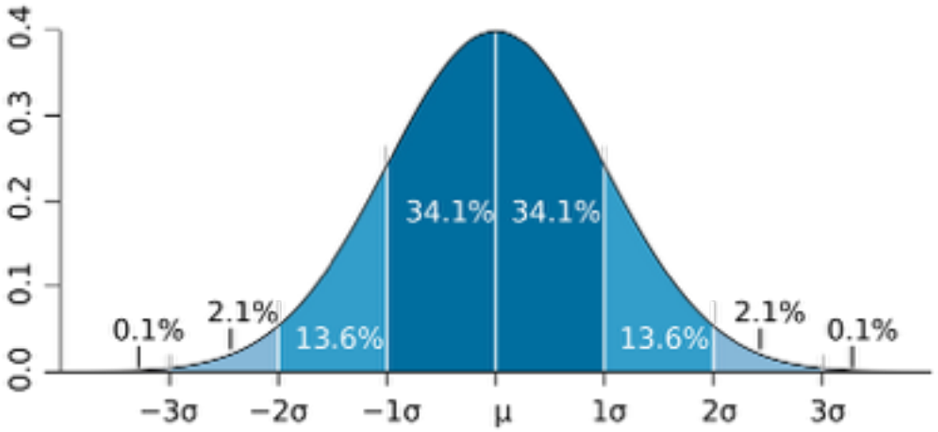
Decomposition and Denoising for moment sequences using convex optimization

Badri Narayan Bhaskar

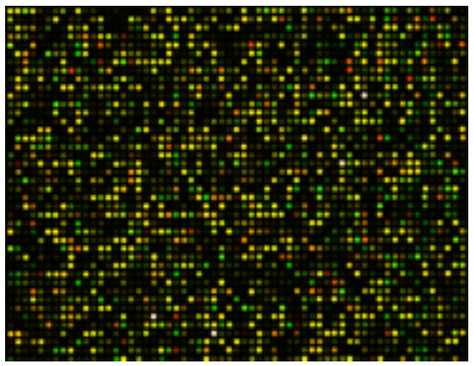
Advisor: Prof. Benjamin Recht

High Dimensional Inference

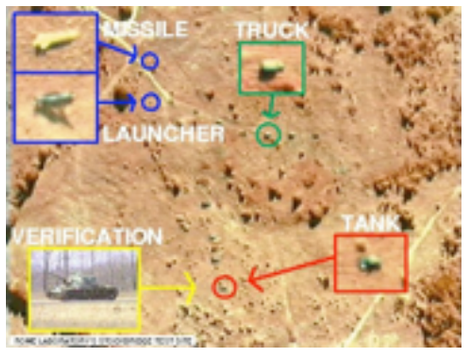


$$n \gg p$$

Statistics



Microarray Data



Hyperspectral Imaging

NETFLIX

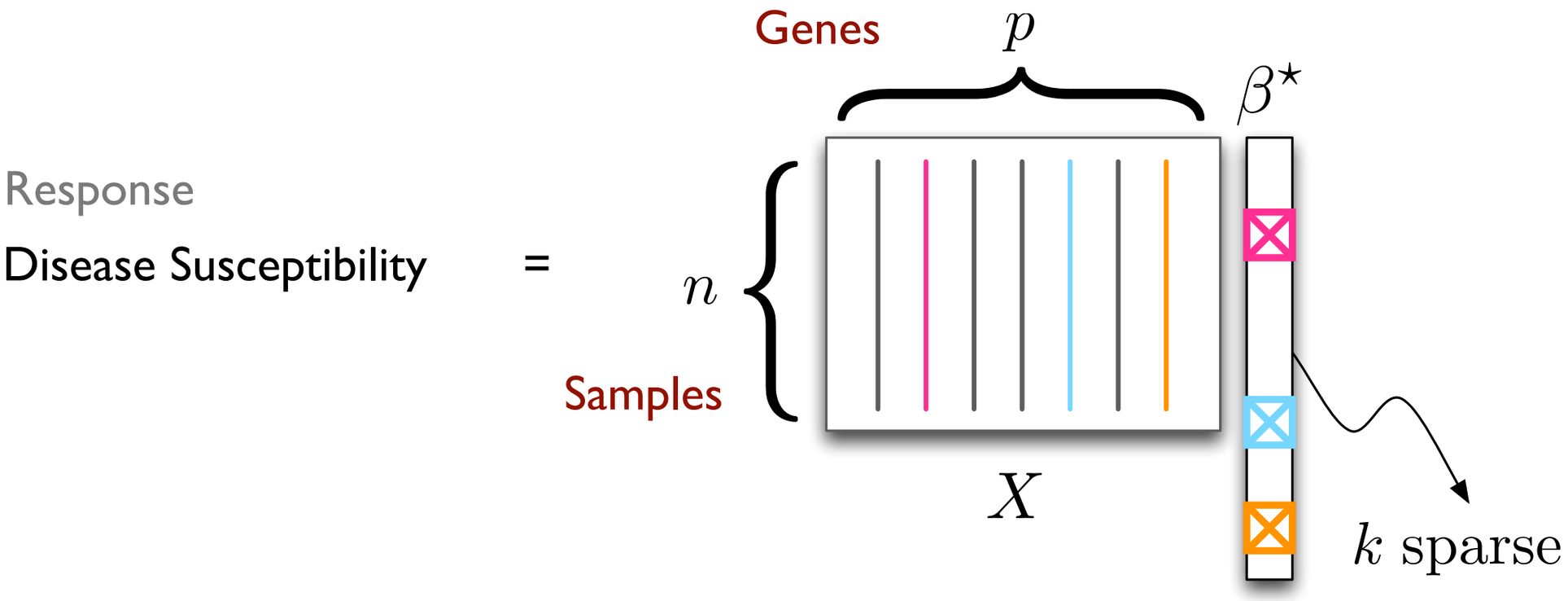
Predict Ratings

Sparse Vector Estimation

Gene Expressions

Response

Disease Susceptibility =



Few Active Genes

Low Rank Matrices

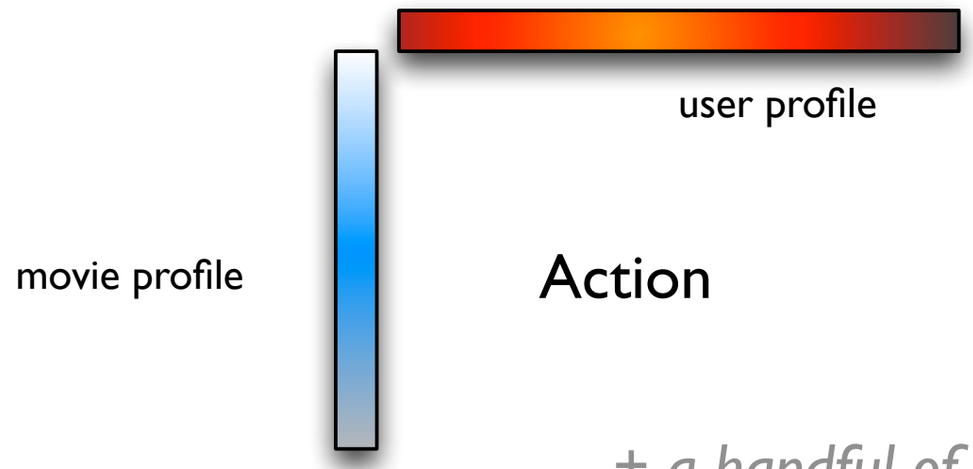
The Netflix Problem

millions of subscribers

thousands of movies

Movie Ratings

Decomposition:



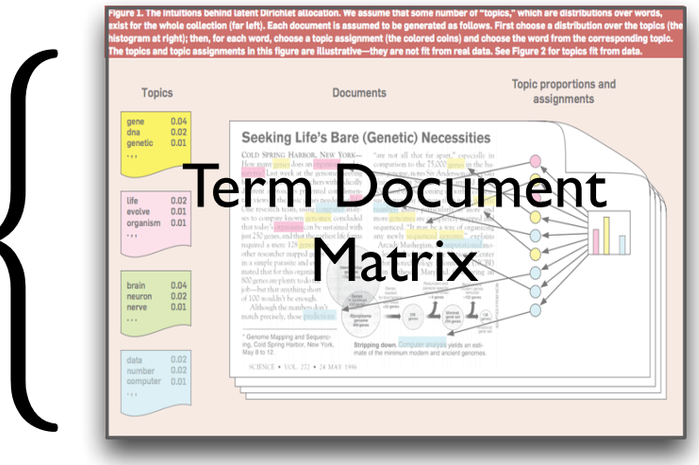
+ a handful of other features...

Low Rank Matrices

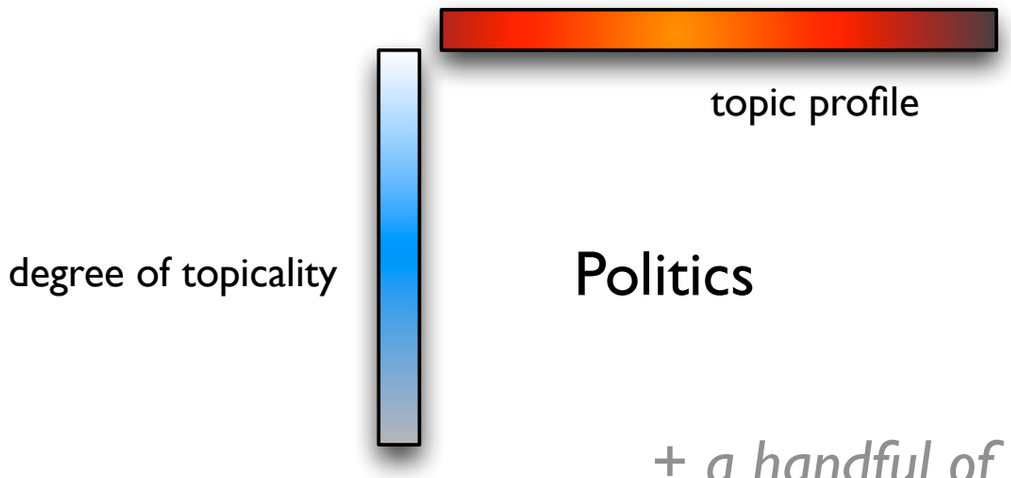
Topic Models

millions of words

thousands of documents

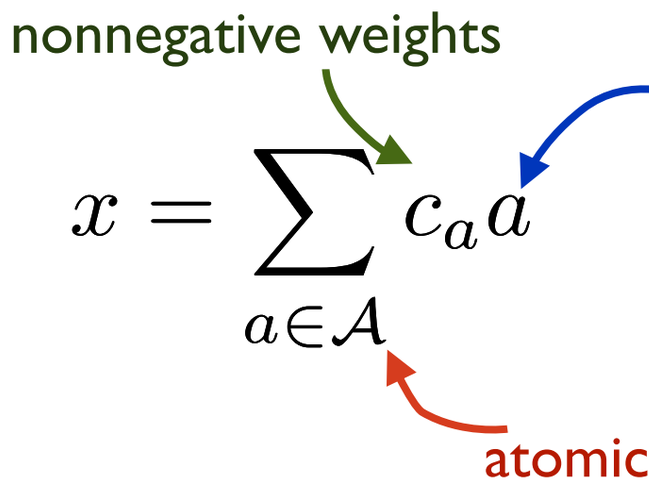


Decomposition:



+ a handful of other features...

Simple objects



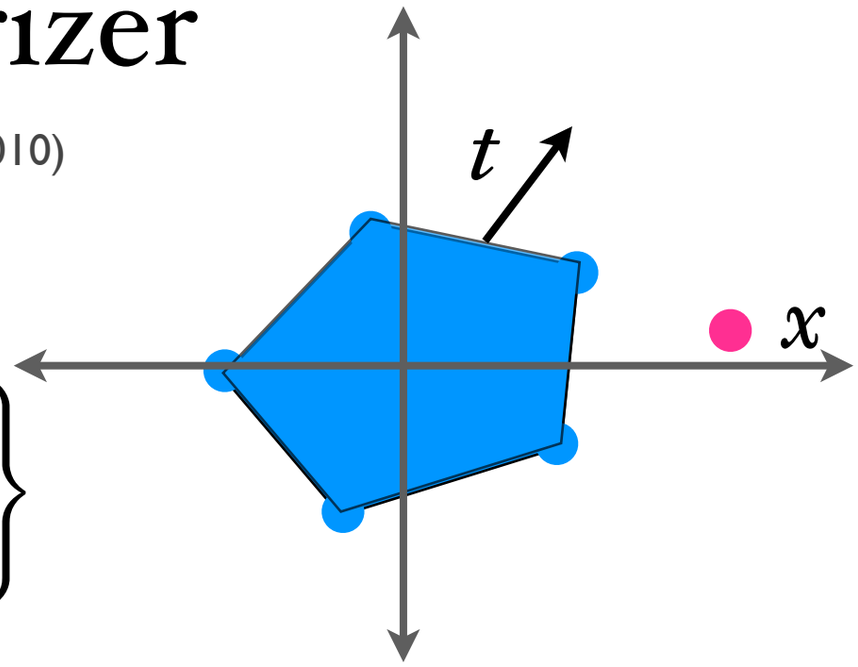
Objects	Notion of Simplicity	Atoms
Vectors	Sparsity	Canonical unit vectors
Matrices	Rank	Rank-1 matrices
Bandlimited signals	No. of frequencies	Complex sinusoids
Linear systems	Number of poles	Single pole systems

Universal Regularizer

Chandrasekaran, Recht, Parillo, Wilsky (2010)

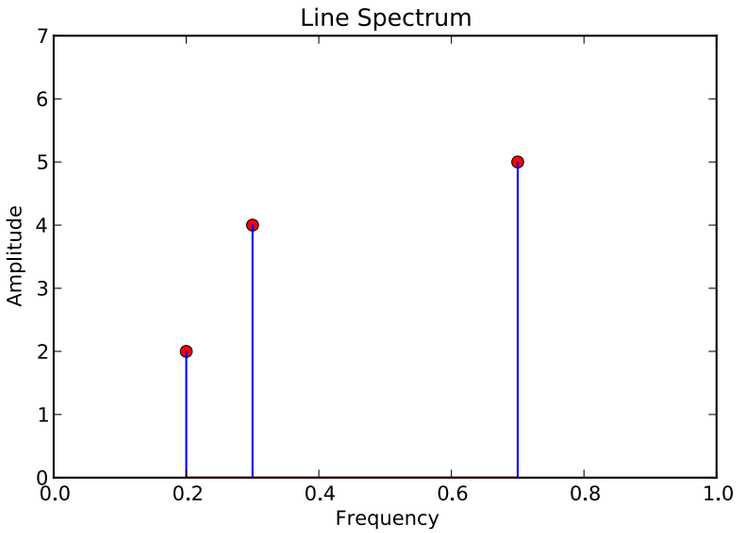
$$\|x\|_{\mathcal{A}} = \inf\{t \geq 0 : x \in t\,\mathbf{conv}(\mathcal{A})\}$$

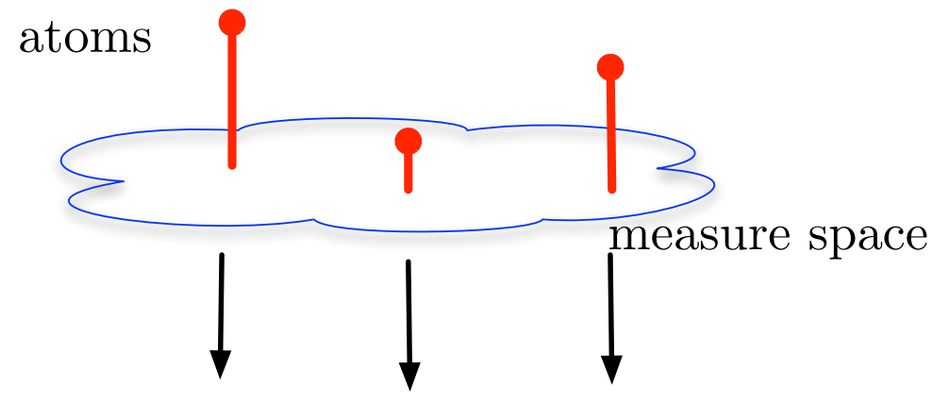
$$= \inf\left\{\sum_a c_a : x = \sum_{a \in \mathcal{A}} c_a a, c_a > 0\right\}$$

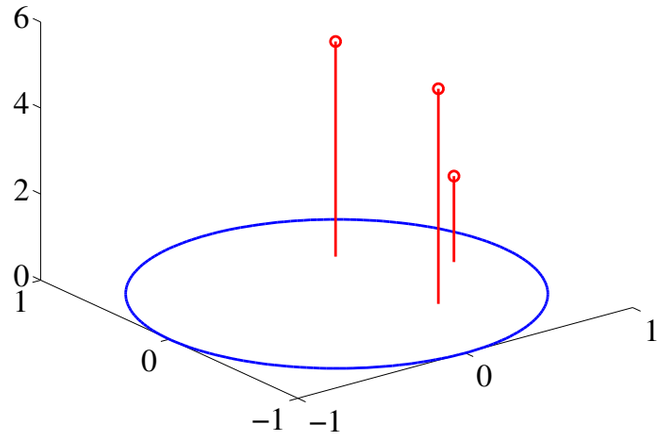


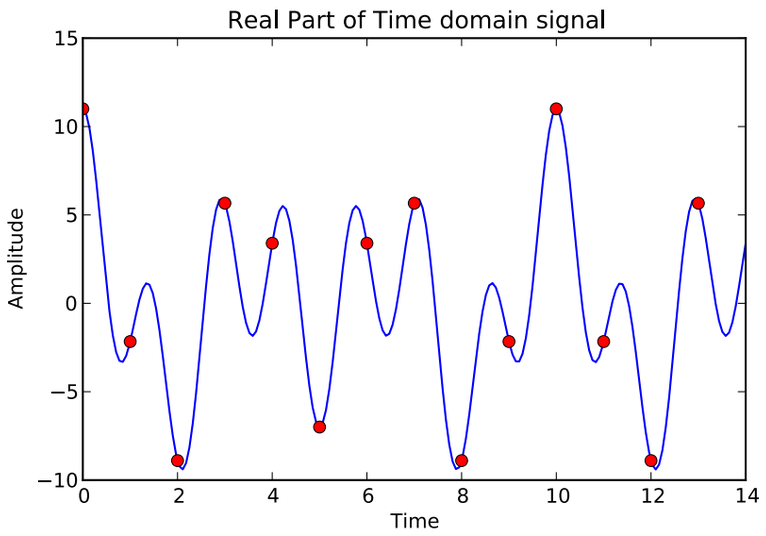
Objects	Notion of Simplicity	Atomic Norm
Vectors	Sparsity	ℓ_1 norm
Matrices	Rank	nuclear norm
Bandlimited signals	No. of frequencies	?
Linear systems	McMillan degree	?

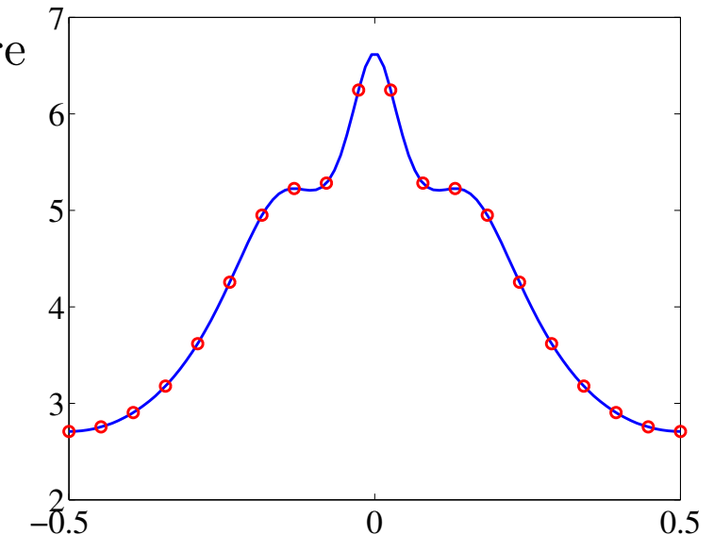
Moment Problems







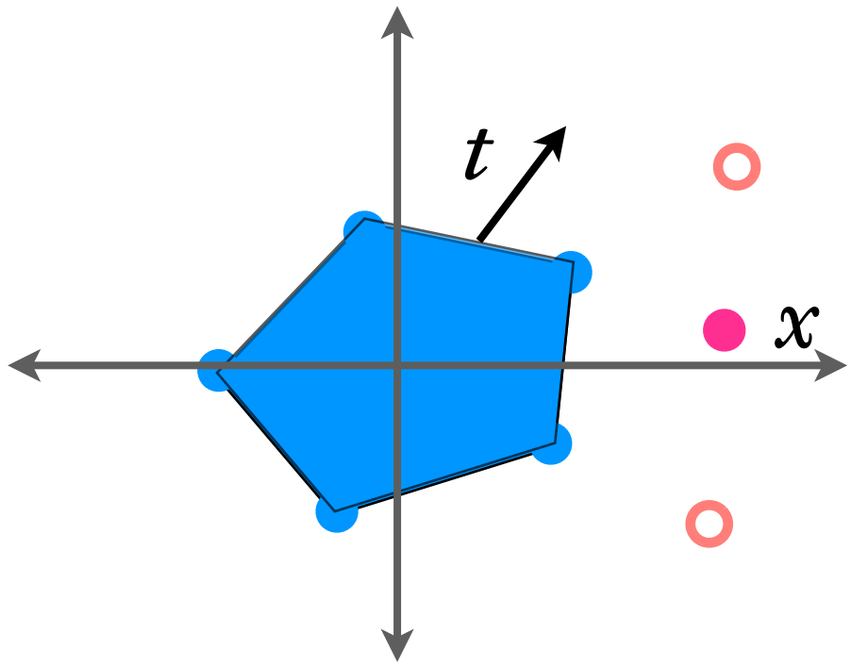




Line Spectral Estimation

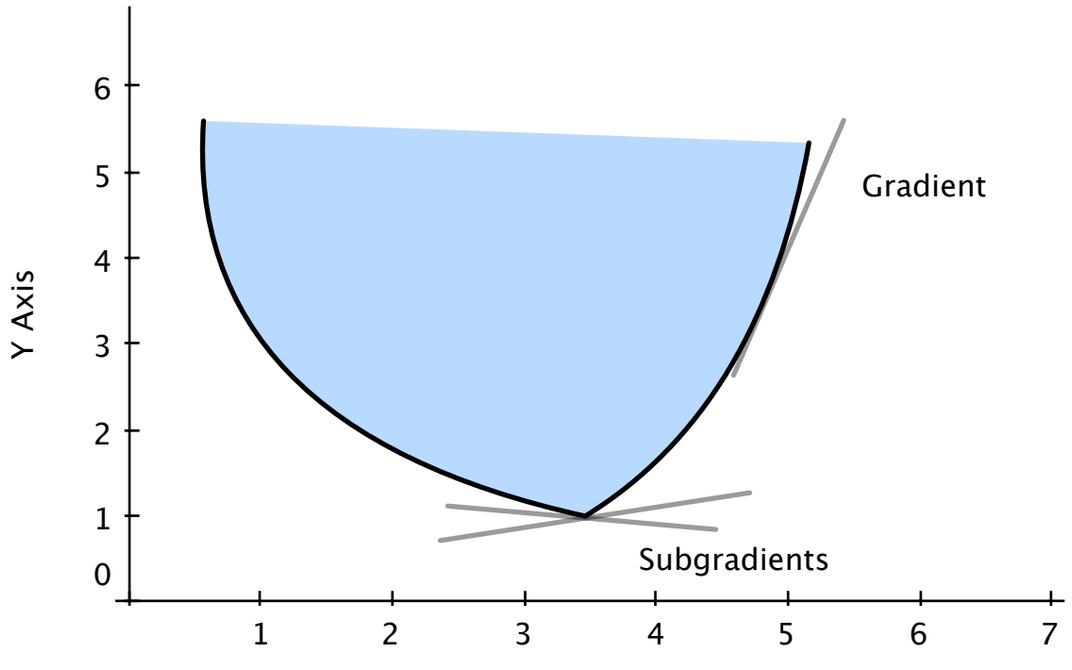
System Identification

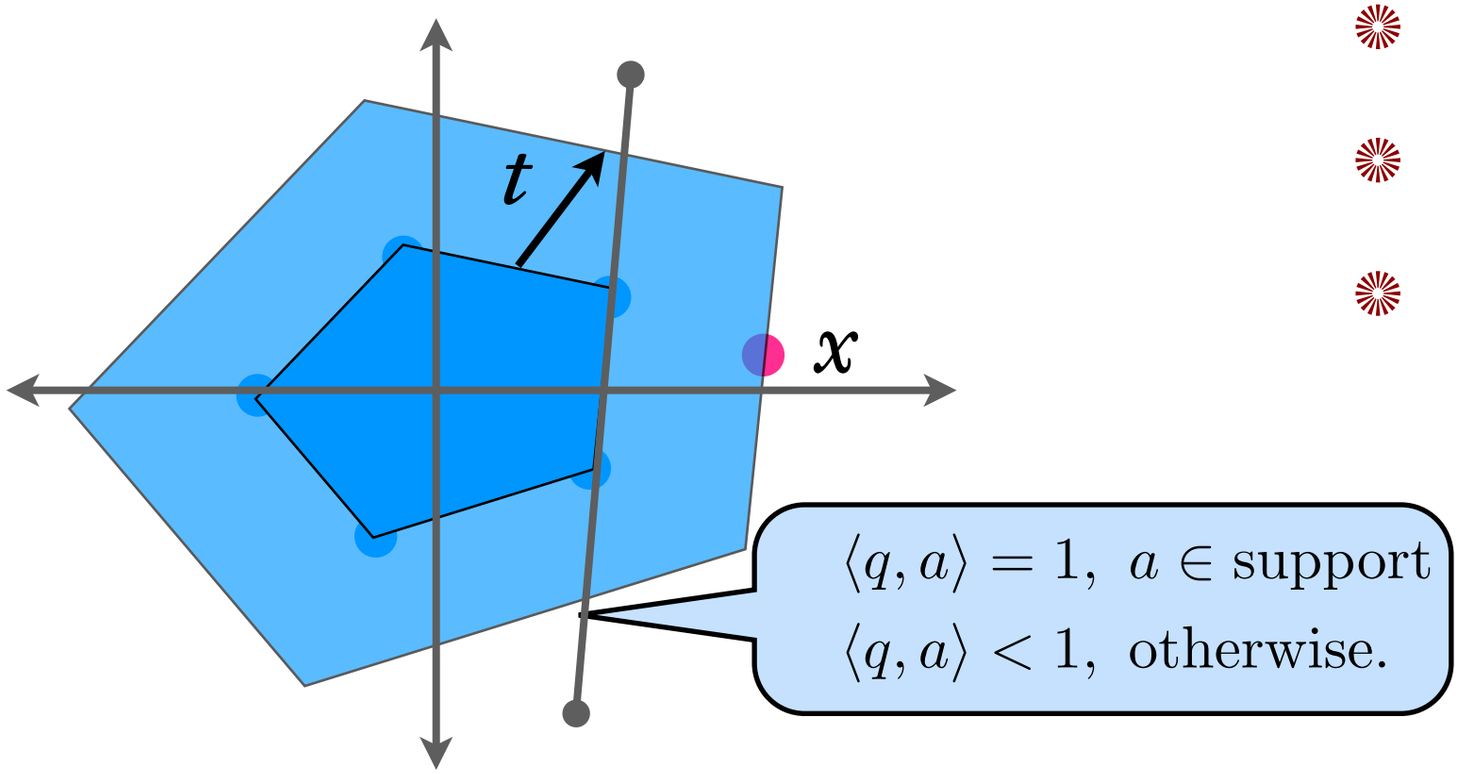
Decomposition



- How many atoms do you need to represent x?
- *composing atoms are on an exposed face*: decomposition is easy to find.
- find a supporting hyperplane to certify
- concentrate on recovering "good" objects with atoms on exposed faces

Dual Certificate





Dual Characterization of atomic norm

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

- optimum value is atomic norm.
- solutions are subgradients at x
- semi-infinite problem

$$\|q\|_{\mathcal{A}}^* = \sup_{a \in \mathcal{A}} \langle q, a\rangle$$

(under mild conditions, Bonsall)

Denoising

$$y = x^\star + w$$

$x^\star$: in $\mathbb{C}^n$

w: noise $\mathcal{N}(0, \sigma^2 I_n)$

AST (Atomic Soft Thresholding)

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

τ: Tradeoff parameter

Dual AST

$$\begin{aligned} &\underset{q}{\text{maximize}} \;\; \langle q, y \rangle - \tau \|q\|_2^2 \\ &\text{subject to } \|q\|_{\mathcal{A}}^* \leq 1 \end{aligned}$$

AST Results

Optimality Conditions

$$\tau\hat{q} + \hat{x} = y$$

$$\|\hat{q}\|_{\mathcal{A}}^{*} \leq 1$$

$$\langle \hat{q}, \hat{x} \rangle = \|\hat{x}\|_{\mathcal{A}}$$

means

$$\hat{q} \in \partial \|\hat{x}\|_{\mathcal{A}}$$

Can find composing atoms!

suggests

$$\tau\hat{q} \approx w \Rightarrow \tau \approx \mathbb{E}\|w\|_{\mathcal{A}}^{*}$$

Tradeoff = Gaussian Width

- No cross validation needed - estimate Gaussian width
- Gaussian width also important to characterize convergence rates

AST Convergence Rates

Choose appropriate tradeoff

$$\tau = \eta \mathbb{E}\left(\|w\|_{\mathcal{A}}^*\right), \eta \geq 1$$

Minimal assumptions on atoms

Universal, but slow rate

$$\boxed{\frac{1}{n}\|\hat{x} - x^\star\|_2^2 \leq \frac{\tau}{n}\|x^\star\|_{\mathcal{A}}}$$

With a "cone condition", faster rates are possible (for large enough η)

Weaker than RIP, Coherence, RSC, RE

$$\boxed{\frac{1}{n}\|\hat{x} - x^\star\|_2^2 \leq C\frac{\tau^2}{\phi^{1/2}n}}$$

Lasso

Slow Rate	Fast Rate
$\frac{1}{n}\|X\hat{\beta} - X\beta^\star\|_2^2 \leq \sigma\sqrt{\frac{\log(p)}{n}}\|\beta^\star\|_1$	$\frac{1}{n}\|X\hat{\beta} - X\beta^\star\|_2^2 \leq C\sigma^2\frac{s\log(p)}{n}$

$\frac{\tau}{n}$

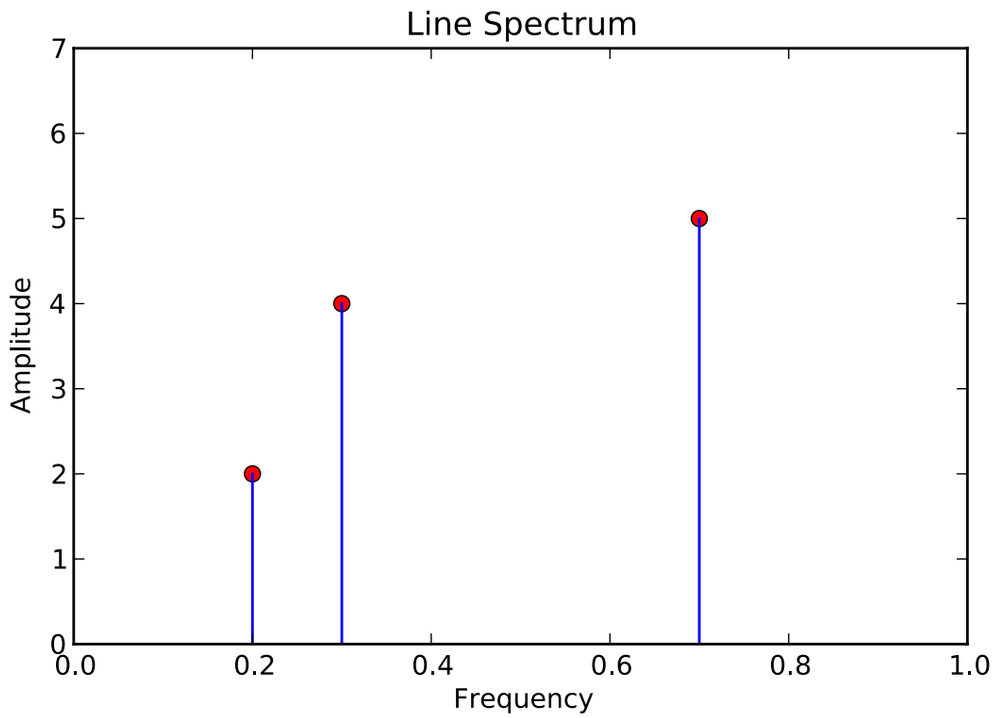
s sparse

Summary

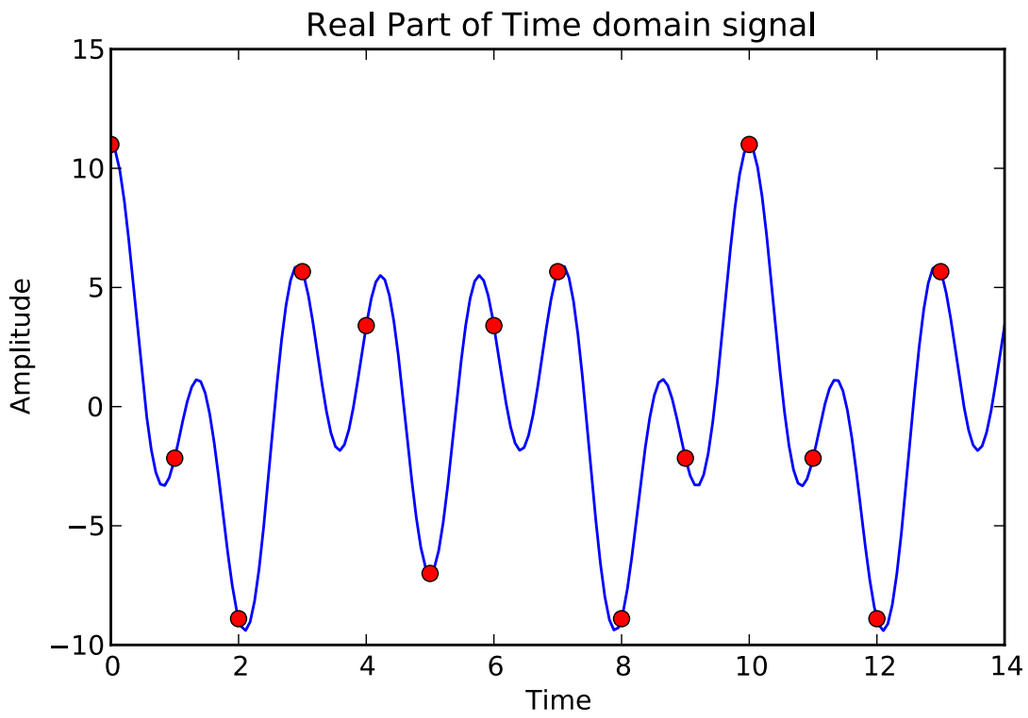
- Notion of simple models
- Atomic norm unifies treatment of many models
- Recovers tight published bounds
- Bounds depend on Gaussian width
- Faster rates with a local cone condition

Line Spectral Estimation

$$\mu = \sum_j c_j \delta_{f_j} \longleftrightarrow \begin{bmatrix} x_0^\star \\ \vdots \\ x_k^\star \\ \vdots \\ x_{n-1}^\star \end{bmatrix} = \sum_{j=1}^{s} c_j e^{i2\pi f_j k}$$



Line Spectrum



Time Samples

Extrapolate the remaining moments!

Super resolution

Time frequency dual

- Extrapolate high frequency information from low frequency samples
- Just exchange time and frequency: same as line spectral estimation

Classical Line Spectrum Estimation

(a crash course)

Nonlinear Parameter Estimation

$$x_k^\star = \sum_{j=1}^{s} c_j e^{2\pi i f_j k}$$

Define for any vector x,

$$T_n(x) = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ x_2^* & x_1 & \dots & x_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^* & x_{n-1}^* & \dots & x_1 \end{bmatrix}$$

FACT:

$$T_n(x^\star) = \sum_{j=1}^{s} c_j^\star \begin{bmatrix} 1 \\ e^{i2\pi f_j} \\ \vdots \\ e^{i2\pi(n-1)f_j} \end{bmatrix} \begin{bmatrix} 1 & e^{-i2\pi f_j} & \dots & e^{-i2\pi(n-1)f_j} \end{bmatrix}$$

Low rank and Toeplitz

Classical techniques

FACT:

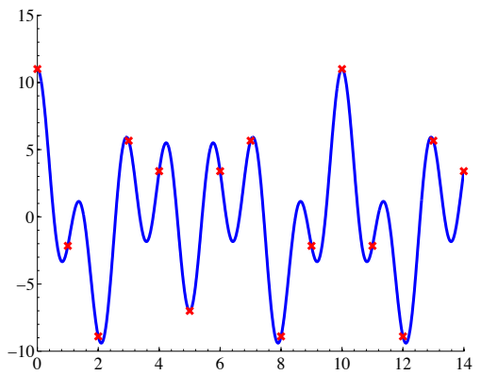
$$T_n(x^\star) = \sum_{j=1}^{s} c_j^\star \begin{bmatrix} 1 \\ e^{i2\pi f_j} \\ \vdots \\ e^{i2\pi(n-1)f_j} \end{bmatrix} \begin{bmatrix} 1 & e^{-i2\pi f_j} & \dots & e^{-i2\pi(n-1)f_j} \end{bmatrix}$$

Low rank and Toeplitz

- **PRONY** estimate s, root finding
- **CADZOW** alternating projections
- **MUSIC** low rank, plot pseudospectrum

- Sensitive to model order
- No rigorous theory
- Only optimal asymptotically

Convex Perspective



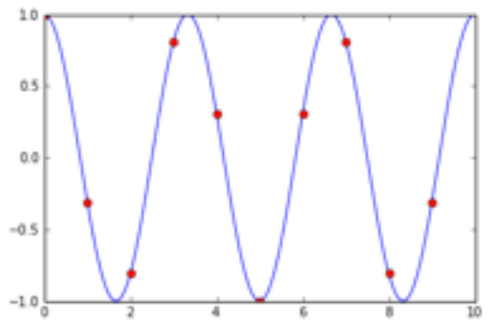
$$\begin{bmatrix} x_0^\star \\ \vdots \\ x_k^\star \\ \vdots \\ x_{n-1}^\star \end{bmatrix} = \sum_{j=1}^{s} c_j e^{i2\pi f_j k}$$

Line Spectrum

frequencies

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

amplitudes



$$\begin{bmatrix} 1 \\ \vdots \\ e^{i2\pi f_j k} \\ \vdots \\ e^{i2\pi f_j (n-1)} \end{bmatrix}$$

Atoms

Positive Amplitudes: $\mathcal{A}_+ = \{a(f) \mid f \in [0,1]\}$
Complex: $\mathcal{A} = \left\{a(f)e^{i\phi} \;\middle|\; f \in [0,1], \phi \in [0,2\pi]\right\}$

Convergence rates

Choose tradeoff as $\tau \approx \sigma\sqrt{n\log(n)}$

With high probability, for n>4,

$$\sigma\sqrt{n\log(n) - \tfrac{n}{2}\log(4\pi\log(n))} \leq \|w\|_{\mathcal{A}}^{*} \leq \sigma\left(1 + \frac{1}{\log(n)}\right)\sqrt{n\log(n) + n\log(16\pi^3/2\log(n))}$$

Dudley's inequality
Bernstein's polynomial theorem

For a signal with n samples of the form,

$$y_k = \sum_{j=1}^{s} c_j e^{2\pi i f_j k} + w_k$$

Using the global AST guarantee, we get,

$$\mathbb{E}\left(\frac{1}{n}\|\hat{x} - x^{\star}\|_2^2\right) \leq C\sigma\sqrt{\frac{\log(n)}{n}}\sum_{j=1}^{s}|c_j|$$

Fast Rates

If every two frequencies are far enough apart (minimum separation condition)

$$\min_{p \neq q} d(u_p, u_q) \geq \frac{4}{n}$$

and the tradeoff parameter is chosen correctly,

$$\tau = \eta\sigma\sqrt{n\log(n)}, \eta \in (1, \infty) \text{ large enough.}$$

For s frequencies and n samples, MSE is

$$\mathbb{E}\left(\frac{1}{n}\|\hat{x} - x^\star\|_2^2\right) \leq C\sigma^2 s\frac{\log(n)}{n}$$

- Not a global condition on the dictionary, local to the signal
- Proof uses many properties of trigonometric polynomials and Fejer kernels derived by Candes and Fernandez Granda.

Can't do much better

Our rate:

$$\mathbb{E}\left(\frac{1}{n}\|\hat{x} - x^\star\|_2^2\right) \leq C\sigma^2 s\frac{\log(n)}{n}$$

Only logarithmic factor from "**oracle rate**"

$$\mathbb{E}\left(\frac{1}{n}\|\hat{x} - x^\star\|_2^2\right) \leq C\sigma^2\frac{s}{n}$$

Also nearly matches **minimax bound** on best prediction rate for signals with 4/n minimum separation.

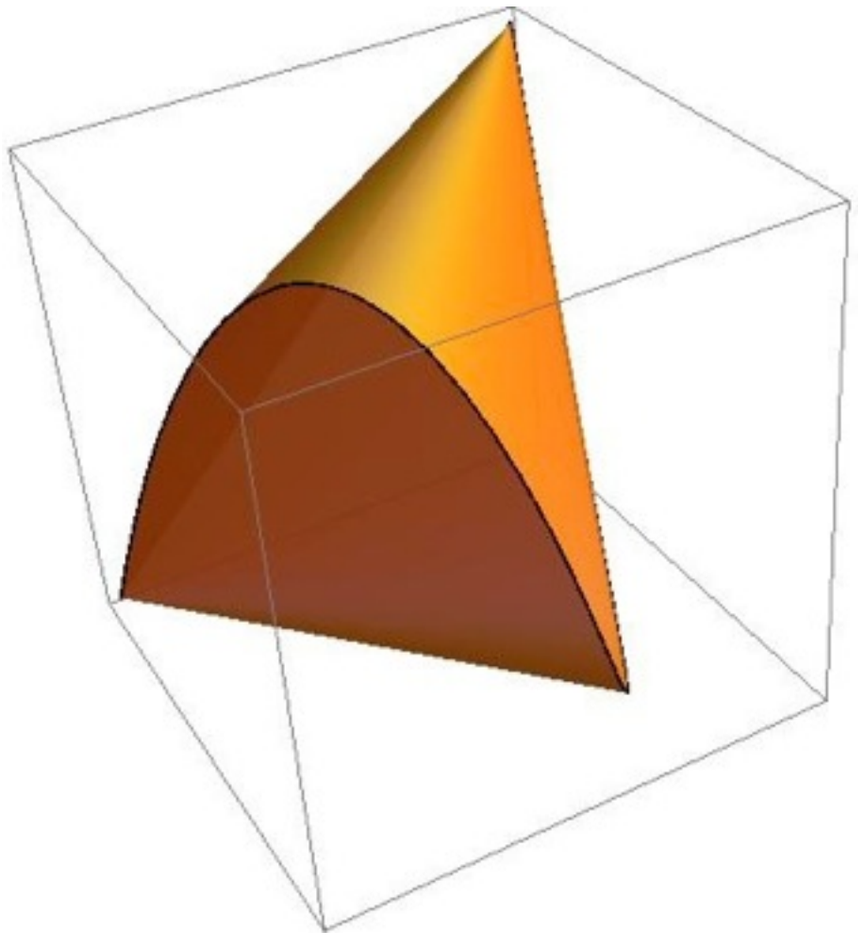
$$\mathbb{E}\left(\frac{1}{n}\|\hat{x} - x^\star\|_2^2\right) \geq C\sigma^2\frac{s\log(n/4s)}{n}$$

Minimum Separation and Exact Recovery

Result of Candes and Fernandez Granda

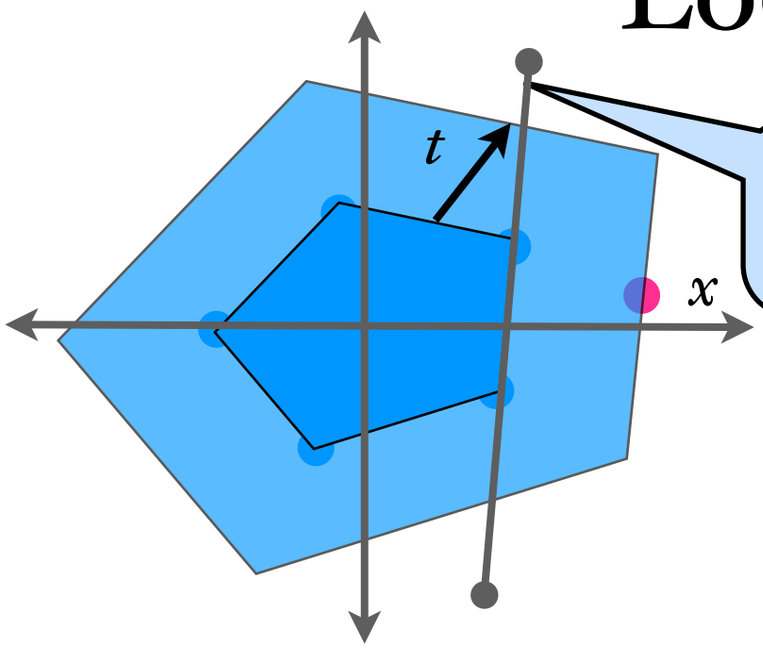
- 4/n separation = can construct explicit dual certificate of exact recovery.
- Convex hull of far enough vertices are exposed faces.
- Prony does not have this limitation. (Then why bother? Robustness guarantee)
- Minimum separation is necessary

except for the positive case...



- Convex hull of every collection of n/2 vertices is an exposed face
- Infinite dimensional version of cyclic polytope, which is neighbourly
- No separation condition needed.

Localizing Frequencies

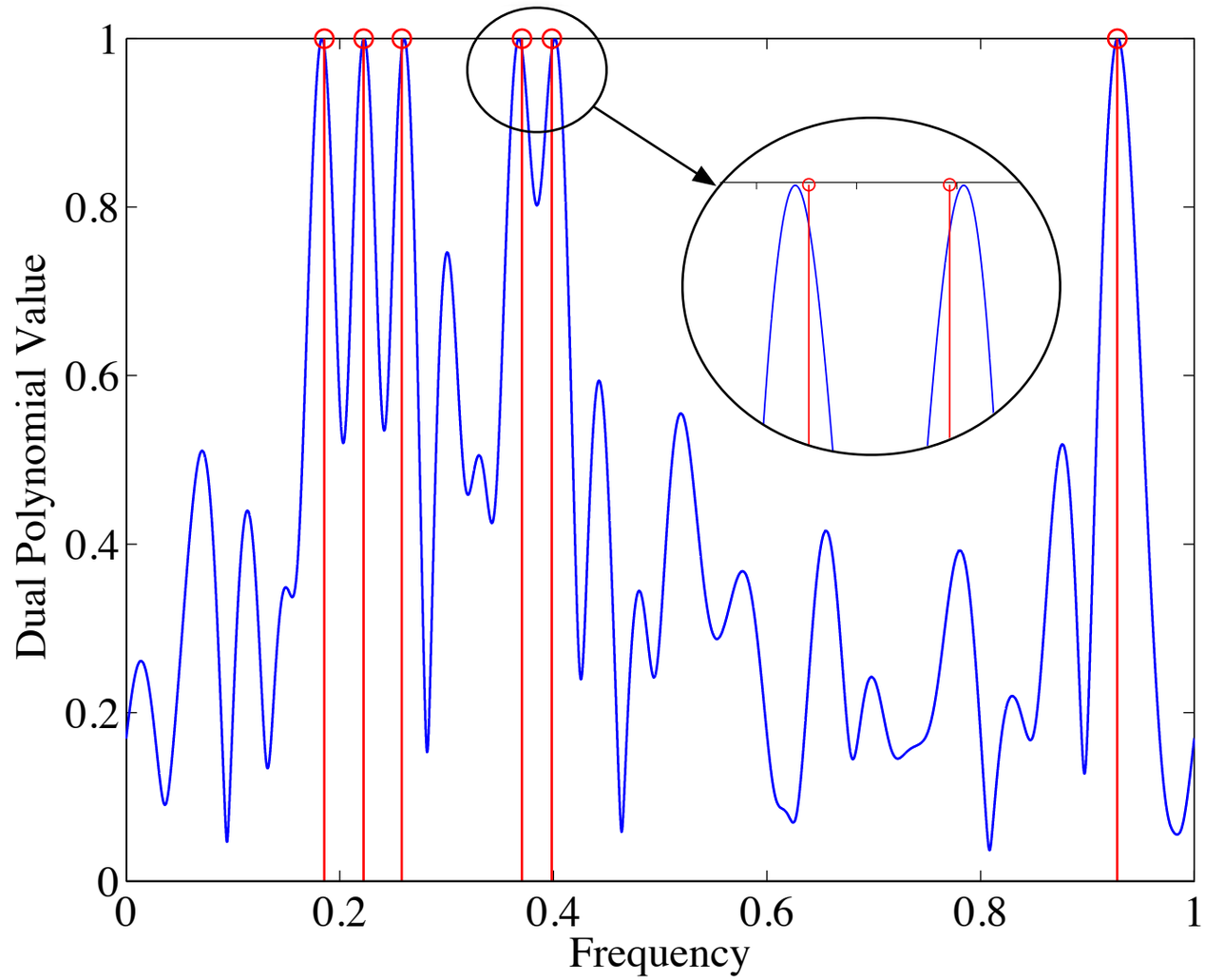


$$\langle q, a \rangle = 1, \ a \in \text{support}$$
$$\langle q, a \rangle < 1, \ \text{otherwise.}$$

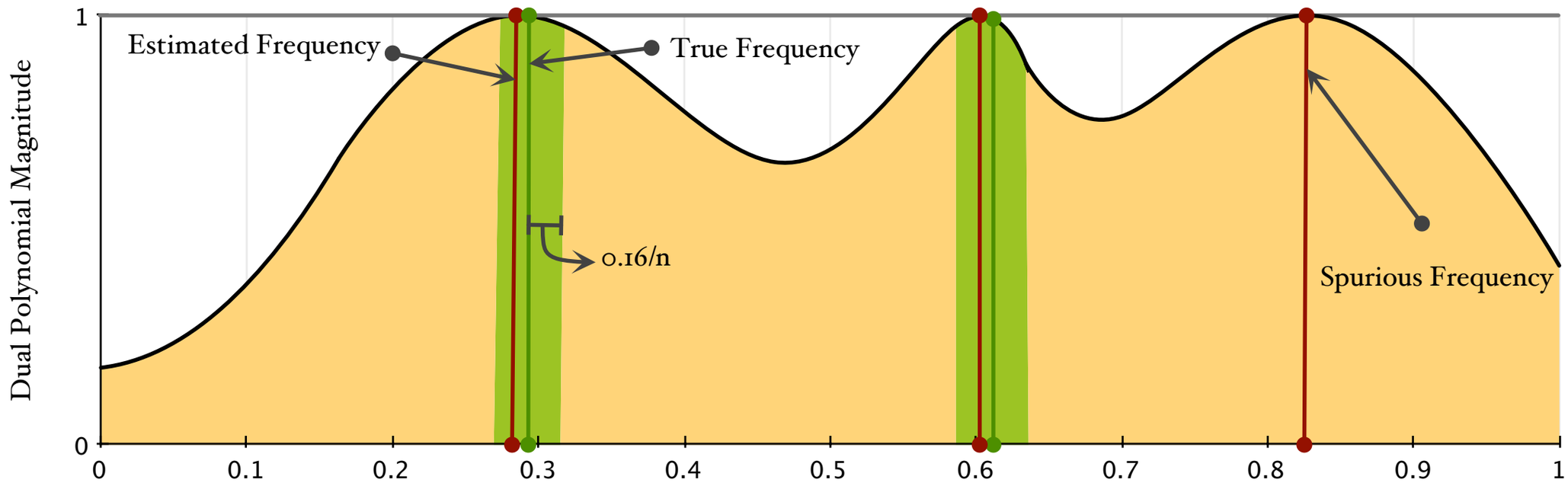
$$y = \hat{x} + \tau \hat{q}$$

Dual Polynomial

$$\hat{Q}(f) = \sum_{j=0}^{n-1} \hat{q}_j e^{-i2\pi j f}$$



Localization Guarantees



$$\sum_{l:\hat{f}_l \in F} |\hat{c}_l| \leq C_1 \sigma \sqrt{\frac{k^2 \log(n)}{n}}$$

Spurious Amplitudes

Weighted Frequency Deviation

$$\sum_{l:\hat{f}_l \in N_j} |\hat{c}_l| \left\{ \min_{f_j \in T} d(f_j, \hat{f}_l) \right\}^2 \leq C_2 \sigma \sqrt{\frac{k^2 \log(n)}{n}}$$

$$\left| c_j - \sum_{l:\hat{f}_l \in N_j} \hat{c}_l \right| \leq C_3 \sigma \sqrt{\frac{k^2 \log(n)}{n}}$$

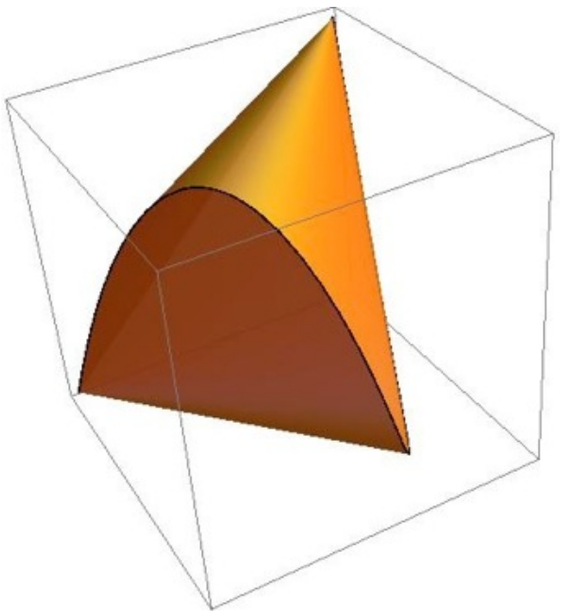
Near region approximation

How to actually solve AST

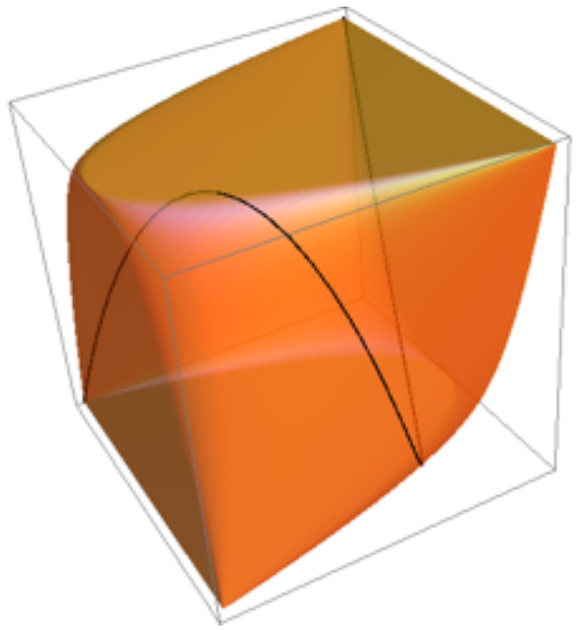
Semidefinite Characterizations

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

$\text{conv}(\mathcal{A}_+)$



$\text{conv}(\mathcal{A})$



Positive Amplitudes: $\mathcal{A}_+ = \{a(f) \mid f \in [0,1]\}$

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

$$\|x\|_{\mathcal{A}_+} = \begin{cases} x_0, & T_n(x) \succeq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

$$\|x\|_{\mathcal{A}} = \inf\left\{ \frac{1}{2n}\operatorname{tr}(T_n(u)) + \frac{1}{2}t \;\middle|\; \begin{bmatrix} T_n(u) & x \\ x^* & t \end{bmatrix} \succeq 0 \right\}$$

Bounded Real Lemma
Caratheodory Toeplitz

Alternating Direction Method of Multipliers

AST SDP Formulation

$$\begin{aligned} \text{minimize}_{t,u,x,Z} \quad & \tfrac{1}{2}\|x-y\|_2^2 + \tfrac{\tau}{2}(t+u_1) \\ \text{subject to} \quad & Z = \begin{bmatrix} T(u) & x \\ x^* & t \end{bmatrix} \\ & Z \succeq 0. \end{aligned}$$

Form Augmented Lagrangian

Momentum Parameter

$$\mathcal{L}_\rho(t,u,x,Z,\Lambda) = \frac{1}{2}\|x-y\|_2^2 + \frac{\tau}{2}(t+u_1) + \left\langle \Lambda, Z - \begin{bmatrix} T(u) & x \\ x^* & t \end{bmatrix} \right\rangle + \frac{\rho}{2}\left\| Z - \begin{bmatrix} T(u) & x \\ x^* & t \end{bmatrix} \right\|_F^2$$

Lagrangian Term

Augmented Lagrangian

Alternating Direction Iterations

$$\begin{aligned} (t^{l+1}, u^{l+1}, x^{l+1}) &\leftarrow \arg\min_{t,u,x} \mathcal{L}_\rho(t,u,x,Z^l,\Lambda^l) && \rightarrow \text{quadratic objective, linear} \\ Z^{l+1} &\leftarrow \arg\min_{Z\succeq 0} \mathcal{L}_\rho(t^{l+1},u^{l+1},x^{l+1},Z,\Lambda^l) && \rightarrow \text{projection on psd cone} \\ \Lambda^{l+1} &\leftarrow \Lambda^l + \rho\left(Z^{l+1} - \begin{bmatrix} T(u^{l+1}) & x^{l+1} \\ x^{l+1*} & t^{l+1} \end{bmatrix}\right). && \rightarrow \text{linear dual update} \end{aligned}$$

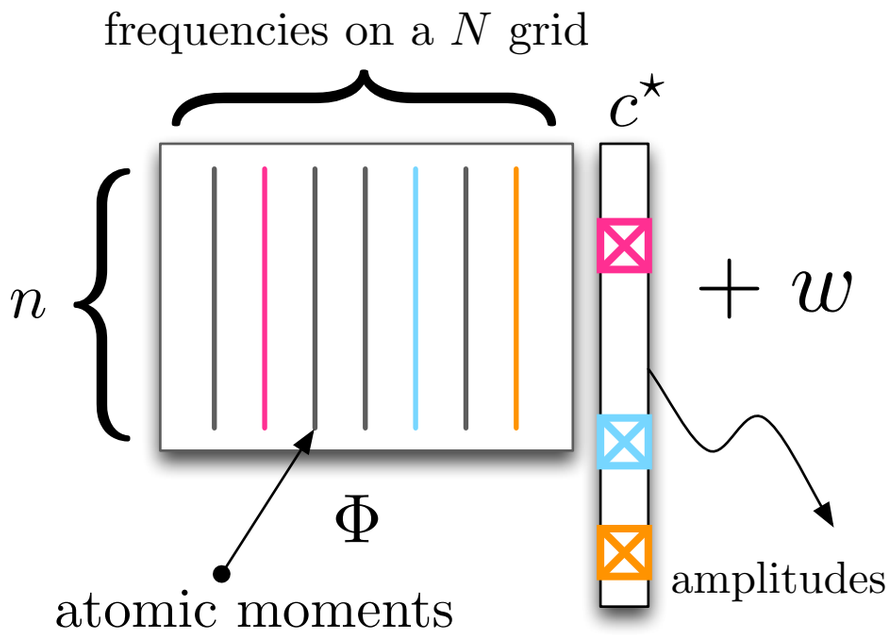
Alternative Discretization Method

Can use a grid $\mathcal{A}_N = \{a(f,\phi) : f = j/N, j = 1,\ldots,N, \phi \in [0,2\pi)\}$

We show $(1 - \frac{2\pi n}{N})\|x\|_{\mathcal{A}_N} \leq \|x\|_{\mathcal{A}} \leq \|x\|_{\mathcal{A}_N}$

Via Bernstein's Polynomial Theorem or Fritz John
Works for general epsilon nets

AST is approximated by Lasso



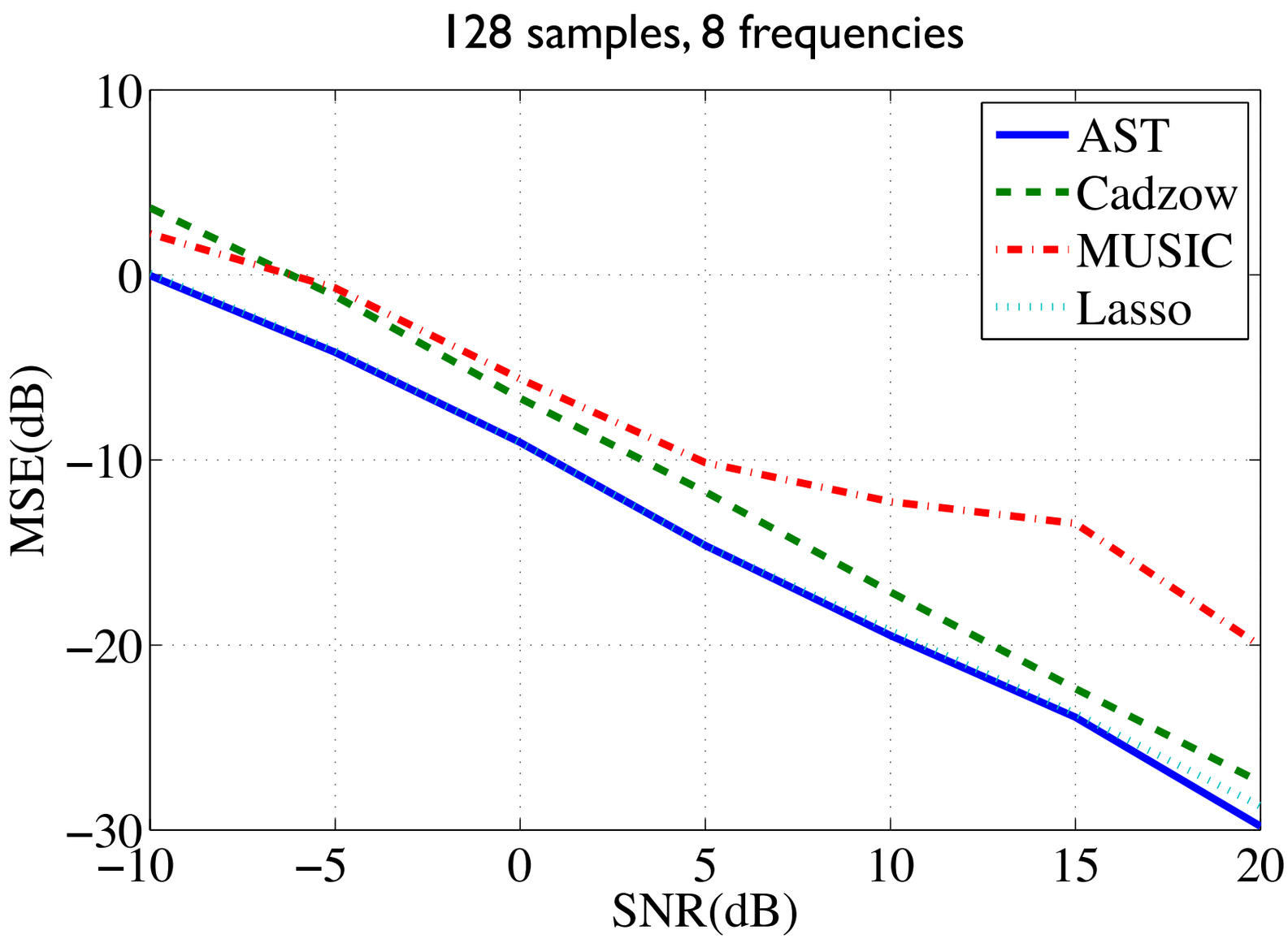
- Fast shrinkage algorithms
- Fourier projections are cheap
- Coherence is irrelevant

Experimental setup

- n=64 or 128 or 256 samples
- no. of frequencies: n/16, n/8, n/4
- SNR: -5 dB, 0 dB, ... , 20 dB
- (Root) MUSIC, Matrix pencil, Cadzow: *All fed true model order.*
- Empirical estimated (not fed) noise variance for AST
- Compared mean-squared-error and localization error metrics, averaged over 20 trials.

Mean Squared Error

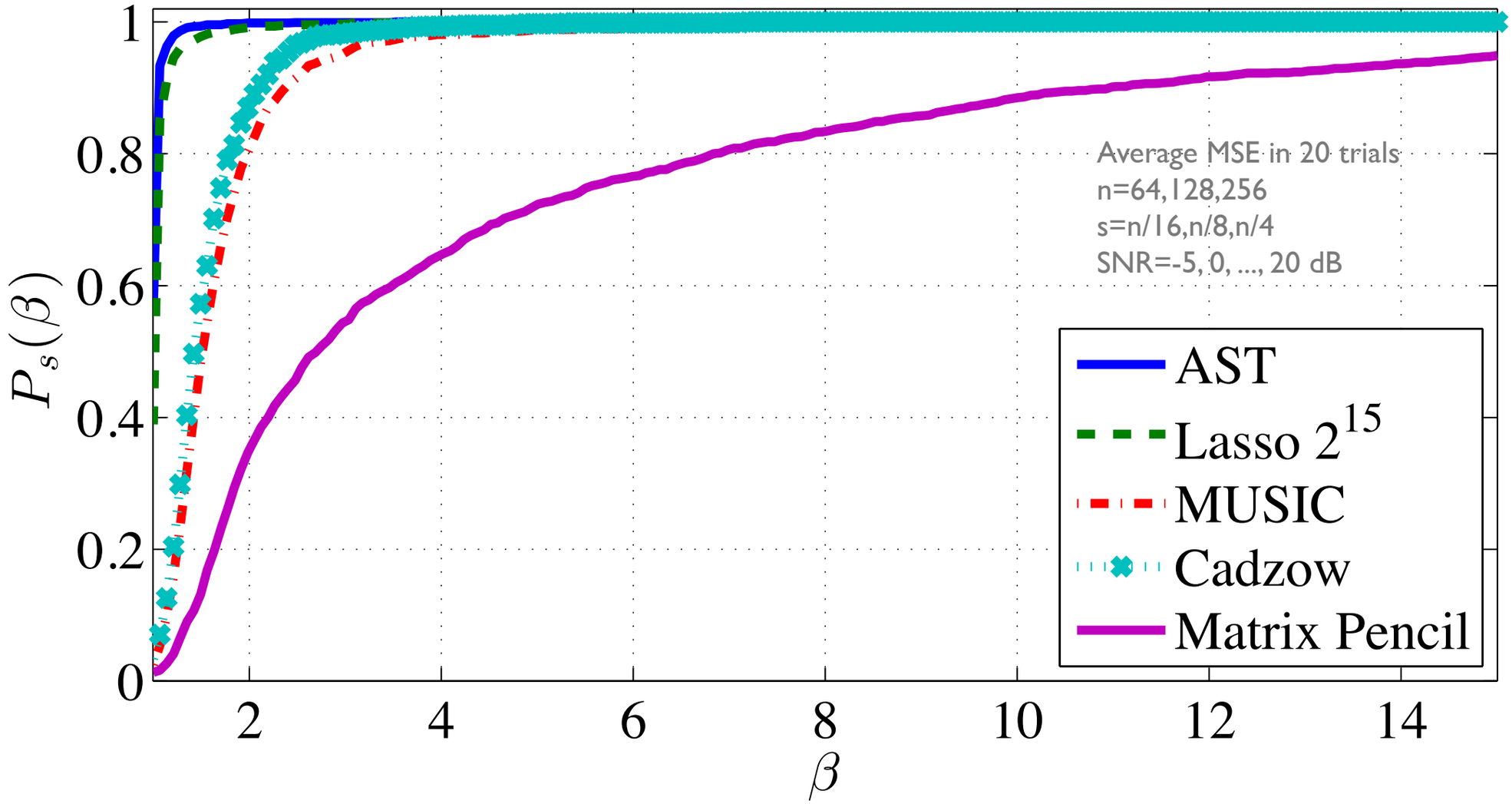
MSE vs SNR



Mean Squared Error

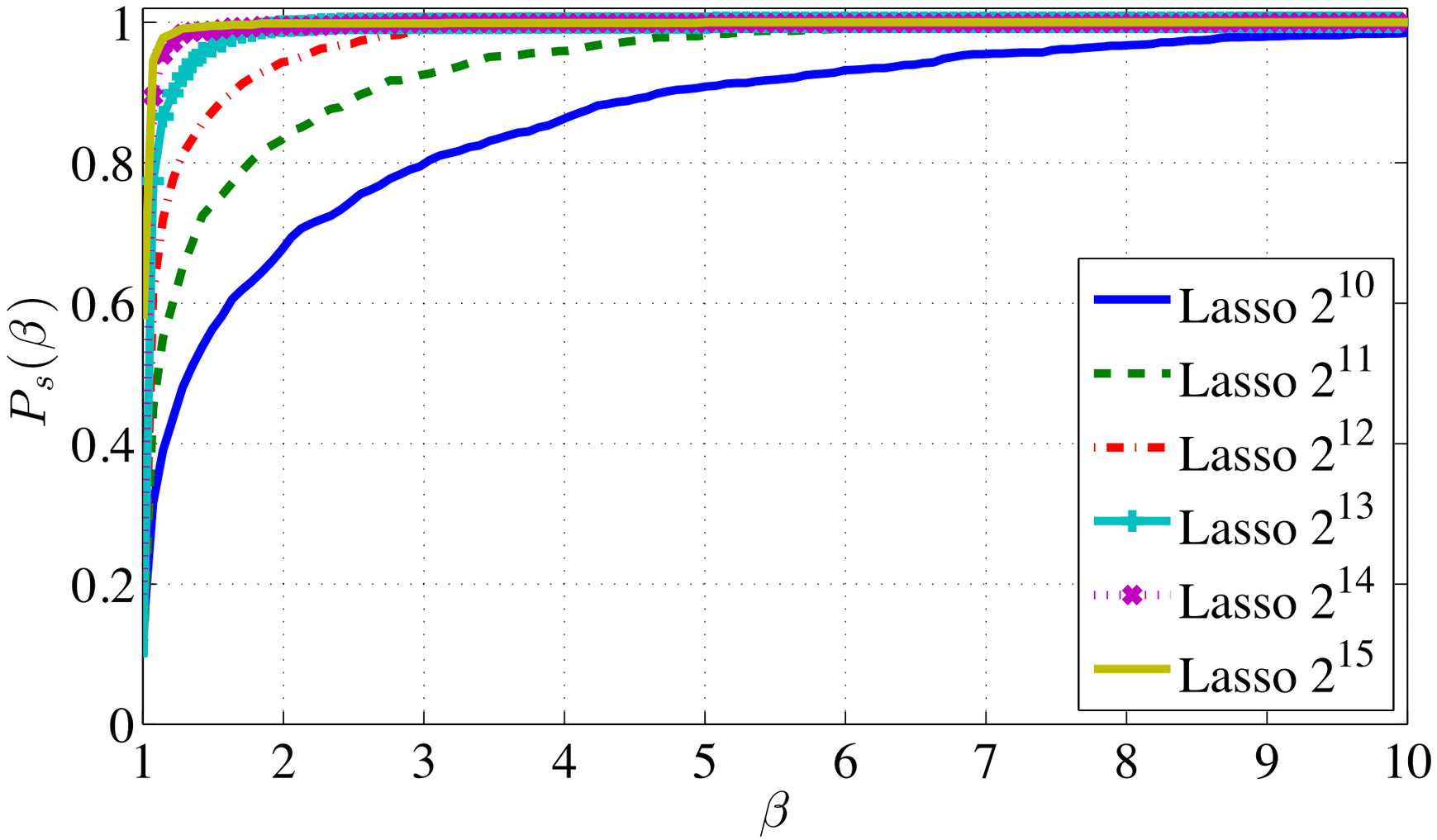
Performance Profiles

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



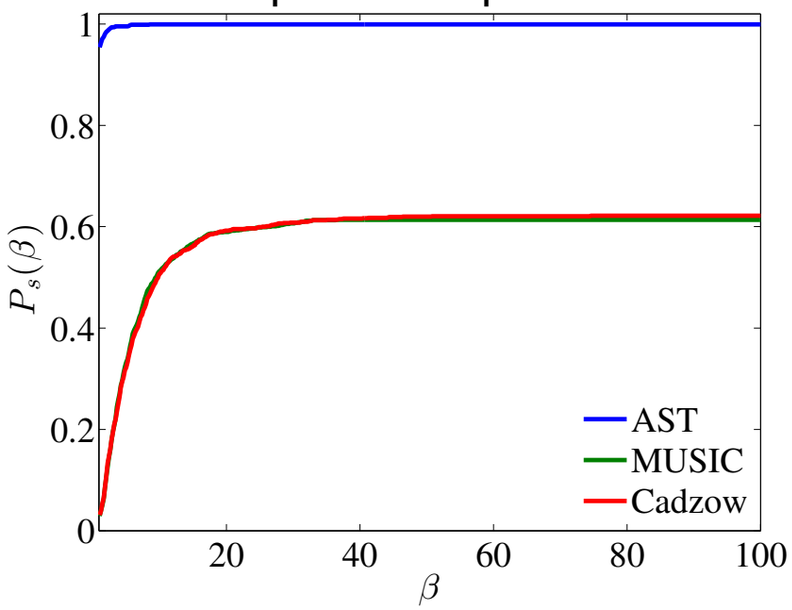
Mean Squared Error

Effect of Gridding

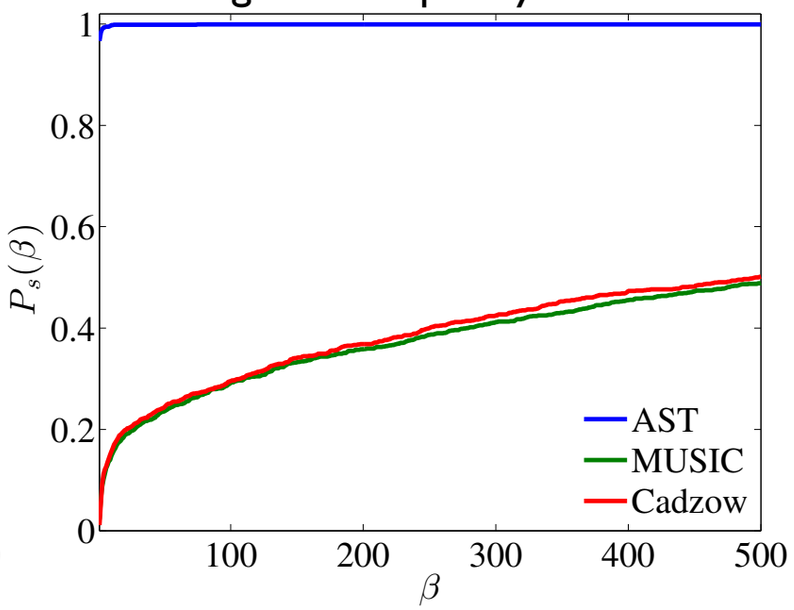


Frequency Localization

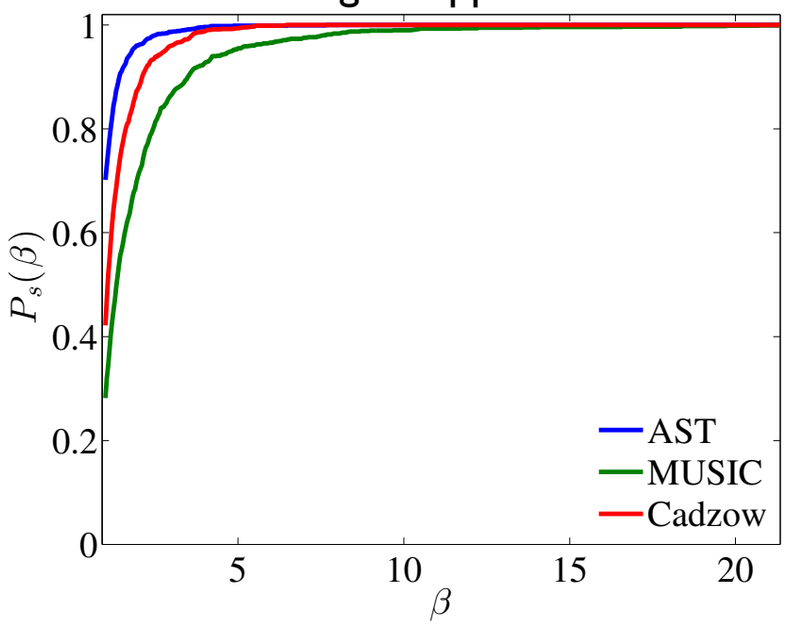
Spurious Amplitudes

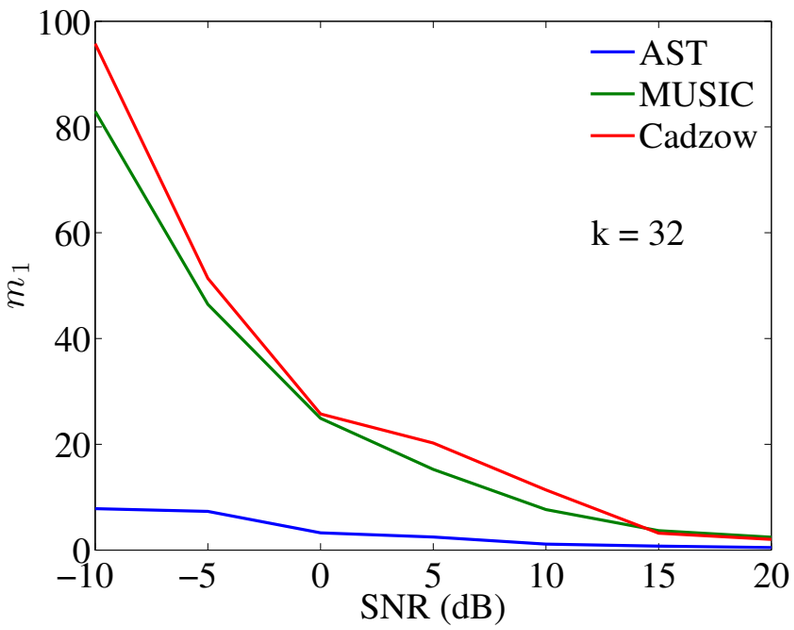


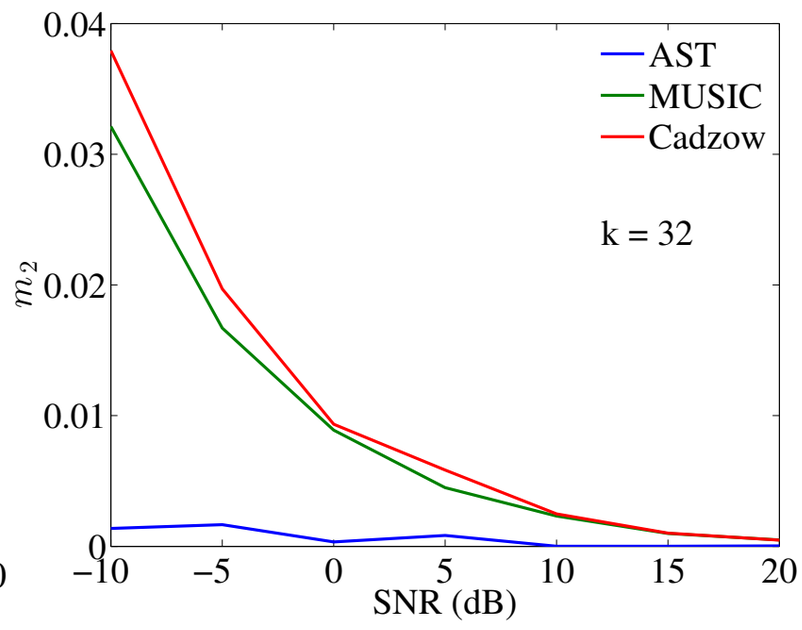
Weighted Frequency Deviation

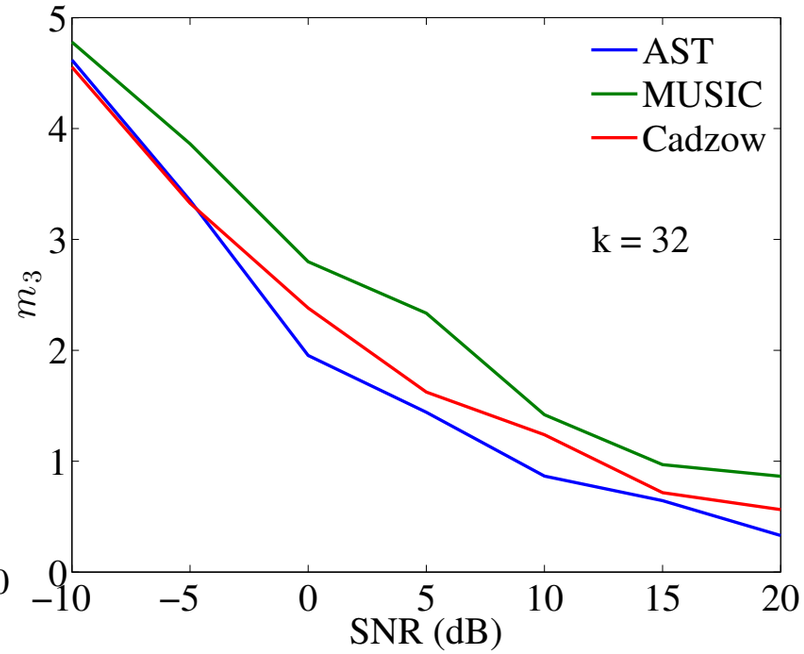


Near region approximation

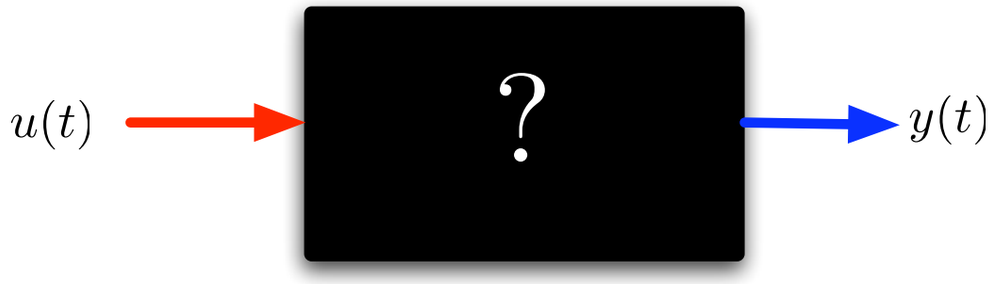








Simple LTI Systems



State Space

$$x[t+1] = Ax[t] + Bu[t]$$
$$y[t] = Cx[t] + Du[t]$$

- Find a simple linear system from time series data
- Nonlinear Kalman filter based estimation
- Parametric methods based on Hankel Low rank formulations

Subspace Methods

With a little computation, Observability

$$\underbrace{\begin{bmatrix} y_1 & y_2 & \cdots & y_k \\ y_2 & y_3 & \cdots & y_{k+1} \\ y_3 & y_4 & \cdots & y_{k+2} \\ \vdots & & \ddots & \vdots \\ y_\ell & y_{\ell+1} & \cdots & y_{\ell+k-1} \end{bmatrix}}_{Y} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{\ell-1} \end{bmatrix} \underbrace{\begin{bmatrix} x_1 & x_2 & \dots & x_k \end{bmatrix}}_{X} + \begin{bmatrix} D & 0 & & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ CA^{\ell-2}B & & & \dots & D \end{bmatrix} \underbrace{\begin{bmatrix} u_1 & u_2 & \cdots & u_k \\ u_2 & u_3 & \cdots & u_{k+1} \\ u_3 & u_4 & \cdots & u_{k+2} \\ \vdots & & \ddots & \vdots \\ u_\ell & u_{\ell+1} & \cdots & u_{\ell+k-1} \end{bmatrix}}_{U}$$

Output

State

System Matrix T

Input

$$Y = \mathcal{O}X + TU$$

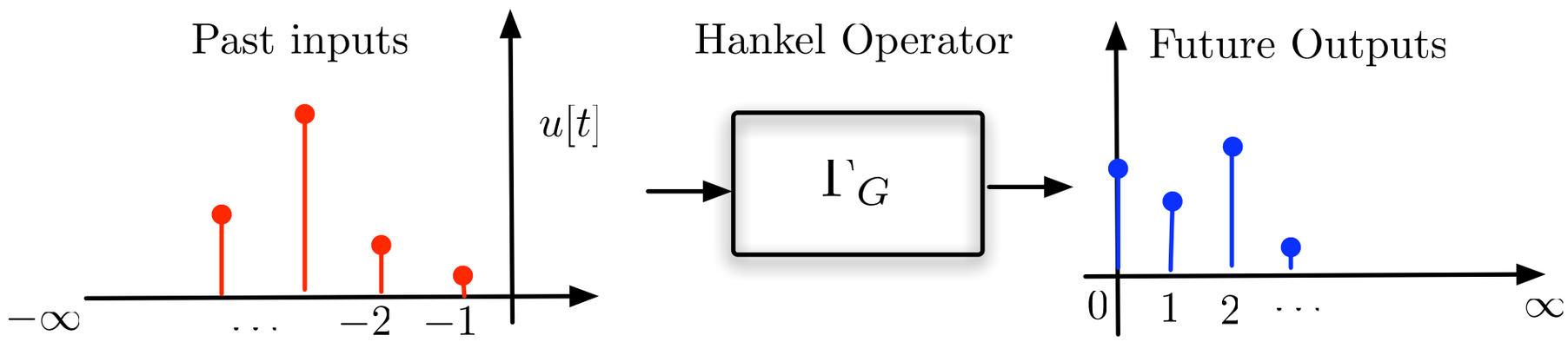
$$YU^\perp = \mathcal{O}XU^\perp$$

→ **SVD** Subspace ID (n4sid)

Low Rank

Nuclear Norm (Vandenberghe et al)

SISO

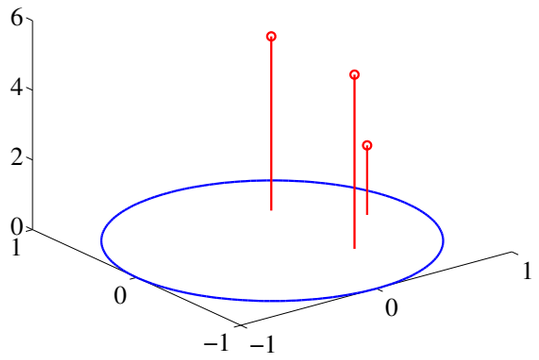


$$\Gamma_G : \ell_2(-\infty, 0) \to \ell_2[0, \infty)$$

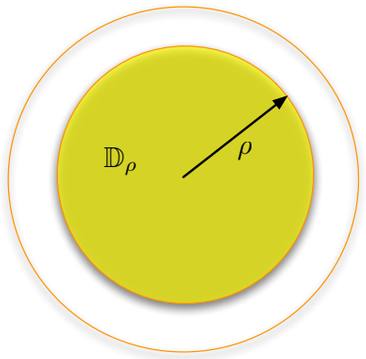
- **Fact**: Rank of Hankel Operator = McMillan Degree
- Relax and use Hankel nuclear norm?
- Not clear how to compute...

Convex Perspective

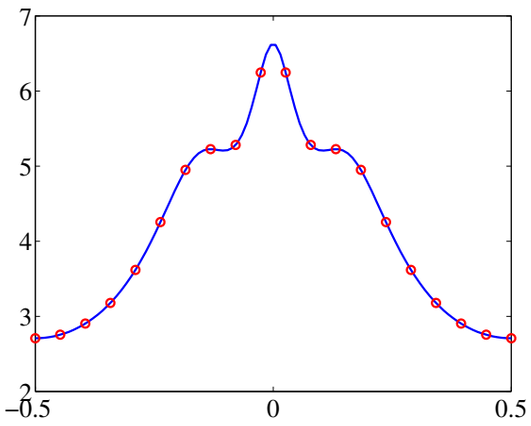
SISO Systems



$$G(z) = \frac{2(1-0.7^2)}{z-0.7} + \frac{5(1-0.5^2)}{z-(0.3+0.4i)} + \frac{5(1-0.5^2)}{z-(0.3-0.4i)}$$



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



$$y = \mathcal{L}\left(G(z)\right) + w$$

Regularize with atomic norms:

$$\underset{G}{\text{minimize}} \; \|y - \mathcal{L}(G)\|_2^2 + \mu \|G\|_{\mathcal{A}}$$

Theorem:

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

Atomic Norm Regularization

How to solve this?

$$\boxed{\underset{G}{\text{minimize}} \, \|y - \mathcal{L}(G)\|_2^2 + \mu \|G\|_{\mathcal{A}}}$$

Equivalent to:

$$\begin{array}{ll} \text{minimize}_{c,x} & \|y - x\|_2^2 + \mu \sum_{w \in \mathbb{D}_\rho} \frac{|c_w|}{1-|w|^2} \\ \text{subject to} & x = \sum_{w \in \mathbb{D}_\rho} c_w \mathcal{L}\left(\frac{1}{z-w}\right) \end{array}$$

Discretize and Set

$$\Phi_{k\ell} := \mathcal{L}_k \left(\frac{1 - |w|^2}{z - w_\ell} \right)$$

$$\underset{c}{\text{minimize}} \, \tfrac{1}{2} \|y - \Phi c\|_2^2 + \mu \|c\|_1$$

Lasso

DAST Analysis

Measure uniform spaced samples of frequency response:

$$\Phi_{k\ell} = \frac{1 - |w_\ell|^2}{e^{2\pi i k/n} - w_\ell}$$

Solve: $\hat{x} = \arg\min_x \frac{1}{2}\|y - x\|_2^2 + \mu\|x\|_{\mathcal{A}_\epsilon}$ over a net on the disk

Set: $\hat{G}(z) = \sum_\ell \hat{c}_\ell \frac{1 - |w_\ell|^2}{z - w_\ell}$ where the coefficients correspond to the support of $\hat{x}$

Then we have,

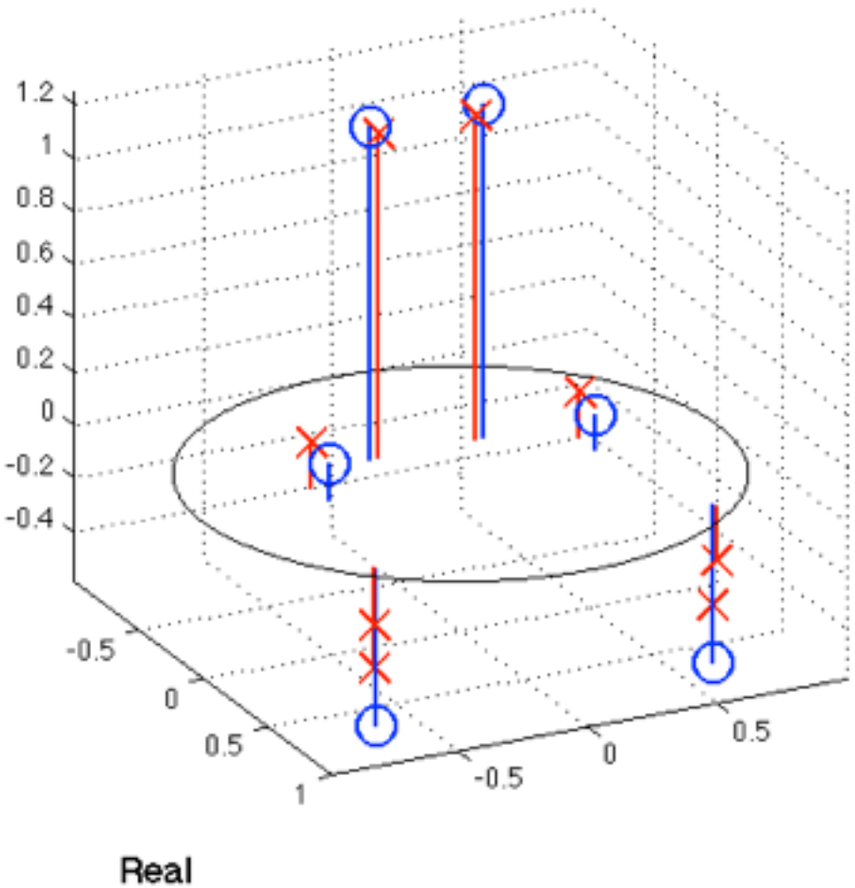
$$\|G - \hat{G}\|_{\mathcal{H}_2}^2 \leq \frac{\kappa_0 \|\Gamma_G\|_1}{(1-\rho)\sqrt{n}}$$

constant

Hankel nuclear norm

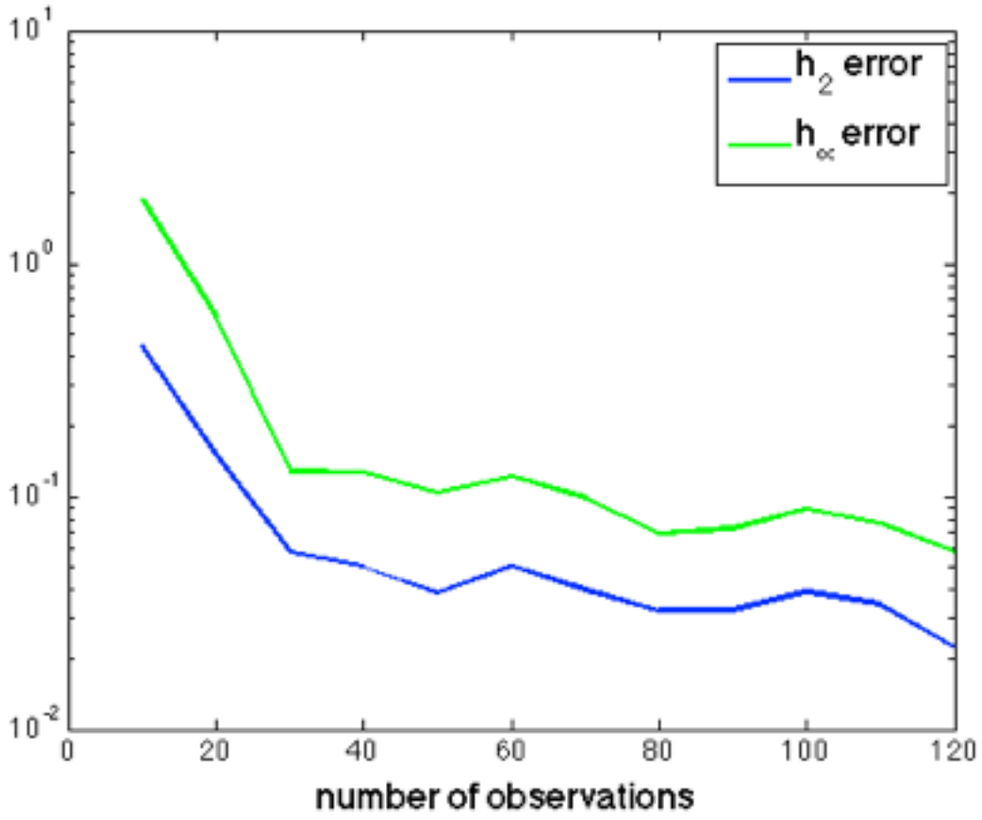
stability radius

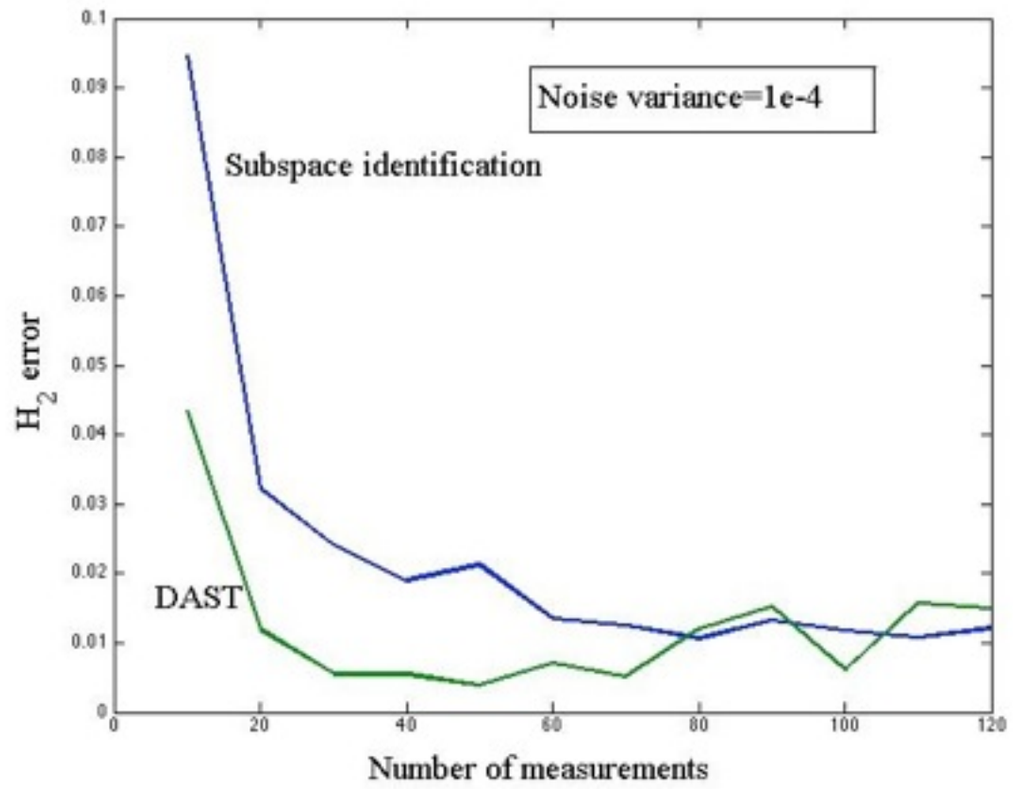
number of measurements



6 poles
$\rho = 0.95$
$\sigma = 1e\text{-}1$
n=30

H_2 err = 2e-2





Summary

- Optimal bounds for denoising using convex methods.
- Better empirical performance than classical approaches
- Local conditions on signals instead of global dictionary conditions
- Discretization works!
- *Acknowledgements*: Results derived in collaboration with Tang, Shah and Prof. Recht.

Future Work

- Better convergence results for discretization.
- Why does weighted mean heuristic work so well for discretization?
- Better algorithms for System Identification
- Greedy techniques
- More atomic sets. General localization guarantees.

Referenced Publications

- B, Recht. "Atomic Norm Denoising for Line Spectral Estimation" in *49th Allerton Conference on Controls and Systems, 2011*

- B, Tang, Recht. "Atomic Norm Denoising for Line Spectral Estimation", submitted to *IEEE Transactions on Signal Processing*

- Tang, B, Recht. "Near Minimax Line Spectral Estimation", *To be submitted*

- Shah, B, Tang, Recht. "Linear System Identification using Atomic Norm Minimization", *Control and Decision Conference, 2012*

Other Publications

- Tang, B, Shah, Recht. “Compressed Sensing off the grid”, submitted to *IEEE Transactions on Information Theory*

- Dasarathy, Shah, B, Nowak. “Covariance Sketching”, *50th Allerton Conference on Controls and Systems, 2012*

- Gubner, B, Hao “Multipath-Cluster Channel Models”, *IEEE Conference on Ultrawideband Systems 2012*

- Wittenberg, Alimadhi, B, Lau “ei. RxC: Hierarchical Multinomial-Dirichlet Ecological Inference Model” *in Zelig: Everyone’s Statistical Software.*

- Tang, B, Recht. “Sparse recovery over continuous dictionaries: Just discretize” *Submitted to Asilomar Conference 2013*

Cone Condition

Inflated Descent Cone $(0 \leq \gamma \leq 1)$

$$C_\gamma(x^\star) = \mathbf{cone}\left\{z \mid \|x^\star + z\|_{\mathcal{A}} \leq \|x^\star\|_{\mathcal{A}} + \gamma\|z\|_{\mathcal{A}}\right\}$$

Descent Cone = C_0

No direction in C_γ with zero atomic norm

$$\phi_\gamma = \inf_{z \neq 0} \left\{ \frac{\|z\|_2}{\|z\|_{\mathcal{A}}} \;\middle|\; z \in C_\gamma(x^\star) \right\}$$

Minimum separation is necessary

Using two sinusoids:

$$\|a(f_1) - a(f_2)\|_2 = 2\pi n^{3/2}|f_1 - f_2|$$

So,

$$\begin{aligned}\|a(f_1) - a(f_2)\|_{\mathcal{A}} &\leq n^{1/2}\|a(f_1) - a(f_2)\|_2 \\ &\leq 2\pi n^2|f_1 - f_2|\end{aligned}$$

When separation is lower than $1/4n^2$ atomic norm can't recover it.

Can empirically show failure with less than $1/n$ separation and adversarial signal