Optimal Spectral Estimation via Atomic Norm Minimization

Gongguo Tang

EECS, Colorado School of Mines

London Workshop on Sparse Signal Processing

September 15, 2016

Motivations and Basic Ideas

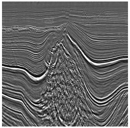
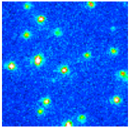
Parameter Estimation in Inverse Problems I

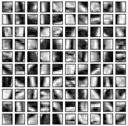
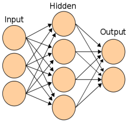
Signal Model: Superposition of parameterized building-block signals

$$x(t) = \sum_{k=1}^{r} c_k a(t; \boldsymbol{\theta}_k), t = t_0, t_1, \ldots, t_{n-1}$$

- Radar/MRI/Microscope/Seismology/Ultrasound imaging
- Parameter estimation in array signal processing
- Matrix and tensor factorization, dictionary learning
- Neural networks



Parameter Estimation in Inverse Problems II

- **Signal Model:** $\mathbf{x} = \sum_{k=1}^{r} c_k \mathbf{a}(\boldsymbol{\theta}_k).$
- DOA and line spectrum estimation:

$$\mathbf{a}(\theta) = [e^{j2\pi t_0 \theta}, e^{j2\pi t_1 \theta}, \ldots, e^{j2\pi t_{n-1}\theta}]^T, \theta \in [0, 1).$$

- Lidar/Single-molecule imaging/neural spike sorting/communication:

$$\mathbf{a}(\tau) = [w(t_0 - \tau), w(t_1 - \tau), \ldots, w(t_{n-1} - \tau)]^T, \tau \in [\tau_{\min}, \tau_{\max}]$$

- Radar and sonar:

$$\mathbf{a}(\boldsymbol{\theta}) = [w(t_0 - \tau)e^{j\omega t_0}, w(t_1 - \tau)e^{j\omega t_1}, \ldots, w(t_{n-1} - \tau)e^{j\omega t_{n-1}}]^T, \boldsymbol{\theta} = (\tau, \omega)$$

- Tensor decomposition:

$$\mathbf{a}(\boldsymbol{\theta}) = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}, \boldsymbol{\theta} = (\mathbf{u}, \mathbf{v}, \mathbf{w}) \in \mathbb{S}^{n-1} \times \mathbb{S}^{m-1} \times \mathbb{S}^{p-1}$$

Line Spectral Estimation I

- Find a combination of sinusoids agreeing with data

$$x(t) = \sum_{k=1}^{r} c_k e^{i2\pi\theta_k t} = \int_0^1 e^{i2\pi\theta t} \mathrm{d}(\sum_{k=1}^{r} c_k \delta(\theta - \theta_k)), \theta_k \in [0,1), c_k \in \mathbb{C}.$$

- Classical signal processing problem with a lot of applications
- New interpretation: super-resolution from low-frequency measurements.

Classical	Contemporary
Prony, MUSIC, Matrix Pencil, ESPIRIT	Sparse recovery
SVD + root finding	gridding + L1 minimization
grid free	flexible, robust model selection quantitative theory
need to know model order lack of quantitative theory not flexible	discretization error basis mismatch numerical instability

Can we bridge the gap?

Y. Chi, et al. "Sensitivity to basis mismatch in compressed sensing."
M. Herman, T. Strohmer. "General deviants: An analysis of perturbations in compressed sensing."

Exploit simplicity/sparsity of the signal, but work directly with the continuously parameterized dictionaries!

Inspirations for Atomic Minimization I

- 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 ℓ_1 norm enforces sparsity w.r.t. the canonical basis vectors.
- The unit ℓ_1 norm ball is $\mathrm{conv}\{\pm\mathbf{e}_k\}$, the (symmetric) convex hull of the basis vectors.

- In **matrix completion**, 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 an infinite number of building-block signals.
- We enforce low-rankness using the nuclear norm:

$$\|X\|_* = \min\{\|\boldsymbol{\sigma}\|_1 : X = \sum_k \sigma_k \mathbf{u}_k \mathbf{v}_k^T\}$$

- The nuclear norm ball is the convex hull of unit-norm, rank-one matrices.

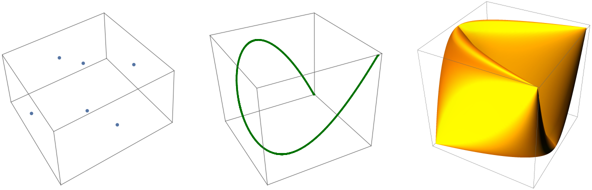
Atomic Norms I

Convex geometry.

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

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

- 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

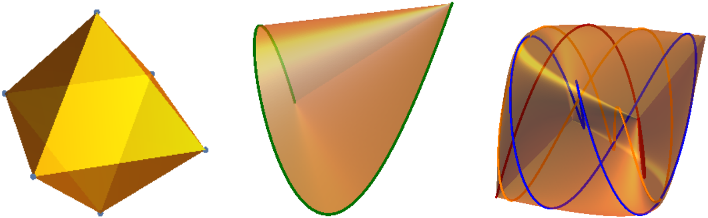
- Prior information: the signal is simple w.r.t. $\mathcal{A}$— it has a parsimonious decomposition using atoms in $\mathcal{A}$

$$\mathbf{x} = \sum_{k=1}^{r} c_k \mathbf{a}(\boldsymbol{\theta}_k)$$

- The atomic norm of any $\mathbf{x}$ is defined as

$$\|\mathbf{x}\|_{\mathcal{A}} = \inf\{\|\boldsymbol{c}\|_1 : \mathbf{x} = \sum_k c_k \mathbf{a}(\boldsymbol{\theta}_k)\} = \inf\{t > 0 : \mathbf{x} \in t\,\mathrm{conv}(\pm\mathcal{A})\}$$

- The unit ball of the atomic norm is the convex hull of the symmetrized atomic set $\pm\mathcal{A}$.



V. Chandrasekaran, B. Recht, P. Parrilo, A Willsky. "The convex geometry of linear inverse problems."

Atomic Norms III

Dual atomic norm.

- The dual atomic norm is defined as

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

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

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

Atoms	Atomic Norm	Dual Atomic Norm
canonical basis vectors	ℓ_1 norm	ℓ_∞ norm
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 IV

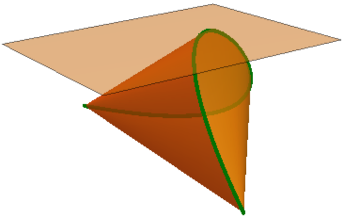
Atomic norm minimization (ANM) problems.

- 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:

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

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

$$\underline{\text{Demixing:}}\ \text{minimize}\ \|\mathbf{x}\|_{\mathcal{A}_1} + \lambda\|\mathbf{z}\|_{\mathcal{A}_2}\ \text{ subject to }\ \mathbf{y} = \mathbf{x} + \mathbf{z}.$$



Problems I

- Atomic decomposition: Given a signal, which decompositions achieve the atomic norm? Put it another way, estimate parameters given full, noise-free data.
- Sampling complexity: how many linear 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/parameter estimation: how well can we approximately recover the active parameters from noisy data?
- Demixing: how well can we separate signals with two different structures?
- Blind atomic decomposition: how to solve the problem when the form of the atoms are not known precisely?
- Computational methods: how shall we solve atomic norm minimization problems?

Optimality of ANM in Line Spectral Estimation

Atomic Decomposition I

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

$$\mathbf{x} = \sum_{k=1}^{r} c_k^\star \mathbf{a}(\boldsymbol{\theta}_k^\star),$$

under what conditions on the parameters $\{c_k^\star, \boldsymbol{\theta}_k^\star\}$, we have

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

- For $\mathcal{A} = \{\mathbf{e}_k\}$, 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 $\mathcal{A} = \{\mathbf{d}_k\}$, a sufficient condition is that the dictionary matrix D satisfies restricted isometry property.

Atomic Decomposition II

Line spectral decomposition.

Theorem (Candès & Fernandez-Granda, 2012)

For line spectal atoms $\mathbf{a}(\theta) = \begin{bmatrix} 1 & e^{j2\pi\theta} & \cdots & e^{j2\pi n\theta} \end{bmatrix}^T$*, if the true paramters* $\{\theta_k^\star\}$ *are separated by* $\frac{4}{n}$*, the atomic norm* $\|\mathbf{x}\|_{\mathcal{A}} = \sum_{k=1}^{r} |c_k^\star|$*.*

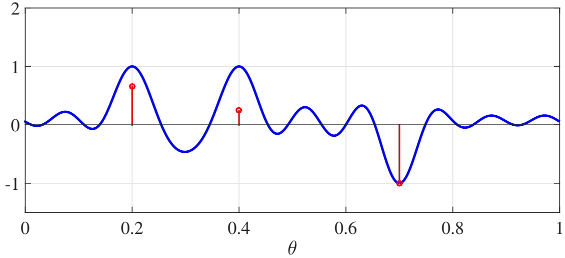
- Compare with Prony's method etc. Foundation for handling noise, outliers, missing data.
- The critical separation was improved to $\frac{2.52}{n}$ (Fernandez-Granda, 2015).
- The separation condition is in a flavor similar to the restricted isometry property for finite dictionaries, and the orthogonality condition for singular value decomposition.

E. Candès, C. Fernandez-Granda. "Towards a Mathematical Theory of Super-resolution."
C. Fernandez-Granda. "Super-resolution of point sources via convex programming."

Atomic Decomposition III

- We can extract the decomposition from the dual optimal solution $\mathbf{q}$.
- The trigonometric polynomial $q(\theta) = \langle \mathbf{a}(\theta), \mathbf{q} \rangle$ satisfies

$$\|q(\theta)\|_{L_\infty} \leq 1$$
$$q(\theta_k^\star) = \text{sign}(c_k^\star), k \in [r]$$



- We identify the true parameters by solving $|q(\theta)| = 1$.

Recovery with Missing Data I

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

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

- Atomic completion for line spectral signal approaches information theoretic limit:

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

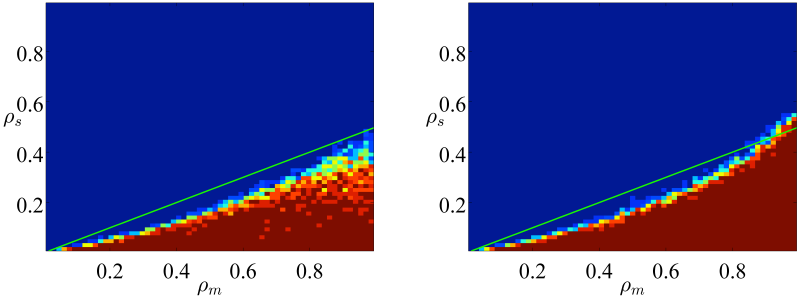
If we observe $\mathbf{x}^\star = \sum_{k=1}^{r} c_k^\star \mathbf{a}(\theta_k^\star)$ on a random subset of $\{0, 1, \ldots, n-1\}$ of size $O(r \log(r) \log(n))$ and the true parameters are separated by $\frac{4}{n}$, then atomic norm minimization successfully completes the signal.

Theorem (Chi and Chen, 2013)

Similar results hold for 2d spectral signals.

G. Tang, B. Bhaskar, P. Shah, B. Recht. "Compressed sensing off the grid."
Y. Chi, and Y. Chen. "Compressive two-dimensional harmonic retrieval via atomic norm minimization."

Recovery with Missing Data II



Denoising I

- Observe noisy measurements: $\mathbf{y} = \mathbf{x}^\star + \mathbf{w}$ with $\mathbf{w}$ white Gaussian noise.
- Denoise $\mathbf{y}$ to obtain (Choose $\lambda \approx 2\mathbb{E}\|\mathbf{w}\|_{\mathcal{A}}^*$)

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

Theorem (Tang, Bhaskar & Recht, 2013)

$$\frac{1}{n}\|\hat{\mathbf{x}} - \mathbf{x}^\star\|_2^2 \leq \frac{C\sigma^2 r \log(n)}{n}$$ *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.

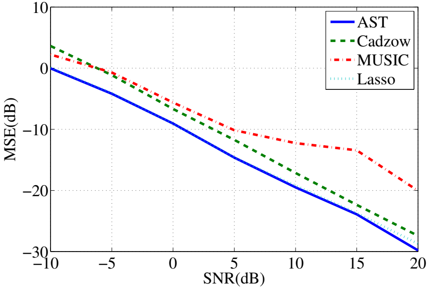
No algorithm can do better than

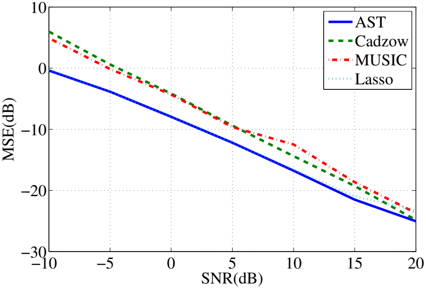
$$\frac{1}{n}\|\hat{\mathbf{x}} - \mathbf{x}^\star\|_2^2 \geq \frac{C'\sigma^2 r}{n}$$

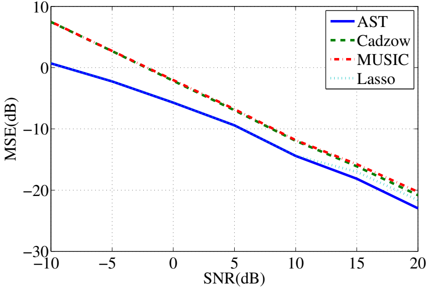
even if we know a priori the well-separated parameters.

G. Tang, Gongguo, B. Bhaskar, B. Recht. "Near minimax line spectral estimation."
B. Bhaskar, G. Tang, B. Recht. "Atomic norm denoising with applications to line spectral estimation."

Denoising II





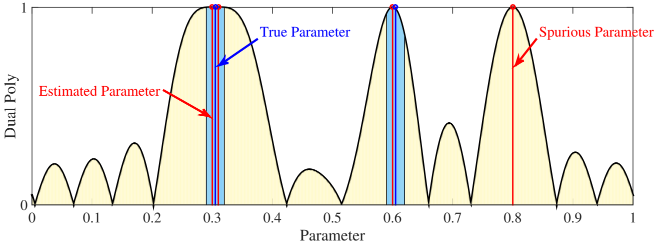


Noisy Support Recovery/Parameter Estimation I

- How well can we localize the frequencies by solving a denoising problem and extracting parameters from its solution?

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

- Here W is a diagonal weighting matrix of size $O(1/n)$.
- The dual optimal solution $\hat{\mathbf{q}} = W(\mathbf{y}-\hat{\mathbf{x}})/\lambda$ is a scaled version of the noise estimator.
- The places where $|\langle \hat{\mathbf{q}}, \mathbf{a}(\theta)\rangle| = 1$ correspond to identified parameters.



V. Duval, G. Peyré. "Exact support recovery for sparse spikes deconvolution." C. Fernandez-Granda. "Support detection in super-resolution."
G. Tang, Gongguo, B. Bhaskar, B. Recht. "Near minimax line spectral estimation."

Noisy Support Recovery/Parameter Estimation II

- Noise is Gaussian with variance σ^2.
- Noise level is measured by $\gamma_0 := \sigma\sqrt{\frac{\log n}{n}}$.

Theorem (Li & Tang, 2016)

Suppose

- *SNR as measured by* $|c_{\min}|/\gamma_0$ *is large.*
- *The dynamic range of the coefficients is small*
- *Regularization parameter* λ *is large compared to* γ_0*.*
- *The frequencies are well-separated.*

Then w.h.p. we can extract exactly r *parameters from* $\hat{\mathbf{x}}$ *or* $\hat{\mathbf{q}}$*, which satisfy*

$$\max |c_k^\star||\hat{\theta}_k - \theta_k^\star| = O(\gamma_0/n) = O(\frac{\sqrt{\log n}}{n^{3/2}}\sigma)$$

$$\max |\hat{c}_k - c_k^\star| = O(\lambda) = O(\sqrt{\frac{\log n}{n}}\sigma)$$

Q. Li, G. Tang. "Approximate support recovery of atomic line spectral estimation: A tale of resolution and precision."

Noisy Support Recovery/Parameter Estimation III

Comparison with CRB, MUSIC, and MLE.

- Only asymptotic bounds available when the number of snapshots T tends to ∞.
- Our algorithm and analysis work for single snapshot, i.e., $T = 1$.

- CRB: $O(\frac{\sigma^2}{T|c|^2 n^3})$
- Atomic: $O(\frac{\sigma^2 \log n}{|c|^2 n^3})$
- MLE: $O(\frac{\sigma^2}{T|c|^2 n^3} + \frac{\sigma^4}{T|c|^4 n^4})$
- MUSIC: $O(\frac{\sigma^2}{T|c|^2 n^3} + \frac{\sigma^4}{T|c|^4 n^4})$

P. Stoica, A. Nehorai. "MUSIC, maximum likelihood, and Cramer-Rao bound."

Noisy Support Recovery/Parameter Estimation IV

Primal-Dual Witness Construction

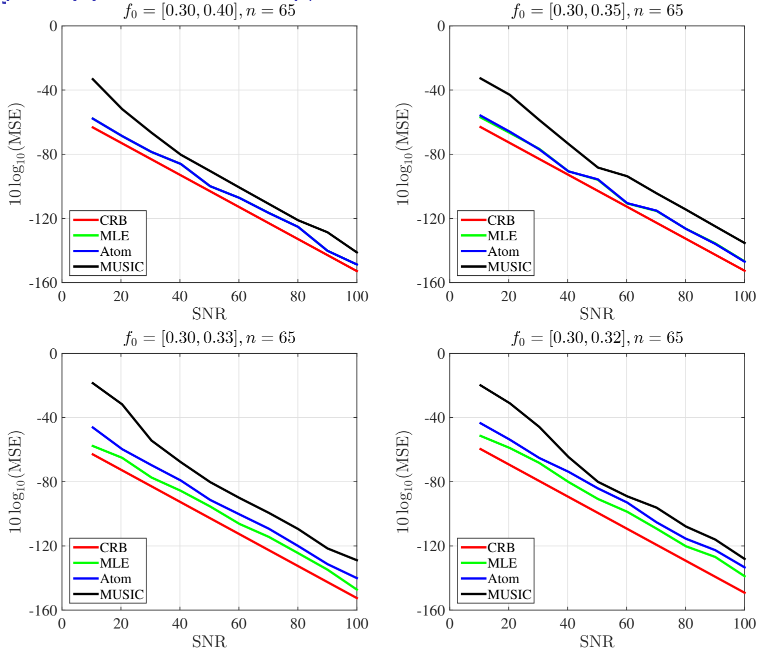
- The unique primal optimal solution $\hat{\mathbf{x}} = \sum_{k=1}^{\hat{r}} \hat{c}_k \mathbf{a}(\hat{\theta}_k)$.
- The unique dual optimal solution $\hat{\mathbf{q}}$ satisfies $\sup_\theta |\langle \mathbf{q}, \mathbf{a}(\theta)\rangle| \leq 1$ and $\langle \mathbf{q}, \mathbf{a}(\hat{\theta}_k)\rangle = \text{sign}(\hat{c}_k), k \in [\hat{r}]$.
- They certify the optimality of each other and are related by $\hat{\mathbf{q}} = W(\mathbf{y} - \hat{\mathbf{x}})/\lambda$.
- To construct $\hat{\mathbf{x}}$ and $\hat{\mathbf{q}}$, fix $\hat{r} = r$ and find $(\hat{\theta}_k, \hat{c}_k)$ by solving

$$\underset{\theta_k, c_k}{\text{minimize}} \frac{1}{2}\|\mathbf{y} - \sum_{k=1}^{r} c_k \mathbf{a}(\theta_k)\|_W^2 + \lambda \sum_{k=1}^{r} |c_k|$$

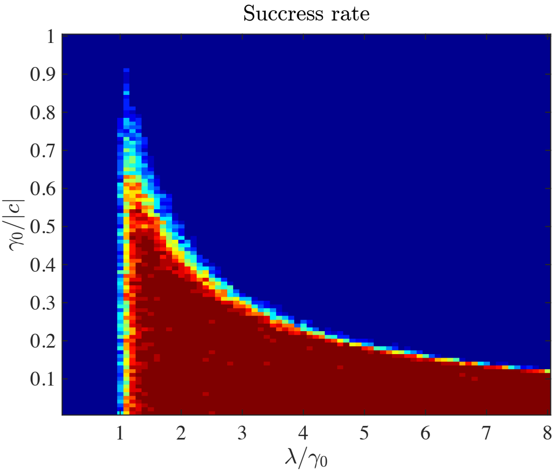
- Run gradient descent initialized by the true frequencies and argue that such a local minimum generates a valid primal-dual optimal solution.
- The atomic solution is the same as the ℓ_1 regularized least-squares solution.

M. Wainwright. "Sharp thresholds for high-dimensional and noisy sparsity recovery using constrained quadratic programming (Lasso)"

Noisy Support Recovery/Parameter Estimation V



Noisy Support Recovery/Parameter Estimation VI



Setup:

- $n = 257$
- $|c_k^\star| = 1$
- Separation $\geq 5/n$

Success means:

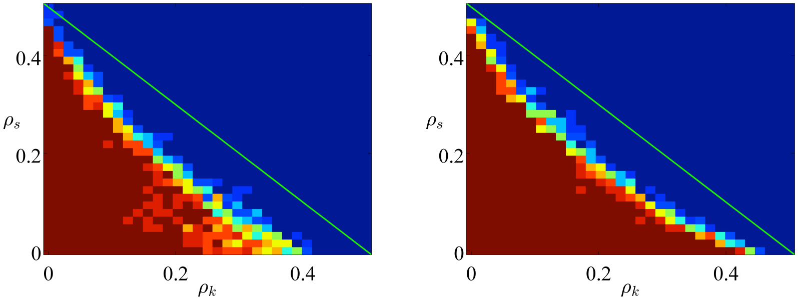
- $\max_k |c_k^\star||\hat{\theta}_k - \theta_k^\star| \leq \frac{\gamma_0}{2n}$
- $\max_k |\hat{c}_k - c_k^\star| \leq 2\lambda$

Demixing/Outlier Detection and Removal I

- Observe corrupted data: $\mathbf{y} = \mathbf{x}^\star + \mathbf{w}^\star$ with $\mathbf{w}^\star$ sparse.
- $\text{minimize}_{\mathbf{w},\mathbf{x}} \|\mathbf{w}\|_{\ell_1} + \|\mathbf{x}\|_{\mathcal{A}}$ subject to $\mathbf{y} = \mathbf{x} + \mathbf{w}$.

Theorem

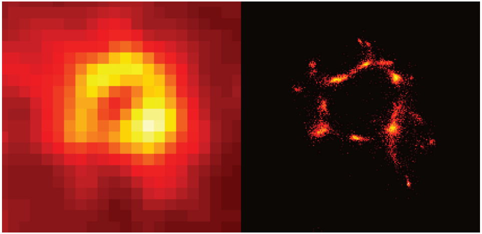
If the frequencies are well-separated, the locations of the outliers are uniformly random, and $r + s \leq \frac{n}{\log^2(n)}$, then we could recover both the freqs and the outliers exactly with high prob.

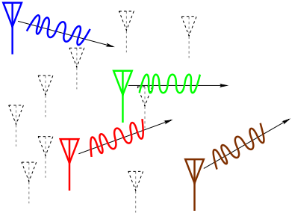


G. Tang, P. Shah, B. Bhaskar, B. Recht, "Robust line spectral estimation."
C. Fernandez-Granda, G. Tang, X. Wang, L. Zheng, "Demixing sines and spikes: robust spectral super-resolution in the presence of outliers".

Non-stationary Blind Super-resolution I

- Single molecule microscopy: Localize point sources from their convolution with point spread functions
- Multi-user communication systems: Estimate the transmitted waveforms





observation model: $y(t) = \sum_{j=1}^{J} \alpha_j g_j(t - t_j)$

Blind Super-resolution: The PSFs or the transmitted waveforms might be unknown/partially known.

Non-stationary Blind Super-resolution II

- Frequency domain model:

$$\mathbf{y} = \sum_{j=1}^{J} c_j \mathbf{a}(\tau_j) \odot \mathbf{g}_j \in \mathbb{C}^N.$$

- Goal: recover $\{\tau_j\}$, $\{c_j\}$ and samples of the unknown waveforms $\{\mathbf{g}_j(n)\}$ from the observations $\mathbf{y}$.
- Assumption: all the $\mathbf{g}_j$ live in the same subspace of dimension K spanned by the columns of B, i.e., $\mathbf{g}_j = B\mathbf{h}_j$.
- $\mathbf{y}$ becomes linear observations of a structured low-rank matrix $\mathbf{X}_o = \sum_{j=1}^{J} c_j \mathbf{h}_j \mathbf{a}(\tau_j)^H$: $\mathbf{y} = \mathcal{B}(X_o)$.

Non-stationary Blind Super-resolution III

- Define the atomic set

$$\mathcal{A} = \left\{ \mathbf{h}\mathbf{a}(\tau)^H : \ \tau \in [0,1), \|\mathbf{h}\|_2 = 1, \mathbf{h} \in \mathbb{C}^{K \times 1} \right\}$$

- Solve an atomic norm minimization problem to recovery X_o

$$\text{minimize} \, \|\mathbf{X}\|_{\mathcal{A}} \ \text{subject to} \ \mathbf{y} = \mathcal{B}(\mathbf{X}). \tag{1}$$

- Once $\mathbf{X}_o$ is recovered, the point sources can be localized by checking the dual polynomial.

Theorem (Chi, 2015)

Assume isotropic and incoherent $\mathbf{b}_n$, separation between τ_k, and stationary $\mathbf{h}_j$, i.e., $\mathbf{h}_j = \mathbf{h}$ for all j. Guaranteed recovery if $N = O(J^2 K^2)$.

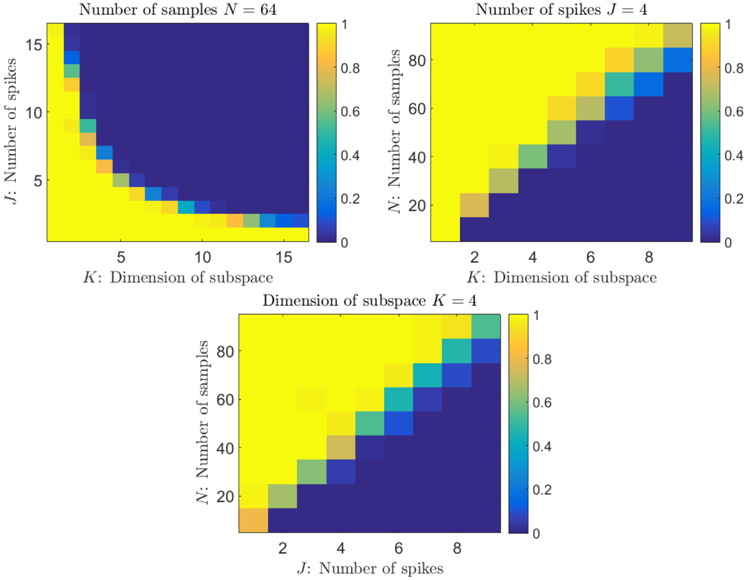
Theorem (Yang, Tang & Wakin, 2016)

Assume isotropic and incoherent $\mathbf{b}_n$, separation between τ_k, and randomness of $\mathbf{h}_j$. Guaranteed recovery if $N = O(JK)$.

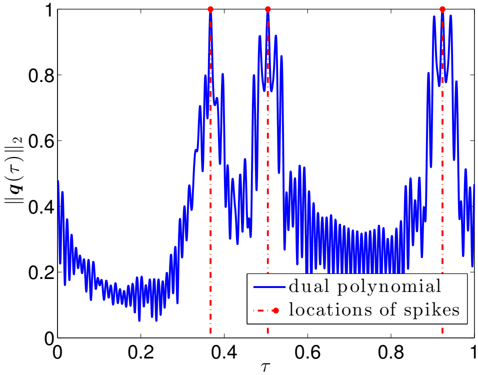
Y. Chi. "Guaranteed blind sparse spikes deconvolution via lifting and convex optimization."

D. Yang, G. Tang, and M. Wakin, " Super-resolution of complex exponentials from modulations with unknown waveforms."

Non-stationary Blind Super-resolution IV



A more practical example I



- Components of $\mathbf{g}$ are samples of Gaussian waveform $g_{\sigma^2}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}$ with unknown variance $\sigma^2 \in [0.1, 1]$
- B: left singular vectors of the best rank-5 approximation of $\mathbf{D_g}$

$$\mathbf{D_g} = \left[\mathbf{g}_{\sigma^2=0.1} \quad \mathbf{g}_{\sigma^2=0.11} \quad \mathbf{g}_{\sigma^2=0.99} \quad \cdots \quad \mathbf{g}_{\sigma^2=1}\right]$$

Computational Methods I

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

$$\|\mathbf{z}\|_{\mathcal{A}}^{*} \leq 1 \Leftrightarrow |\langle \mathbf{z}, \mathbf{a}(\boldsymbol{\theta}) \rangle| \leq 1 \ \forall \boldsymbol{\theta} \in \Theta$$

- Line spectral atoms:

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

- The latter states that the magnitude of a complex trigonometric polynomial is bounded by 1 everywhere.
- Bounded real lemma (Dumitrescu, 2007):

$$\begin{aligned} & |\sum_{k=0}^{n-1} \mathbf{z}_k e^{j2\pi\theta k}| \leq 1 \ \forall \theta \in [0, 1] \\ \Leftrightarrow & \begin{bmatrix} Q & \mathbf{z} \\ \mathbf{z}^H & 1 \end{bmatrix} \succeq 0, \\ & \mathrm{trace}(Q, j) = \delta(j = 0), j = 0, \ldots, n-1. \end{aligned}$$

Computational Methods II

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

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

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

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

$$\Leftrightarrow$$

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

B. Bhaskar, G. Tang, B. Recht. "Atomic norm denoising with applications to line spectral estimation."

ANM for Tensors

Atomic Tensor Decomposition. I

- For tensor problems, the atomic set

$$\mathcal{A} = \{\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} : \|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = \|\mathbf{w}\|_2 = 1\}.$$

- The tensor atomic norm is the tensor nuclear norm.
- When can we extract the rank-one factors of a tensor by observing its entries?

$$T = \textstyle\sum_{k=1}^{r} c_k^\star \mathbf{u}_k^\star \otimes \mathbf{v}_k^\star \otimes \mathbf{w}_k^\star$$

- Equivalently, when

$$\|T\|_* = \|T\|_{\mathcal{A}} = \sum_{k=1}^{r} |c_k^\star|$$

Atomic Tensor Decomposition. II

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

- *Incoherence:* $\max_{k\neq l}\{|\langle \mathbf{u}_k^\star, \mathbf{u}_l^\star\rangle|, |\langle \mathbf{v}_k^\star, \mathbf{v}_l^\star\rangle|, |\langle \mathbf{w}_k^\star, \mathbf{w}_l^\star\rangle|\} \leq \frac{\mathrm{polylog}(n)}{\sqrt{n}}$
- *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):* $r = O(n^{17/16}/\,\mathrm{polylog}(n))$

guarantees atomic tensor decomposition.

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

Random $\{\mathbf{u}_k^\star\}, \{\mathbf{v}_k^\star\}$ *and* $\{\mathbf{w}_k^\star\}$ *satisfy the conditions with high probability.*

Li, Prater, Shen, Tang. "Overcomplete tensor decomposition via convex optimization."

Computational Methods I

SOS relaxations for tensors.

- 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.

Computational Methods II

Theorem (Tang & Shah, 2015)

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

G. Tang, P. Shah. "Guaranteed tensor decomposition: A moment approach."

Concluding Remarks

- ANM provides a universal framework for constructing convex regularizers that promote certain notion of simplility/parsimony/sparsity.
- ANM is optimal in many senses: approaching information-theoretic limit in sampling complexity, minimax optimal in denoising, approaching CRB in parameter estimation, etc.
- When ANM works, the corresponding optimization is usually also computationally feasible/easy.
- ANM applies naturally to inverse problems in signal processing and machine learning.

Acknowledgements

- Thanks to B. Bhaskar, C. Fernandez-Granda, Q. Li, A. Prater, B. Recht, P. Shah, and L. Shen, M. Wakin, D. Yang, Z. Zhu for collaborations and help with materials reported in this talk.
- This work is partly supported by National Science Foundation under grants CCF-1464205, and the Air Force Summer Faculty Fellowship.