consider the following matrix enhancement  $\mathcal{H}(X)$  [Cadzow, Hua]. Choose two pencil parameters  $k_1 = \Theta(n_1)$  and  $k_2 = \Theta(n_2)$ .



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

$$\mathcal{H}(m{X}) = \left[egin{array}{cccccc} m{X}_0 & m{X}_1 & \cdots & m{X}_{n_1-k_1} \ m{X}_1 & m{X}_2 & \cdots & m{X}_{n_1-k_1+1} \ dots & dots & dots & dots \ m{X}_{k_1-1} & m{X}_{k_1} & \cdots & m{X}_{n_1-1} \end{array}
ight],$$

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

$$\boldsymbol{X}_{l} = \mathcal{H}(\boldsymbol{X}[l,:]) = \begin{bmatrix} x_{l,0} & x_{l,1} & \cdots & x_{l,n_{2}-k_{2}} \\ x_{l,1} & x_{l,2} & \cdots & x_{l,n_{2}-k_{2}+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{l,k_{2}-1} & x_{l,k_{2}} & \cdots & x_{l,n_{2}-1} \end{bmatrix}$$

#### Structured Low-Rank Matrix

- The dimensionality of  $\mathcal{H}(\boldsymbol{X})$  is proportional to  $n_1n_2 \times n_1n_2$ .
- The enhanced matrix can be decomposed as follows [Hua 1992]:

$$\mathcal{H}(\boldsymbol{X}) = \sum_{k=1}^{K} a_{k} \underbrace{\left( \begin{bmatrix} 1\\z_{1,k}\\ \vdots\\ z_{1,k}^{k-1} \end{bmatrix} \otimes \begin{bmatrix} 1\\z_{2,k}\\ \vdots\\ z_{2,k}^{k-1} \end{bmatrix} \right)}_{\boldsymbol{u}(\boldsymbol{z}_{k})} \underbrace{\left( \begin{bmatrix} 1\\z_{1,k}\\ \vdots\\ z_{1,k}^{n_{1}-k_{1}} \end{bmatrix} \otimes \begin{bmatrix} 1\\z_{2,k}\\ \vdots\\ z_{2,k}^{n_{2}-k_{2}} \end{bmatrix} \right)^{T}}_{\boldsymbol{v}(\boldsymbol{z}_{k})^{T}},$$

- The enhanced form  $\mathcal{H}(X)$  is decomposed into a sum of rank-one parametric atoms.
  - $\operatorname{rank}(\mathcal{H}(\boldsymbol{X})) \leq K$
  - Spectral Sparsity  $\Rightarrow$  Structured Low-Rank



#### Promoting Parsimony via Convex Relaxations

$$x = \sum_{k=1}^{K} c_k a_k, \quad a_k \in \mathcal{A}$$

- The atomic set  $\mathcal{A}$  can be finite, countably infinite, or continuous.
- **Decompose** the signal x into the *fewest* number of *atoms* in an atomic set A: combinatorial!

$$\|x\|_{\mathcal{A},0} = \min\{\mathbf{K} : x = \sum_{k=1}^{K} c_k a_k, \ a_k \in \mathcal{A}\}$$

• Relax by the convex surrogate, yielding the atomic norm:

$$\|x\|_{\mathcal{A}} = \inf \{t > 0 : x \in t \operatorname{conv}(\mathcal{A})\}$$
$$= \inf \left\{ \sum_{i} |c_i| \left| x = \sum_{i} c_i a_i, \ a_i \in \mathcal{A} \right\} \right\}$$

Chandrasekaran, V., B. Recht, P. A. Parrilo, and A. S. Willsky. "The convex geometry of linear inverse problems." Foundations of Computational Mathematics 12, no. 6 (2012): 805-849.

#### Consequence for Localization

**Identify** activated atoms (source localization) via the dual solution q:

 $\max \langle x,q \rangle \quad \text{subject to} \quad \|q\|_{\mathcal{A}}^* \leq 1$ 

• Relaxation is tight (recover the decomposition), when:

strict boundeness:  $|\langle a,q\rangle| < 1$ ,  $q \in \mathcal{A} \setminus \{a_k\}$ interpolation:  $\langle a_k,q \rangle = \text{sign}(c_k)$ ,



Example if  $\mathcal{A}$  is parameterized over [0,1]

#### Choosing atoms as rank-1 matrices: Nuclear Norm

$$\mathcal{H}(\boldsymbol{X}) = \sum_{k=1}^{K} a_{k} \underbrace{\begin{pmatrix} 1\\z_{1,k}\\\vdots\\z_{1,k}^{k_{1}-1}\\z_{2,k}^{k_{2}-1} \end{pmatrix}}_{\boldsymbol{u}(\boldsymbol{z}_{k})} \underbrace{\begin{pmatrix} 1\\z_{2,k}\\\vdots\\z_{1,k}^{k_{2}-1}\\z_{2,k}^{k_{2}-1} \end{pmatrix}}_{\boldsymbol{v}(\boldsymbol{z}_{k})^{T}} \underbrace{\begin{pmatrix} 1\\z_{1,k}\\\vdots\\z_{2,k}^{n_{2}-k_{2}}\\z_{2,k}^{n_{2}-k_{2}} \end{pmatrix}^{T},$$

• Set  $\mathcal{A} = \{ \text{rank-one matrices} \} = \{ uv^T, ||u|| = ||v|| = 1 \}$  leads to the nuclear norm:

$$\|\mathcal{H}(\boldsymbol{X})\|_{*} = \inf\left\{\sum_{i} |c_{i}| \Big| \mathcal{H}(\boldsymbol{X}) = \sum_{i} c_{i} \boldsymbol{u} \boldsymbol{v}^{T}, \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1\right\}$$

which is a semidefinite program that promotes structured low-rank.

• The large atomic set handles not only damping modes, but more general LTI systems identification.

#### Choosing atoms as sinusoids: Atomic Norm

$$\mathcal{H}(\boldsymbol{X}) = \sum_{k=1}^{K} a_{k} \underbrace{\left( \begin{bmatrix} 1\\z_{1,k}\\ \vdots\\ z_{1,k}^{k_{1}-1} \end{bmatrix} \otimes \begin{bmatrix} 1\\z_{2,k}\\ \vdots\\ z_{2,k}^{k_{2}-1} \end{bmatrix} \right)}_{\boldsymbol{u}(\boldsymbol{z}_{k})} \underbrace{\left( \begin{bmatrix} 1\\z_{1,k}\\ \vdots\\ z_{1,k}^{n_{1}-k_{1}} \end{bmatrix} \otimes \begin{bmatrix} 1\\z_{2,k}\\ \vdots\\ z_{2,k}^{n_{2}-k_{2}} \end{bmatrix} \right)^{T}}_{\boldsymbol{v}(\boldsymbol{z}_{k})^{T}},$$

• Set  $A = \{\text{complex sinusoids}\} = \{u(e^{-j2\pi\tau})v(e^{-j2\pi\tau})^T\}$  leads to the atomic norm:

$$\|\boldsymbol{X}\|_{\mathcal{A}} = \inf\left\{\sum_{i} |c_{i}| \Big| \mathcal{H}(\boldsymbol{X}) = \sum_{i} c_{i} \boldsymbol{u} (e^{-j2\pi\tau}) \boldsymbol{v} (e^{-j2\pi\tau})^{T}, \ \tau \in [0,1)\right\}$$

which is a semidefinite program that promotes structured low-rank.

 The smaller atomic set allows a tighter convex relaxation for complex sinusoids.

E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," Communications on Pure and Applied Mathematics, vol. 67, no. 6, pp. 906-956, 2014.

Y. Chi, and Y. Chen, "Compressive Two-DimensionalHarmonic Retrieval via Atomic Norm Minimization," IEEE Trans. on Signal Processing, vol. 63, pp. 1030-1042, 2015.

# **Case Studies**

#### Robustness to Missing Data

**Missing data:** a subset of entries is observed in an index set  $\Omega$ , where  $m = |\Omega| \ll n = n_1 n_2$ .

$$\boldsymbol{Y}_{i,j} = \boldsymbol{X}_{i,j}, \quad (i,j) \in \Omega.$$

#### EMaC (Enhanced Matrix Completion)

 $\min_{\boldsymbol{M} \in \mathbb{C}^{n_1 \times n_2}} \quad \left\| \mathcal{H}(\boldsymbol{M}) \right\|_* \quad \textit{subject to} \quad \boldsymbol{M}_{i,j} = \boldsymbol{Y}_{i,j}, \quad \forall (i,j) \in \Omega$ 



#### Success Conditions



• Gram matrix of the source via Dirichlet kernel:

$$\boldsymbol{G}_{i,j} = \mathcal{D}(\boldsymbol{z}_i - \boldsymbol{z}_j)$$

- Incoherence is defined as smallest  $\mu$  that  $\sigma_{\min} (\boldsymbol{G}) \geq \frac{1}{\mu}.$
- $\mu = \Theta(1)$  for many scenarios.



### Performance Guarantee of EMaC

#### Theorem (Chen and Chi, TIT 2014)

If  $\Omega$  is sampled uniform at random, EMaC recovers  ${\pmb X}$  perfectly with high probability if

 $m \gtrsim \mu K \log^4 n.$ 

- Near-optimal sample complexity as long as  $\mu = \Theta(1)$  up to logarithmic factors.
- RHS: Phase transition when  $n_1 = n_2 = 15$ .
- Reconstruction is robust to additional bounded noise and sparse outliers, by adding additional regularizations.



Y. Chen and Y. Chi, "Robust Spectral Compressed Sensing via Structured Matrix Completion," IEEE Trans. on Information Theory, vol. 60, pp. 6576-6601, 2014.

#### Robustness to Multi-Modalities

Multi-Modalities (two PSFs):

$$y(t) = \sum_{k=1}^{K_1} \alpha_{1,k} g_1(t - \tau_{1,k}) + \sum_{k=1}^{K_2} \alpha_{2,k} g_2(t - \tau_{2,k})$$

or equivalently in the frequency domain:

$$\boldsymbol{y} = \boldsymbol{g}_1 \odot \boldsymbol{x}_1 + \boldsymbol{g}_2 \odot \boldsymbol{x}_2,$$

where  $oldsymbol{x}_1$  and  $oldsymbol{x}_2$  are spectrally-sparse.



#### AtomicDemix

$$\min_{\boldsymbol{x}_1,\boldsymbol{x}_2} \quad \|\boldsymbol{x}_1\|_{\mathcal{A}} + \|\boldsymbol{x}_2\|_{\mathcal{A}} \quad \textit{subject to} \qquad \boldsymbol{y} = \boldsymbol{g}_1 \odot \boldsymbol{x}_1 + \boldsymbol{g}_2 \odot \boldsymbol{x}_2.$$

Y. Li and Y. Chi, "Stable Separation and Super-Resolution of MixtureModels," accepted to Applied and Computational Harmonic Analysis, in press.

### Success Conditions

• Separation condition (source): for each component, define the minimum separation between point sources as

$$\Delta_i = \min_{k \neq j} |\tau_{ik} - \tau_{ij}| \ge \frac{1}{M};$$

- Random signs (source): The signs of the coefficients α<sub>ik</sub>'s are i.i.d. generated from a symmetric distribution on the complex unit circle.
- Incoherence condition (PSF): Each entry of the PSF sequence  $g_i$ 's is generated i.i.d. from a uniform distribution on the complex unit circle.



• The PSF functions should be incoherent across components

#### Performance Guarantees of AtomicDemix

#### Theorem (Li and Chi, ACHA 2017+)

Let  $M \ge 4$ . With probability at least  $1 - \delta$ , AtomicDemix recovers  $\{x_1, x_2\}$  perfectly as long as

$$M \gtrsim (K_1 + K_2) \log\left(\frac{K_1 + K_2}{\delta}\right) \log\left(\frac{M}{\delta}\right).$$

- Near-optimal sample complexity up to logarithmic factors.
- The reconstruction is also stable in the presence of noise.





#### Robustness to Unknown PSF

**Blind deconvolution:** estimating the source when g(t) is *unknown*:

$$y(t) = \sum_{k=1}^{K} \alpha_k g(t - \tau_k).$$



• In the frequency domain: bilinear form

$$oldsymbol{y} = oldsymbol{g} \odot oldsymbol{x} = \mathsf{diag}(oldsymbol{g})oldsymbol{x}$$

No. of unknowns > No. of equations!

#### Bilinear Inverse Problem and Lifting

• **Subspace assumption:** We assume the sequence *g* lies in some *known* low-dimensional subspace:

$$g = Bh \in \mathbb{C}^{4M+1}$$

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

• The lifting trick:

$$egin{aligned} egin{aligned} egi$$

Translates a bilinear problem to an underdetermined linear one!

#### AtomicLift

• Z can be regarded as a **spectrally-sparse ensemble** with the same set of frequencies:

$$\boldsymbol{Z} = \boldsymbol{x}\boldsymbol{h}^T = [h_1\boldsymbol{x}, h_2\boldsymbol{x}, \dots, h_L\boldsymbol{x}].$$

whose structure can be motivated by the *atomic norm* using the multiple measurement vector (MMV) model [Li and Chi, TSP 2016]:

$$\|\boldsymbol{Z}\|_{\mathcal{A}} = \min_{\boldsymbol{s}, \boldsymbol{W}} \left\{ \frac{1}{2} \mathsf{Tr}(\mathsf{Toep}(\boldsymbol{s})) + \frac{1}{2} \mathsf{Tr}(\boldsymbol{W}) \middle| \begin{bmatrix} \mathsf{Toep}(\boldsymbol{s}) & \boldsymbol{Z} \\ \boldsymbol{Z}^{H} & \boldsymbol{W} \end{bmatrix} \succeq \boldsymbol{0} \right\}$$

#### AtomicLift

$$\hat{Z} = \operatorname*{argmin}_{Z} \|Z\|_{\mathcal{A}}$$
 subject to  $y = \mathcal{X}(Z)$ .

Y. Li and Y. Chi, "Off-the-Grid Line Spectrum Estimation and Denoising with Multiple Measurement Vectors," IEEE Trans. on Signal Processing, vol. 64, pp. 1257-1269, 2016.

#### Success Conditions of AtomicLift

• Separation condition (source): define the minimum separation between point sources

$$\Delta = \min_{k \neq j} |\tau_k - \tau_j| \ge \frac{1}{M};$$

- Random signs (source): The signs of the coefficients α<sub>ik</sub>'s are i.i.d. generated from a symmetric distribution on the complex unit circle.
- Incoherence condition (PSF): Each row of the subspace B is i.i.d. from a population F, i.e. b<sub>n</sub> ~ F:
  - Isometry property:

$$\mathbb{E}\boldsymbol{b}\boldsymbol{b}^{H}=\boldsymbol{I}_{L}, \quad \boldsymbol{b}\sim F.$$

Spreadness property: for b = [b<sub>1</sub>,..., b<sub>L</sub>]<sup>T</sup>
 μ is the smallest number that

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



### Performance Guarantee of AtomicLift

#### Theorem (Chi, JSTSP 2016)

Let  $M \ge 4$ . With probability at least  $1 - \delta$ , AtomicLift recovers Z perfectly, as long as

$$M \gtrsim \mu \frac{KL}{\delta} \log^2\left(\frac{M}{\delta}\right).$$

- When the coherence parameter  $\mu=\Theta(1)$  , O(KL) samples suffice.
  - O(K) samples suffice when PSF is known.
- This requirement is larger than O(K + L).
- The reconstruction is robust to noise.



Y. Chi, "Guaranteed Blind Sparse SpikesDeconvolution via Lifting and Convex Optimization," IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 4, pp. 782-794, 2016.

# **Applications**

With my student Jiaqing Huang (PhD 2016), in collaboration with OSU Davis Heart and Lung Research Institute, we have developed several algorithms for super-resolution imaging.









**3DMACS (BOE 2015):** a multi-camera approach that leverages diversity to improve resolution in CSSTORM.



J. Huang, M. Sun, K. Gumpper, Y. Chi, and J.Ma, "3D Multifocus Astigmatism and Compressed Sensing (3DMACS) Based Supper-resolution Reconstruction," Biomedical Optics Express, vol. 6, pp. 902-917, 2015.

**MempSTORM (OL 2015):** a fast and cheap method for 2D super-resolution imaging using truncated SVD-based spectral estimation that is much faster than state-of-the-art CSSTORM.



J. Huang, K. Gumpper, Y. Chi, M. Sun, and J. Ma, "Fast Two-dimensional Super-resolution Image Reconstruction Algorithm for Ultra-high Emitter Density," Optics Letters, vol. 40, pp. 2989-2992, 2015.

**TVSTORM (TCI 2017, in press):** measure optimization for 3-D image reconstruction under Poisson noise that achieves better performance than state-of-the-art CSSTORM.



J. Huang, M. Sun, J.Ma and Y. Chi, "Super-Resolution Image Reconstruction for High-Density 3D Single-Molecule Microscopy," IEEE Transactions on Computational Imaging, in press.

**Geometric Representations:** embed both sparsity and physically-meaningful constraints for source localization.

**Convex Relaxations:** enable provable and robust source localization in the presence of missing data, outliers, multiple modalities, and mis-calibrations.

#### Imaging Applications: fast

algorithms are developed for super-resolution image reconstruction in single-molecule fluorescence microscopy.



#### Acknowledgement

Our group has been generously supported by ongoing and past support from a diverse set of sponsors:



# Thank you for listening!

www.ece.osu.edu/~chi