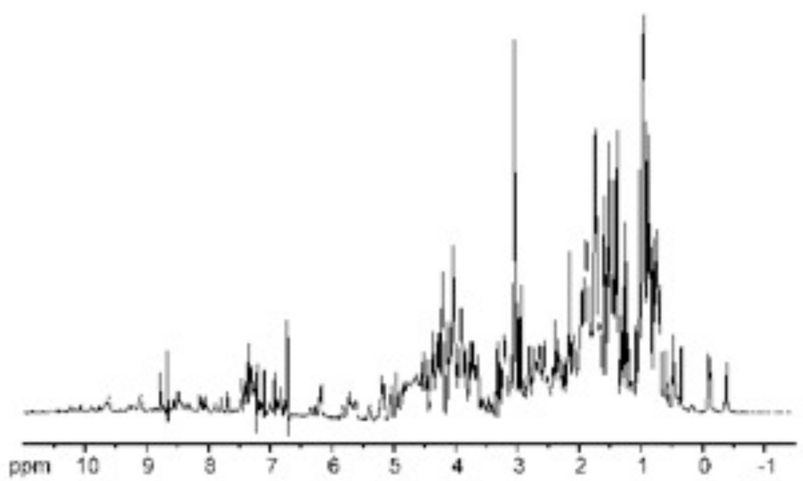
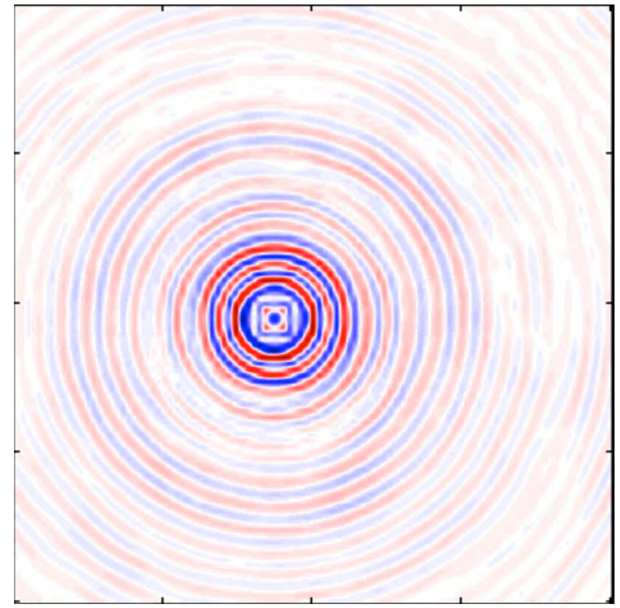
Going off the grid

Benjamin Recht
University of California, Berkeley

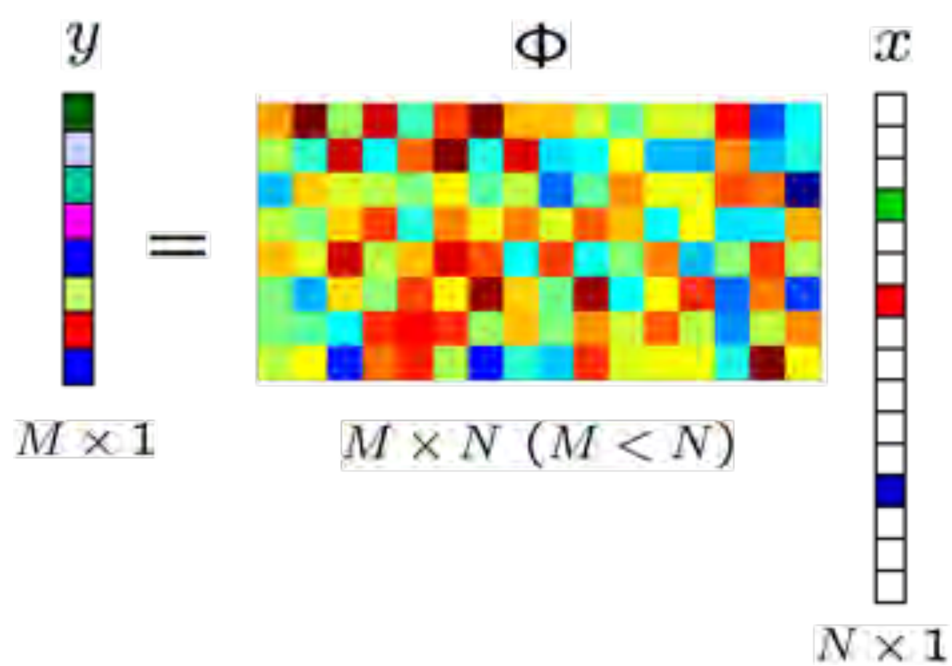
Joint work with
Badri Bhaskar
Parikshit Shah
Gonnguo Tang

We live in a continuous world...

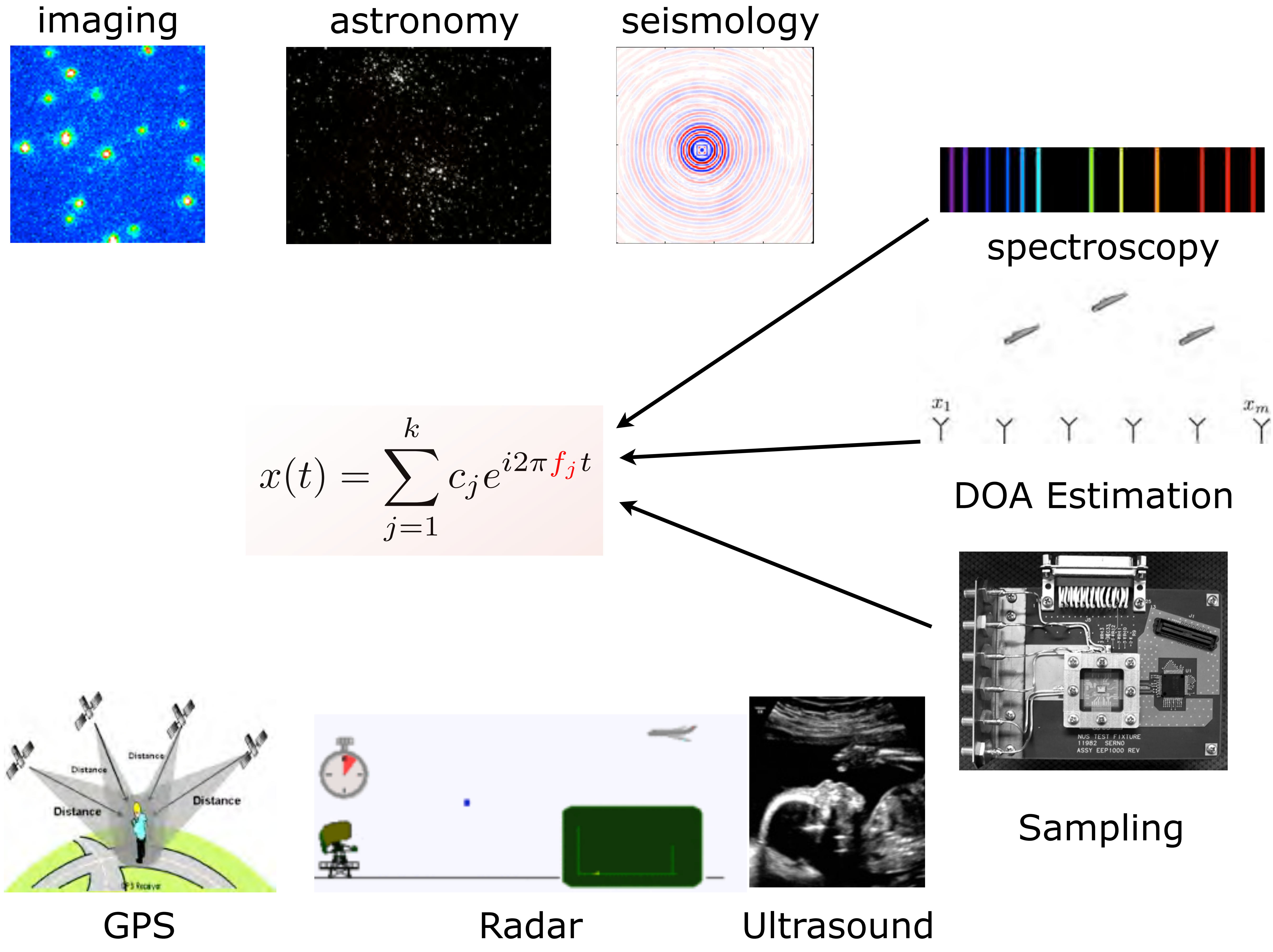


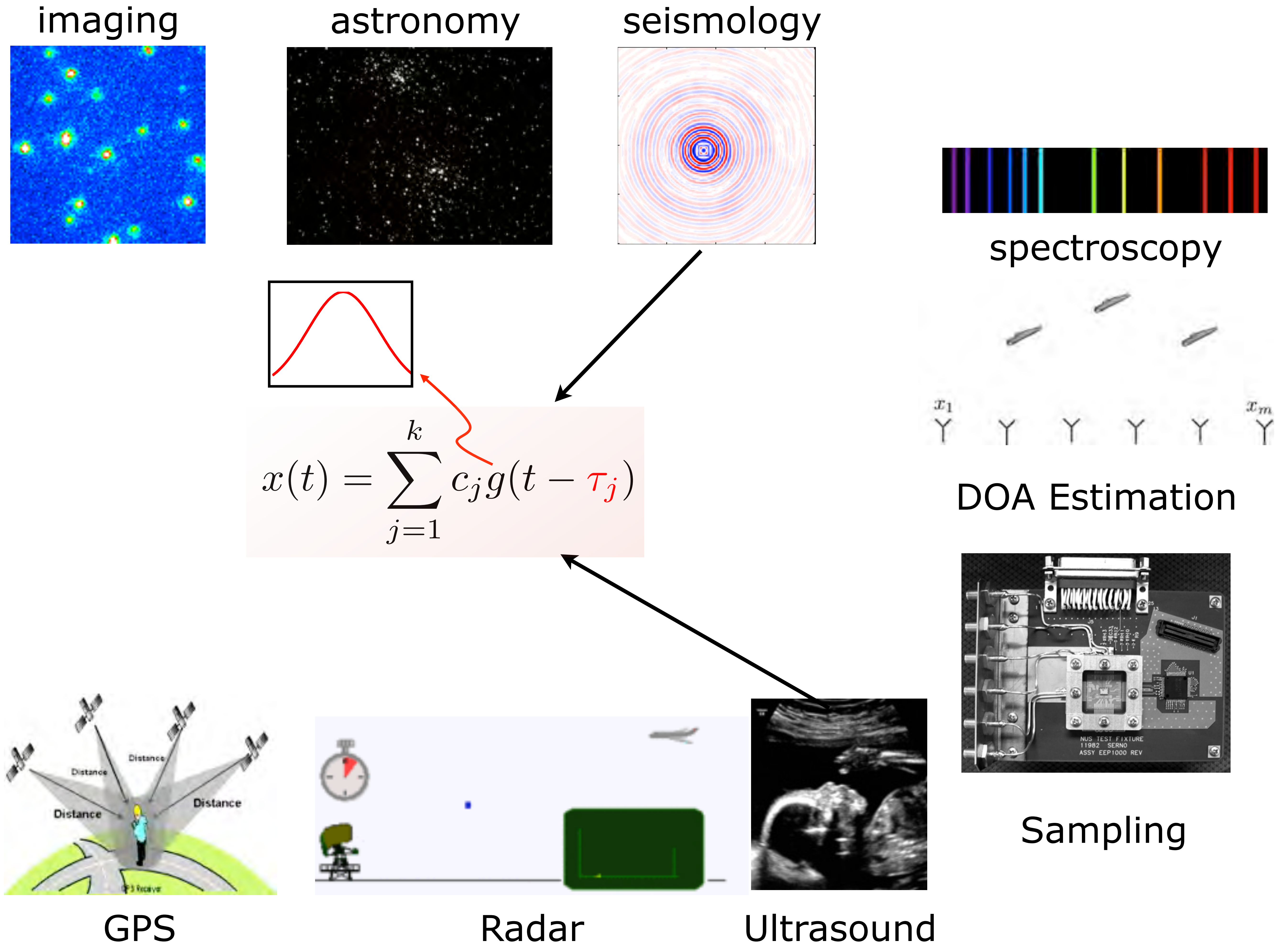


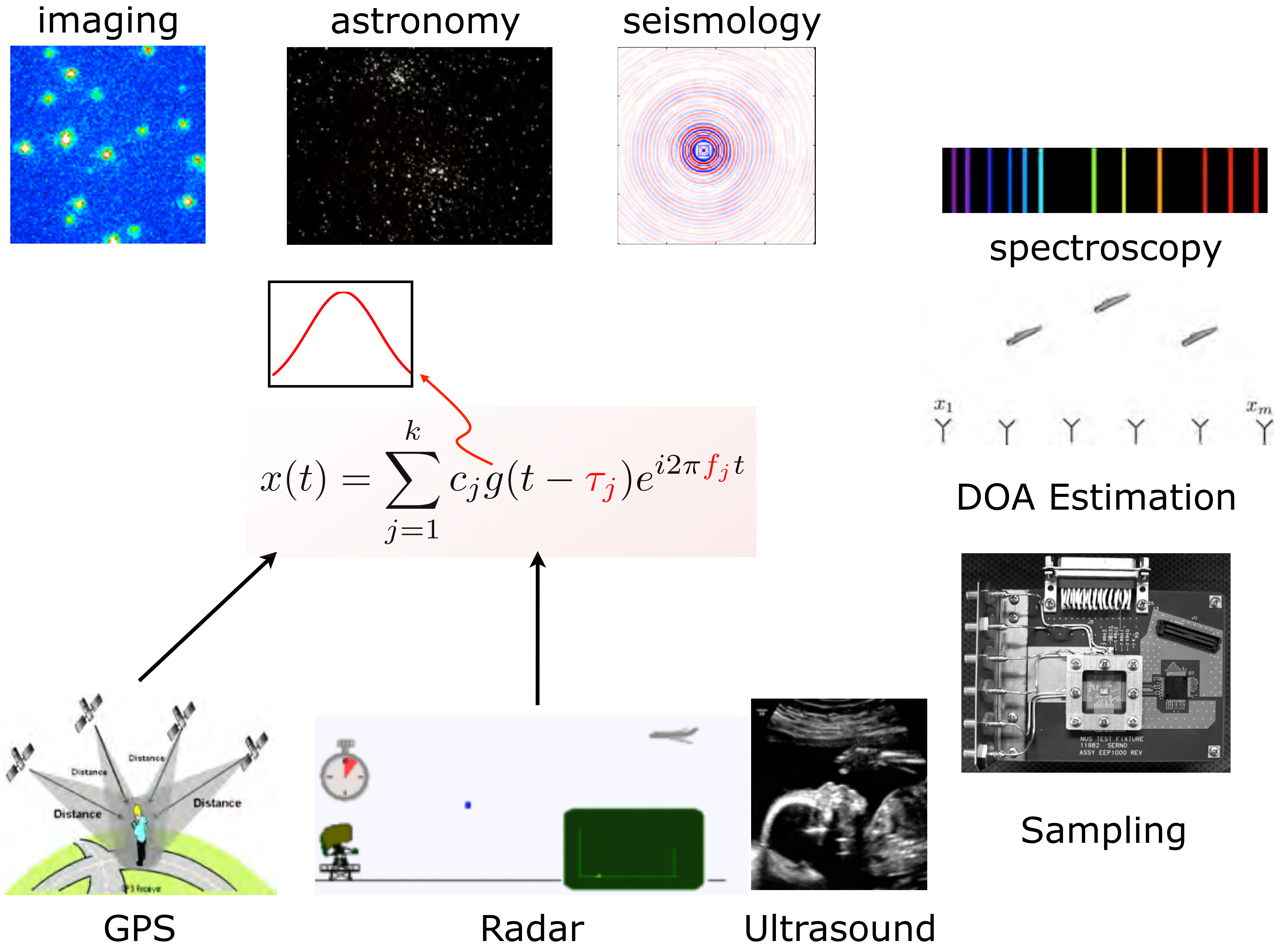
But we work with digital computers...

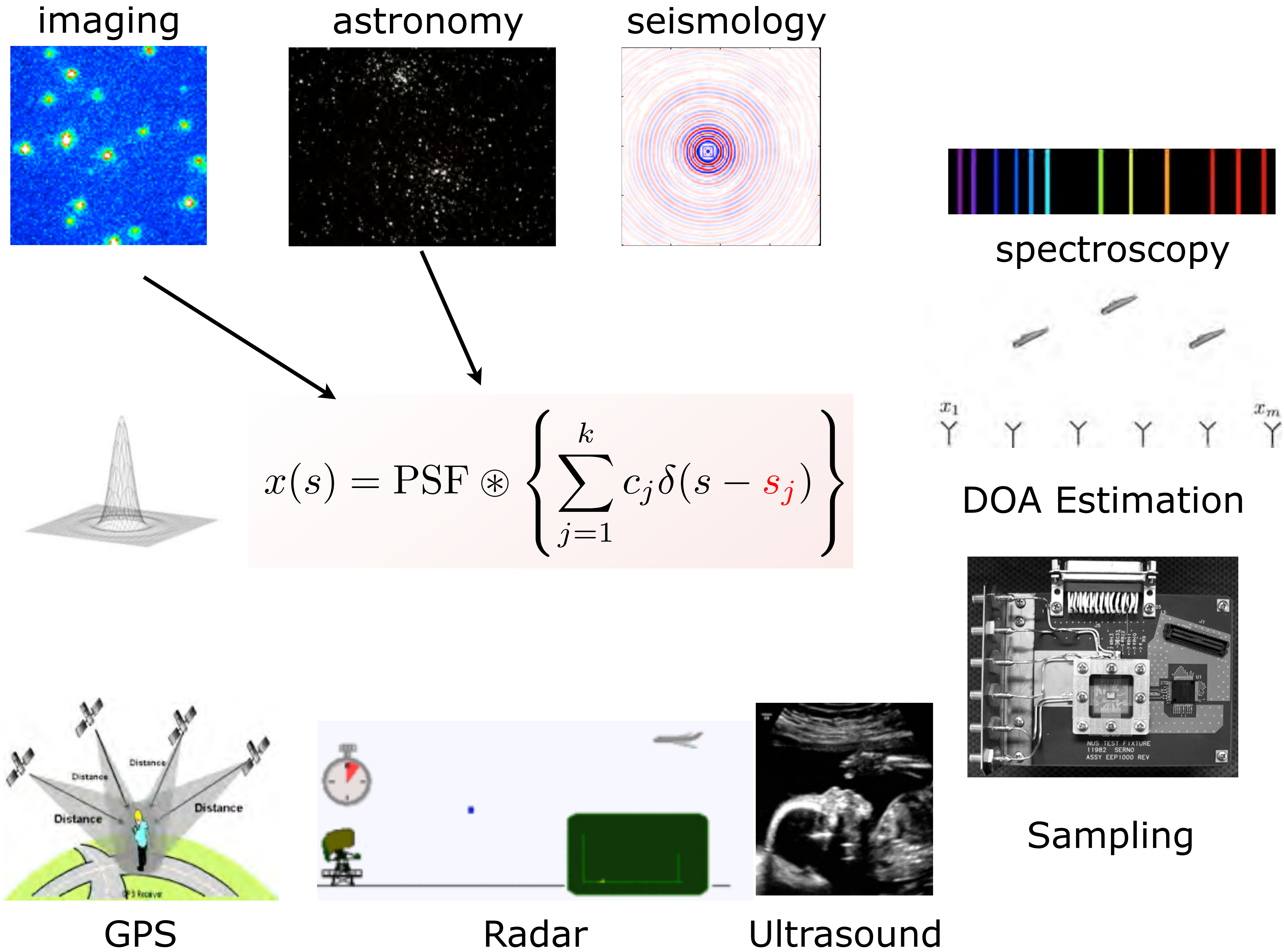


What is the price of living on the grid?

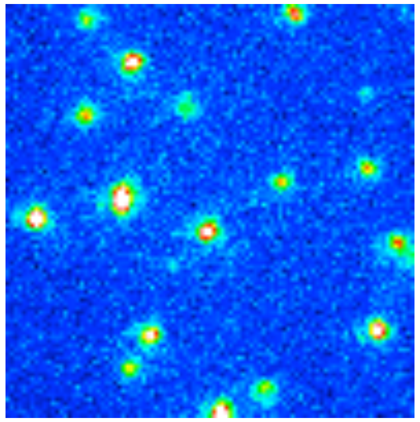

imaging
astronomy
seismology
spectroscopy
x_1
x_m
DOA Estimation
$$x(t) = \sum_{j=1}^{k} c_j e^{i2\pi f_j t}$$
Sampling
Distance
Distance
Distance
Distance
GPS
Radar
Ultrasound


imaging
astronomy
seismology
spectroscopy
x_1
x_m
DOA Estimation
$$x(t) = \sum_{j=1}^{k} c_j g(t - \tau_j)$$
Sampling
Distance
Distance
Distance
Distance
GPS
Radar
Ultrasound

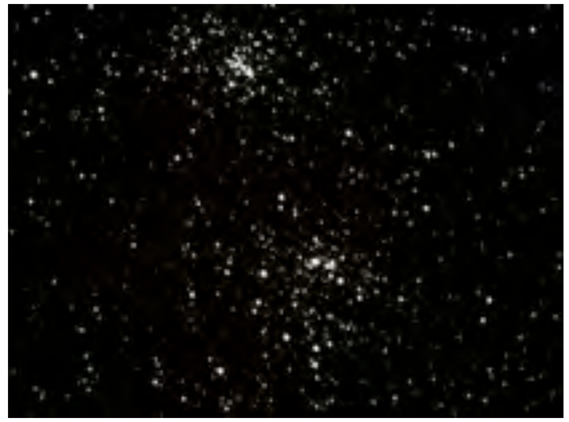

imaging
astronomy
seismology
spectroscopy
$x(t) = \sum_{j=1}^{k} c_j g(t - \tau_j) e^{i2\pi f_j t}$
x_1
x_m
DOA Estimation
Sampling
Distance
Distance
Distance
Distance
GPS
Radar
Ultrasound


imaging
astronomy
seismology
spectroscopy
x_1
x_m
DOA Estimation
$$x(s) = \mathrm{PSF} \circledast \left\{ \sum_{j=1}^{k} c_j \delta(s - s_j) \right\}$$
Sampling
Distance
Distance
Distance
Distance
GPS
Radar
Ultrasound

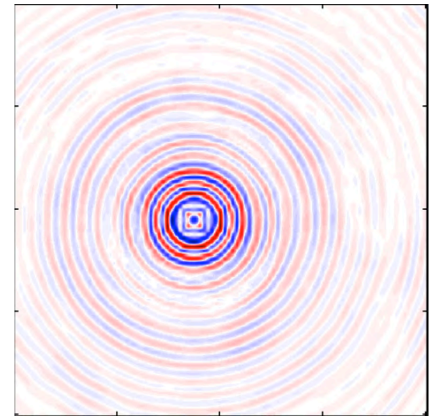
imaging

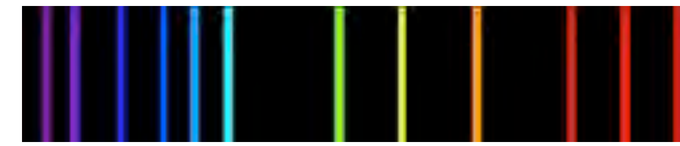


astronomy

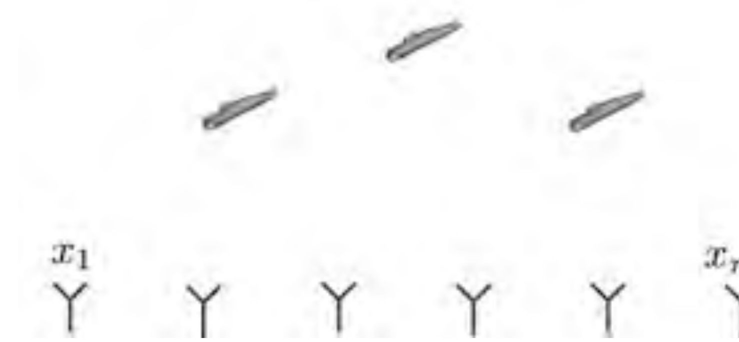


seismology

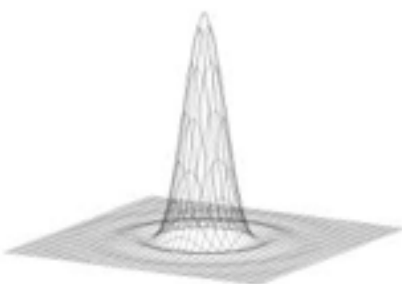




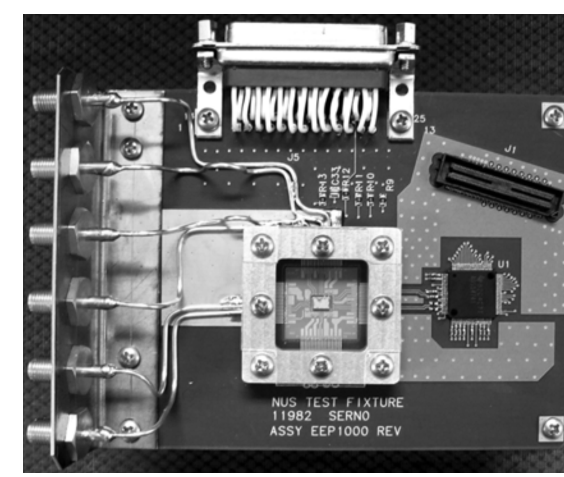
spectroscopy



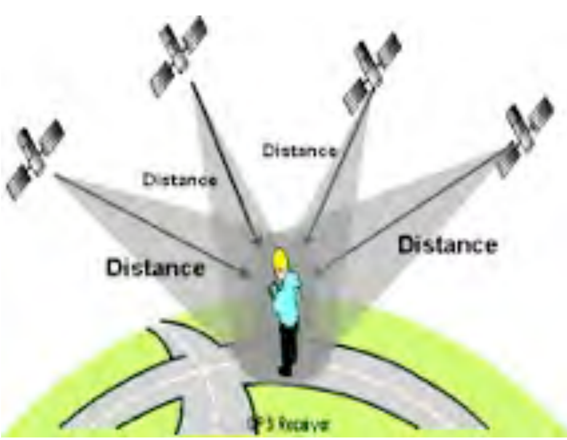
DOA Estimation



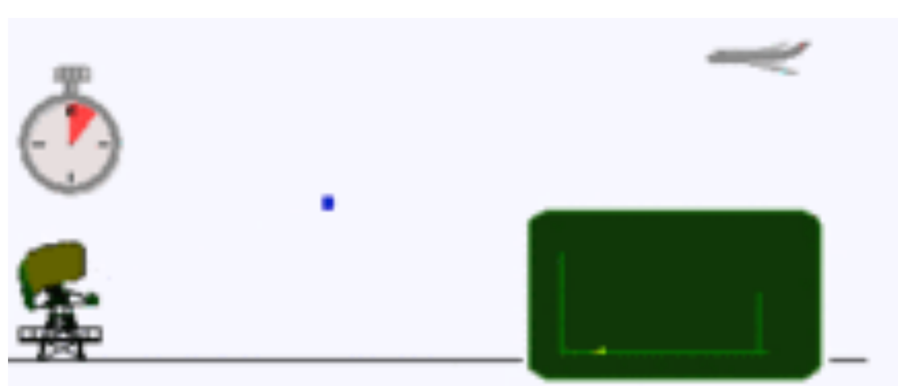
$$\hat{x}(\omega) = \widehat{\mathrm{PSF}}(\omega) \sum_{j=1}^{k} c_j e^{-i2\pi\omega s_j}$$



Sampling



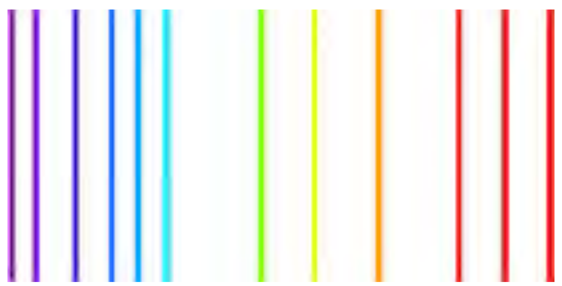
GPS



Radar

Ultrasound

Observe a sparse combination of sinusoids



$$x_m = \sum_{k=1}^{s} c_k e^{i2\pi m u_k} \quad \text{for some} \quad u_k \in [0,1)$$

Spectrum Estimation: find a combination of sinusoids agreeing with time series data

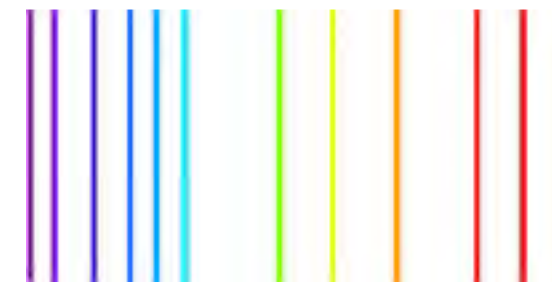
Classic (1790...): Prony's method

Assume coefficients are positive for simplicity

$$\sum_{k=1}^{r} c_k \begin{bmatrix} 1 \\ e^{i2\pi u_k} \\ e^{i4\pi u_k} \\ e^{i6\pi u_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{i2\pi u_k} \\ e^{i4\pi u_k} \\ e^{i6\pi u_k} \end{bmatrix}^* = \begin{bmatrix} x_0 & \bar{x}_1 & \bar{x}_2 & \bar{x}_3 \\ x_1 & x_0 & \bar{x}_1 & \bar{x}_2 \\ x_2 & x_1 & x_0 & \bar{x}_1 \\ x_3 & x_2 & x_1 & x_0 \end{bmatrix} =: \mathrm{toep}(x)$$

- toep(x) is positive semidefinite, and any null vector corresponds to a polynomial that vanishes at $e^{i2\pi u_k}$
- MUSIC, ESPRIT, Cadzow, etc.

Observe a sparse combination of sinusoids



$$x_m = \sum_{k=1}^{s} c_k e^{i2\pi m u_k} \quad \text{for some} \quad u_k \in [0,1)$$

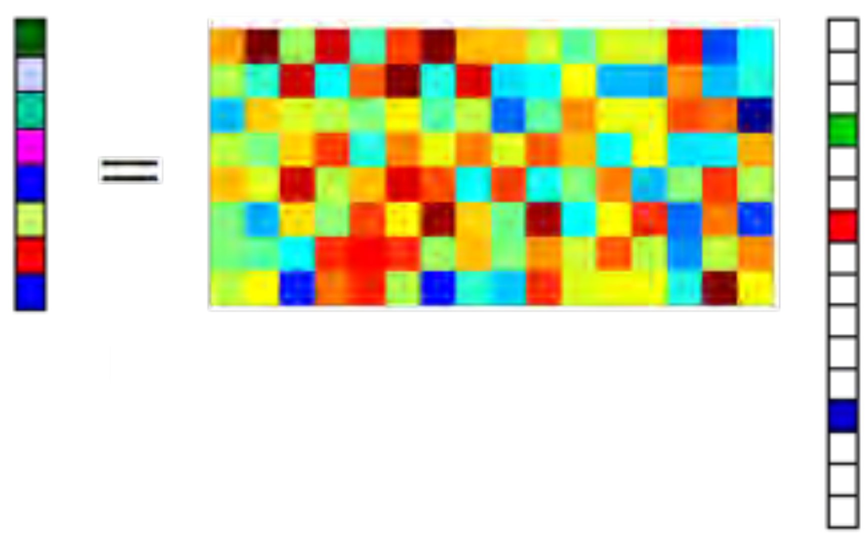
Spectrum Estimation: find a combination of sinusoids agreeing with time series data

Contemporary: $x \approx Fc$

sparse

$n \times N$

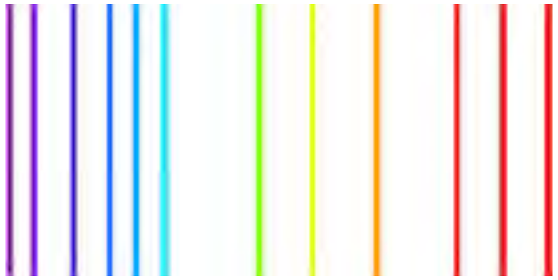
$F_{ab} = \exp(i2\pi ab/N)$



Solve with LASSO:

$$\text{minimize} \quad \|x - Fc\|_2^2 + \mu\|c\|_1$$

Observe a sparse combination of sinusoids



$$x_m = \sum_{k=1}^{s} c_k e^{i2\pi m u_k} \quad \text{for some} \quad u_k \in [0, 1)$$

Spectrum Estimation: find a combination of sinusoids agreeing with time series data

Classic	Contemporary
SVD	gridding+L1 minimization
grid free	robust model selection quantitative theory
need to know model order lack of quantitative theory unstable in practice	discretization error basis mismatch numerical instability

Can we bridge the gap?

Linear Inverse Problems

- Find me a solution of

$$y = \Phi x$$

- Φ n x p, n<p

- Of the infinite collection of solutions, which one should we pick?

- Leverage structure:

Sparsity | Rank | Smoothness | Symmetry

- How do we design algorithms to solve underdetermined systems problems with priors?

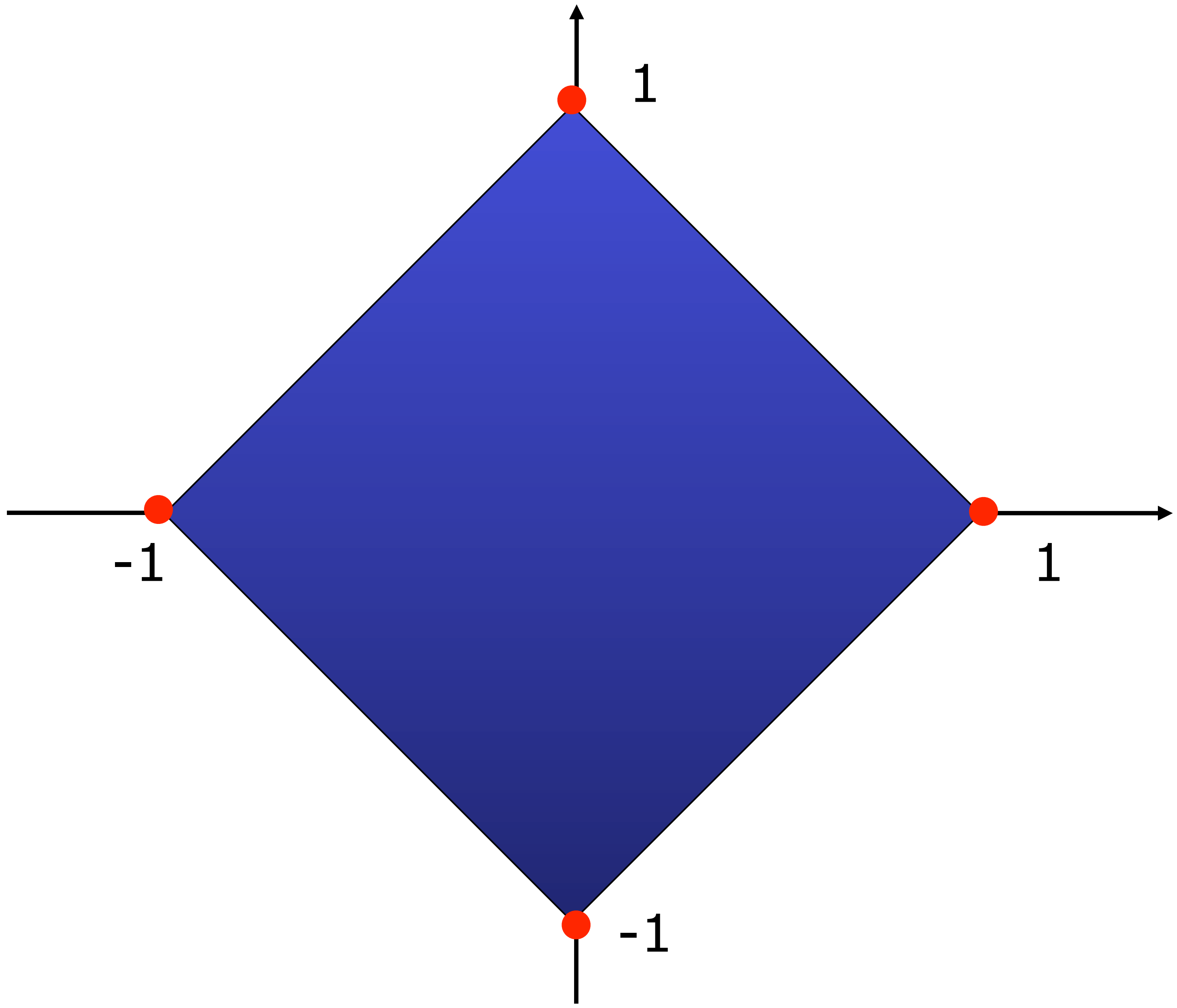
Sparsity

- 1-sparse vectors of Euclidean norm 1

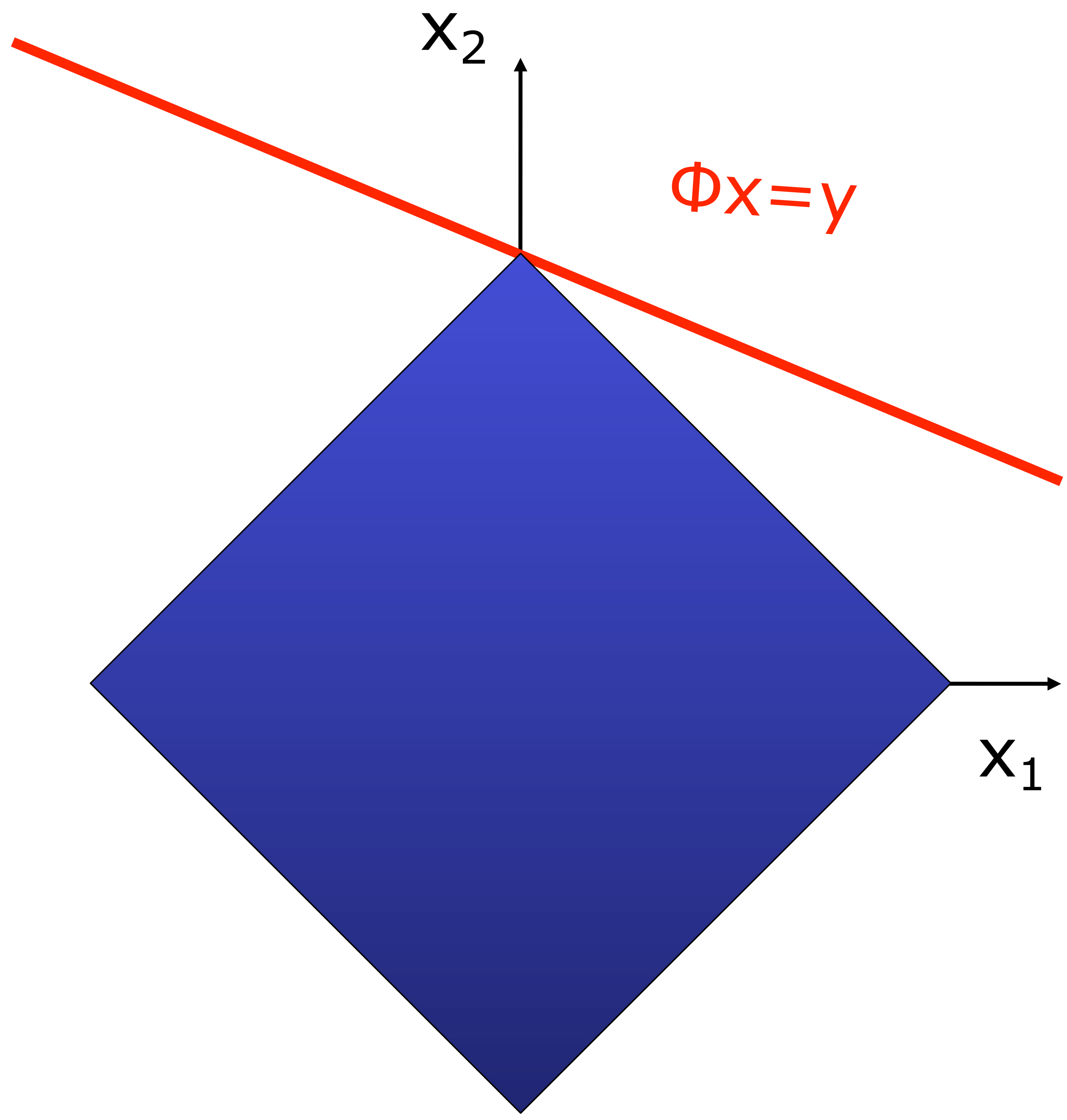
- Convex hull is the unit ball of the l_1 norm

$$\{x \ : \ \|x\|_1 \leq 1\}$$

$$\|x\|_1 = \sum_{i=1}^{p} |x_i|$$



$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & \Phi x = y \end{array}$$



Compressed Sensing: Candes, Romberg, Tao, Donoho, Tanner, Etc...

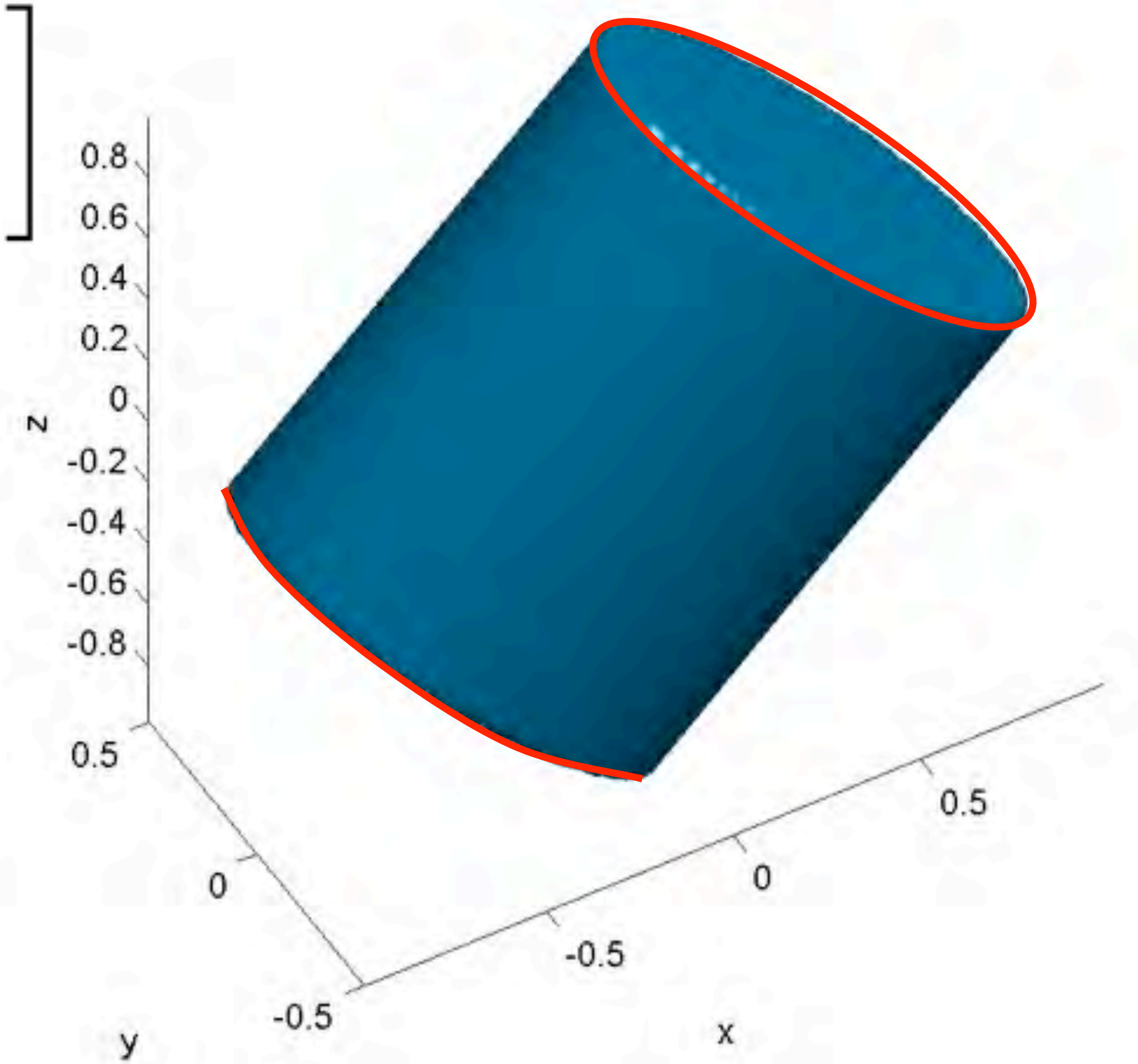
Rank

- 2x2 matrices
- plotted in 3d

$$\begin{bmatrix} x & y \\ y & z \end{bmatrix}$$

— rank 1

$x^2 + z^2 + 2y^2 = 1$



Convex hull:

$$\{X \;:\; \|X\|_* \leq 1\}$$

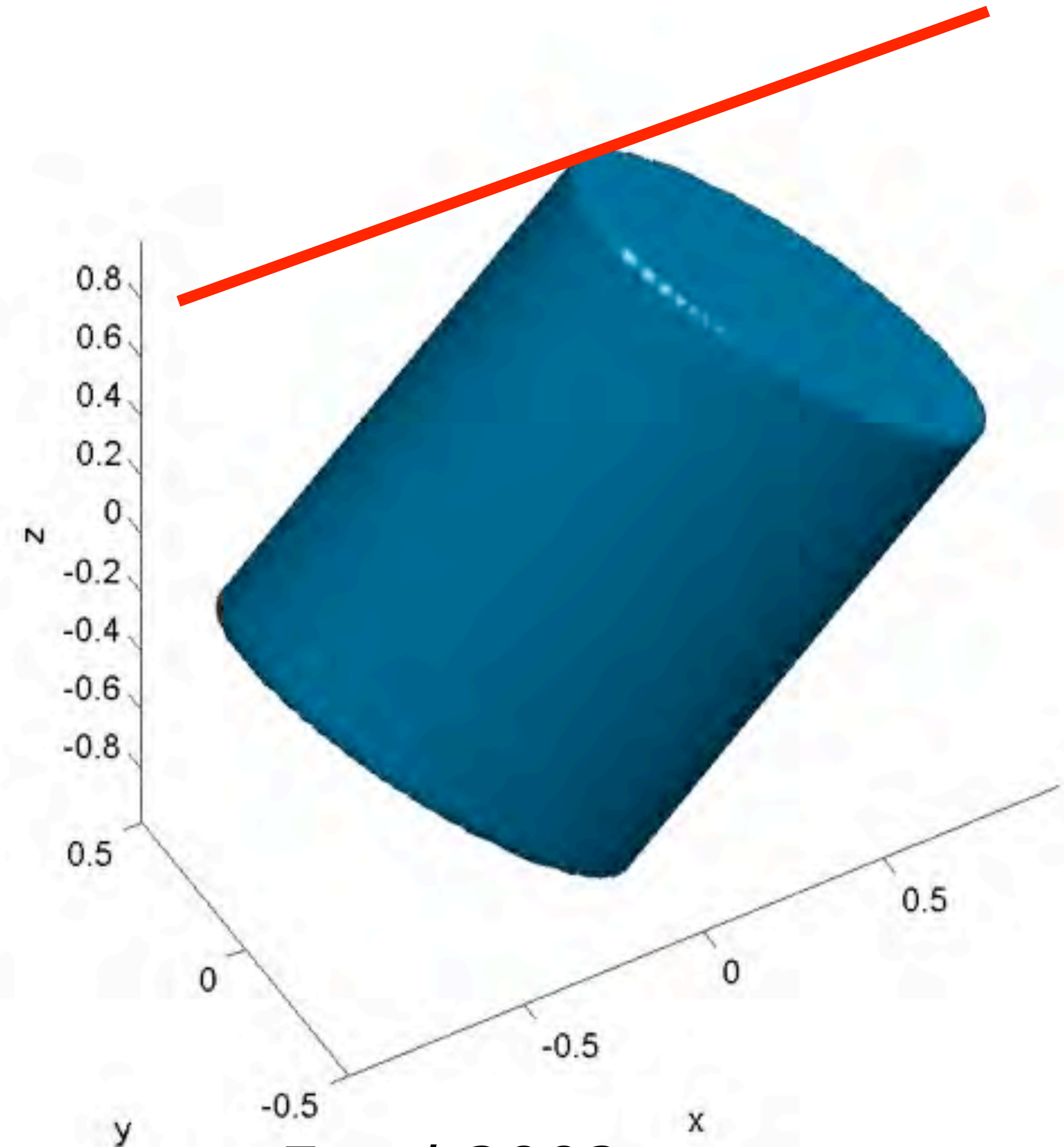
$$\|X\|_* = \sum_i \sigma_i(X)$$

- 2x2 matrices
- plotted in 3d

$$\left\| \begin{bmatrix} x & y \\ y & z \end{bmatrix} \right\|_* \leq 1$$

$$\|X\|_* = \sum_i \sigma_i(X)$$

Nuclear Norm Heuristic



Fazel 2002.
R, Fazel, and Parillo 2007
Rank Minimization/Matrix Completion

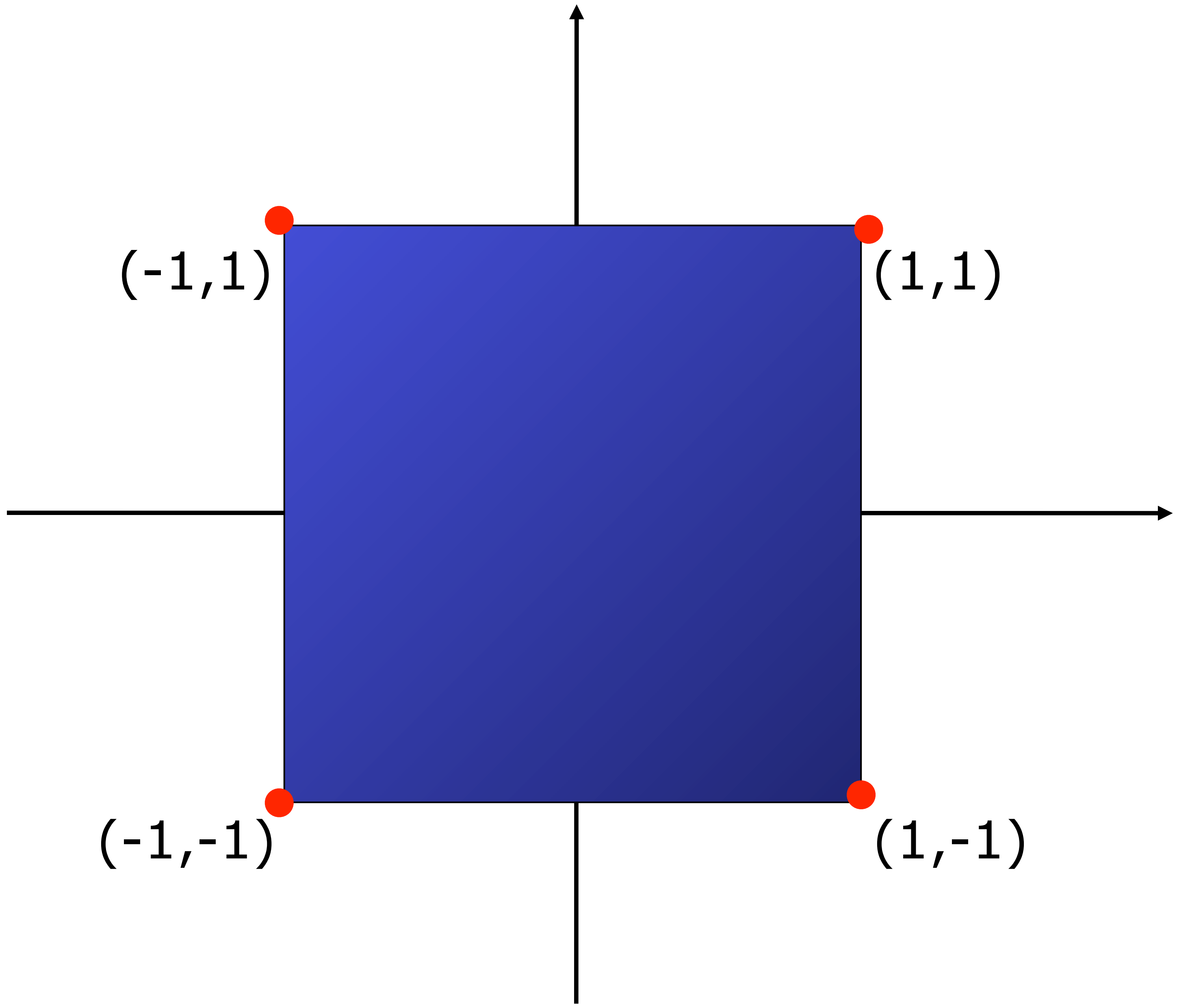
Integer Programming

- Integer solutions:
  all components of x are ±1

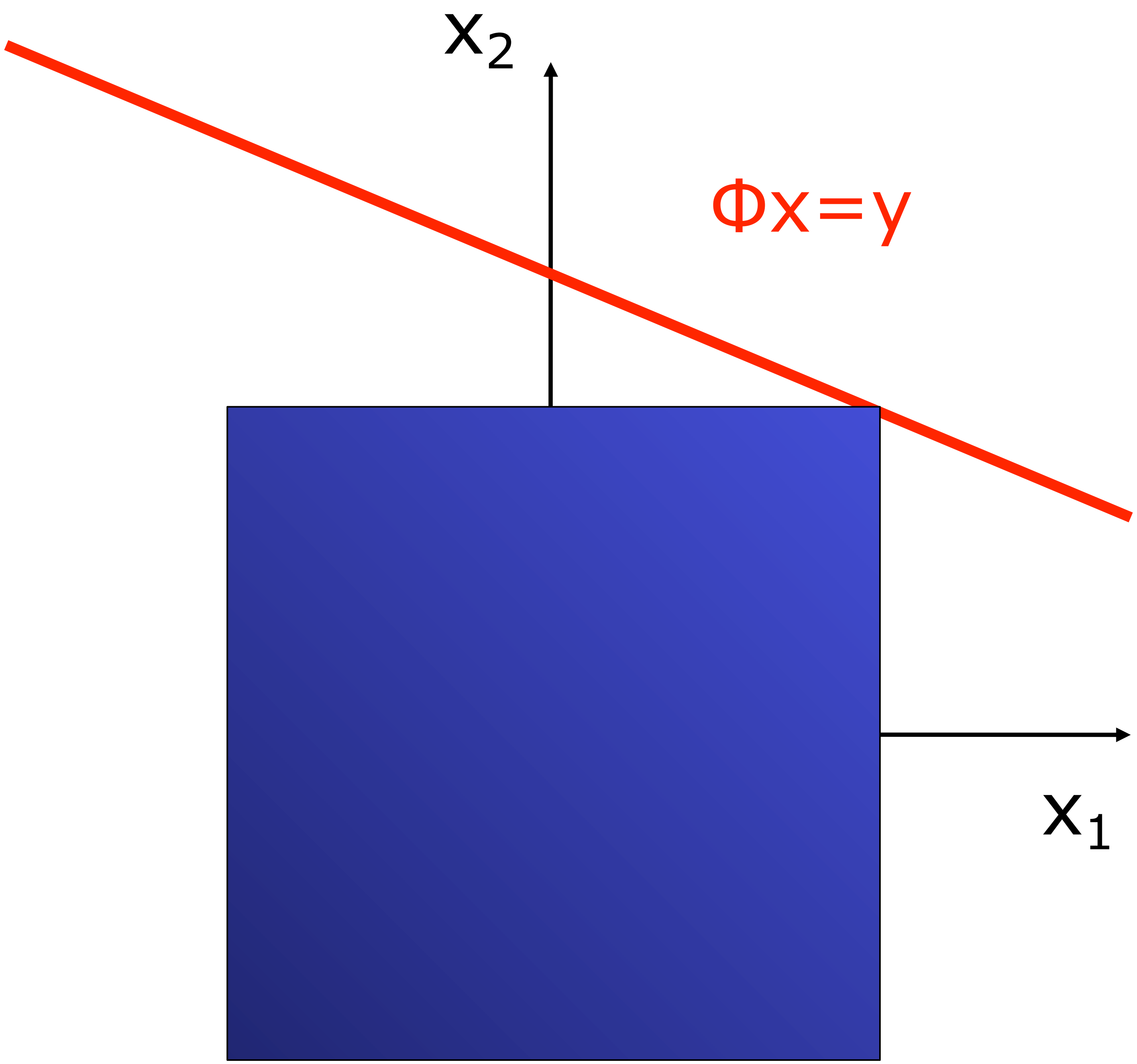
- Convex hull is the unit ball of the l_1 norm

$$\{x \ : \ \|x\|_\infty \leq 1\}$$

$$\|x\|_\infty = \max_i |x_i|$$



$$\begin{array}{ll} \text{minimize} & \|x\|_\infty \\ \text{subject to} & \Phi x = y \end{array}$$



Donoho and Tanner 2008
Mangasarian and R 2009.

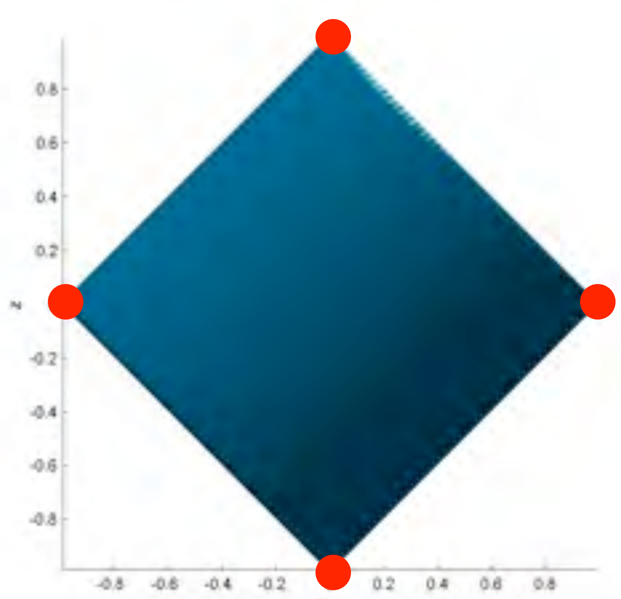
Parsimonious Models

rank

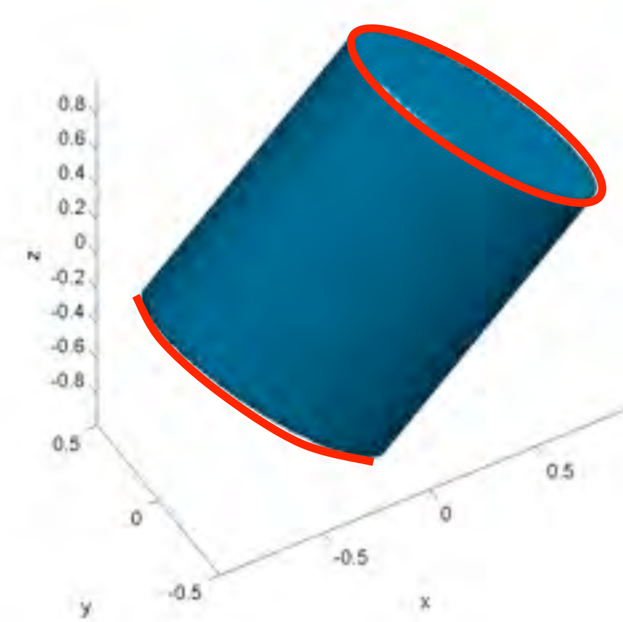
$$x = \sum_{k=1}^{r} w_k \alpha_k$$

model　　weights　　atoms

- Search for best linear combination of fewest atoms
- “rank” = fewest atoms needed to describe the model



$$\|x\|_{\mathcal{A}} \equiv \inf_{(w,\alpha)} \sum_{k=1}^{r} |w_k|$$



Atomic Norms

- Given a basic set of *atoms,* $\mathcal{A}$, define the function

$$\|x\|_{\mathcal{A}} = \inf\{t > 0 \ : \ x \in t\mathrm{conv}(\mathcal{A})\}$$

- When $\mathcal{A}$ is centrosymmetric, we get a norm

$$\|x\|_{\mathcal{A}} = \inf\{\sum_{a\in\mathcal{A}} |c_a| \ : \ x = \sum_{a\in\mathcal{A}} c_a a\}$$

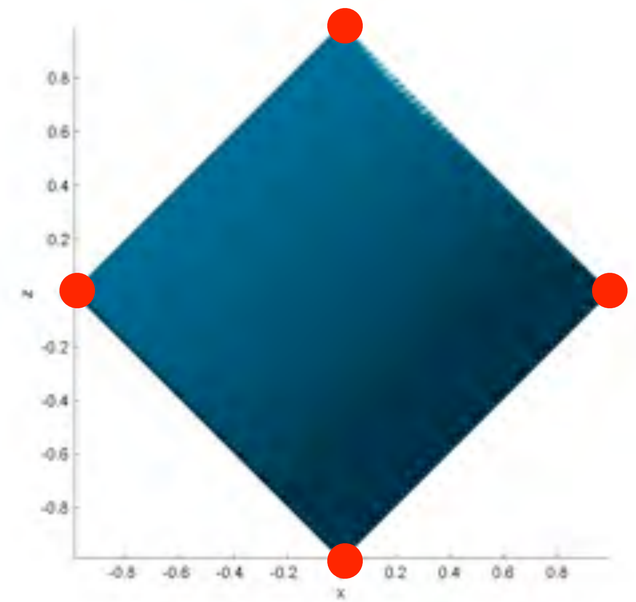
IDEA: $\begin{array}{ll} \text{minimize} & \|z\|_{\mathcal{A}} \\ \text{subject to} & \Phi z = y \end{array}$

- When does this work?

Atomic Norm Minimization

IDEA:

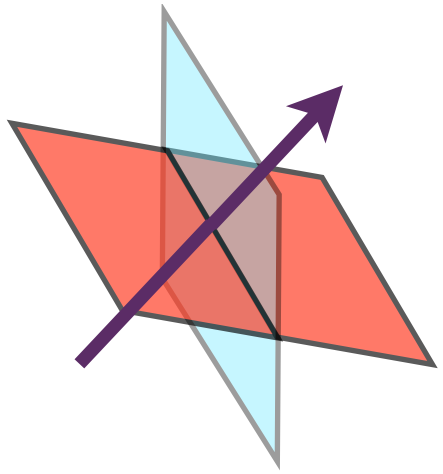
$$\begin{array}{ll} \text{minimize} & \|z\|_{\mathcal{A}} \\ \text{subject to} & \Phi z = y \end{array}$$

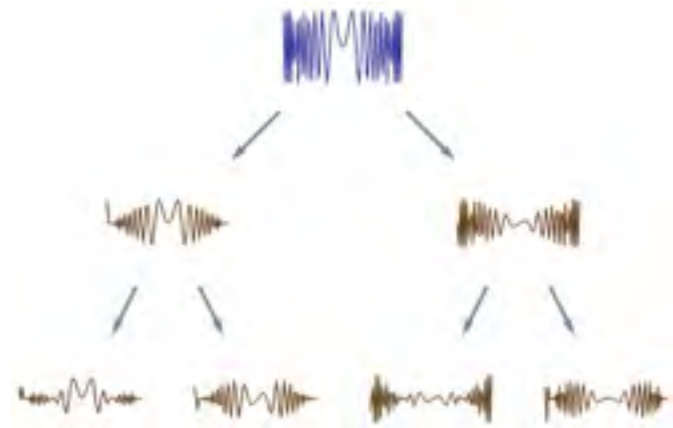


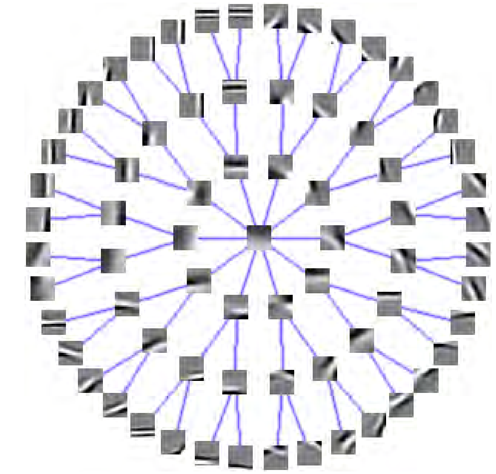
- Generalizes existing, powerful methods
- Rigorous formula for developing new analysis algorithms
- Precise, tight bounds on number of measurements needed for model recovery
- One algorithm prototype for a myriad of data-analysis applications

Chandrasekaran, R, Parrilo, and Willsky

Union of Subspaces





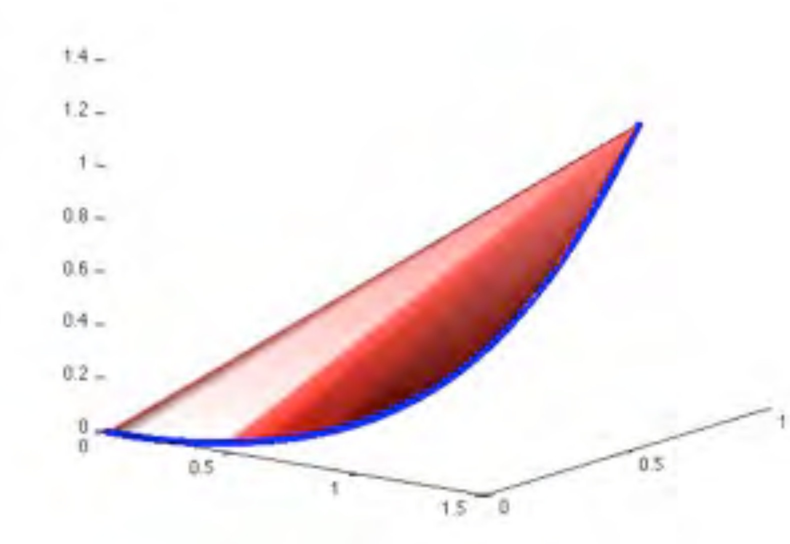


- X has structured sparsity: linear combination of elements from a set of subspaces $\{U_g\}$.
- Atomic set: unit norm vectors living in one of the U_g

Permutations and Rankings

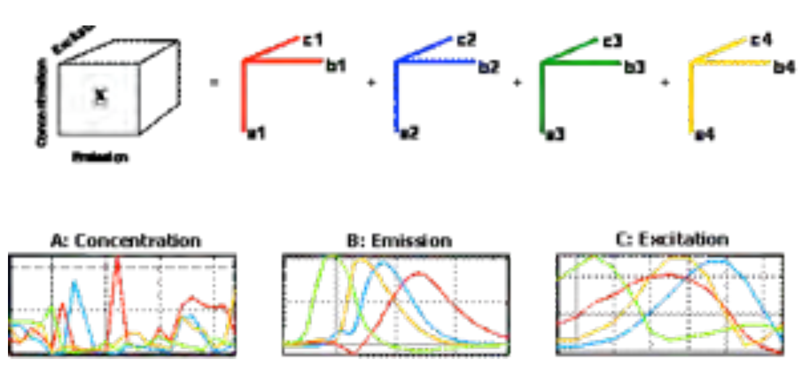
- X a sum of a few permutation matrices
- Examples: Multiobject Tracking, Ranked elections, BCS
- Convex hull of permutation matrices: doubly stochastic matrices.

- Moments: convex hull of of $[1,t,t^2,t^3,t^4,...]$, $t \in T$, some basic set.
- System Identification, Image Processing, Numerical Integration, Statistical Inference
- **Solve with semidefinite programming**



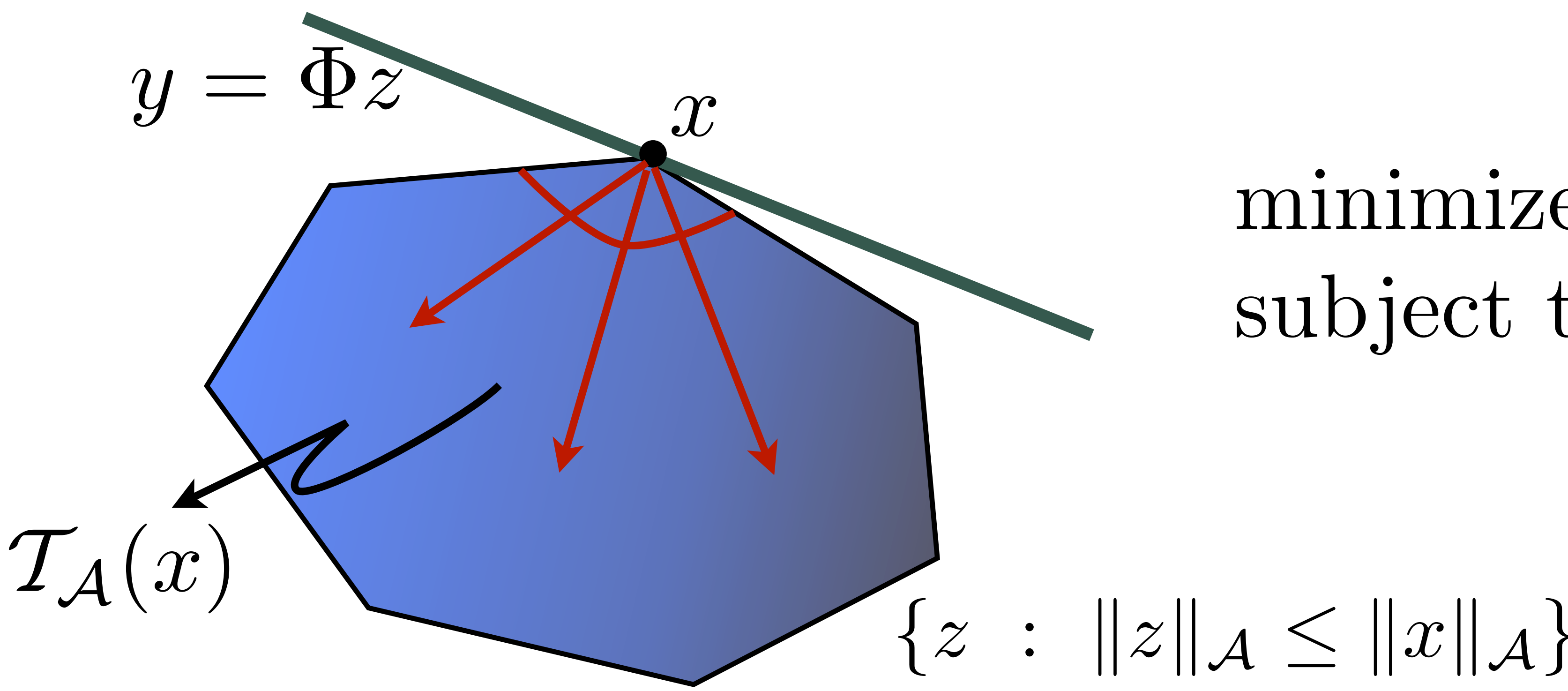
- Cut-matrices: sums of rank-one sign matrices.
- Collaborative Filtering, Clustering in Genetic Networks, Combinatorial Approximation Algorithms
- **Approximate with semidefinite programming**

- Low-rank Tensors: sums of rank-one tensors
- Computer Vision, Image Processing, Hyperspectral Imaging, Neuroscience
- **Approximate with alternating least-squares**



Tangent Cones

- Set of directions that decrease the norm from x form a cone:
  $\mathcal{T}_{\mathcal{A}}(x) = \{d \ : \ \|x + \alpha d\|_{\mathcal{A}} \leq \|x\|_{\mathcal{A}} \text{ for some } \alpha > 0\}$

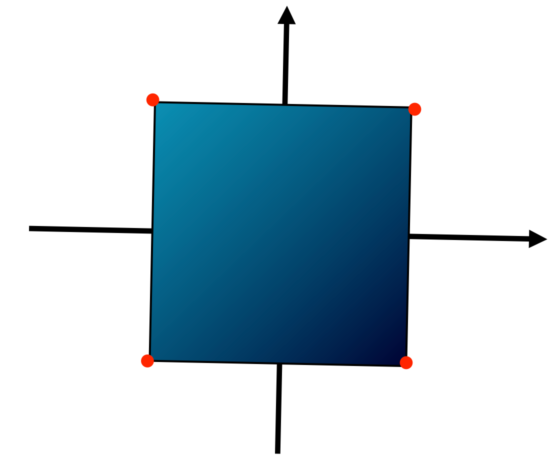


$$\begin{array}{ll} \text{minimize} & \|z\|_{\mathcal{A}} \\ \text{subject to} & \Phi z = y \end{array}$$

- x is the unique minimizer if the intersection of this cone with the null space of Φ equals {0}

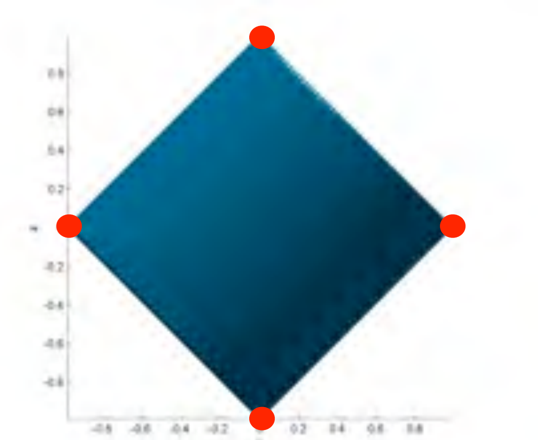
Rates

- Hypercube: $n \geq p/2$



- Sparse Vectors, p vector, sparsity s

$$n \geq 2s \log\left(\frac{p}{s}\right) + \frac{5s}{4}$$

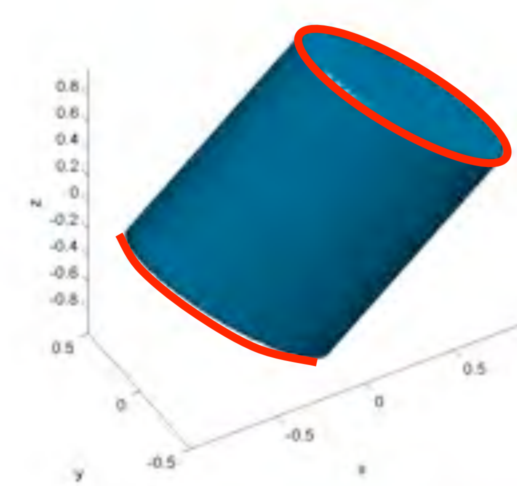


- Block sparse, M groups (possibly overlapping), maximum group size B, k active groups

$$n \geq k\left(\sqrt{2\log\left(M-k\right)} + \sqrt{B}\right)^2 + kB$$

- Low-rank matrices: p_1 x p_2, (p_1<p_2), rank r

$$n \geq 3r(p_1 + p_2 - r)$$



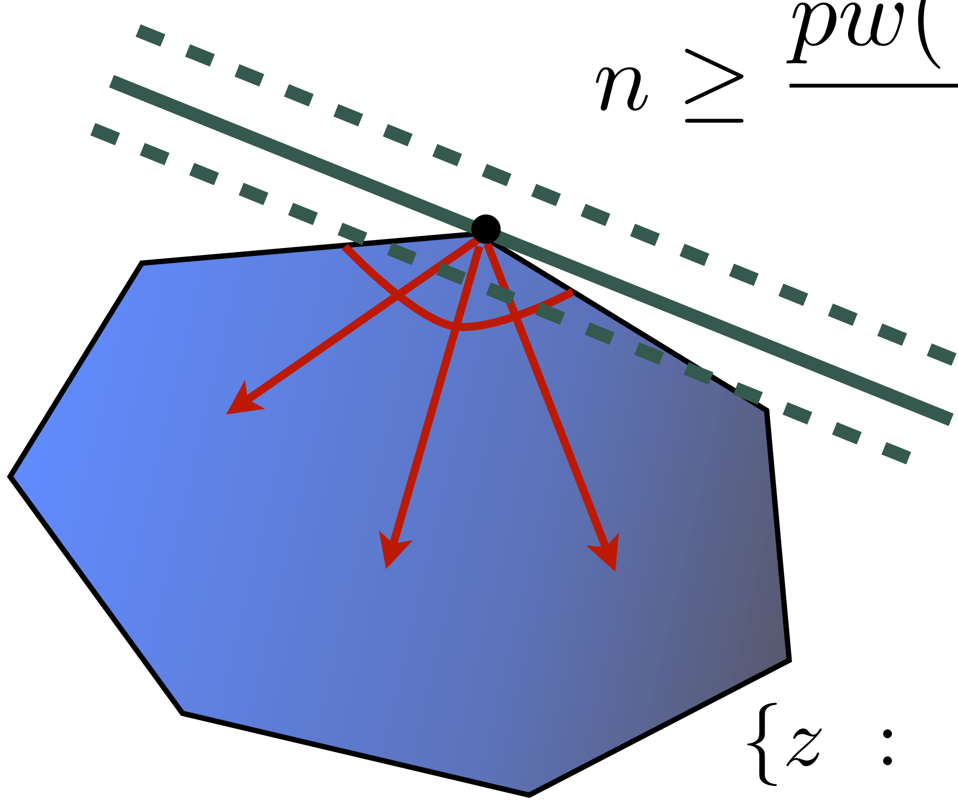
Robust Recovery (deterministic)

- Suppose we observe $y = \Phi x + w$ $\qquad \|w\|_2 \leq \delta$

$$\begin{array}{ll} \text{minimize} & \|z\|_{\mathcal{A}} \\ \text{subject to} & \|\Phi z - y\| \leq \delta \end{array}$$

- If $\hat{x}$ is an optimal solution, then $\|x - \hat{x}\| \leq \dfrac{2\delta}{\epsilon}$ provided that

$$n \geq \frac{p w(\mathcal{T}_{\mathcal{A}}(x) \cap \mathbb{S}^{p-1})^2}{(1-\epsilon)^2}$$

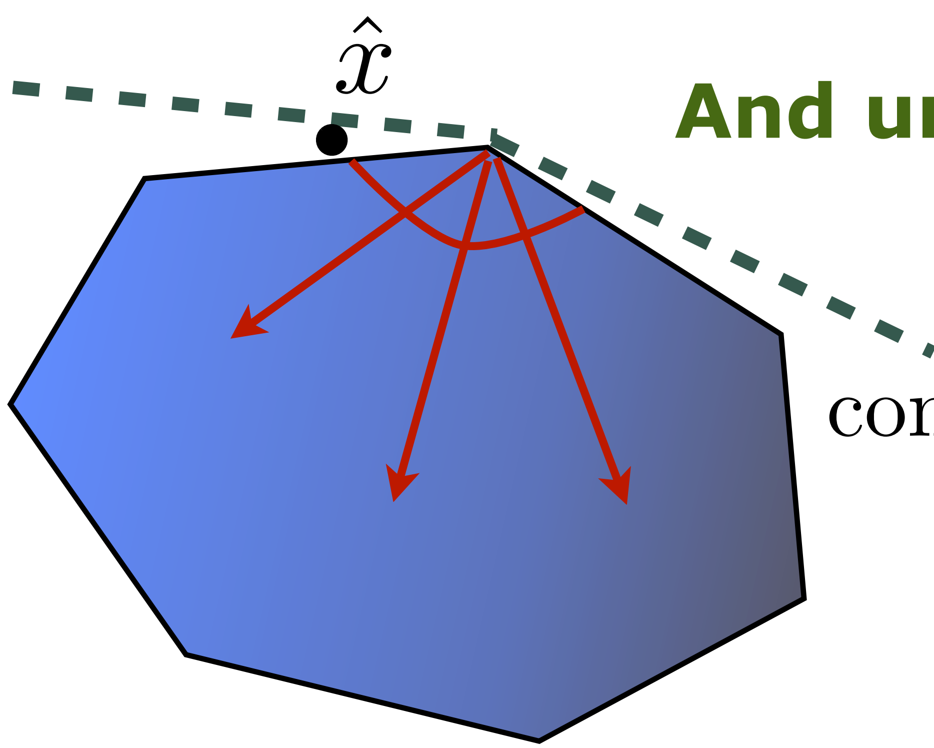


Robust Recovery (statistical)

- Suppose we observe $y = \Phi x + w$

$$\text{minimize} \quad \|\Phi z - y\|^2 + \mu \|z\|_{\mathcal{A}}$$

- If $\hat{x}$ is an optimal solution, then $\|\Phi x - \Phi \hat{x}\|_2 \leq \sqrt{\mu \|x\|_{\mathcal{A}}}$ provided that $\mu \geq \mathbb{E}_w[\|\Phi^* w\|_{\mathcal{A}}^*]$



And under an additional "cone condition"

$$\|x - \hat{x}\|_2 \leq \eta(x, \mathcal{A}, \Phi, \gamma)\mu$$

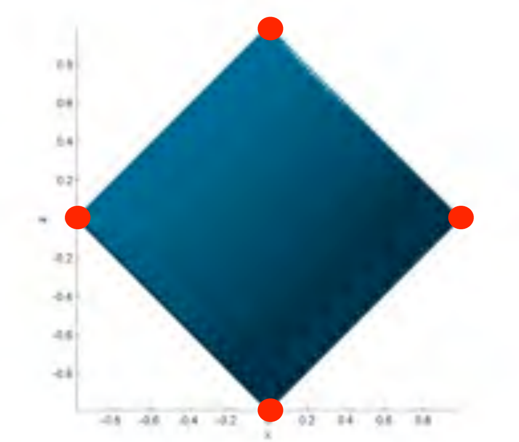
$$\text{cone}\{u \ : \ \|x + u\|_{\mathcal{A}} \leq \|x\|_{\mathcal{A}} + \gamma \|u\|\}$$

Bhaskar, Tang, and R 2011

Denoising Rates (re-derivations)

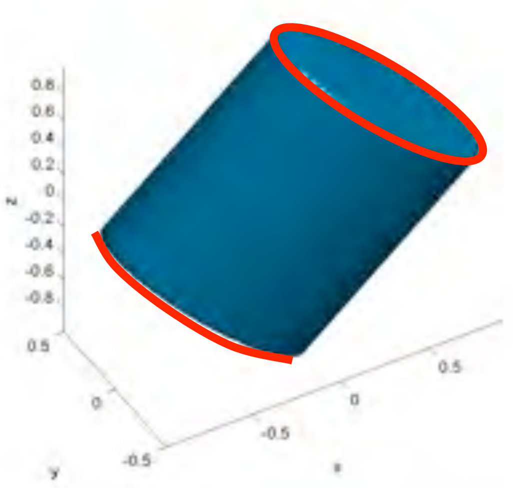
- Sparse Vectors, p vector, sparsity s

$$\frac{1}{p}\|\hat{x} - x^\star\|_2^2 = O\left(\frac{\sigma^2 s \log(p)}{p}\right)$$

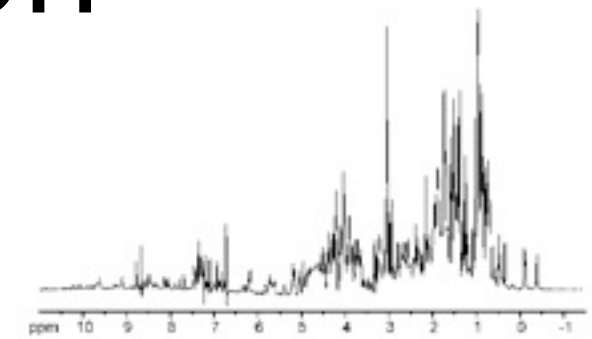


- Low-rank matrices: p_1 x p_2, (p_1<p_2), rank r

$$\frac{1}{p_1 p_2}\|\hat{x} - x^\star\|_F^2 = O\left(\frac{\sigma^2 r}{p_1}\right)$$



Spectrum Estimation



Observe a sparse combination of sinusoids

$$x_m^\star = \sum_{k=1}^{s} c_k^\star e^{i2\pi m u_k^\star} \quad \text{for some} \quad u_k^\star \in [0,1)$$

Observe: $y = x^\star + \omega$ (signal plus noise)

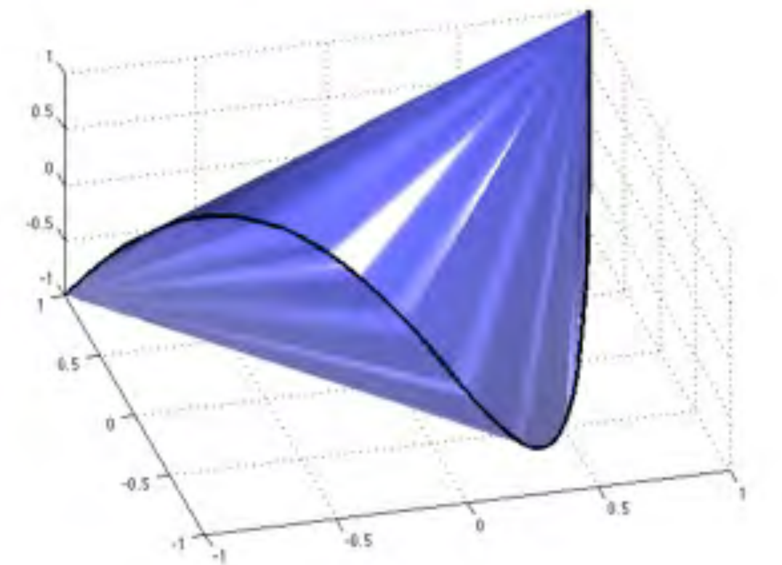
Atomic Set

$$\mathcal{A} = \left\{ \begin{bmatrix} e^{i\theta} \\ e^{i2\pi\phi+i\theta} \\ e^{i4\pi\phi+i\theta} \\ \dots \\ e^{i2\pi\phi n+i\theta} \end{bmatrix} : \begin{array}{l} \theta \in [0,2\pi), \\ \phi \in [0,1) \end{array} \right\}$$

Classical techniques (Prony, Matrix Pencil, MUSIC, ESPRIT, Cadzow), use the fact that noiseless moment matrices are low-rank:

$$\sum_{k=1}^{r} \alpha_k \begin{bmatrix} 1 \\ e^{\phi_k i} \\ e^{2\phi_k i} \\ e^{3\phi_k i} \end{bmatrix} \begin{bmatrix} 1 \\ e^{\phi_k i} \\ e^{2\phi_k i} \\ e^{3\phi_k i} \end{bmatrix}^* = \begin{bmatrix} \mu_0 & \mu_1 & \mu_2 & \mu_3 \\ \bar{\mu}_1 & \mu_0 & \mu_1 & \mu_2 \\ \bar{\mu}_2 & \bar{\mu}_1 & \mu_0 & \mu_1 \\ \bar{\mu}_3 & \bar{\mu}_2 & \bar{\mu}_1 & \mu_0 \end{bmatrix} \succeq 0$$

Atomic Norm for Spectrum Estimation



IDEA:

$$\begin{array}{ll} \text{minimize} & \|z\|_{\mathcal{A}} \\ \text{subject to} & \|\Phi z - y\| \leq \delta \end{array}$$

Atomic Set:

$$\mathcal{A} = \left\{ \begin{bmatrix} e^{i\theta} \\ e^{i2\pi\phi + i\theta} \\ e^{i4\pi\phi + i\theta} \\ \vdots \\ e^{i2\pi\phi n + i\theta} \end{bmatrix} : \begin{array}{l} \theta \in [0, 2\pi), \\ \phi \in [0, 1) \end{array} \right\}$$

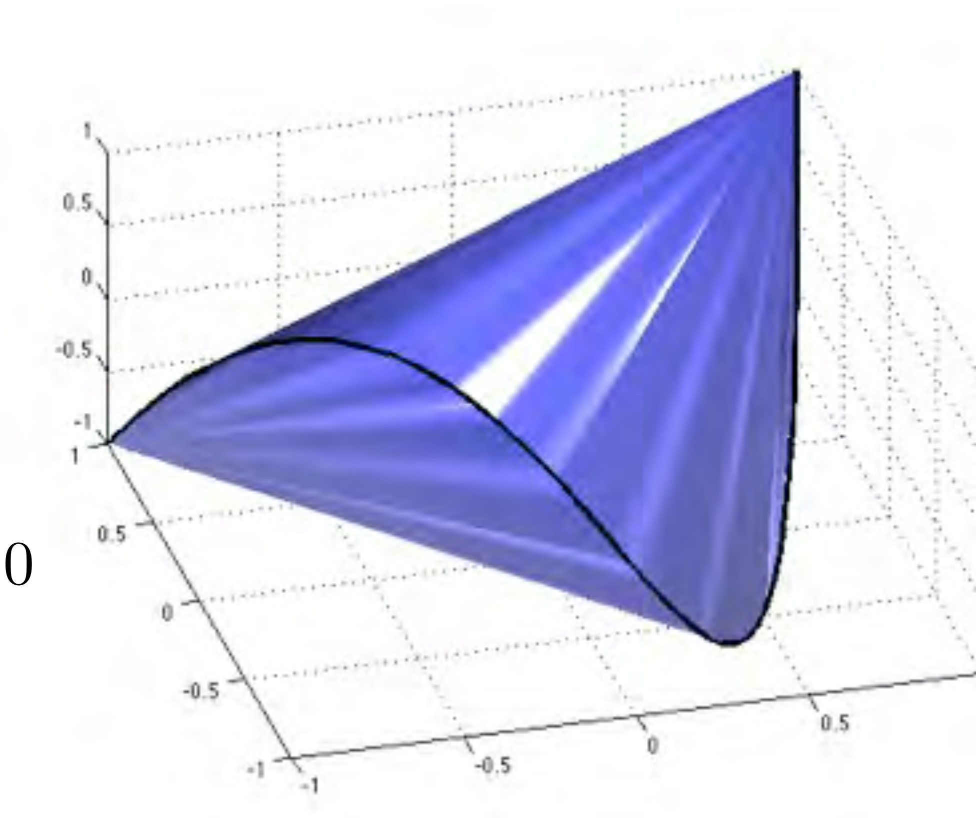
- How do we solve the optimization problem?
- Can we approximate the true signal from partial and noisy measurements?
- Can we estimate the frequencies from partial and noisy measurements?

Which atomic norm for sinusoids?

$$x_m = \sum_{k=1}^{s} c_k e^{i2\pi m u_k} \quad \text{for some} \quad u_k \in [0,1)$$

Assume c_k are positive for simplicity

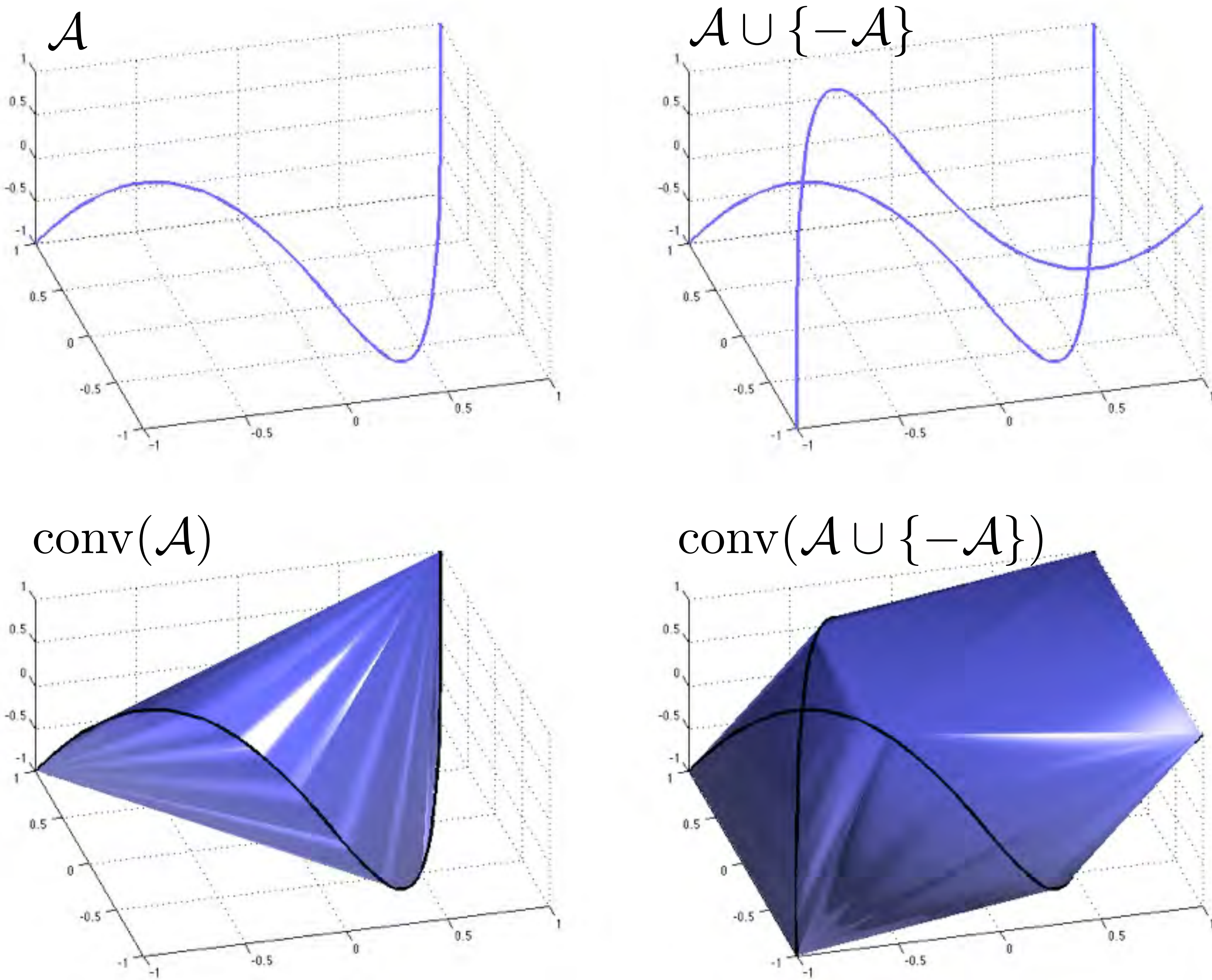
$$\sum_{k=1}^{r} c_k \begin{bmatrix} 1 \\ e^{i2\pi u_k} \\ e^{i4\pi u_k} \\ e^{i6\pi u_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{i2\pi u_k} \\ e^{i4\pi u_k} \\ e^{i6\pi u_k} \end{bmatrix}^* = \begin{bmatrix} x_0 & \bar{x}_1 & \bar{x}_2 & \bar{x}_3 \\ x_1 & x_0 & \bar{x}_1 & \bar{x}_2 \\ x_2 & x_1 & x_0 & \bar{x}_1 \\ x_3 & x_2 & x_1 & x_0 \end{bmatrix} \succeq 0$$



- Convex hull is characterized by linear matrix inequalities (Toeplitz positive semidefinite)

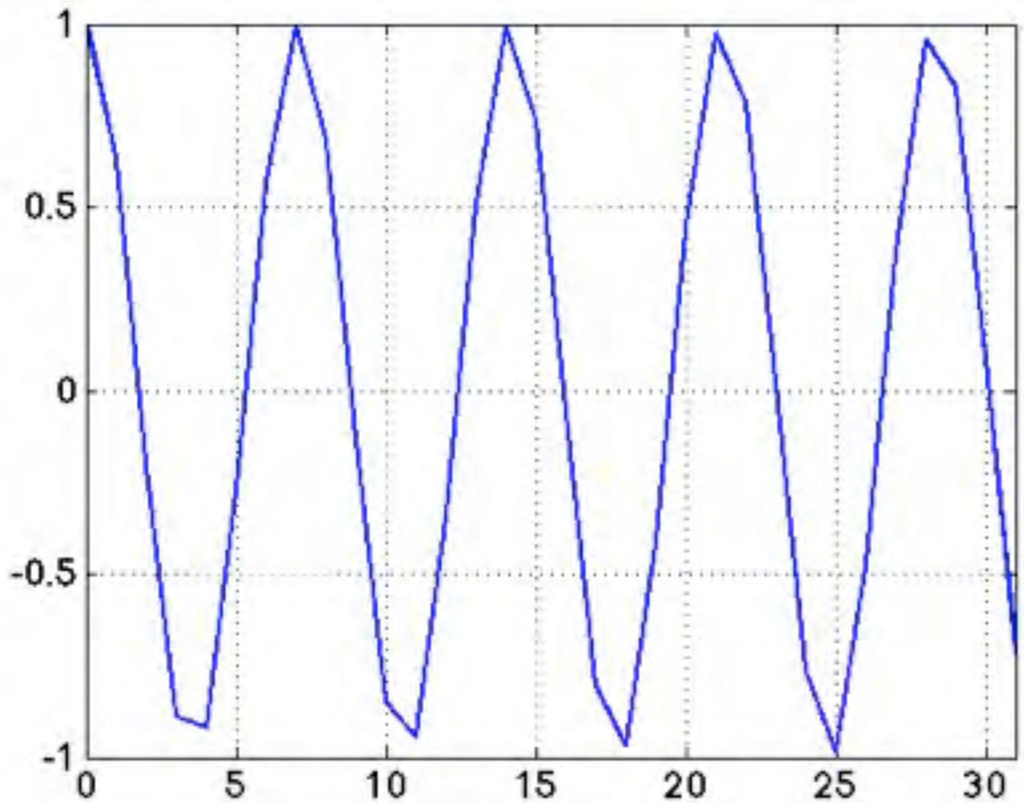
$$\|x\|_{\mathcal{A}} = \inf \left\{ \tfrac{1}{2} t + \tfrac{1}{2} w_0 \ : \ \begin{bmatrix} t & x^* \\ x & \operatorname{toep}(w) \end{bmatrix} \succeq 0 \right\}$$

- *Moment Curve* of [t,t^2,t^3,t^4,...], $t \in \mathbb{S}^1$

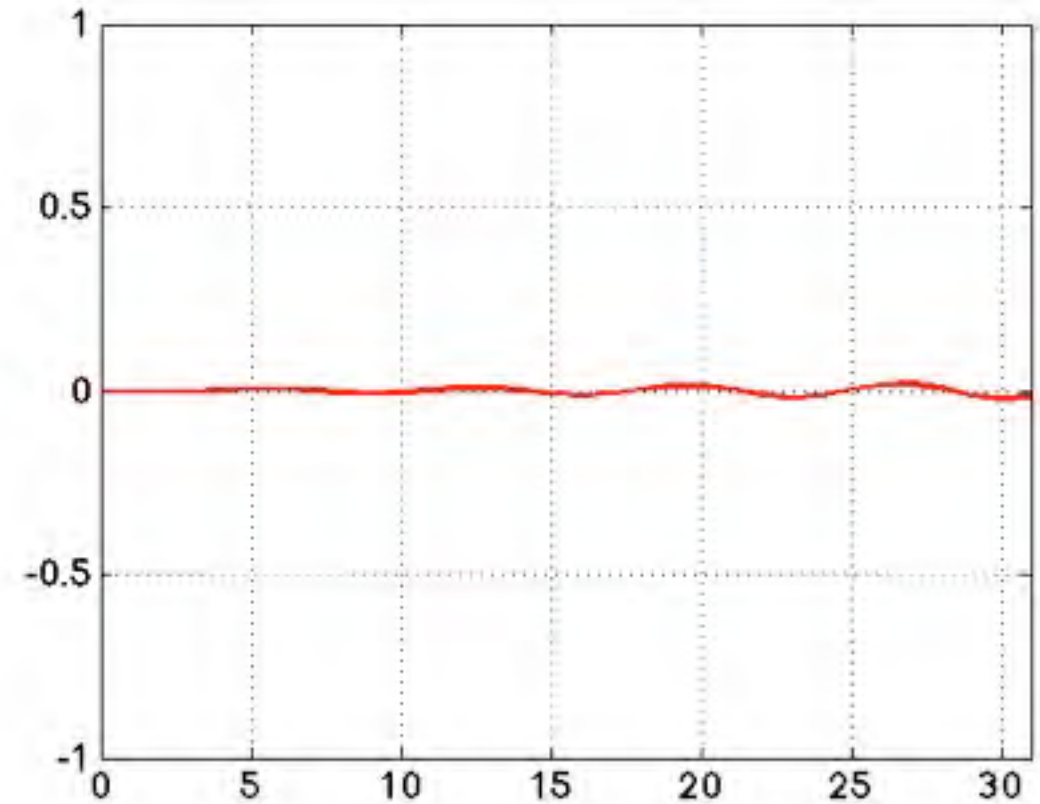


$\mathcal{A}$
$\mathcal{A} \cup \{-\mathcal{A}\}$
$\mathrm{conv}(\mathcal{A})$
$\mathrm{conv}(\mathcal{A} \cup \{-\mathcal{A}\})$

$$\cos(2\pi u_1 k)/2 + \cos(2\pi u_2 k)/2$$



$$\cos(2\pi u_1 k)/2 - \cos(2\pi u_2 k)/2$$

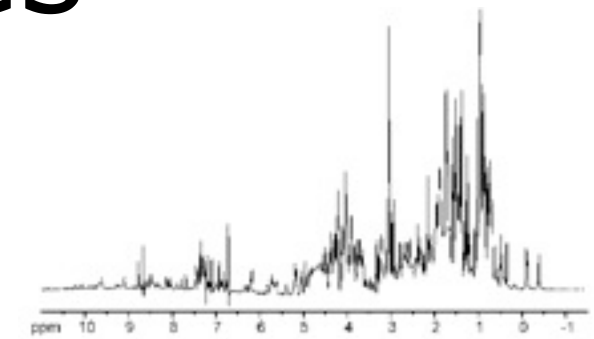


$$u_1 = 0.1413$$
$$u_2 = 0.1411$$

Nearly optimal rates

Observe a sparse combination of sinusoids

$$x_m^\star = \sum_{k=1}^{s} c_k^\star e^{i2\pi m u_k^\star} \quad \text{for some} \quad u_k^\star \in [0,1)$$



Observe: $y = x^\star + \omega$ (signal plus noise)

Assume frequencies are far apart: $\min_{p \neq q} d(u_p, u_q) \geq \frac{4}{n}$

Solve:

$$\hat{x} = \arg\min_x \tfrac{1}{2}\|x - y\|_2^2 + \mu\|x\|_{\mathcal{A}}$$

Error Rate:

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

No algorithm can do better than

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

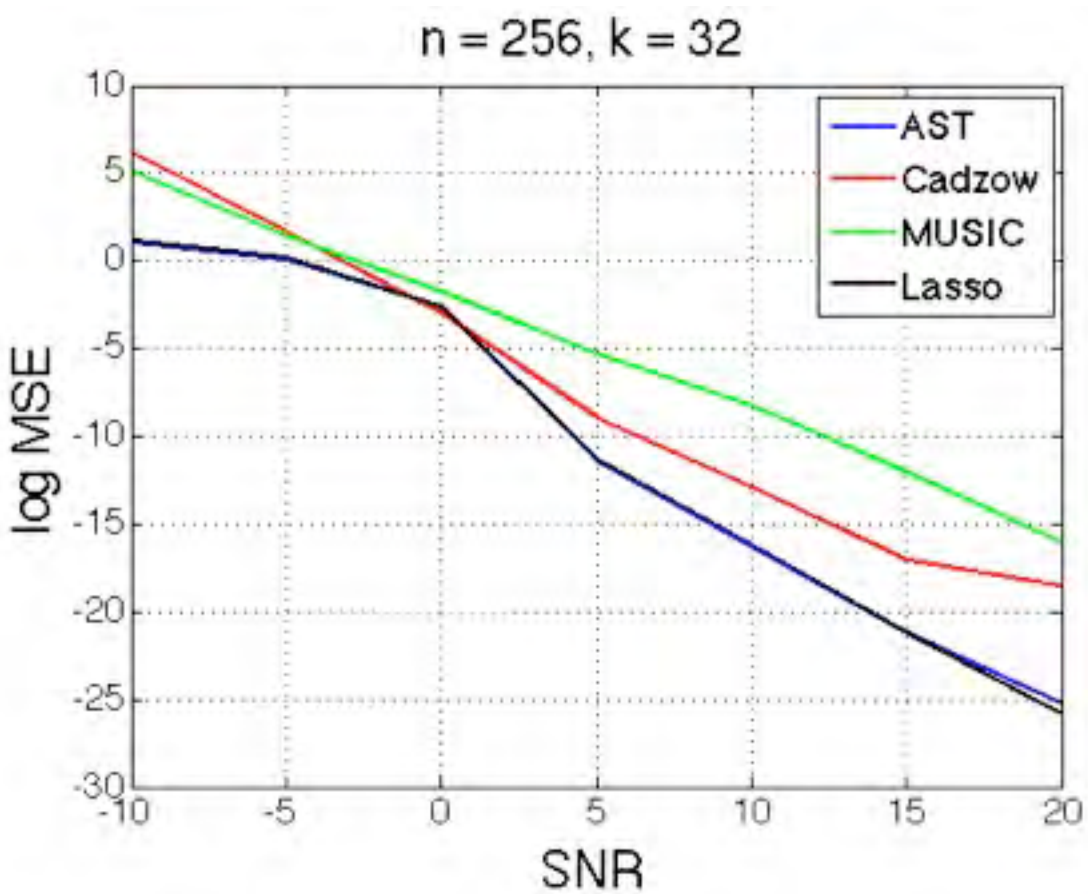
even if the frequencies are well-separated

No algorithm can do better than

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

even if we knew all of the frequencies (u_k^*)

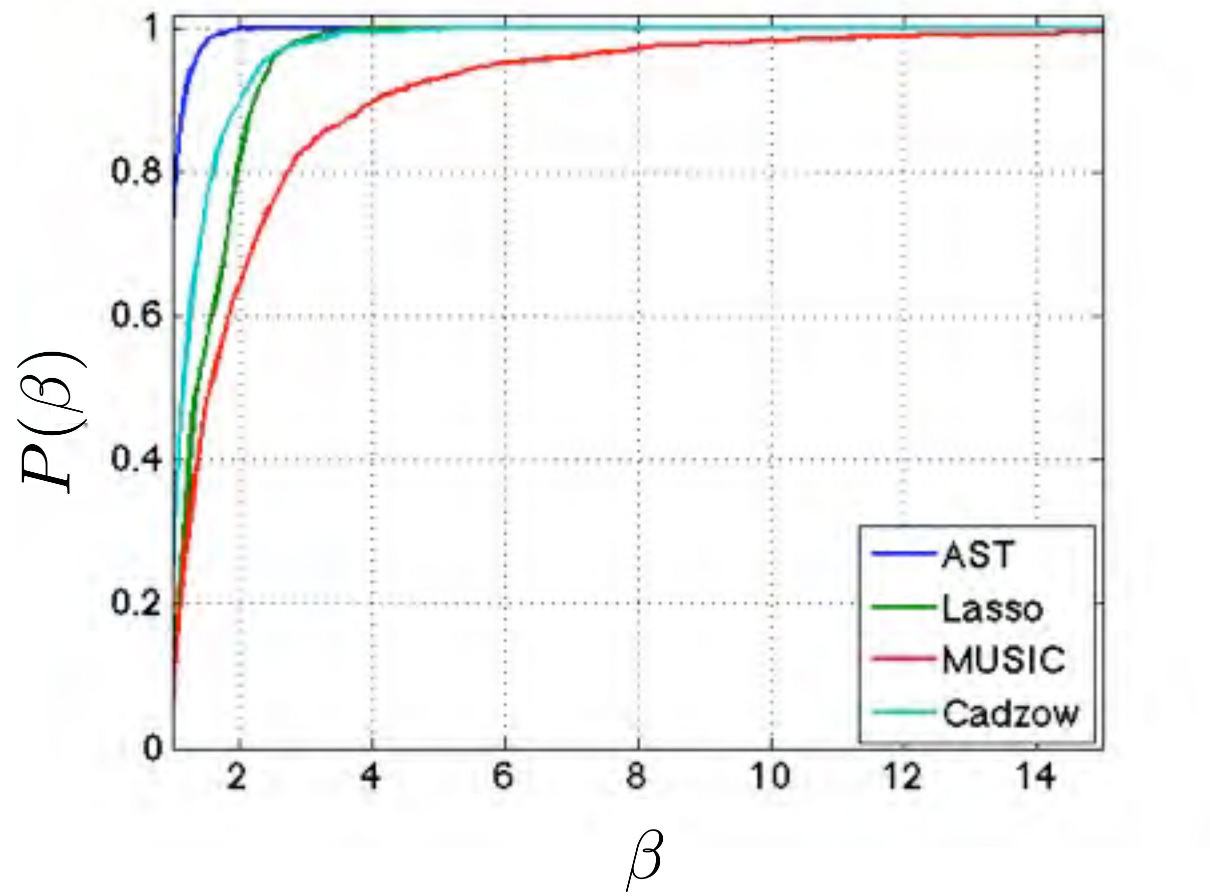
Mean Square Error



- Frequencies generated at random with 1/n separation. Random phases, fading amplitudes.
- Cadzow and MUSIC provided model order. AST and LASSO estimate noise power.

Lower is better

Performance Profile

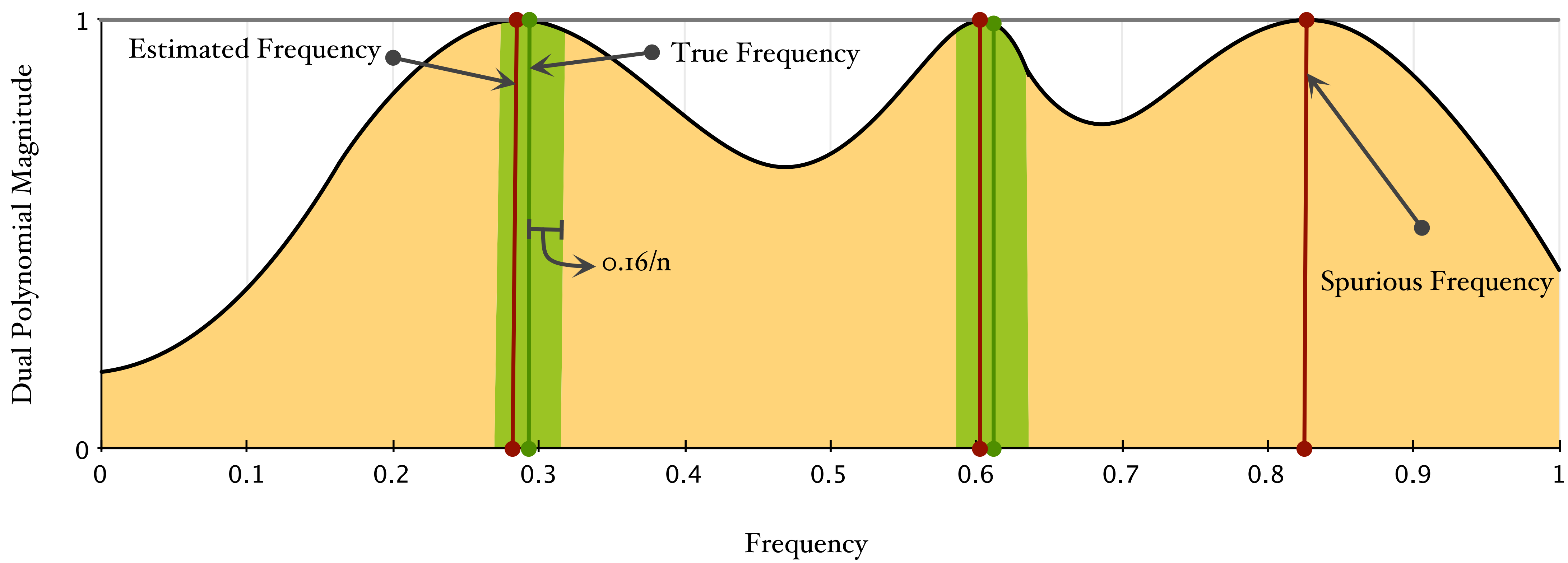


- Performance profile over all parameter values and settings
- For algorithm s:

$$P_s(\beta) = \frac{\#\left\{p \in \mathcal{P} \ : \ \mathrm{MSE}_s(p) \leq \beta \min_s \mathrm{MSE}_s(p)\right\}}{\#(\mathcal{P})}$$

Higher is better

Localization Guarantees



Spurious Amplitudes

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

Weighted Frequency Deviation

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

Near region approximation

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

Frequency Localization

Spurious Amplitudes

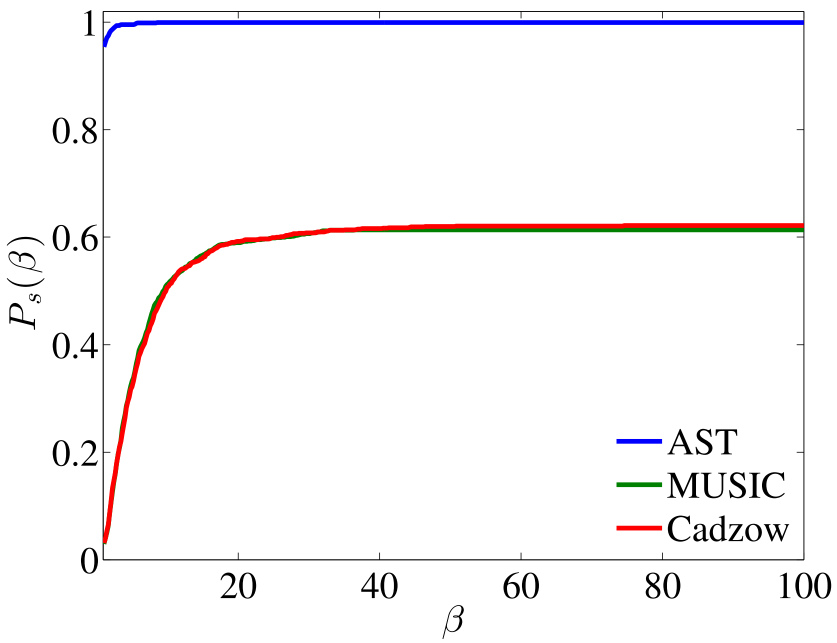
$$\sum_{l:\hat{f}_l \in F} |\hat{c}_l| \leq C_1 \sigma \sqrt{\frac{k^2 \log(n)}{n}}$$

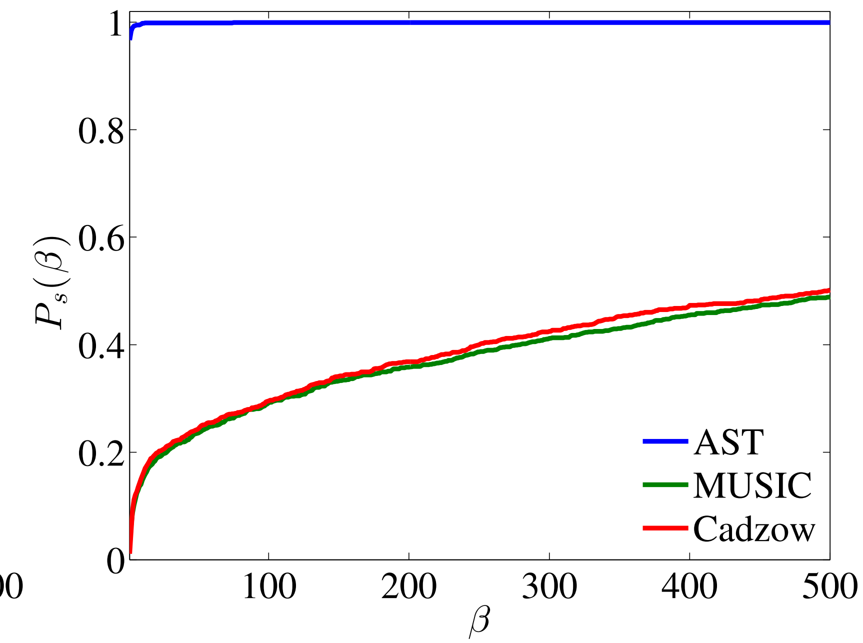
Weighted Frequency Deviation

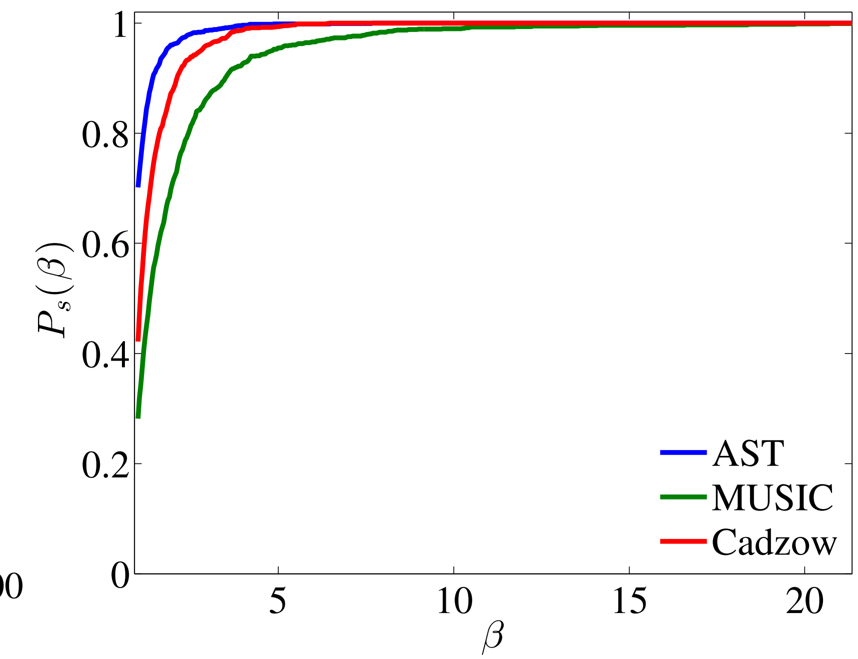
$$\sum_{l:\hat{f}_l \in N_j} |\hat{c}_l| \left\{ \min_{f_j \in T} d(f_j, \hat{f}_l) \right\}^2 \leq C_2 \sigma \sqrt{\frac{k^2 \log(n)}{n}}$$

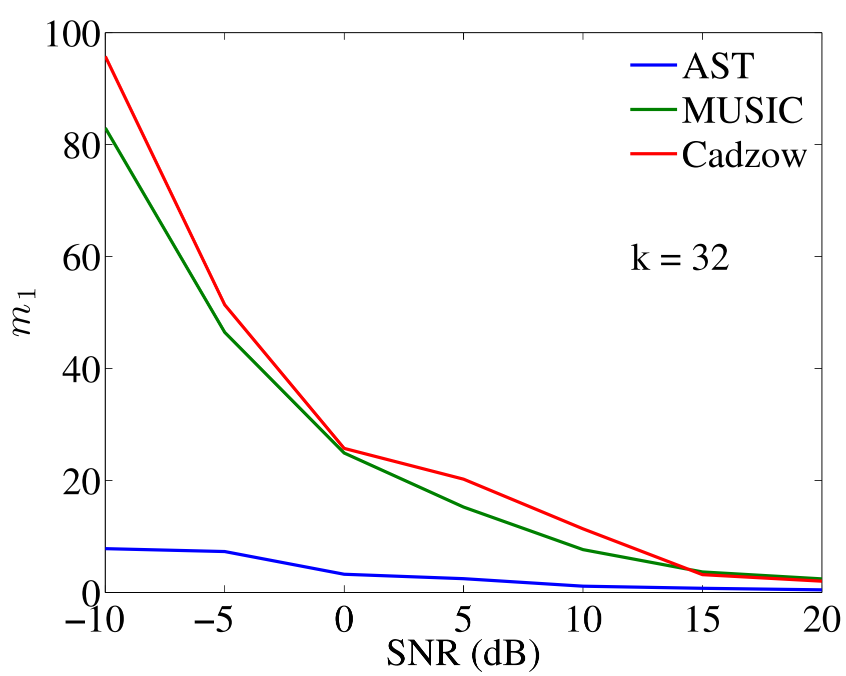
Near region approximation

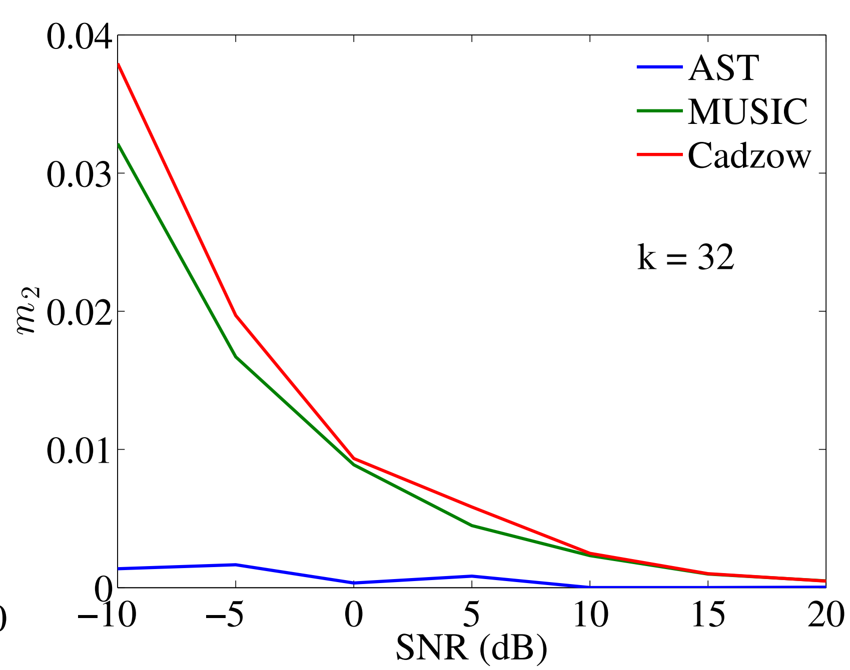
$$\left| c_j - \sum_{l:\hat{f}_l \in N_j} \hat{c}_l \right| \leq C_3 \sigma \sqrt{\frac{k^2 \log(n)}{n}}$$

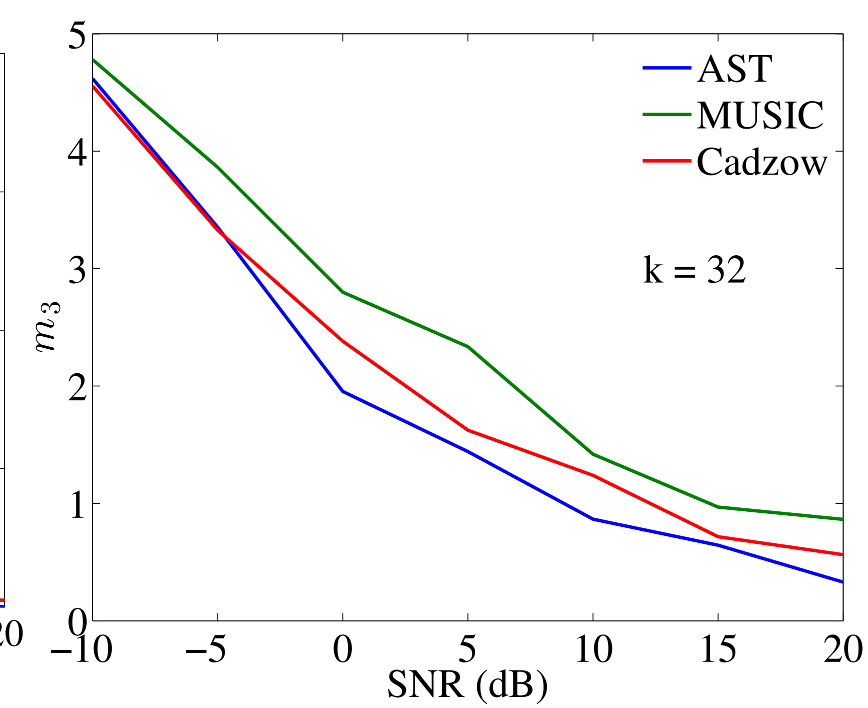












Incomplete Data/Random Sampling

- Observe a random subset of samples

$$T \subset \{0, 1, \ldots, n-1\}$$

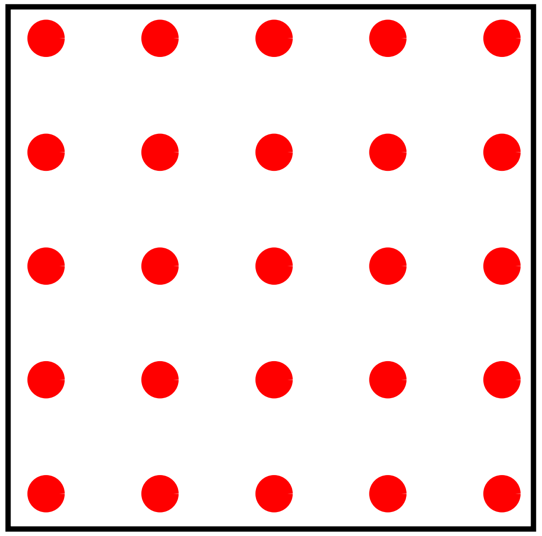
- On the grid, Candes, Romberg & Tao (2004)

- Off the grid, new, Compressed Sensing extended to the wide continuous domain

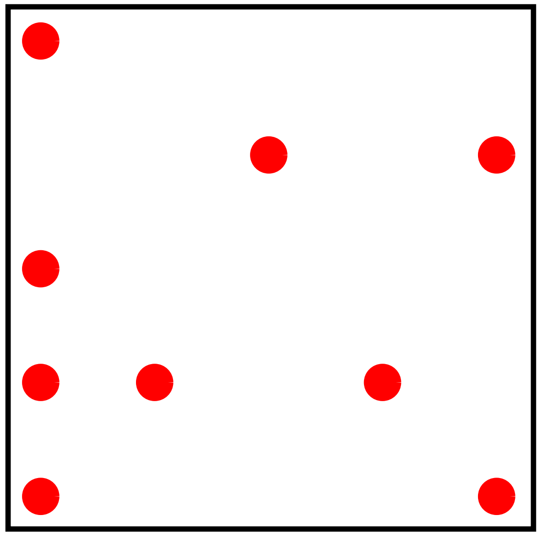
- Recover the missing part by solving

$$\begin{aligned} \underset{z}{\text{minimize}} \ & \|z\|_{\mathcal{A}} \\ \text{subject to} \ & z_T = x_T. \end{aligned}$$

- Extract frequencies from the dual optimal solution

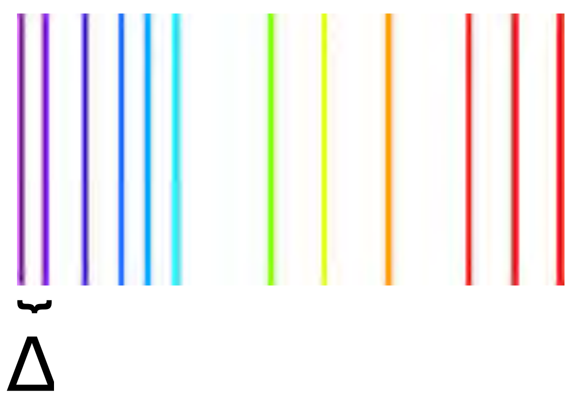


Full observation



Random sampling

Exact Recovery (off the grid)



$$x_m = \sum_{k=1}^{s} c_k \exp(2\pi i m u_k)$$

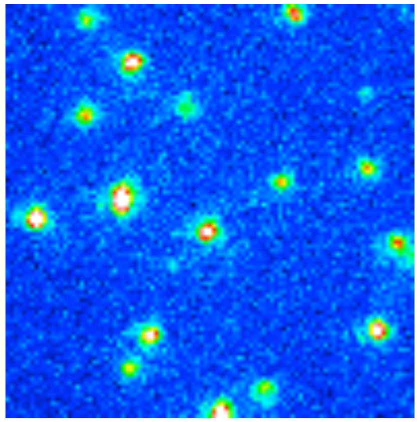
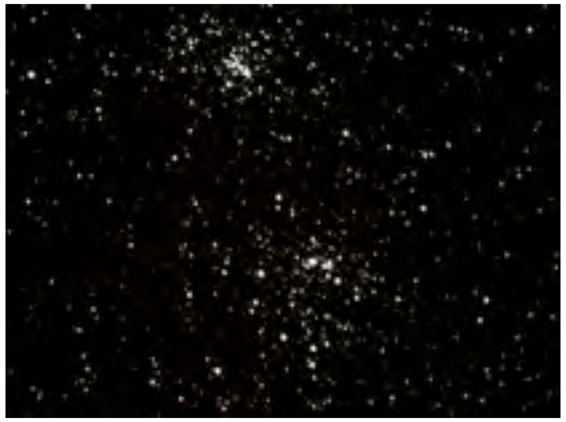
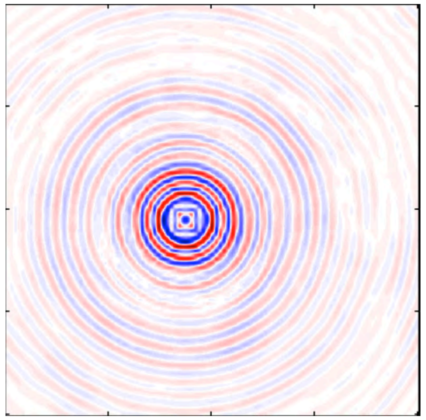
WANT $\{u_k, c_k\}$

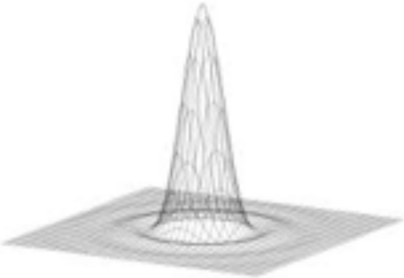
Theorem: (Candès and Fernandez-Granda 2012) A line spectrum with minimum frequency separation Δ > 4/s can be recovered from the first 2s Fourier coefficients via atomic norm minimization.

Theorem: (Tang, Bhaskar, Shah, and R. 2012) A line spectrum with minimum frequency separation Δ > 4/n can be recovered from most subsets of the first n Fourier coefficients of size at least O(s log(s) log(n)).

- s random samples are better than s equispaced samples.
  - On a grid, this is just compressed sensing
  - Off the grid, this is new
- No balancing of coherence as n grows.

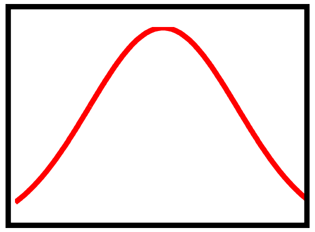
imaging | astronomy | seismology

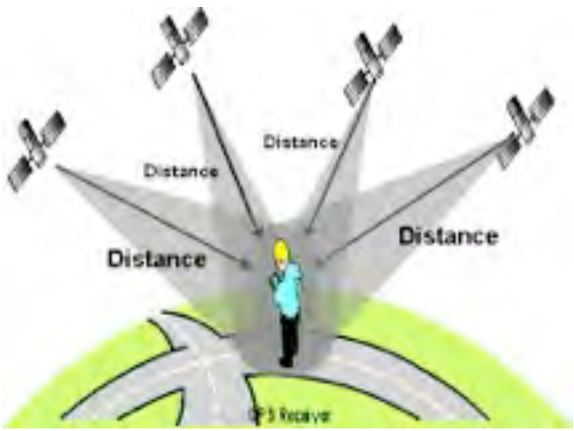




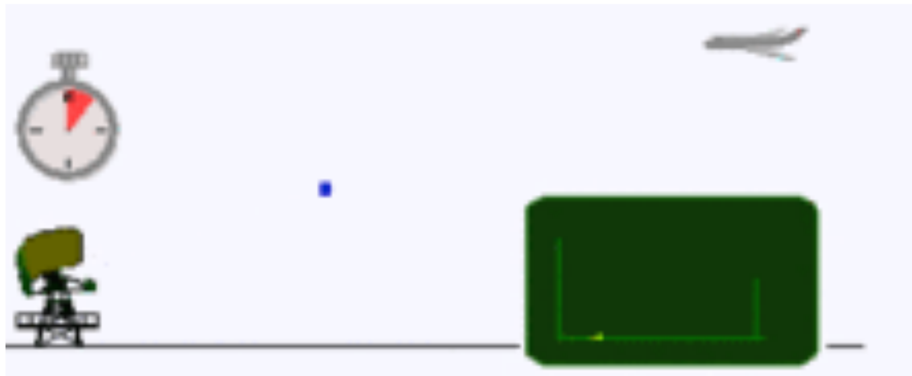


$$\hat{x}(\omega) = \widehat{\mathrm{PSF}}(\omega) \sum_{j=1}^{k} c_j e^{-i2\pi\omega s_j}$$

$$x(t) = \sum_{j=1}^{k} c_j g(t - \tau_j) e^{i2\pi f_j t}$$





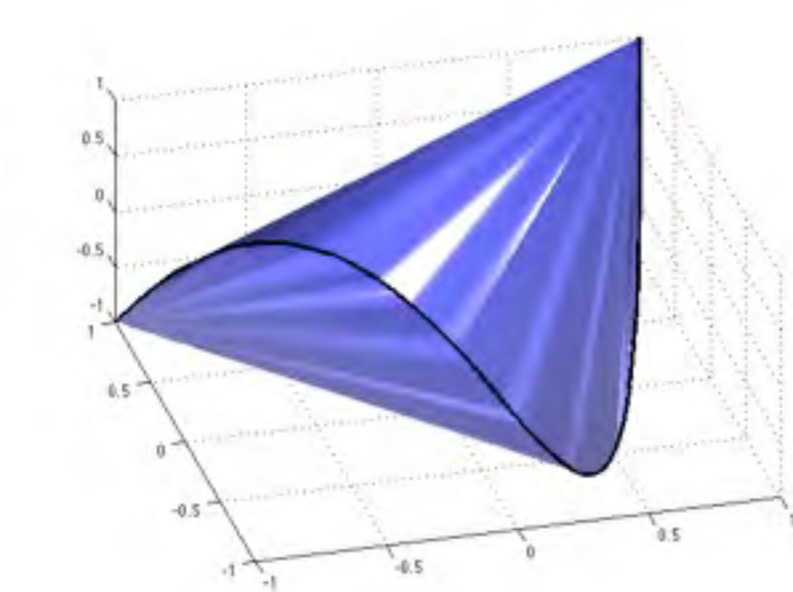



GPS | Radar | Ultrasound

Summary

- Unified approach for continuous problems in estimation.
- Obviates incoherence and basis mismatch
- New insight into bounds on estimation error and exact recovery through convex duality and algebraic geometry
- State-of-the-art results in classic fields of spectrum estimation and system identification (ask me later!)

- Recovery of general sparse signal trains
- Scaling atomic norm algorithms
- More atomization in signal processing... (moments, PDEs, deep atoms)



Acknowledgements

Work developed with Venkat Chandrasekaran, Babak Hassibi (Caltech), Weiyu Xu (Iowa), Pablo A. Parrilo, Alan Willsky (MIT), Maryam Fazel (Washington) Badri Bhaskar, Rob Nowak, Nikhil Rao, Gongguo Tang (Wisconsin).





For all references, see:

`http://pages.cs.wisc.edu/~brecht/publications.html`

References

- Atomic norm denoising with applications to line spectral estimation. Badri Narayan Bhaskar, Gongguo Tang, and Benjamin Recht. Submitted to *IEEE Transactions on Signal Processing*. 2012.

- Compressed sensing off the grid. Gongguo Tang, Badri Bhaskar, Parikshit Shah, and Benjamin Recht. Submitted to *IEEE Transactions on Information Theory. 2012.*

- Near Minimax Line Spectral Estimation. Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Submitted to *IEEE Transactions on Information Theory, 2*013.

- The convex geometry of inverse problems. Venkat Chandrasekaran, Benjamin Recht, Pablo Parrilo, and Alan Willsky. To Appear in *Foundations on Computational Mathematics.* 2012.

- All references can be found at

`http://pages.cs.wisc.edu/~brecht/publications.html`