Denoising, Decomposition and Demixing of Moment Sequences by Convex Optimization

Badri Narayan Bhaskar

Advisor: Prof. Benjamin Recht

Collaborators: Gongguo Tang, Parikshit Shah

Department of Electrical Engineering,
University of Wisconsin-Madison

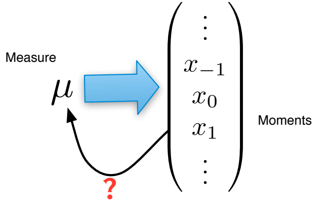
Preliminary Examination
July 26, 2012

Moments

The moments of a measure μ are given by

$$x_m = \int z^m d\mu(z)$$

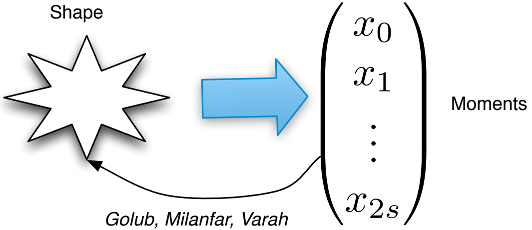
Moment Problem



Trigonometric Moments μ on $\mathbb{T}$

Power Moments μ on $\mathbb{R}$

Shape from Moments

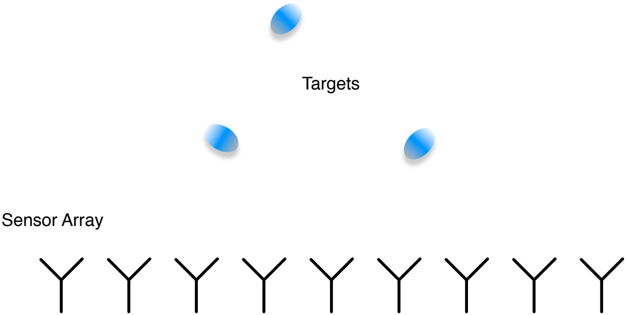


Theorem (Motzkin-Schoenberg)

Suppose $\mathcal{P}$ has vertices $z_1, \ldots, z_s$.

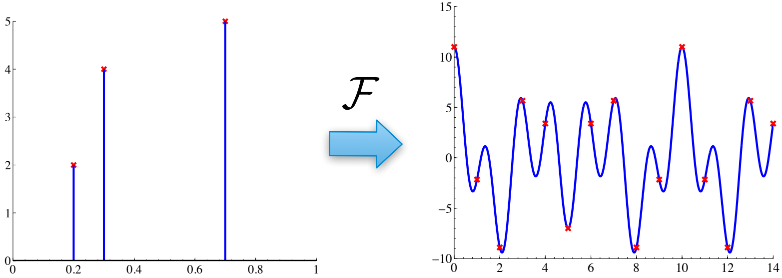
$$k(k-1)\int_{\mathcal{P}} z^{k-2} dx dy = \sum_{j=1}^{k} a_j z_j^k = \int z^k \left(\sum_{j=1}^{k} a_j \delta(z - z_k) \right) dx\, dy$$

Array Signal Processing



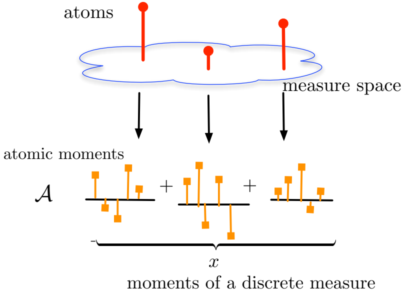
- Estimate location and the direction of arrival
- Deploy as few sensors as possible

Line Spectrum Estimation



- Discrete spectrum $=$ Mixture of sinusoids

Strcture and Simplicity



- Nonlinear estimation problem.
- Ill-posed – need structure.
- Simple objects: few atoms.
- Continuously many atoms.

"weights" "features"

$$x = \sum_{a \in T} c_a a$$

simple $\Rightarrow T \subset \mathcal{A}$ is small

Examples of Simple Objects

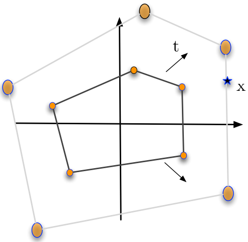
Object	Notion of Parsimony	Atoms ($\mathcal{A}$)
Vectors	Sparsity	1-sparse vectors
Matrices	Rank	Rank-1 matrices
Bandlimited Signals	No. of Frequencies	Complex sinusoids
Linear Systems	McMillan Degree	Single pole systems

Atomic Norm

(*Chandrasekaran, Recht, Parillo, Willsky*, 2010)

- The atomic norm at x is given by:

$$\|x\|_{\mathcal{A}} = \inf \{t > 0 \mid x \in t\,\mathrm{conv}(\mathcal{A})\}$$
$$= \inf \left\{ \sum_a c_a \;\middle|\; x = \sum_{a \in \mathcal{A}} c_a a,\; c_a > 0 \right\}$$



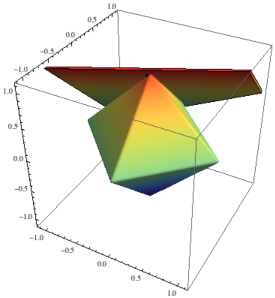
- The dual norm is the support function of $\mathcal{A}$:

$$\|z\|_{\mathcal{A}}^{*} = \sup_{x \in \mathcal{A}} \langle z, x \rangle.$$

Examples

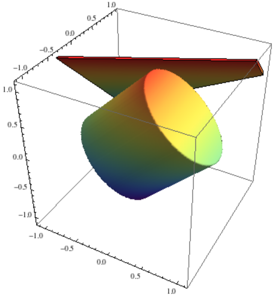
- $x^\star$ is sparse. $y = \Phi x^\star$.
- minimize$_x$ $\|x\|_1$
  subject to $y = \Phi x$

Figure: ℓ_1 norm ball $\left\| \begin{bmatrix} x & y & z \end{bmatrix}^T \right\|_1 \leq 1$

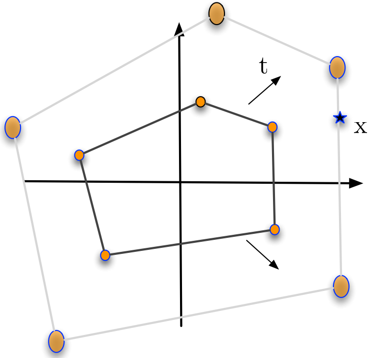


- $X^\star$ is low rank. $Y = \mathcal{A}(X^\star)$.
- minimize$_x$ $\|X\|_*$
  subject to $Y = \mathcal{A}(X)$.

Figure: Nuclear Norm ball $\left\| \begin{bmatrix} x & y \\ y & z \end{bmatrix} \right\|_* \leq 1$



Denoising with Atomic Norms



Atomic Soft Thresholding

Suppose $x^\star$ is simple and we observe $y = x^\star + w$.

Primal Problem

$$\underset{x}{\text{minimize}} \frac{1}{2}\|y - x\|_2^2 + \tau\|x\|_{\mathcal{A}}$$

Atomic Soft Thresholding

Suppose $x^\star$ is simple and we observe $y = x^\star + w$.

Primal Problem

$$\underset{x}{\text{minimize}} \; \frac{1}{2}\|y - x\|_2^2 + \tau\|x\|_{\mathcal{A}}$$

Dual Problem

$$\begin{aligned} &\underset{z}{\text{minimize}} && \|y - z\|_2 \\ &\text{subject to} && \|z\|_{\mathcal{A}}^* \leq \tau \end{aligned}$$

Atomic Soft Thresholding

Suppose $x^\star$ is simple and we observe $y = x^\star + w$.

Primal Problem

$$\underset{x}{\text{minimize}}\ \frac{1}{2}\|y - x\|_2^2 + \tau\|x\|_{\mathcal{A}}$$

Dual Problem

$$\begin{aligned} &\underset{z}{\text{minimize}} && \|y - z\|_2 \\ &\text{subject to} && \|z\|_{\mathcal{A}}^* \leq \tau \end{aligned}$$

- The primal-dual solution pair $(\hat{x}, \hat{z})$ satisfy *(i)* $y = \hat{x} + \hat{z}$, *(ii)* $\langle \hat{x}, \hat{z} \rangle = \tau\|\hat{x}\|_{\mathcal{A}}$, *(iii)* $\|\hat{z}\|_{\mathcal{A}}^* \leq \tau$.

Atomic Soft Thresholding

Suppose $x^\star$ is simple and we observe $y = x^\star + w$.

Primal Problem

$$\underset{x}{\text{minimize}} \frac{1}{2}\|y - x\|_2^2 + \tau\|x\|_{\mathcal{A}}$$

Dual Problem

$$\begin{aligned} \underset{z}{\text{minimize}} \ & \|y - z\|_2 \\ \text{subject to} \ & \|z\|_{\mathcal{A}}^* \leq \tau \end{aligned}$$

- The primal-dual solution pair $(\hat{x}, \hat{z})$ satisfy *(i)* $y = \hat{x} + \hat{z}$, *(ii)* $\langle \hat{x}, \hat{z} \rangle = \tau\|\hat{x}\|_{\mathcal{A}}$, *(iii)* $\|\hat{z}\|_{\mathcal{A}}^* \leq \tau$.
- Correct choice of regularization parameter. $\tau = \mathbb{E}\,\|w\|_{\mathcal{A}}^*$

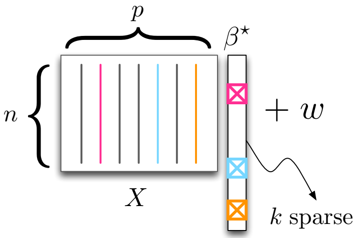
Proposition (MSE)

If $\tau \geq \tau_0 \geq \mathbb{E}\left(\|w\|_{\mathcal{A}}^*\right)$, then $\mathbb{E}\left(\|x^\star - \hat{x}\|_2^2\right) \leq \tau\|x^\star\|_{\mathcal{A}}$.

Proposition (MSE)

If $\tau \geq \tau_0 \geq \mathbb{E}\left(\|w\|_{\mathcal{A}}^*\right)$, then $\mathbb{E}\left(\|x^\star - \hat{x}\|_2^2\right) \leq \tau\|x^\star\|_{\mathcal{A}}$.

Application to sparse combination of columns



- $\frac{1}{n}\mathbb{E}\left(\|X\hat{\beta} - X\beta^\star\|_2^2\right) \leq \sigma\sqrt{\frac{k\log(p)}{n}}\|\beta^\star\|_2.$

- No assumptions on $X \implies$ Slow rate.

Fast Rates

Conditions on approximate descent cone of $\mathcal{A}$:

- $T_\gamma(x^\star, \mathcal{A}) = \operatorname{cone}\left(\{z \mid \|x^\star + z\|_{\mathcal{A}} \leq \|x^\star\|_{\mathcal{A}} + \gamma\|z\|_{\mathcal{A}}\}\right).$
- $\phi_\gamma(x^\star, \mathcal{A}) = \inf\left\{\frac{\|z\|_2}{\|z\|_{\mathcal{A}}} \;\middle|\; z \in T_\gamma(x^\star, \gamma)\right\} > 0$

Theorem

If $\tau \geq 2\tau_0$ and $\phi_{1/2}(x^\star, \mathcal{A}) > 0$, then with high probability,

$$\|x^\star - \hat{x}\|_2^2 = O\left(\frac{\tau^2}{\phi_{1/2}^2}\right).$$

Fast Rates

Conditions on approximate descent cone of $\mathcal{A}$:

- $T_\gamma(x^\star, \mathcal{A}) = \text{cone}\left(\{z \mid \|x^\star + z\|_\mathcal{A} \leq \|x^\star\|_\mathcal{A} + \gamma\|z\|_\mathcal{A}\}\right).$
- $\phi_\gamma(x^\star, \mathcal{A}) = \inf\left\{ \frac{\|z\|_2}{\|z\|_\mathcal{A}} \;\middle|\; z \in T_\gamma(x^\star, \gamma)\right\} > 0$

Theorem

If $\tau \geq 2\tau_0$ and $\phi_{1/2}(x^\star, \mathcal{A}) > 0$, then with high probability,

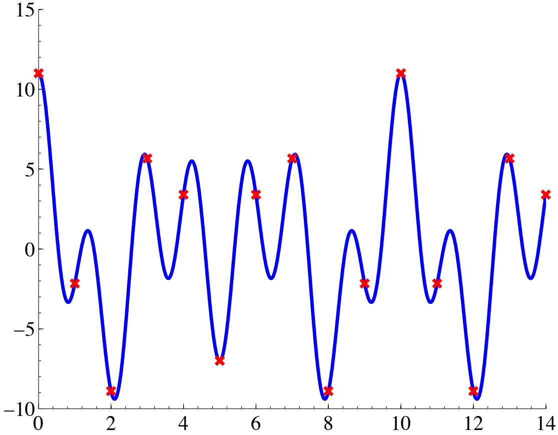
$$\|x^\star - \hat{x}\|_2^2 = O\left(\frac{\tau^2}{\phi_{1/2}^2}\right).$$

For the sparse combination of columns,

- With no assumptions on X, MSE $= O\left(\sqrt{\frac{k\log(p)}{n}}\right)$.
- If $\phi_{1/2}(X\beta^\star) > 0$, MSE $= O\left(\frac{k\log(p)}{n}\right)$.

Atomic norm denoising with applications to line spectrum estimation

B, Gongguo Tang, Ben Recht



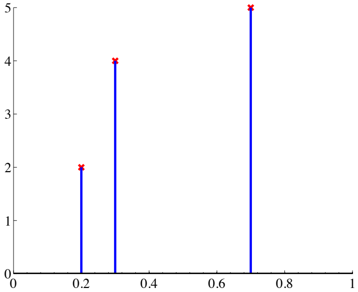
Trigonometric Moments and Line Spectrum

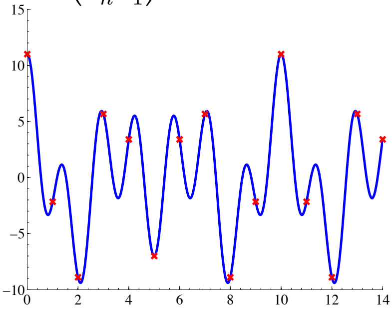
Measure

$$\mu = \sum_{j=1}^{s} c_j^\star \delta_{u_j^\star}$$

Trigonometric Moments

$$\begin{pmatrix} x_0^\star \\ \vdots \\ x_k^\star \\ \vdots \\ x_{n-1}^\star \end{pmatrix} \qquad \boxed{x_k^\star = \sum_{j=1}^{s} c_j^\star e^{i2\pi u_j^\star k}}$$





$$\underbrace{\begin{pmatrix} x_0^\star \\ x_1^\star \\ \vdots \\ x_{n-1}^\star \end{pmatrix}}_{x^\star} = \sum_{j=1}^{s} c_j^\star \underbrace{\begin{pmatrix} 1 \\ \exp(i2\pi u_j^\star) \\ \vdots \\ \exp(i2\pi(n-1)u_j^\star) \end{pmatrix}}_{a(u_j^\star)}$$

$$x^\star = \sum_{j=1}^{s} c_j^\star a(u_j^\star)$$

After a little algebra,

$$Tx^\star = \begin{pmatrix} x_0^\star & x_1^\star & \cdots & x_{n-1}^\star \\ x_{-1}^\star & x_0^\star & \ddots & x_n \\ \vdots & \ddots & \ddots & \vdots \\ x_{-n+1}^\star & \cdots & \cdots & x_0^\star \end{pmatrix} = \sum_{j=1}^{s} c_j^\star a(u_j^\star) a(u_j^\star)^*.$$

- $Tx^\star \in \mathbb{C}^{n\times n}$ is always rank s, as long as $n \geq s$.

Classical Line Spectrum Estimation

- Recall $Tx^\star = \sum_{j=1}^s c_j^\star a(u_j^\star)a(u_j^\star)^*$
- $n = s+1 \Rightarrow$ null vector q satisfies $q^*Tx^\star q = 0$.

$$q^*Tx^\star q = \sum_{j=1}^{s} c_j^\star \left|q^* a(u_j^\star)\right|^2 = \sum_{j=1}^{s} c_j^\star \left|\sum_{k=0}^{n} q_k^\star e^{i2\pi u_j^\star k}\right|^2$$

  Roots of $q \Rightarrow$ frequencies (Prony, 1795).
- Sensitive to noise.

Cadzow's Heuristic

Suppose $y = x^\star + w$. Recall that $Tx^\star$ must have rank s.

Algorithm 1 Cadzow's Alternating Projections

$i \leftarrow 0$
$X^{(i)} \leftarrow Ty$.
while termination conditions aren't met **do**
  $X^{(i+1)} \leftarrow \arg\inf_{\mathrm{rank}(X)=k} \left\| X - X^{(i)} \right\|$
  $X^{(i+1)} \leftarrow \textsc{Toeplitzise}\left(X^{(i+1)}\right)$
  $i \leftarrow i + 1$
end while
return $T^{-1}(X^{(i)})$

Denoising Samples of Line Spectral Signals

frequencies

$$x^\star = \sum_{j=1}^{s} c_j^\star \underbrace{a(u_j^\star)}$$

amplitudes

Atomic Moment Vectors

$$a(u) = \begin{pmatrix} 1 \\ \exp(i2\pi u) \\ \vdots \\ \exp(i2\pi(n-1)u) \end{pmatrix}.$$

Positive Atoms $\mathcal{A}_+ = \{a(u) \mid u \in [0,1]\}$.

Centrosymmetric $\mathcal{A} = \left\{e^{i\phi} a(u) \;\middle|\; u \in [0,1],\ \phi \in [0, 2\pi]\right\}$

Atomic Soft Thresholding

$$\text{minimize}_x \frac{1}{2}\|x - y\|_2^2 + \tau \|x\|_{\mathcal{A}}$$

Semidefinite characterization (known)

Figure: Moment Curve $(\cos(t), \cos(2t), \cos(3t))$

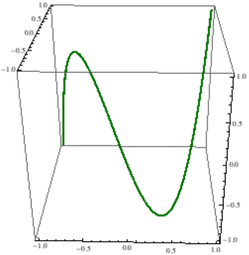
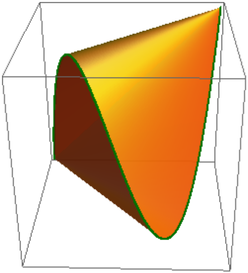


Figure: Convex Hull of the Positive Moment Curve



$$\|x\|_{\mathcal{A}} = \begin{cases} nx_0 & Tx \succeq 0 \\ \infty & \text{otherwise.} \end{cases}$$

Semidefinite characterization (our result)

Figure: Phase Symmetric Moment Curve

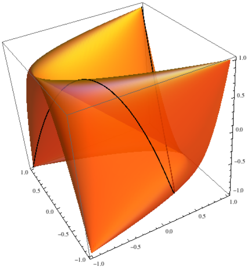
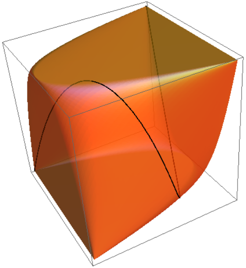


Figure: Convex Hull



$$\|x\|_{\mathcal{A}} = \inf\left\{ \tfrac{1}{2}\operatorname{trace}(Tu + \tfrac{1}{2}t) \;\middle|\; \begin{bmatrix} Tu & x \\ x^* & t \end{bmatrix} \succeq 0 \right\}.$$

Choice of Regularization Parameter

The "correct" choice of regularization parameter is given by the expected dual atomic norm of noise. We have,

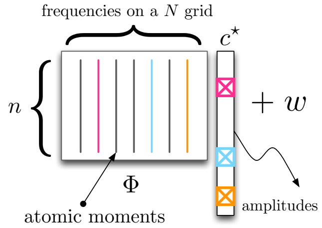
$$\begin{aligned}\mathbb{E}\left(\|w\|_{\mathcal{A}}^{*}\right) &= \mathbb{E}\sup_{u,\phi}\left\langle w, e^{i\phi}a(u)\right\rangle \\ &= \mathbb{E}\sup_{u}\left|\sum_{k=0}^{n-1} w_k e^{i2\pi uk}\right|\end{aligned}$$

which is the expected maximum modulus of a random polynomial.

- We show $\mathbb{E}\left(\|w\|_{\mathcal{A}}^{*}\right) \leq \sigma\left(1+\frac{1}{\log(n)}\right)\sqrt{n\log(n)+n\log(4\pi\log(n))}$.
- We also show $\mathbb{E}\left(\|w\|_{\mathcal{A}}^{*}\right) \geq \sigma\sqrt{n\log(n)-\frac{n}{2}\log(4\pi\log(n))}$.
- Using MSE proposition (slow rate), $\mathbb{E}\|\hat{x}-x^{\star}\|_2^2 \lesssim \sigma\sqrt{\frac{\log(n)}{n}}\sum_{l=1}^{k}|c_l^{\star}|$.

Discretized Atomic Soft Thresholding (DAST)

- For $N > 2\pi n$, put $\mathcal{A}_N = \left\{ a_{m/N,\phi} \;\middle|\; m = 0, \ldots, N-1 \right\}$.
- We show $\left(1 - \frac{2\pi n}{N}\right) \|x\|_{\mathcal{A}_N} \leq \|x\|_{\mathcal{A}} \leq \|x\|_{\mathcal{A}_N}$.
- Now a Lasso problem:

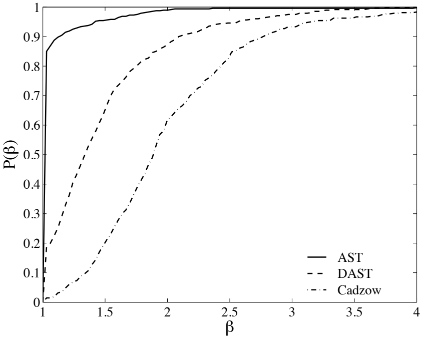


- *Remark:* DAST is $O(n \log N)$ since Φ is a DFT matrix
- Coherence doesn't matter for denoising.

Performance Profiles

$P(\beta)$ = Fraction of experiments with MSE less than $\beta\times$ minimum MSE.

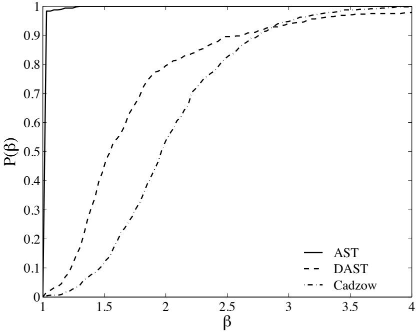
Figure: Random frequencies



Performance Profiles

$P(\beta)$ = Fraction of experiments with MSE less than $\beta\times$ minimum MSE.

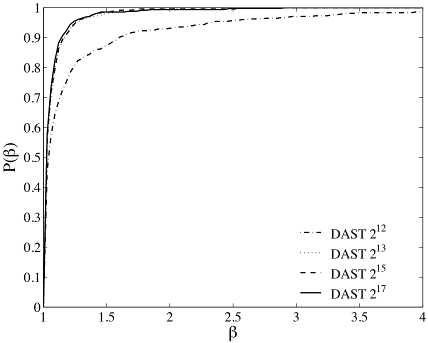
Figure: Equispaced frequencies



Performance Profiles

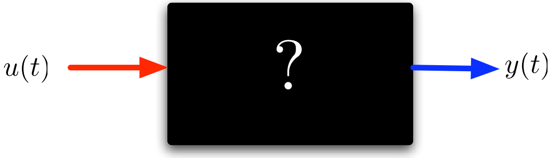
$P(\beta) =$ Fraction of experiments with MSE less than $\beta\times$ minimum MSE.

Figure: Comparing Discretization Levels



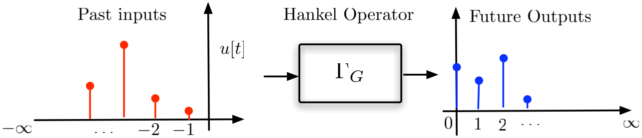
Linear System Identification via Atomic Norm Regularization

Parikshit Shah, B., Gongguo Tang, Ben Recht



Simple LTI systems

$$G(z) = \sum_{k=1}^{\infty} g_k z^{-k} = \sum_{j=1}^{s} \frac{c_j(1-|w_j|^2)}{z-w_j}$$



$\Gamma_G : \ell_2(-\infty, 0) \to \ell_2[0, \infty)$

$$\text{McMillan degree} = \text{number of poles} = \text{rank}(\Gamma_G)$$

Goal: Infer a system of low McMillan degree from linear measurements.

Hankel Nuclear Norm

- Minimize "nuclear norm" in place of rank!
- Spectral Series Representation:

$$\Gamma_G(x) = \sum_j \sigma_j \langle v_j, x \rangle u_j$$

  Hankel Nuclear Norm

$$\|\Gamma_G\|_* = \sum_j \sigma_j$$

- LMI characterization of $\|\Gamma_G\|_*$?

Atomic Norm

Partial Fraction Expansion

$$G(z) = \sum_{j=1}^{s} \frac{c_j(1-|w_j|^2)}{z-w_j}$$

Pick the atomic set

$$\mathcal{A} = \left\{ \frac{1-|w|^2}{z-w} \; : \; w \in \mathbb{D} \right\} .$$

Atomic Norm

$$\|G(z)\|_{\mathcal{A}} = \inf \left\{ \sum_{w\in\mathbb{D}} |c_w| \; : \; G(z) = \sum_{w\in\mathbb{D}} \frac{c_w(1-|w|^2)}{z-w} \right\} ,$$

Theorem (Peller, Coifman, Rochberg, Bonsall, Walsh)

$$\tfrac{\pi}{8}\|G\|_{\mathcal{A}} \leq \|\Gamma_G\|_1 \leq \|G\|_{\mathcal{A}}.$$

Atomic Norm Minimization

- Observe $y = \mathcal{L}(G) + w$.
- Examples
  - FIR coefficients
  - Frequency samples
  - Output to a known input
- Natural Minimization Problem to solve is
$$\text{minimize}_G \ \tfrac{1}{2}\left\|\mathcal{L}(G) - y\right\|_2^2 + \mu\|G\|_{\mathcal{A}} .$$
- We argue it is enough to solve the denoising problem
$$\text{minimize}_x \ \tfrac{1}{2}\|x - y\|_2^2 + \mu\|x\|_{\mathcal{L}(\mathcal{A})} .$$

  where
$$\|x\|_{\mathcal{L}(\mathcal{A})} = \inf\left\{\sum_{w\in\mathbb{D}_\rho} |c_w| \ : \ x_i = \sum_{w\in\mathbb{D}_\rho} c_w \mathcal{L}_i\left(\frac{1-|w|^2}{z-w}\right)\right\} .$$

Our Result

Let us assume that $\hat{G}$ is an approximate solution to the discretized version of the atomic soft thresholding problem.

Theorem

Suppose $G^\star$ contains all poles within $\{z : |z| \le \rho\}$ for some $\rho < 1$. Suppose we observe n samples

$$y_k = G^\star(e^{i2\pi k/n}) + w_k$$

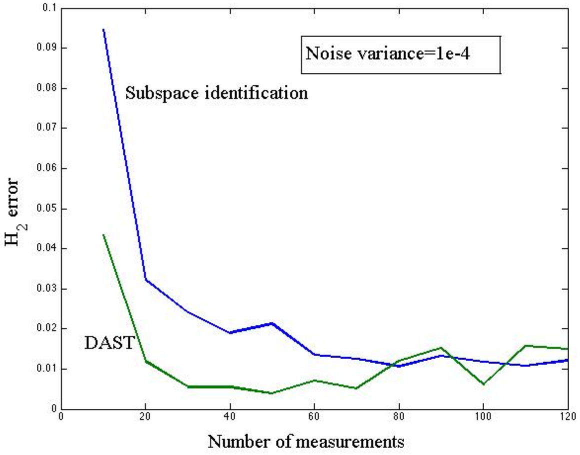
for $k = 0, \ldots, (n-1)$ where w is i.i.d. $\mathcal{N}(0, \sigma^2)$. There is a quantity C depending on ρ and σ such that for sufficiently large n

$$\left\|\hat{G}(z) - G_\star(z)\right\|_{\mathcal{H}_2}^2 \le C\|\Gamma_{G_\star}\|_1 n^{-\frac{1}{2}}$$

with probability exceeding $1 - e^{-o(n)}$.

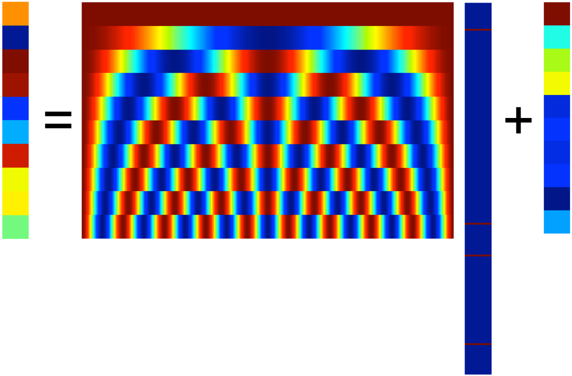
Experiments

- Compares favorably to Subspace ID for system identification.
- Does not need an estimate of model order.



Compressed Sensing off the Grid

Gongguo Tang, B., Parikshit Shah, Ben Recht

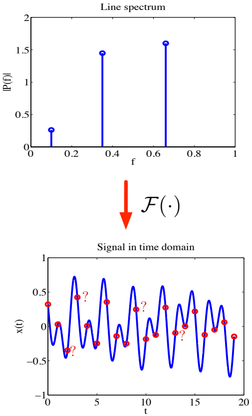


- Recover sparse vectors from a few incoherent measurements.
- CS Problem: s sparse DFT from $O(s \log(n))$ samples.
- *Continous Case*: What if frequencies don't lie on a grid?

Theorem (Gongguo,B,Parikshit,Ben)

Let Δ be the minimum frequency separation in the signal. Exact recovery is possible with just $s\,\mathrm{polylog}(1/\Delta)$ random time samples by atomic norm minimization

- PRONY 1795: $2s$ samples sufficient. noise sensitive
- CANDES, FERNANDES-GRANDA 2012 : If $n > 2s$, convex optimization can recover spectra with $\Delta > \frac{4}{n}$.
- OUR RESULT: $O(s\log(s)\log(n))$ **random** samples are sufficient for $\Delta > \frac{4}{n}$.
    - Missing Data Problem
    - Sub-Nyquist sampling



Observed

Figure: $\mathcal{A}_+ = (\cos(t), \cos(2t), \cos(3t))$

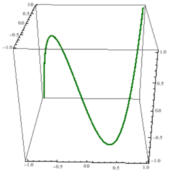


Figure: $\{\mathcal{A}_+, -\mathcal{A}_+\}$

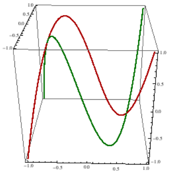


Figure: Convex Hull of $\mathcal{A}_+$

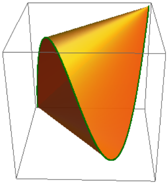
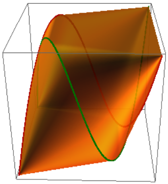


Figure: Convex Hull of $\{\mathcal{A}, -\mathcal{A}\}$



- Interpolate missing entries by

$$\begin{aligned} &\text{minimize}_x \, \|x\|_{\mathcal{A}} \\ &\text{subject to} \;\; x_j = x_j^\star, \;\; \text{for } j \in T. \end{aligned}$$

- Semidefinite formulation

$$\begin{aligned} &\text{minimize}_{u,t} \, \text{trace}(X) \\ &\text{subject to} \;\; X = \begin{pmatrix} Tu & x \\ x^* & t \end{pmatrix} \succeq 0 \\ &\qquad\qquad x_j = x_j^\star, \;\; \text{for } j \in T. \end{aligned}$$

- Relaxation of

$$\begin{aligned} &\text{minimize}_{u,t} \, \text{rank}(X) \\ &\text{subject to} \;\; X = \begin{pmatrix} Tu & x \\ x^* & t \end{pmatrix} \succeq 0 \\ &\qquad\qquad x_j = x_j^\star, \;\; \text{for } j \in T. \end{aligned}$$

Phase Transition Plots

Demixing

- Suppose x is sparse with respect to the atomic set $\mathcal{A}$ and y is sparse with respect to the atomic set $\mathcal{B}$
- We observe $z = x + y$.
- When does

$$\begin{gathered}\text{minimize}_{x,y} \|x\|_{\mathcal{A}} + \|y\|_{\mathcal{B}} \\ \text{subject to } z = x + y.\end{gathered}$$

succeed in recovering x and y?

Uncertainty Principles and Previous Work

- DONOHO AND STARK (1989) If a s sparse signal has a k sparse **DFT**, then

$$k + s \geq 2\sqrt{n}.$$

Uncertainty Principles and Previous Work

- Donoho and Stark (1989) If a s sparse signal has a k sparse **DFT**, then
$$k + s \geq 2\sqrt{n}.$$
- Donoho, Elad, Brucsktein (2002) Suppose x is s sparse, y has k sparse DFT, if
$$k + s \leq \frac{1 + \sqrt{n}}{2},$$
ℓ^1 minimization can separate $x + y$.

Uncertainty Principles and Previous Work

- Donoho and Stark (1989) If a s sparse signal has a k sparse **DFT**, then
$$k + s \geq 2\sqrt{n}.$$
- Donoho, Elad, Brucsktein (2002) Suppose x is s sparse, y has k sparse DFT, if
$$k + s \leq \frac{1+\sqrt{n}}{2},$$
ℓ^1 minimization can separate $x + y$.
- Candes and Romberg (2004) show that for most signals
$$k + s \sim \frac{n}{\sqrt{\log(n)}}$$
and in fact, when
$$k + s \leq C\frac{n}{\sqrt{\log(n)}}$$
ℓ^1 minimization succeeds with high probability.

Demixing Conjecture

Suppose we observe $z = x^\star + y^\star$, where $x^\star \in \mathbb{C}^n$ is a k-sparse combination of atoms from

$$\mathcal{A} = \left\{ [e^{i\phi} \ \cdots e^{i(2\pi(n-1)t+\phi)}]^T, \, t \in [0,1], \phi \in [0, 2\pi] \right\}.$$

with phases chosen uniformly randomly and $y^\star \in \mathbb{C}^n$ is a s-sparse vector. Then, $(x^\star, y^\star)$ is the unique optimal solution to demixing atomic norm minimization if the minimum separation between frequencies in $x^\star$ exceeds $\frac{1}{n}$ and

$$k + s \le \frac{Cn}{\sqrt{\log(n)}}$$

where C is some numerical constant.

Accelerated Convergence Rates

- Experiments suggest much faster rates of convergence than predicted when frequencies are far apart.
- Our bound is pessimistic – Phase transition plots suggest the presence of a fast regime
- We can recover support at least "approximately" when frequencies are well separated.
- Difficult to connect separation with the condition on descent cones.

Fast Rate Conjecture

Suppose we observe noisy samples $y = x^\star + w$ where $w \in \mathcal{N}(0, \sigma^2 I_n)$ is Gaussian noise and for $j \in \{0, \ldots, n-1\}$,

$$x_j^\star = \sum_{l=1}^{k} c_l e^{i2\pi j f_l}$$

where $c = [c_1 \ldots c_k]^T \in \mathbb{C}^k$ are unknown amplitudes, $[f_1 \ldots f_k] \in [0,1]^k$ are normalized frequencies satisfying the separation condition

$$\Delta_{min} = \min_{\substack{p \neq q \\ 1 \leq p,q \leq n}} (f_p - f_q) > \frac{1}{n},$$

the optimal solution $\hat{x}$ to the atomic soft thresholding problem satisfies

$$\mathbb{E} \left\| \hat{x} - x^\star \right\|_2^2 \leq \frac{Ck\sigma^2 \log(n)}{n}.$$

where C is a numerical constant. The same rate holds even if we observe a fraction $s \log^2(n)$ of entries according to a Bernoulli model, instead of all n samples.

Minimax Rates

- Need to understand the fundamental limits of estimation in different regimes.
- Worst case error (or adversarial error) for the best estimator
- Can be information theoretically determined
- Coincides with the "fast rates" for other high dimensional inference problems.

Information Theoretic Determination

Basic Approach:

- Recast into a multiple hypothesis problem.
- Pack as many hypotheses at a critical separation so they are nearly indistinguishable.
- Find a lower bound on the probability of error by Fano's inequality.
- Fano's inequality requires an estimate of mutual information. Find a suitable upper bound for mutual information by a cover of the model space.
- This method provides optimal minimax rates for a number of problems.

Minimax Rate Conjecture for Line Spectrum Estimation

- Let Δ be the minimum separation between the normalized frequencies.
- Define the minimax rate in terms of the mean square error is given by

$$M(\Delta, k) = \min_{\hat{x}} \max_{x^\star \in \mathcal{X}} \frac{1}{n} \mathbb{E} \left\| x^\star - \hat{x} \right\|_2^2 .$$

- *Conjecture*: If $\Delta > \frac{1}{n}$,

$$M(\Delta, k) \geq \frac{C k \sigma^2 \log(n)}{n}$$

  where C is independent of k, σ and n.

Scaling for Big Data

- Examples
  1. Heart rate and gas exchange data
  2. Time Series forecasting
  3. Neuronal spikes from Physiological Data
- Large scale problems - Challenges
- Parallel and Distributed Algorithms
- First order methods?
- Active set methods?

Trigonometric Moments

Theorem (Herglotz)

A sequence of numbers $\{x_k\}_{k=-\infty}^{\infty}$ form the trigonometric moments of a positive measure if and only if the sequence is positive definite. In other words, $Tx \succeq 0$ for every n and the sequence is Hermitian symmetric ($x_{-k} = x_k^*$).

Theorem (Caratheodory, Toeplitz, Fejer (1917))

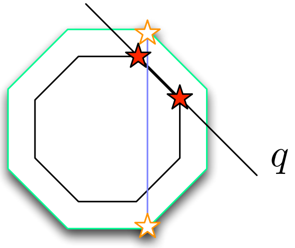
Suppose $x \in C^n$ such that $Tx \succeq 0$. Then, there exists r frequencies $f_1, \ldots, f_r \in [0, 1]$ and positive numbers $c_1, \ldots, c_r$ such that

$$x_k = \sum_{j=1}^{r} c_j \exp(i2\pi f_j k) \text{ for each } k = 0, \ldots, (n-1)$$

$$r = \operatorname{rank}(Tx)$$

Exact Recovery

- Guarantee that a given decomposition has minimum atomic norm amongst all decompositions.
- Show it lies on a simplicial, exposed face of $\text{conv}(\mathcal{A})$



Dual Certificate for Line Spectral Estimation

Definition (Dual Certificate)

A dual certificate is a q satisfying

- $\langle q, a(f_j) \rangle = \operatorname{sign}(c_j)$, for $j = 1, \ldots, k$.
- $\langle q, a(f) \rangle < 1$ if $f \notin \{f_1, \ldots, f_k\}$.

Dual Polynomial

Note that

$$\langle q, a(f) \rangle = \sum_{k=1}^{n} q_k e^{i2\pi k f} = Q(f)$$

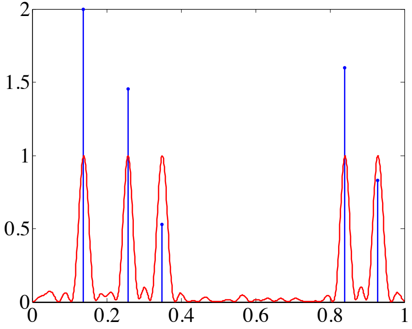
is a trigonometric polynomial. So, we identify the "dual vector" with the "dual polynomial"

Dual Certificate for Line Spectrum Estimation

Necessary conditions for q to be a dual certificate:

$$Q(f_k) = \mathrm{sign}(c_k) \qquad \text{for each } k = 1, \ldots, s$$
$$Q'(f_k) = 0 \qquad \text{for each } k = 1, \ldots, s$$

Figure: Dual Polynomial and Recovered Frequencies



Result of Candes and Fernandes-Granda

Candes and Fernandes-Granda (2012) propose writing

$$Q(f) = \sum_j \alpha_j K(f - f_j) + \beta_j K'(f - f_j)$$

where $K(f)$ is a trigonometric kernel and $\alpha \in \mathbb{C}^k$ and $\beta \in \mathbb{C}^k$ are undetermined coefficients.

Minimum separation:

$$\Delta = \min_{j \neq k} |f_j - f_k| \geq \Delta_{\min} = \frac{2}{n}.$$

If $\Delta \geq \frac{2}{n}$, convex optimization can exactly recover the correct decomposition from $n > 2s$ Nyquist samples.

- Prony's technique does not have a resolution limitation.
- Positive case doesn't either

Convex Hull of $\{\mathcal{A}, -\mathcal{A}\}$ – Insight

Figure: Moment Curve $\mathcal{A}$ and $-\mathcal{A}$

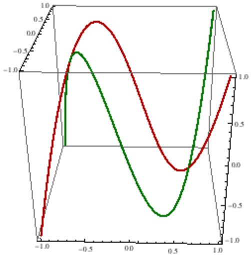
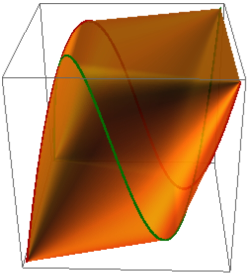


Figure: Convex Hull of $\{\mathcal{A}, -\mathcal{A}\}$



Compressed Sensing

Recall that the trigonometric moment sequence is given by

$$x_j^\star = \sum_{k=1}^{s} c_k^\star \exp(i2\pi u_k^\star j)$$

- Candes and Fernandes-Granda (2012): Explicit dual construction for exact recovery from $x_0, \ldots, x_{n-1}$ if frequencies are at least $4/n$ apart.
- *Sampling Model*: With a probability $p = m/n$, observe each entry independently, so we observe m entries on an average out of n. Let $T \subset \{0, \ldots, n-1\}$ index the observed entries.
- *Uniform Phases*: Assume that the phases of c_k are uniformly distributed in $[0, 2\pi]$.
- Interpolate the missing entries by

$$\begin{aligned} &\text{minimize}_x \; \|x\|_{\mathcal{A}} \\ &\text{subject to} \;\; x_j = x_j^\star, \;\; \text{for } j \in T. \end{aligned}$$

Dual Certificate for missing data

Definition (Dual Certificate)

A dual certificate q must satisfy

- $\langle q, a(f_j) \rangle = \text{sign}(c_j)$, for $j = 1, \ldots, k$.
- $\langle q, a(f) \rangle < 1$ if $f \notin \{f_1, \ldots, f_k\}$.
- $q_{T^c} = 0$.

- Define a random kernel
$$K(f) = \sum_{j=-(n-1)}^{n-1} u_j \delta_j \exp(i2\pi j f)$$
where the Bernoulli variable δ_j only sums observed entries in place of the deterministic kernel
$$\bar{K}(f) = \sum_{j=-(n-1)}^{n-1} u_j \exp(i2\pi j f)$$

- Proof relies on Matrix concentration, polynomial Bernstein inequality and proof techniques in Romberg et al.