Parameter Estimation in Compressive Sensing: The Delay-Doppler Case

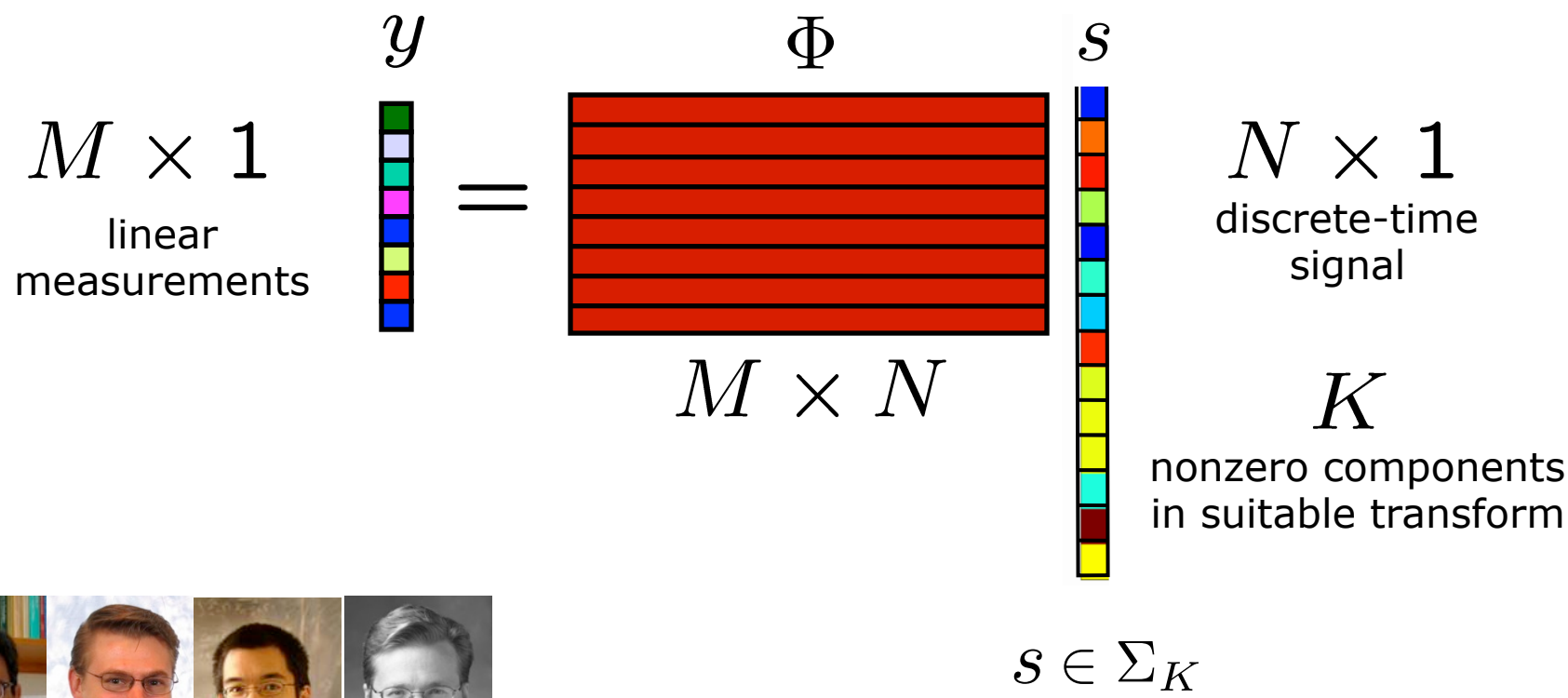
Marco F. Duarte

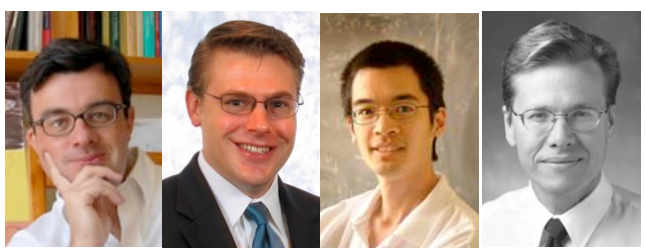
UMASS AMHERST

CoSeRa - Bonn, September 18 2013

Compressive Sensing (CS)

- Integrates linear acquisition with dimensionality reduction





[Candès, Romberg, Tao; Donoho 2006]

Radar: The Basics (1-D)

- Transmit low-autocorrelation signal $f(t)$, (e.g., Alltop seq.), receive reflection from point source at range x and speed v

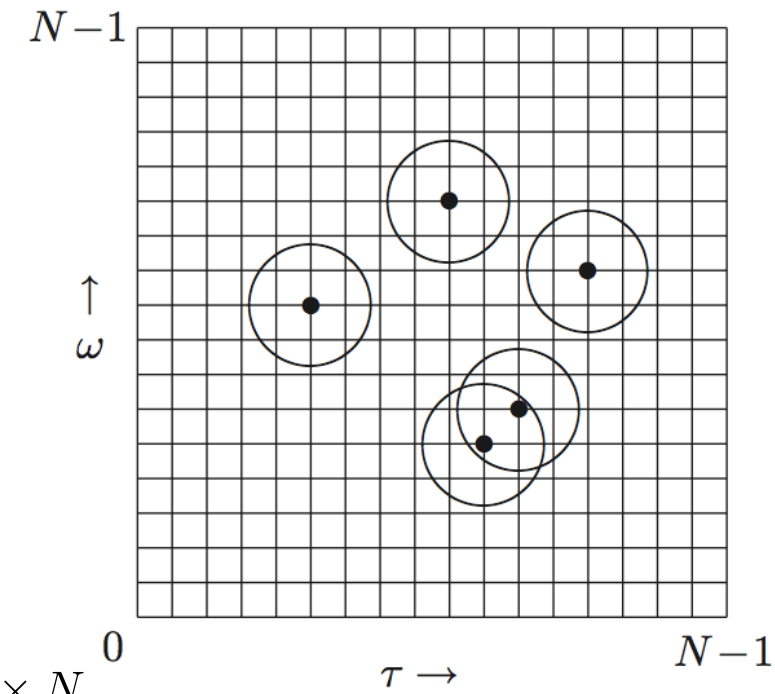
$$r(t) = s_{xv} f(t - \tau_x) e^{j2\pi\omega_v t}$$

- Range and velocity (x,v) of the target can be inferred from time delay-doppler shift (τ_x, ω_v)

CS Radar: The Basics (1-D)

- Time-domain sampling (N_t), ***discretization*** of observable ranges (N_x) and velocities (N_v) (parameter space sampling) provides an easy-to-implement ***Gabor frame*** as a sparsity dictionary for observations:

$$\Psi = [\Psi^{(0)} \ \Psi^{(1)} \ \ldots \ \Psi^{(N_x-1)}]$$



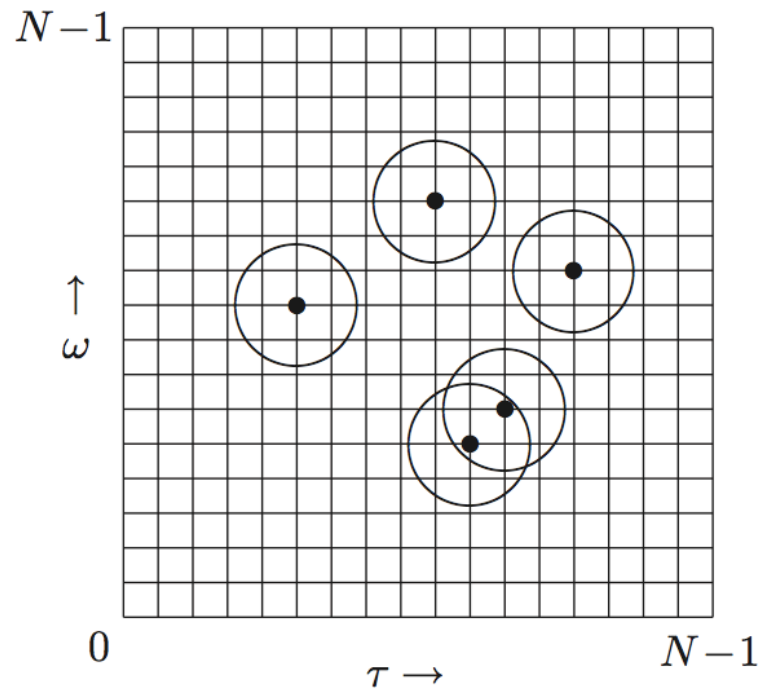
  - N_x Components $\Psi^{(i)} = D_i W_{N_v}$ of size $N_t \times N_v$
  - D_i: Diagonal matrix w/samples of f, shifted by i
  - W_{N_v} : Modulation matrix with entries
    $W_{N_v}[p,q] = e^{j2\pi pq/N_v}$

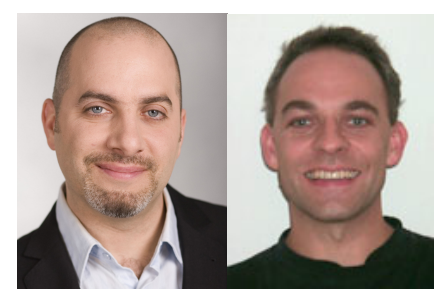
- Each sparsity dictionary element (shift/modulation) linked to ***delay-Doppler pair value***

[Herman and Strohmer 2009]

CS Radar: The Basics (1-D)

- Observations from multiple point sources are ***aggregated*** by the transmission media, providing sampled measurements $r = \Psi s$
- CS measurements: $y = \Phi r = \Phi \Psi s$
- Recover vector s, "read" out delay-Doppler map
- Vector s is sparse ***only*** in the case when the target's delay/doppler is ***among values sampled*** in the Gabor frame





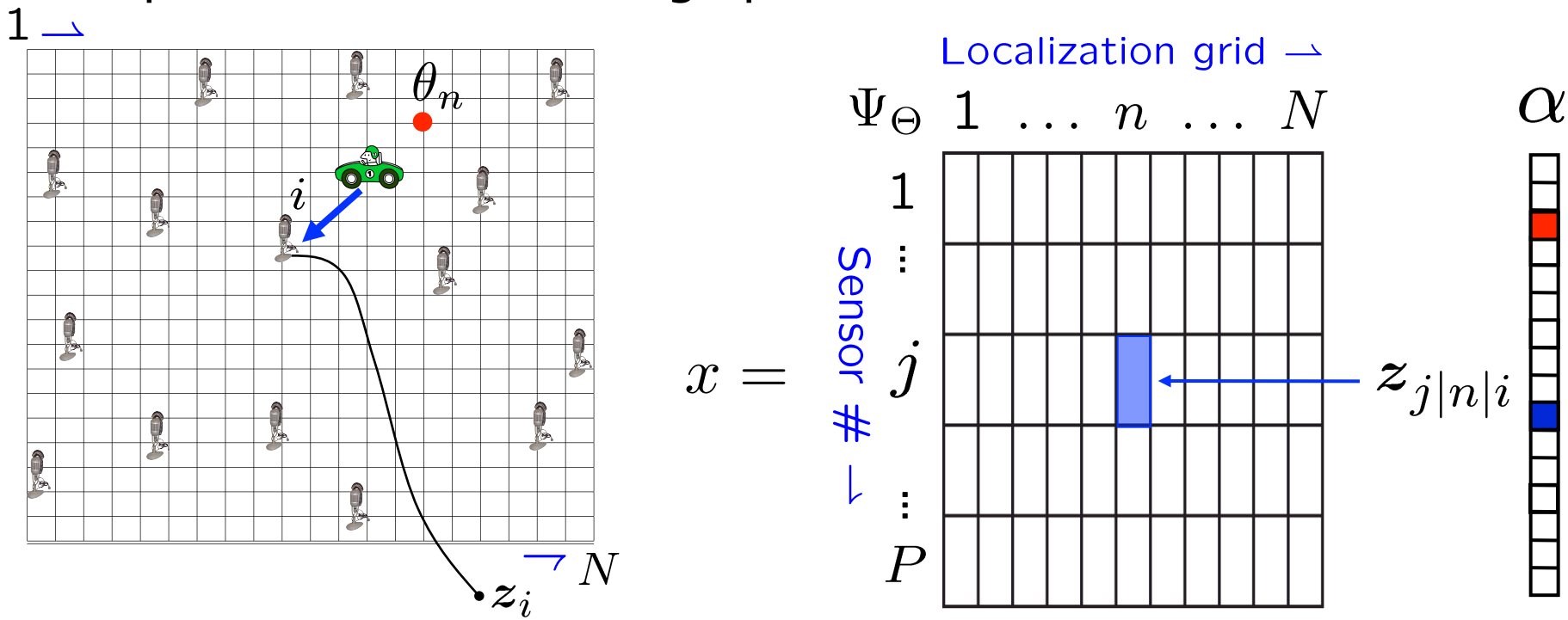
[Herman and Strohmer 2009]

Parametric Dictionaries for Sparsity

- Integrates sparsity/CS with ***parameter estimation***
- ***Parametric dictionaries*** (PDs) collect observations for a set of values of parameter of interest (one per column)

$$\Theta = \{\theta_1, \ldots, \theta_N\}$$

- Simple signals (e.g., few localization targets) can be expressed via PDs using sparse coefficient vectors



[Gorodntisky and Rao 1997] [Malioutov, Cetin, Willsky 2005]
[Cevher, Duarte, Baraniuk 2008] [Cevher, Gurbuz, McClellan, Chellapa 2008][...]

Resolution in Frequency Domain

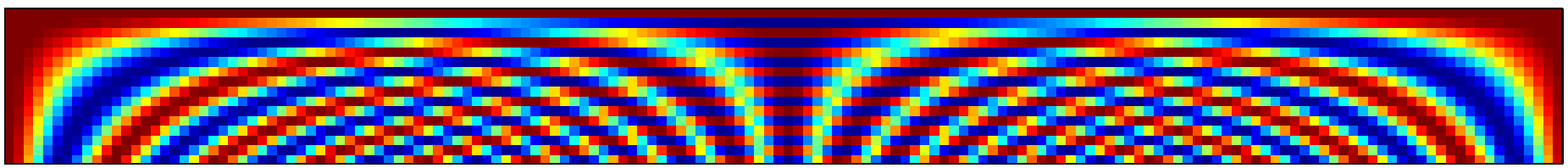
- Redundant Fourier Frame $\Psi(c)$

$$\Psi(c) = \left[e\left(\frac{1}{c}\right) \; e\left(\frac{2}{c}\right) \; \ldots \; e\left(\frac{N - 1/c}{c}\right) \right]$$

$$e(f) = \left[e^{j2\pi f/N} \; e^{j2\pi 2f/N} \; \ldots \; e^{j2\pi(N-1)f/N} \right]^T$$

$$N = 1024$$

$$\Psi(c),\, c = 10$$



- ***Increased resolution*** allows for more scenes to be formulated as sparse in parametric dictionary

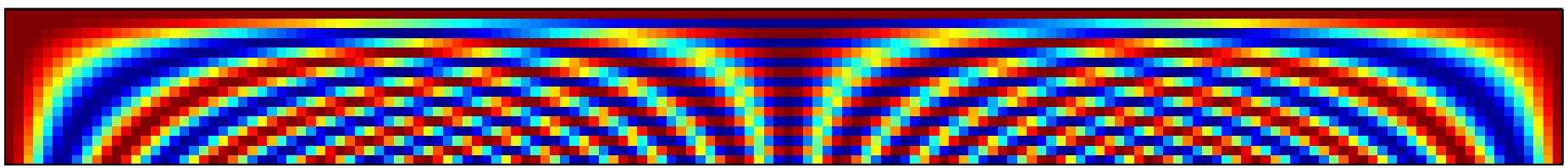
Resolution in Frequency Domain

- Redundant Fourier Frame $\Psi(c)$

$$\Psi(c) = \left[e\left(\frac{1}{c}\right) \; e\left(\frac{2}{c}\right) \; \ldots \; e\left(\frac{N - 1/c}{c}\right) \right]$$

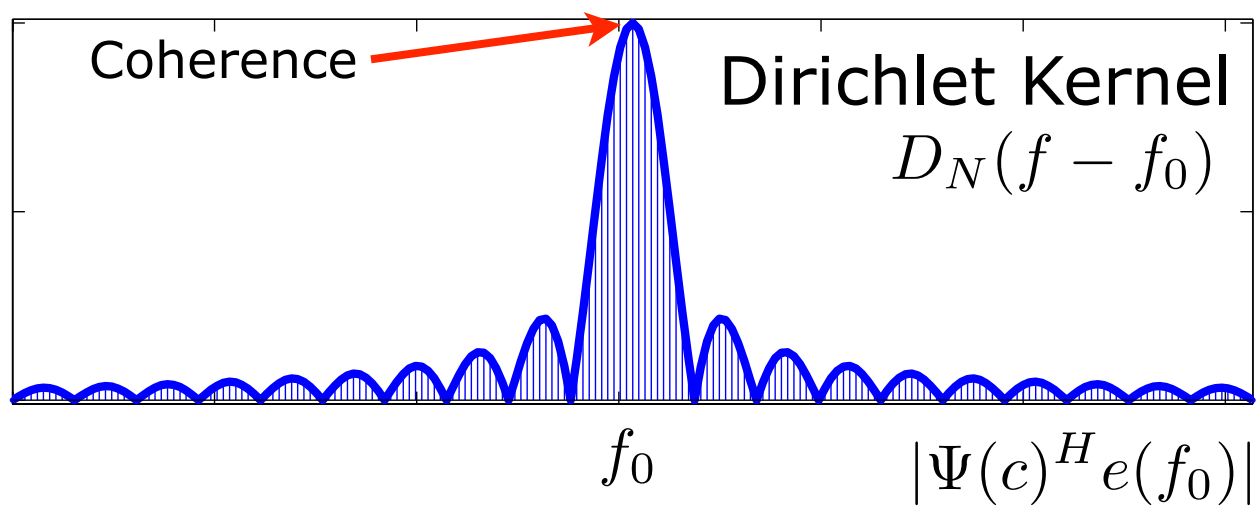
$N = 1024$

$\Psi(c),\ c = 10$



Recovery algorithms operate similarly to ***matched filtering***:

$p = \Psi(c)^H s$



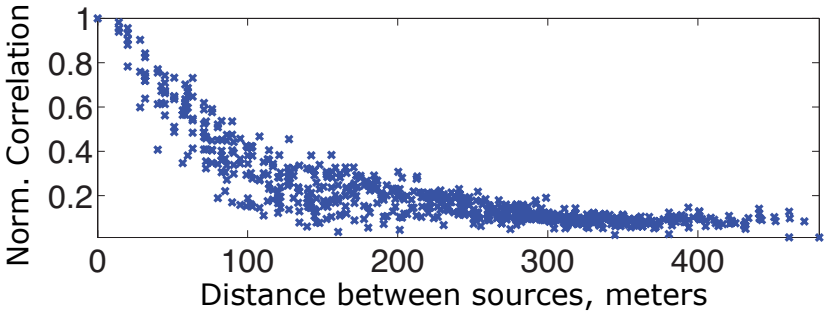
Part 1: Dealing with Coherence

Rich Baraniuk
(Rice U.)

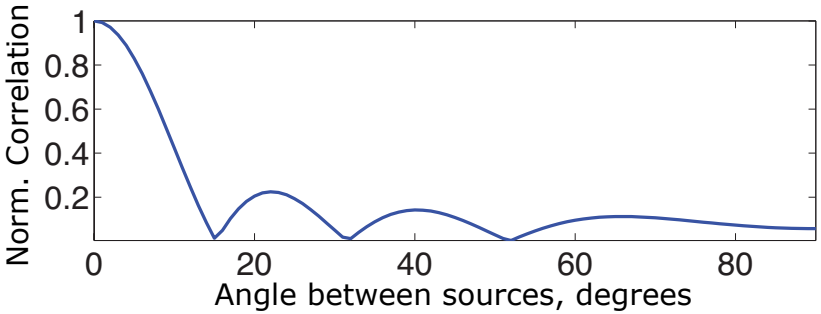
Issues with Parametric Dictionaries

- As parameter resolution ***increases*** (e.g., larger number of grid points), PD becomes ***increasingly coherent***, hampering sparse approximation algorithms
- PD's high coherence is a manifestation of ***resolution issues*** in underlying estimation problem

- ***Structured sparsity models*** can mitigate this issues by preventing PD elements with coherence above target maximum ν from appearing simultaneously in recovered signal



Near-Field Localization

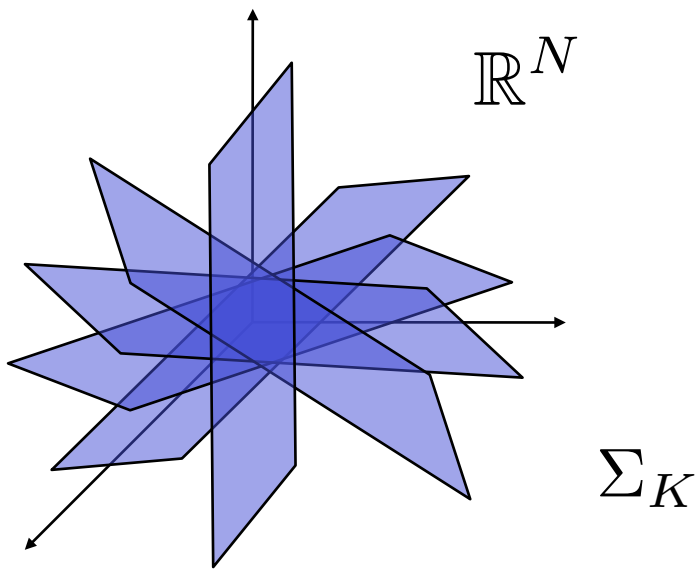


Far-Field Localization

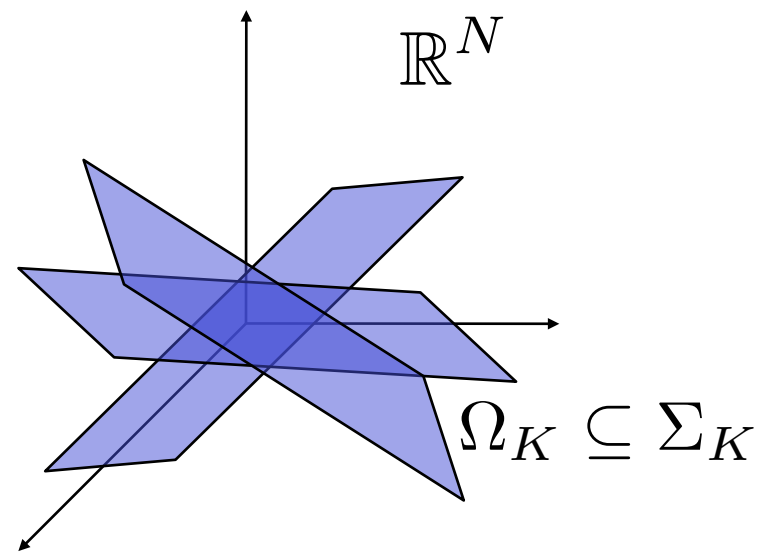
[Duarte, 2012]

Structured Sparse Signals

- A K**-sparse** signal lives on the collection of K-dim subspaces aligned with coordinate axes



- A K**-structured sparse** signal lives on a particular (reduced) collection of K-dimensional canonical subspaces

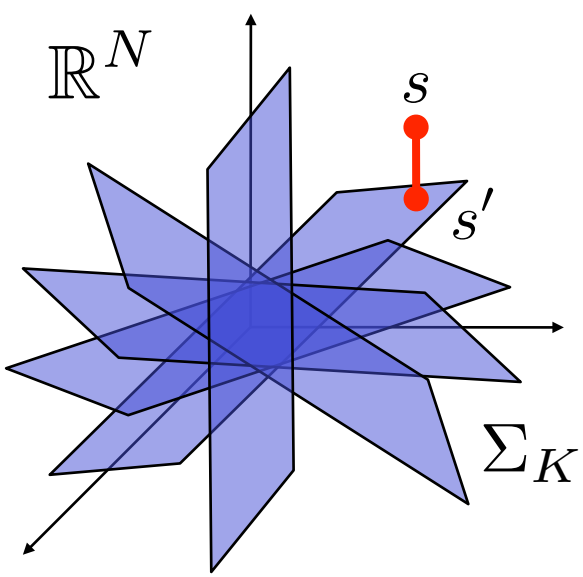


[Baraniuk, Cevher, Duarte, Hegde 2010]

Leveraging Structure in Recovery

Many state-of-the-art sparse recovery algorithms (greedy and optimization solvers) rely on **thresholding** $s' = \mathcal{T}(s, K)$ [Daubechies, Defrise, and DeMol; Nowak, Figueiredo, and Wright; Tropp and Needell; Blumensath and Davies...]

$$s'(n) = \begin{cases} s(n) & \text{if } |s(n)| \text{ is among } K \text{ largest,} \\ 0 & \text{otherwise.} \end{cases}$$

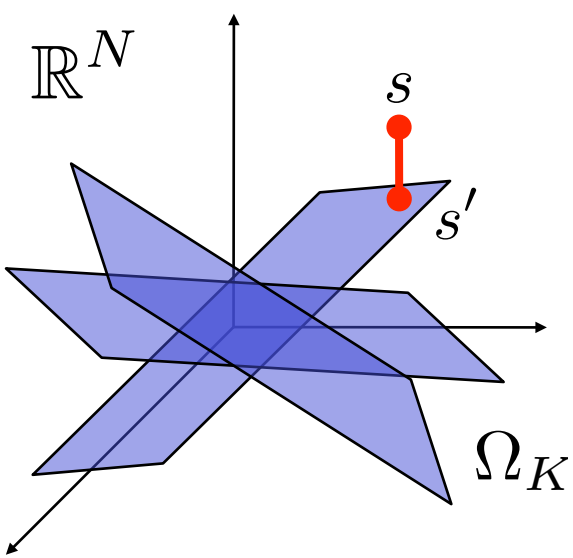


Thresholding provides the **best approximation** of s within Σ_K

$$s' = \arg\min_{\bar{s} \in \Sigma_K} \|s - \bar{s}\|_2$$

Structured Recovery Algorithms

- Modify existing approaches to obtain **structure-aware recovery algorithms**: replace the thresholding step with a **best structured sparse approximation step** that finds the closest point within union of subspaces

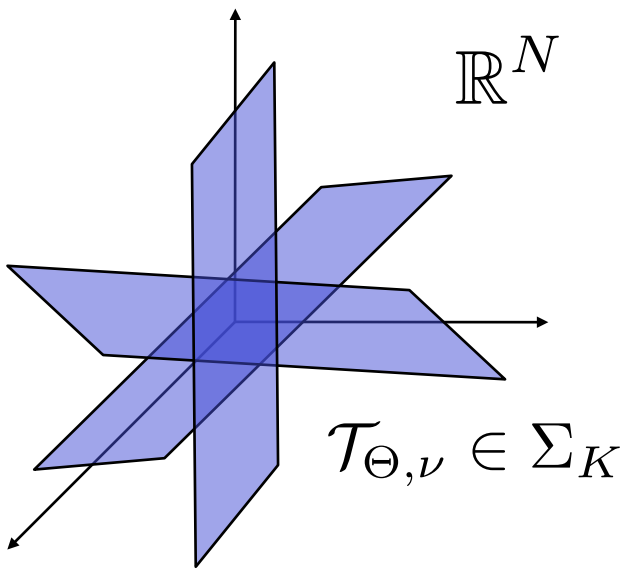


$$s' = \mathbb{M}(s, K) = \arg\min_{\bar{s} \in \Omega_K} \|s - \bar{s}\|_2$$

Greedy structure-aware recovery algorithms **inherit guarantees** of generic counterparts *(even though feasible set may be nonconvex)*

[Baraniuk, Cevher, Duarte, Hegde 2010]

Structured Frequency-Sparse Signals

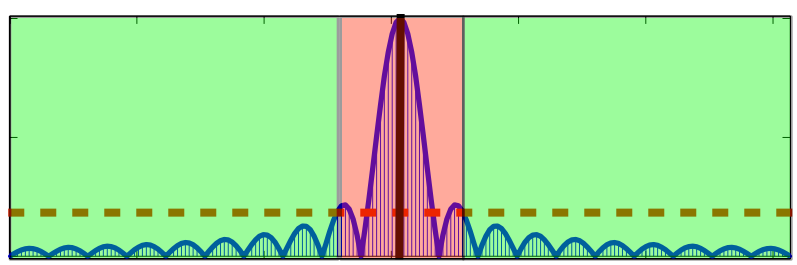


- A **K-structured PD-sparse** signal f consists of K PD elements that are mutually incoherent:

$$s = \sum_{k=1}^{K} a_k \Psi(\theta_k) \in \mathcal{T}_{\Theta,\nu} \quad \text{if}$$

$$\theta_k \in \Theta, \ |\langle \Psi(\theta_k), \Psi(\theta_k') \rangle| \leq \nu \ \forall \ k \neq k'$$

- If x is K-structured frequency-sparse, then there exists a K-sparse vector α such that $s = \Psi_\Theta \alpha$ and the nonzeros in are spaced apart from each other (***band exclusion***).



Standard Sparse Signal Recovery

Iterative Hard Thresholding

Inputs:

- Measurement vector y
- Measurement matrix $\Phi\Psi$
- Sparsity K

Output:

- PD coefficient estimate $\hat{s}$

- Initialize: $\hat{s}_0 = 0, r = y, i = 0$
- While halting criterion false,
  - $i \leftarrow i + 1$
  - $b \leftarrow \hat{s}_{i-1} + \Psi^T \Phi^T r$ *(estimate signal)*
  - $\hat{s}_i \leftarrow \mathcal{T}(b, K)$ *(obtain best sparse approx.)*
  - $r \leftarrow y - \Phi\Psi\hat{s}_i$ *(calculate residual)*
- Return estimate $\hat{s} = \hat{s}_i$

[Blumensath and Davies 2009]

Structured Sparse Signal Recovery

Band-Excluding IHT

Inputs:

- Measurement vector y
- Measurement matrix $\Phi\Psi$
- Structured sparse approx. algorithm $\mathbb{M}(x, K)$

Output:

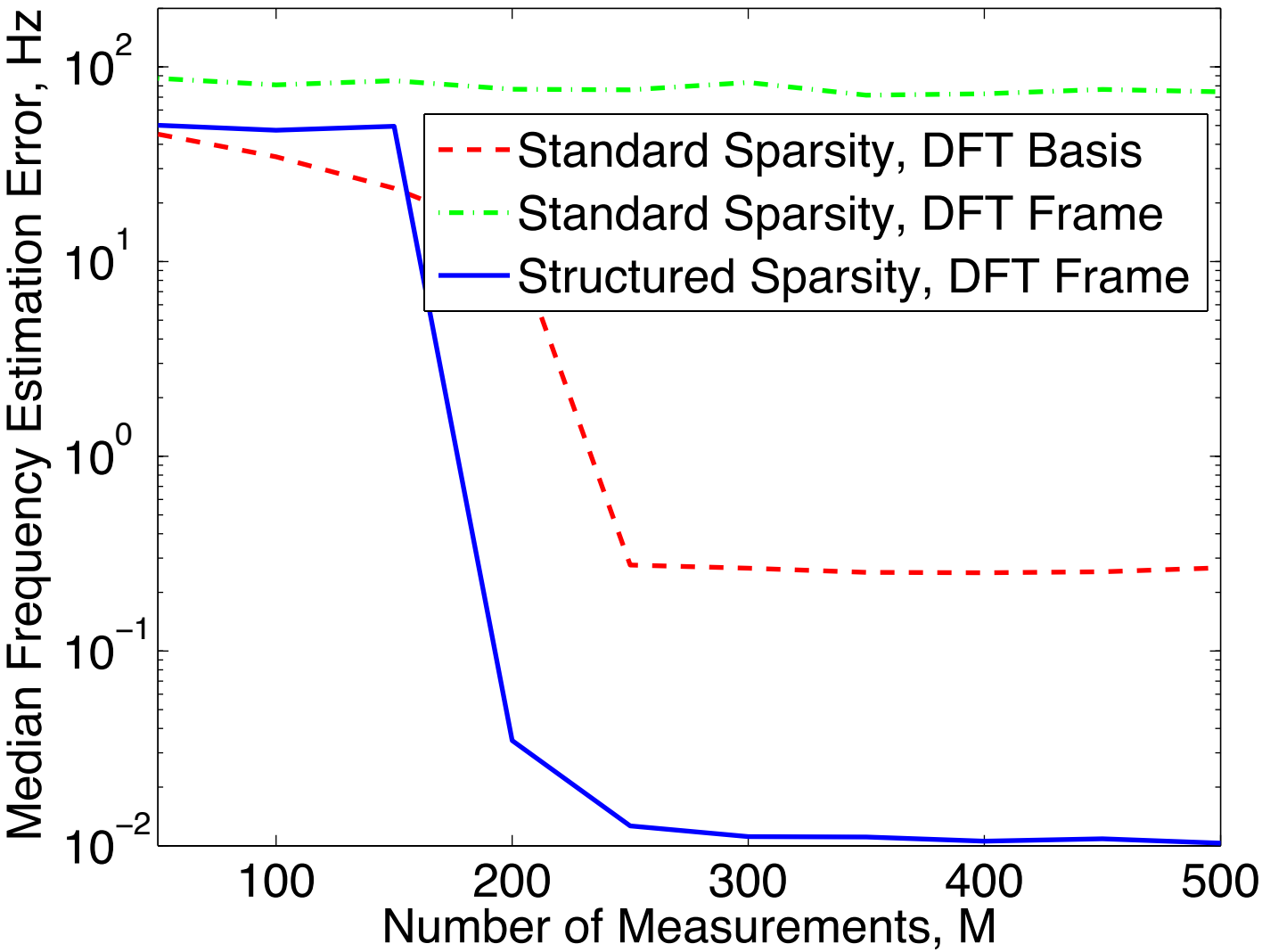
- PD coefficient estimate $\hat{s}$

- Initialize: $\hat{s}_0 = 0, r = y, i = 0$
- While halting criterion false,
  - $i \leftarrow i + 1$
  - $b \leftarrow \hat{s}_{i-1} + \Psi^T \Phi^T r$ *(estimate signal)*
  - $\hat{s}_i \leftarrow \mathbb{M}(b, K)$ *(obtain **band-excluding** sparse approx.)*
  - $r \leftarrow y - \Phi\Psi\hat{s}_i$ *(calculate residual)*
- Return estimate $\hat{s} = \hat{s}_i$

Can be applied to a variety of greedy algorithms (CoSaMP, OMP, Subspace Pursuit, etc.)

[Duarte and Baraniuk, 2012] [Fannjiang and Liao, 2012]

Compressive Line Spectral Estimation: Performance Evaluation



$N = 1024$
$K = 20$
$c = 5$
$\Delta f = 0.1$ Hz
for DFT frame

Recovery via
IHT Algorithm
(or variants)

Part 2: From Discrete to Continuous Models

Hamid Dadkhahi

Karsten Fyhn (Aalborg U.)

S.H. Jensen (Aalborg U.)

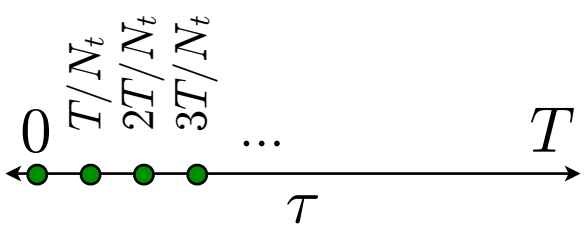
From Discrete to Continuous Models

- PD can be conceived as a **sampling** from an **infinite set** of parametric signals; e.g., TDOA dictionary s_τ contains a discrete set of values for the time delay parameter:

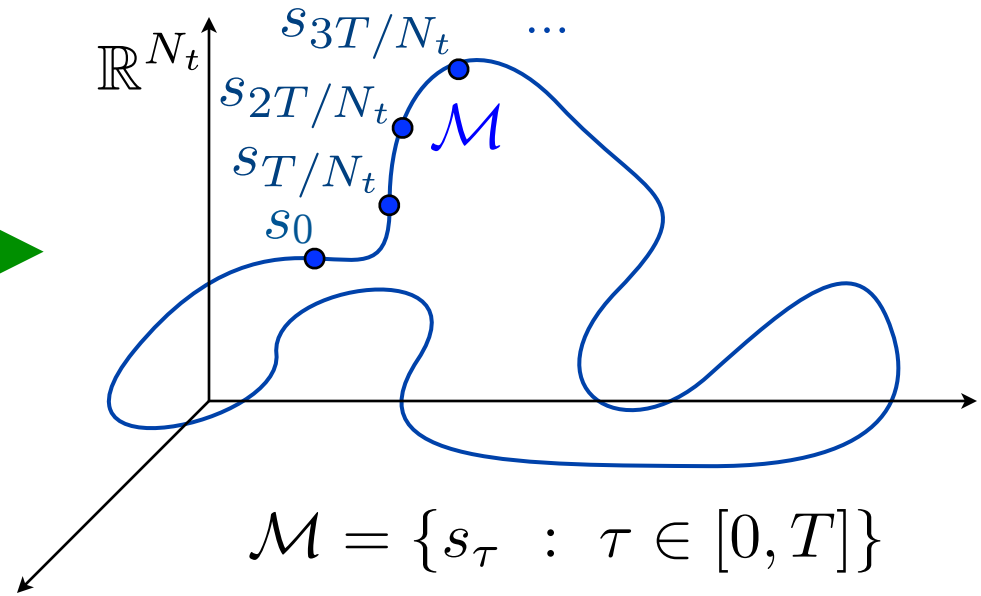
$$\Psi_{\mathrm{TDOA}} = [s_0 \;\; s_{T/N_t} \;\; \ldots \;\; s_{T(N_t-1)/N_t}]$$

$$s_\tau = \left[s(\tau) \;\; s\left(\tau + \frac{T}{N_t}\right) \;\; s\left(\tau + \frac{2T}{N_t}\right) \;\; \ldots \;\; s\left(\tau + \frac{(N-1)T}{N}\right) \right]^*$$

- Vector s_τ varies smoothly in each entry as a function of τ; we can represent the signal set as a one-dimensional **nonlinear manifold**:



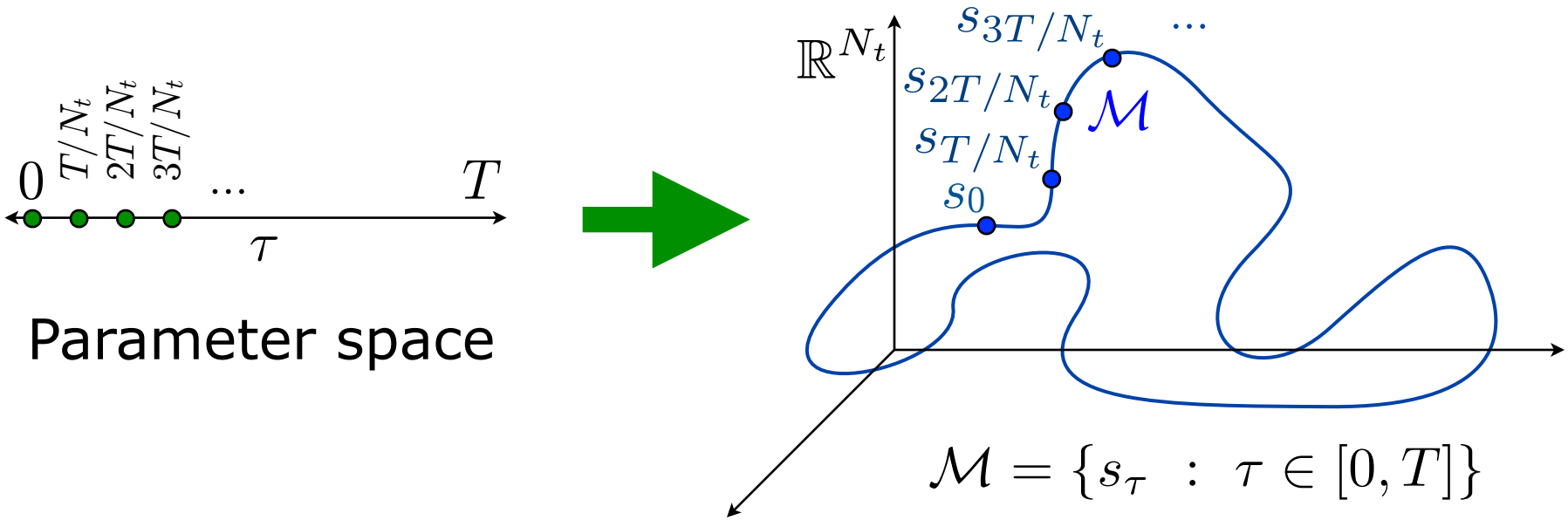
Parameter space



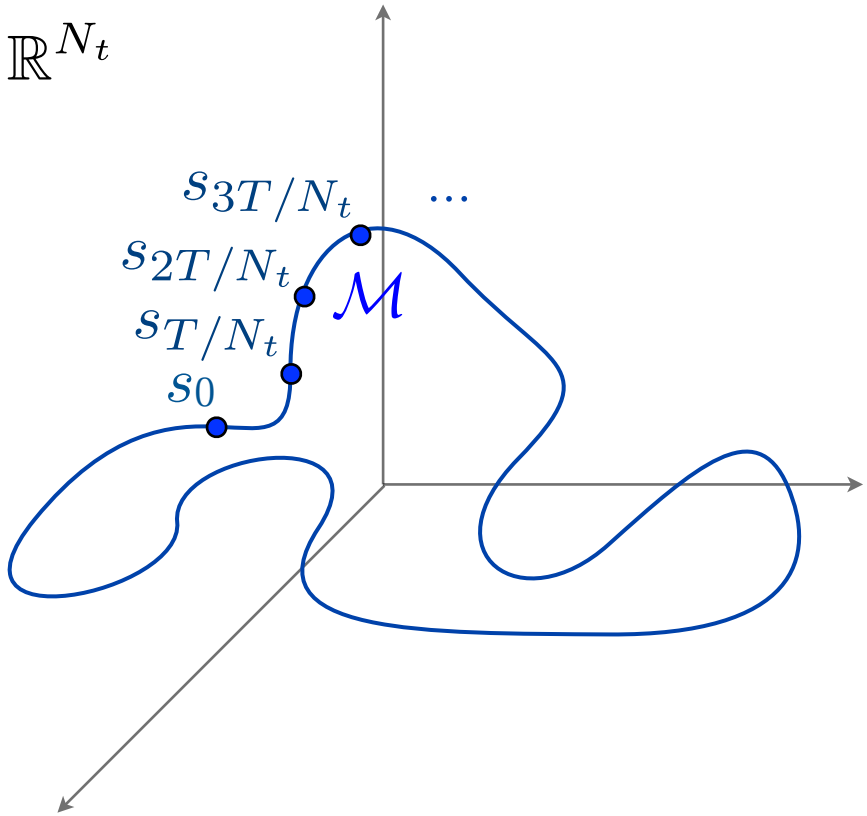
$$\mathcal{M} = \{ s_\tau \;:\; \tau \in [0, T] \}$$

From Discrete to Continuous Models

- For computational reasons, we wish to design methods that allow us to ***interpolate*** the manifold from the samples obtained in the PD to increase the resolution of the parameter estimates.
- An ***interpolation-based*** compressive parameter estimation algorithm obtains projection values for sets of manifold samples and interpolates manifold around location of peak projection to get parameter estimate

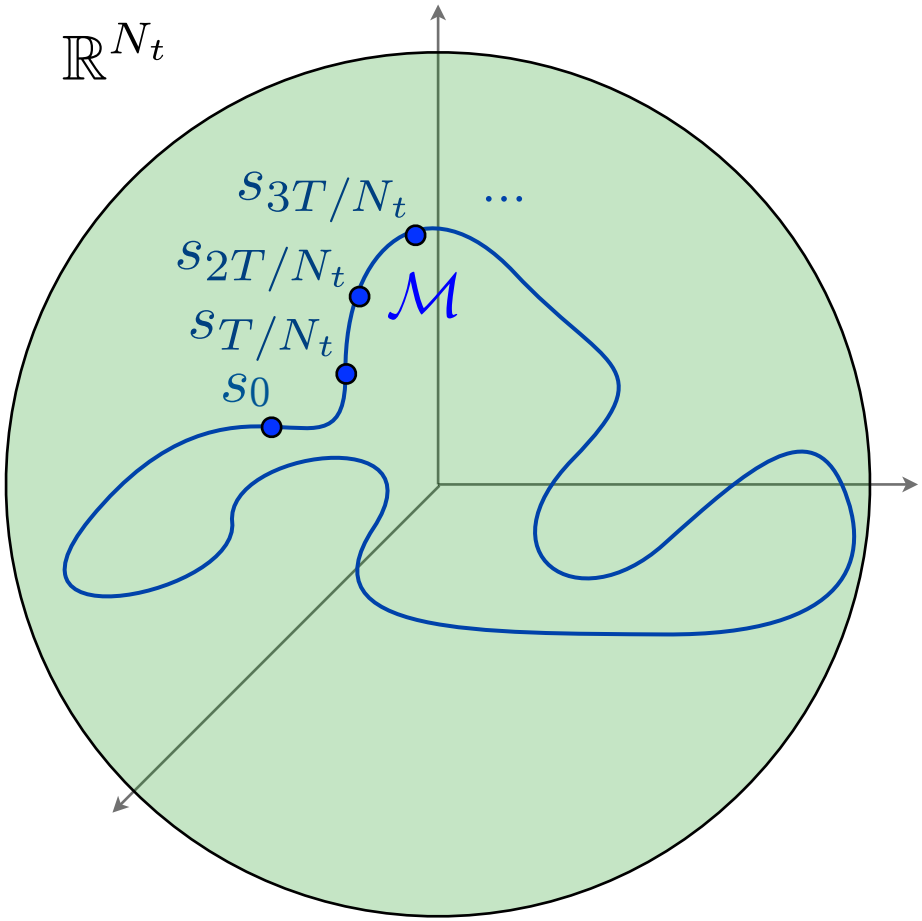


Interpolating the Manifold: Polar Interpolation



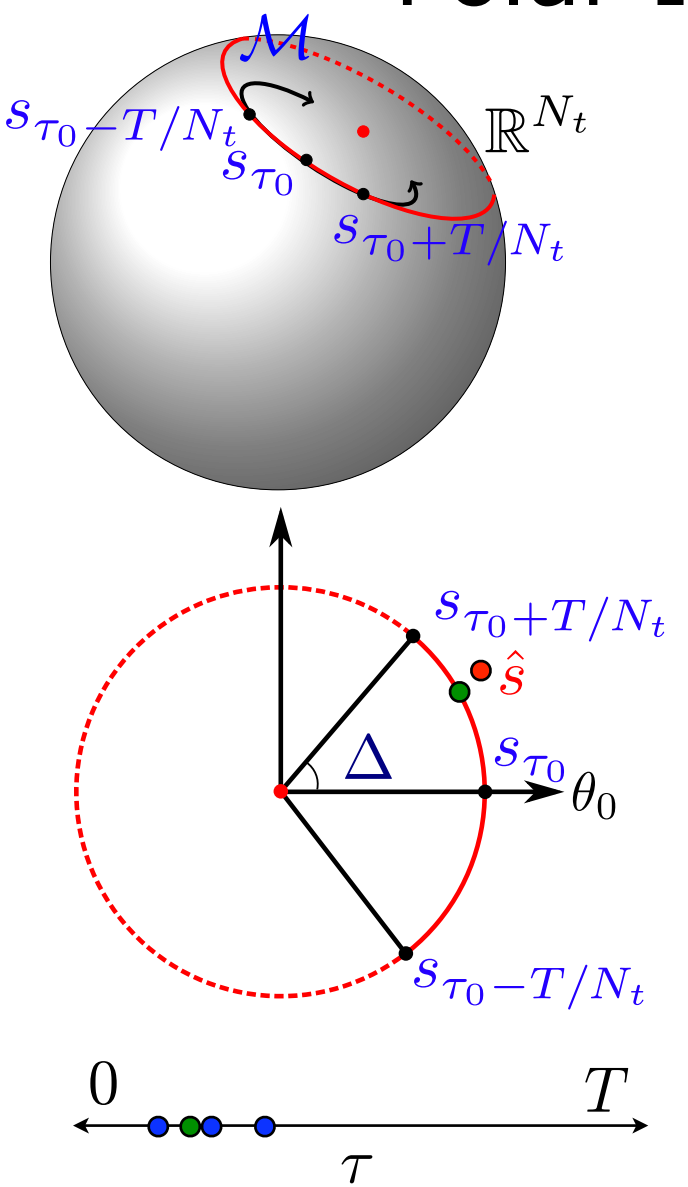
- All points in manifold have ***equal norm*** (delayed versions of fixed waveform)
- Distance between manifold samples is ***uniform*** (depends only on parameter difference, not on parameter values)
- TDOA manifold features these two properties

Interpolating the Manifold: Polar Interpolation



- Manifold must be contained within unit Euclidean ball (***hypersphere***)
- Manifold has ***uniform curvature***, enabling parameter-independent interpolation scheme
- ***Project*** signal estimates into hypersphere
- Find closest point in manifold by ***interpolating*** from closest samples with polar coordinates

Interpolating the Manifold: Polar Interpolation



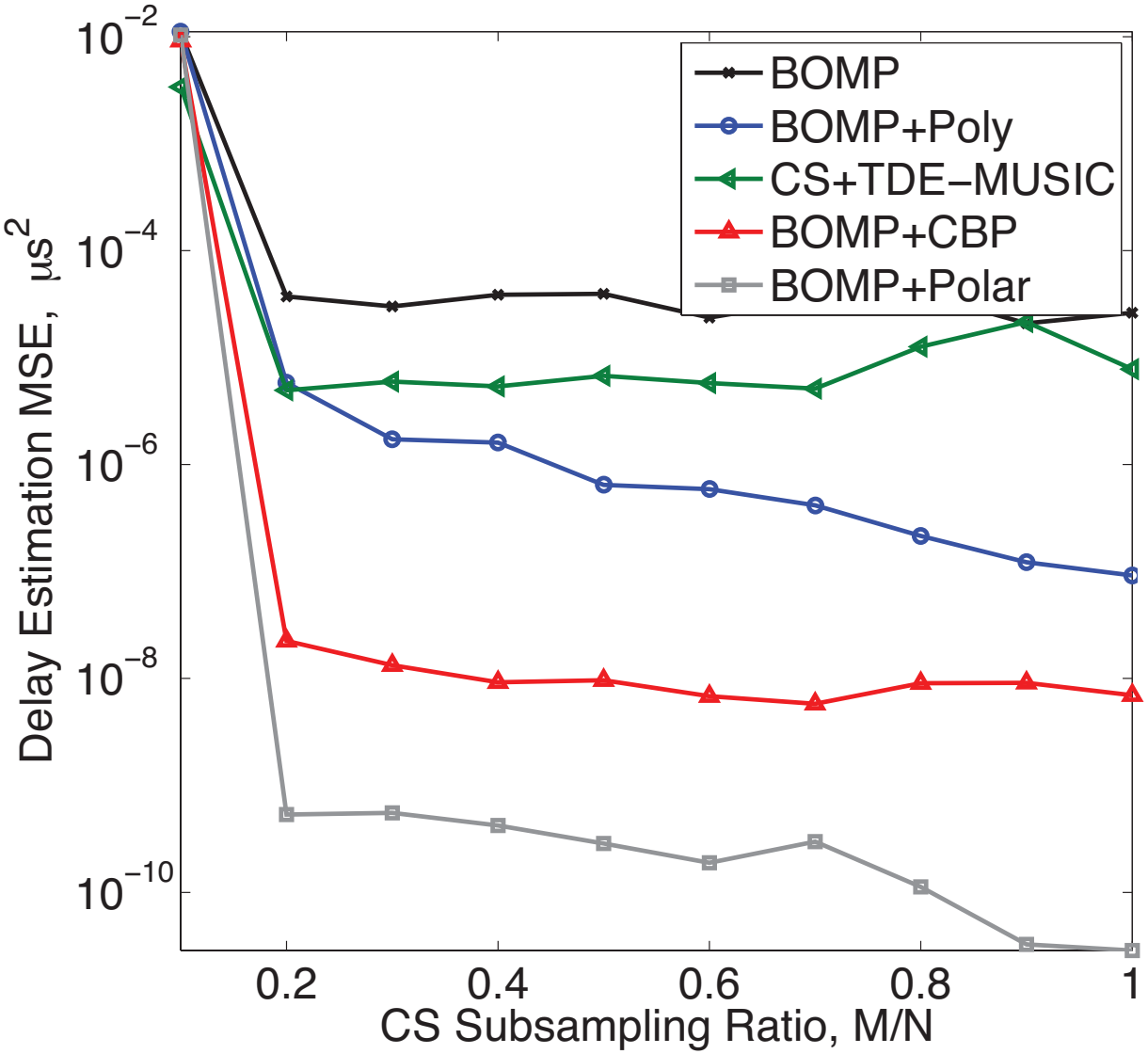
- Find closest point in manifold by interpolating from closest samples with ***polar coordinates***:

$$s_{\tau_0 - T/N_t} \leftrightarrow \angle = \theta_0 - \Delta$$
$$s_{\tau_0} \leftrightarrow \angle = \theta_0$$
$$s_{\tau_0 + T/N_t} \leftrightarrow \angle = \theta_0 + \Delta$$
$$\hat{s} \leftrightarrow \angle = \ ?$$

- Map back from manifold to ***parameter space*** to obtain final parameter estimates

Akin to Continuous Basis Pursuit (CBP)
[Ekanadham, Tranchina, and Simoncelli 2011]

Compressive Time Delay Estimation: Performance Evaluation

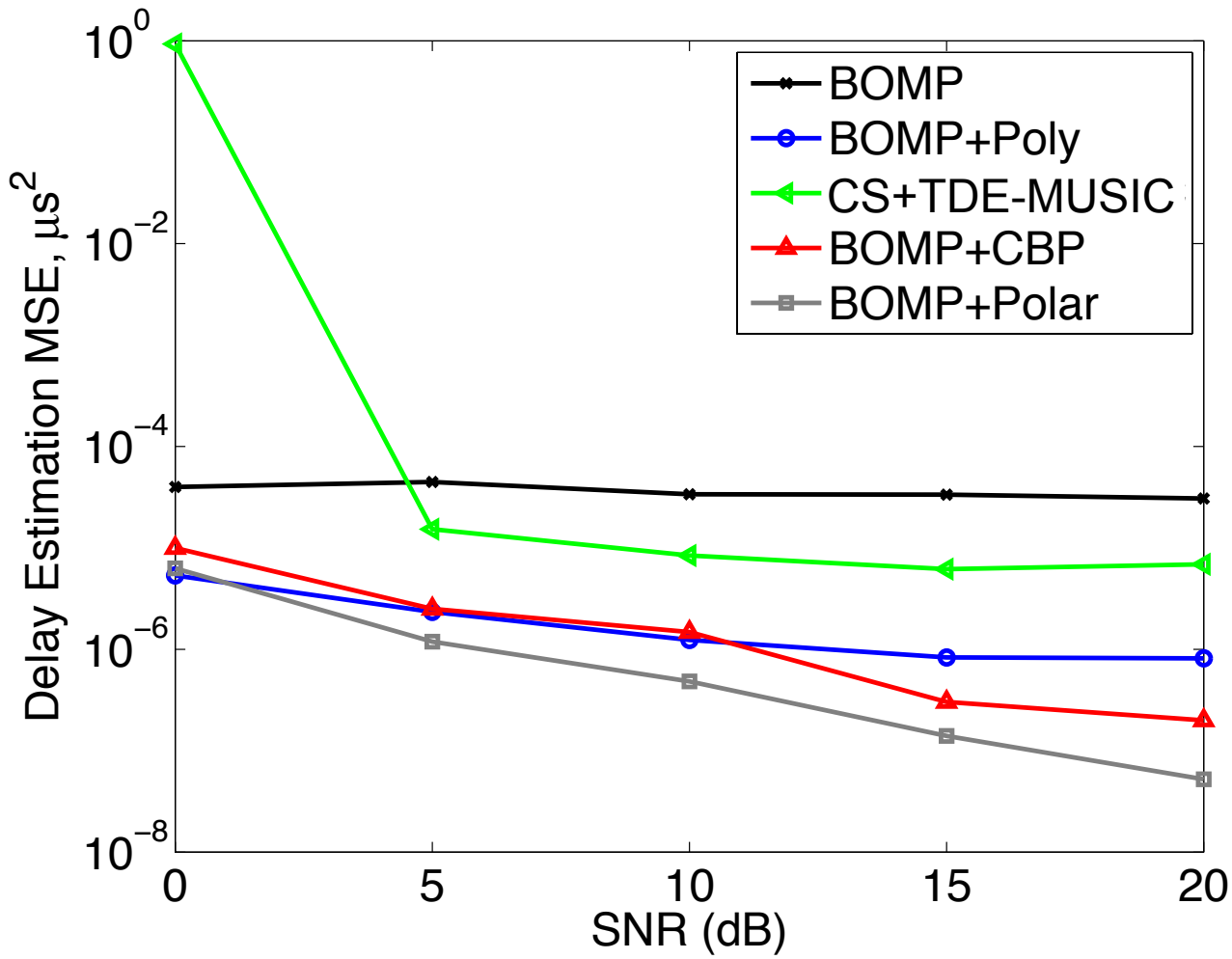


Waveform: chirp of length $1\,\mu\mathrm{s}$

$$N_t = 500,\ K = 3,$$
$$c = 1,\ \Delta\tau = 1\mu\mathrm{s}$$

BOMP [Fannjiang and Liao 2012]

Compressive Time Delay Estimation: Performance Evaluation (Noise)

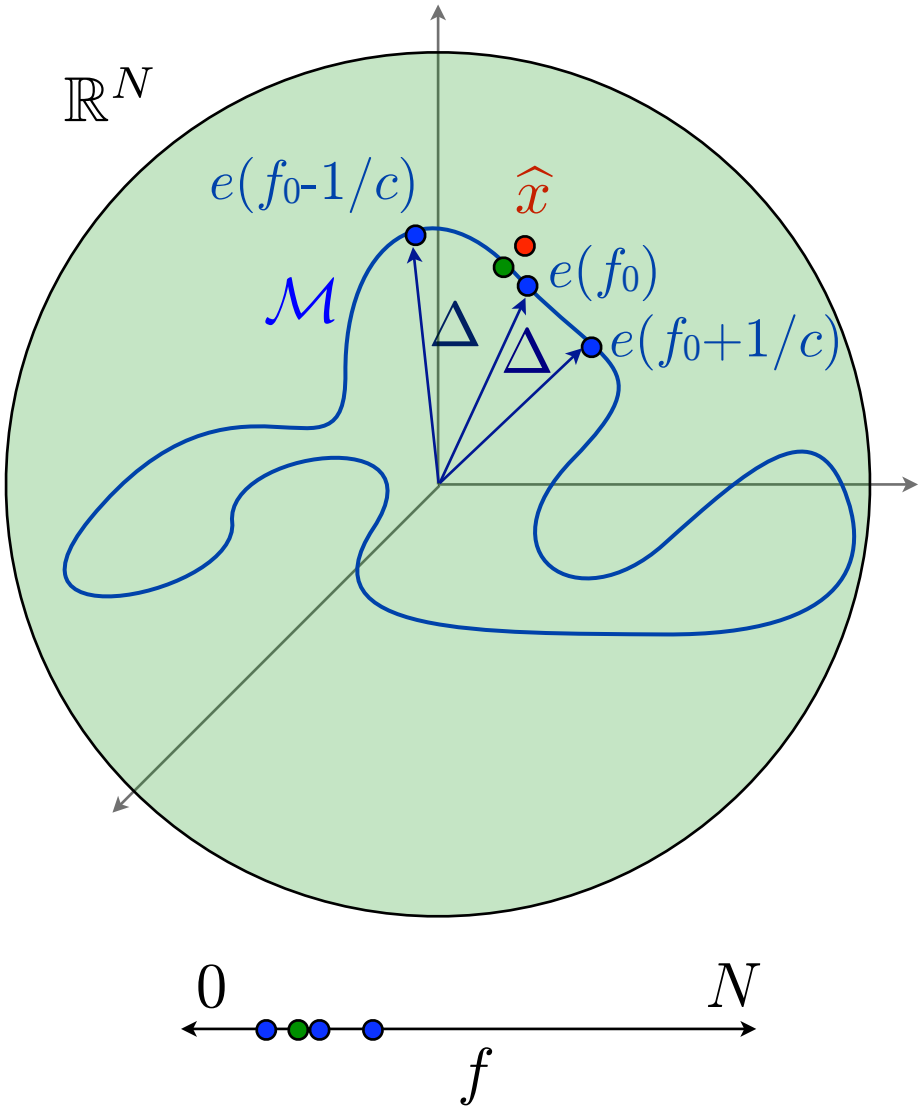


Waveform: chirp of length $1\,\mu\mathrm{s}$

$$N_t = 500,\ K = 3,$$
$$c = 1,\ \Delta\tau = 1\mu\mathrm{s}$$

BOMP [Fannjiang and Liao 2012]

Polar Interpolation for Off-The-Grid Frequency Estimation

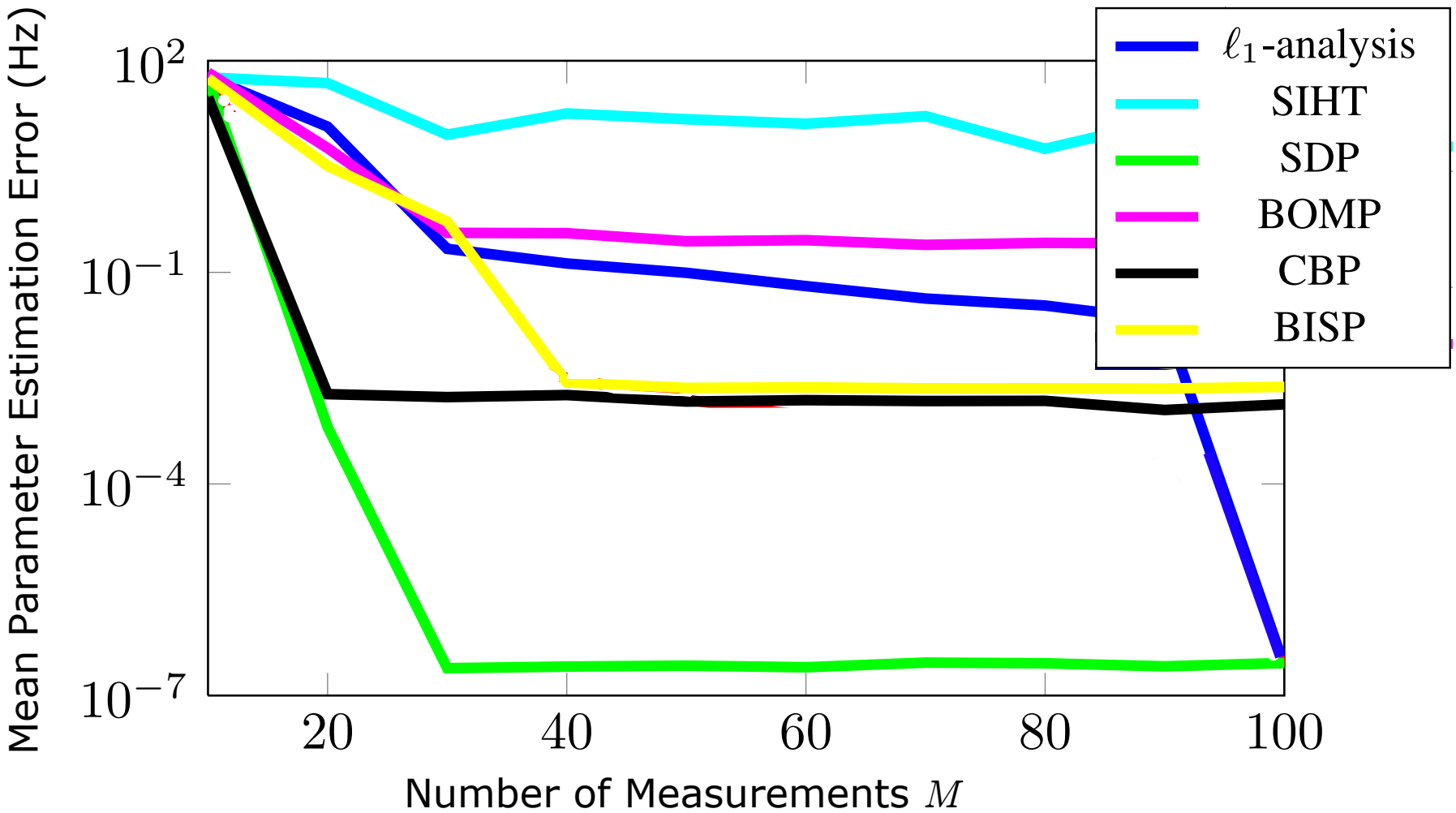


- Same properties from time delay estimation manifold are also present in ***complex exponential manifold***:
  - uniform curvature
  - equal norm

$$e(f_0 - 1/c) \leftrightarrow \angle = \theta_0 - \Delta$$
$$e(f_0) \leftrightarrow \angle = \theta_0$$
$$e(f_0 + 1/c) \leftrightarrow \angle = \theta_0 + \Delta$$
$$\widehat{x} \leftrightarrow \angle = ?$$

- Use polar interpolation
- Map back from manifold to ***frequency estimates***

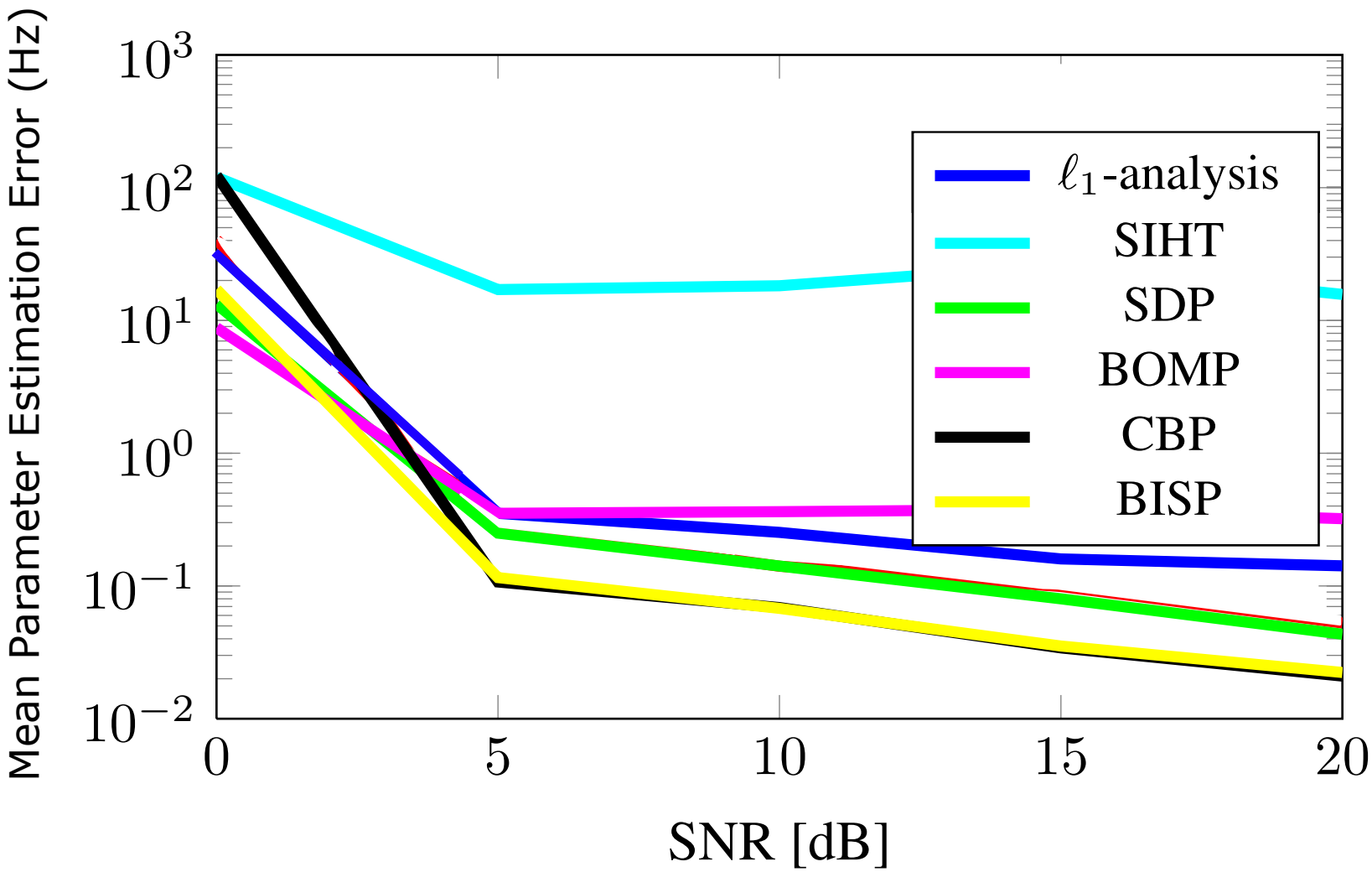
Compressive Line Spectral Estimation: Performance Evaluation



$N = 100,\ K = 4,$
$c = 5,\ \Delta f = 0.2$ Hz

BOMP [Fannjiang and Liao 2012]
SDP [Tang, Rhaskar, Shah, Recht 2012]

Compressive Line Spectral Estimation: Performance Evaluation (Noise)

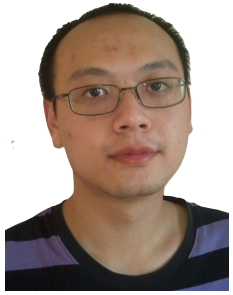


$N = 100,\ K = 4,\ M = 50,\ c = 5,\ \Delta f = 0.2$ Hz

Compressive Line Spectral Estimation: Computational Expense

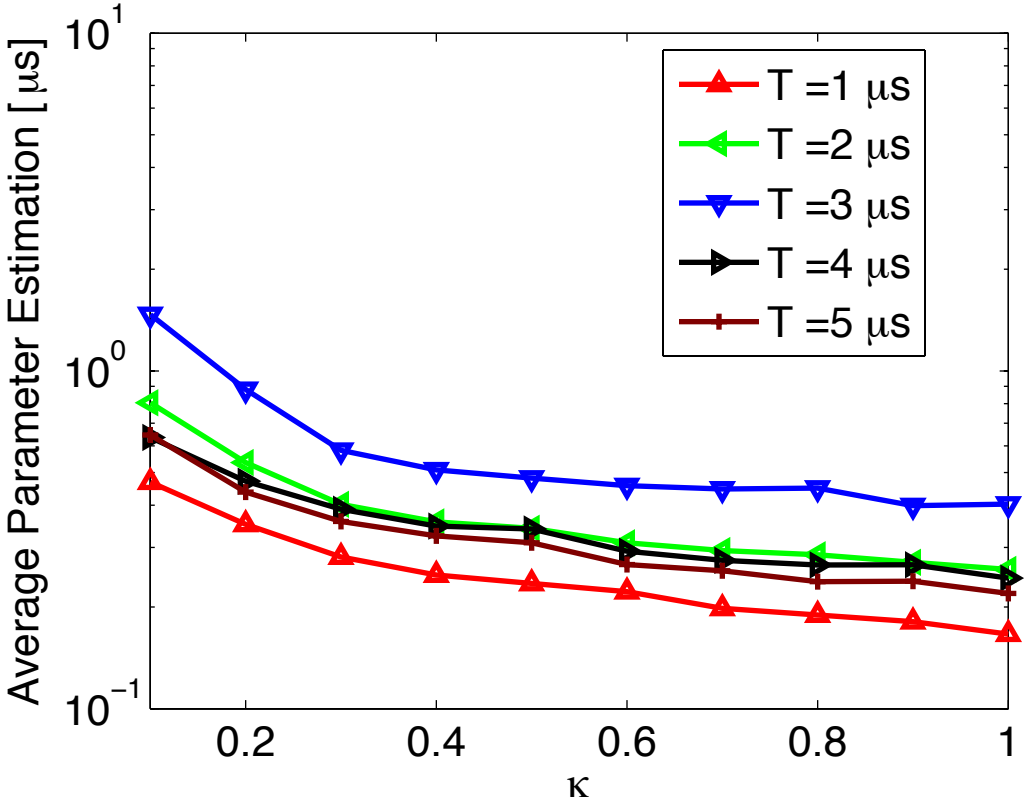
Time (seconds)	Noiseless	Noisy
ℓ_1-analysis	9.5245	8.8222
SIHT	0.2628	0.1499
SDP	8.2355	9.9796
BOMP	0.0141	0.0101
CBP	46.9645	40.3477
BISP	5.4265	1.4060

Part 3: Meaningful Performance Metrics



Dian Mo

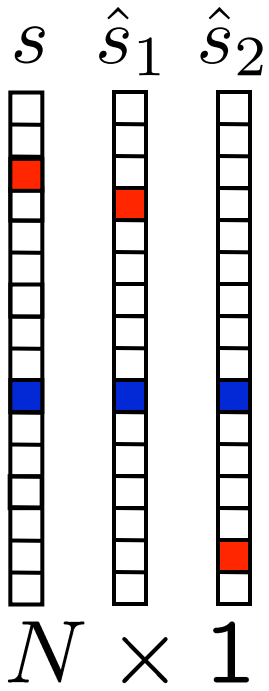
Issues with PDs/Structured Sparsity: Sensitivity to Maximal Coherence Value



- Example: Compressive Time Delay Estimation (TDE) with PD and random demodulator
- Performance depends on measurement ratio $\kappa = M/N$
- Structured sparsity (***band exclusion***) used to enable high-resolution TDE

- Parameter ν set to ***optimal value*** for chirp of length $1\ \mu\mathrm{s}$
- As length of chirp wave increases, performance of compressive TDE ***varies widely***
- ***Shape*** of correlation function dependent on chirp length

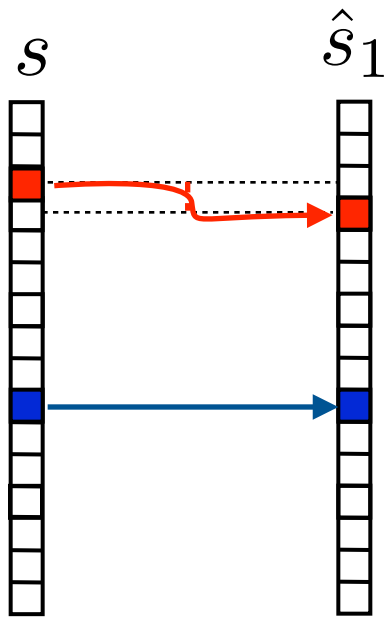
Issues with PDs: Euclidean Norm Guarantees

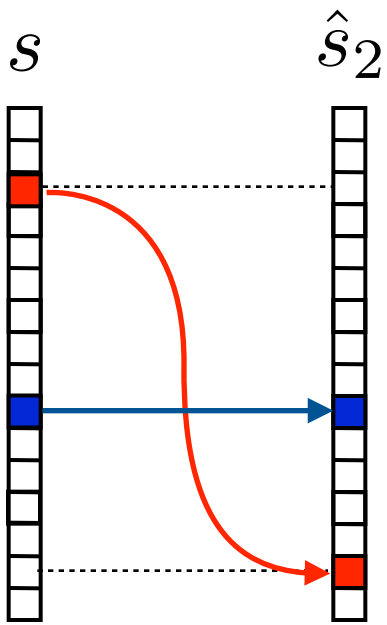


delay grid

- Most recovery methods provide ***guarantees*** to keep Euclidean norm error $\|s - \hat{s}\|_2$ small
- This metric, however, is ***not connected to quality*** of parameter estimates
- ***Example***: both estimates have same Euclidean error $\|s - \hat{s}_1\|_2 = \|s - \hat{s}_2\|_2$ but provide very different location estimates.
- We search for a performance metric ***better suited*** to the use of PD coefficient vectors (i.e., $\|\theta - \widehat{\theta}\|$)

Improved Performance Metric: Earth Mover's Distance





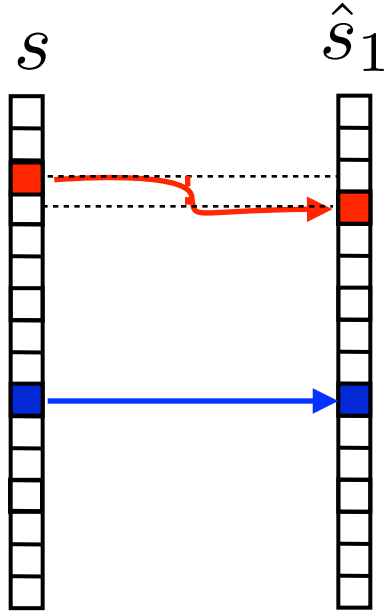
- Earth Mover's Distance (EMD) is based on ***concept of "mass" flowing*** between the entries of first vector in order to match the second
- EMD value is ***minimum "work" needed*** (measured as mass x transport distance) for first vector to match second:

$$\mathrm{EMD}(\alpha, \widehat{\alpha}) := \min_{f} \sum_{i,j=1}^{L} f_{ij} |i - j|$$

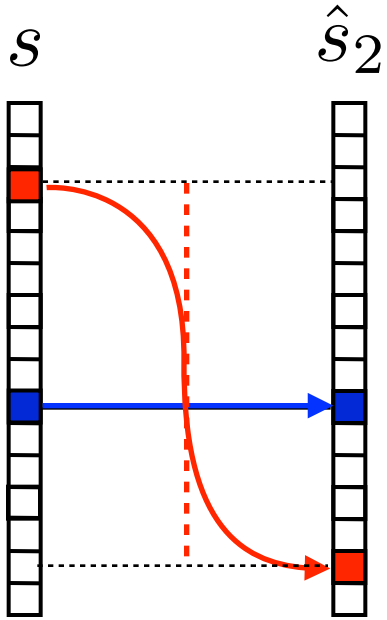
$$\text{s.t.} \sum_{j} f_{ij} = |\alpha_i| \ \forall \ i = 1, \ldots, N,$$

$$\sum_{i} f_{ij} = |\widehat{\alpha}_j| \ \forall \ j = 1, \ldots, N.$$

Improved Performance Metric: Earth Mover's Distance

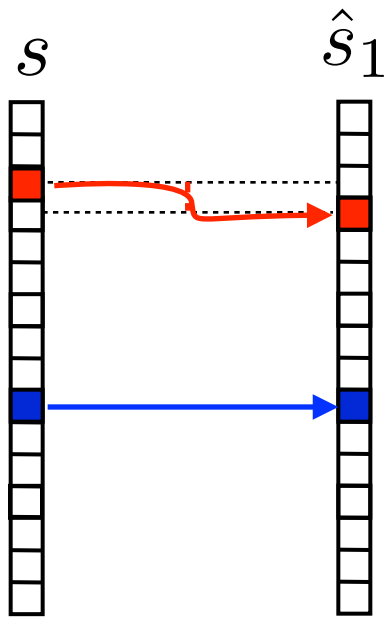


- Earth Mover's Distance (EMD) is based on ***concept of "mass" flowing*** between the entries of first vector in order to match the second
- EMD value is ***minimum "work" needed*** (measured as mass x transport distance) for first vector to match second:



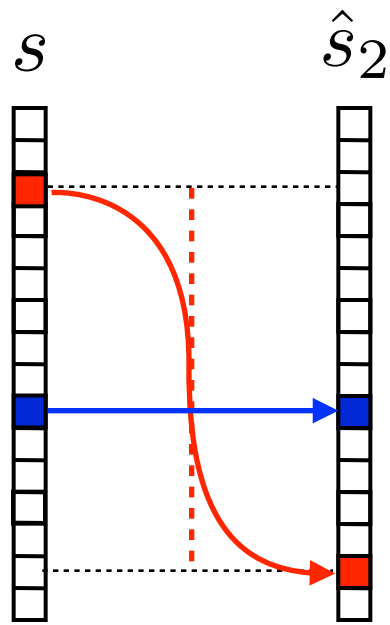
- When PDs are used, EMD captures ***parameter estimation error*** by measuring distance traveled by "mass"
- Parameter values must be ***proportional to indices*** in PD coefficient vector
- How to introduce EMD metric into CS recovery process?

Sparse Approximation with Earth Mover's Distance



- To integrate into greedy algorithms, we will need to solve the ***EMD-optimal K-sparse approximation problem***

$$\widehat{x}_K = \arg\min_{\bar{x} \in \Sigma_K} \mathrm{EMD}(x, \bar{x})$$



- It can be shown that approximation can be obtained by performing ***K-median clustering*** on set of points at locations $\{1, \ldots, N\}$ with respective weights $\{|x[1]|, \ldots, |x[N]|\}$
- Cluster centroids provide ***support*** of $\widehat{x}_K$, values can be easily computed to minimize EMD/estimation error

[Indyk and Price 2009]

Structured Sparse Signal Recovery

Band-Excluding IHT

Inputs:

- Measurement vector y
- Measurement matrix $\Phi\Psi$
- Structured sparse approx. algorithm $\mathbb{M}(x, K)$

Output:

- PD coefficient estimate $\hat{s}$

- Initialize: $\hat{s}_0 = 0, r = y, i = 0$
- While halting criterion false,
  - $i \leftarrow i + 1$
  - $b \leftarrow \hat{s}_{i-1} + \Psi^T \Phi^T r$ *(estimate signal)*
  - $\hat{s}_i \leftarrow \mathbb{M}(b, K)$ *(obtain **band-excluding** sparse approx.)*
  - $r \leftarrow y - \Phi\Psi\hat{s}_i$ *(calculate residual)*
- Return estimate $\hat{s} = \hat{s}_i$

Can be applied to a variety of greedy algorithms (CoSaMP, OMP, Subspace Pursuit, etc.)

[Duarte and Baraniuk, 2012] [Fannjiang and Liao, 2012]

EMD + Sparse Signal Recovery

Clustered IHT

Inputs:

- Measurement vector y
- Measurement matrix $\Phi\Psi$
- Sparsity K

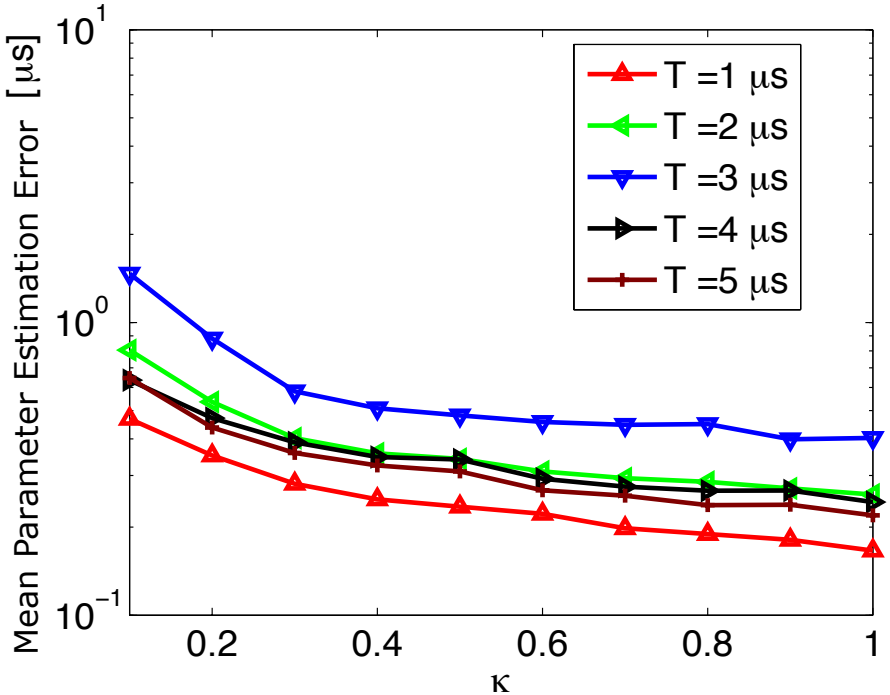
Output:

- PD coefficient estimate $\hat{s}$

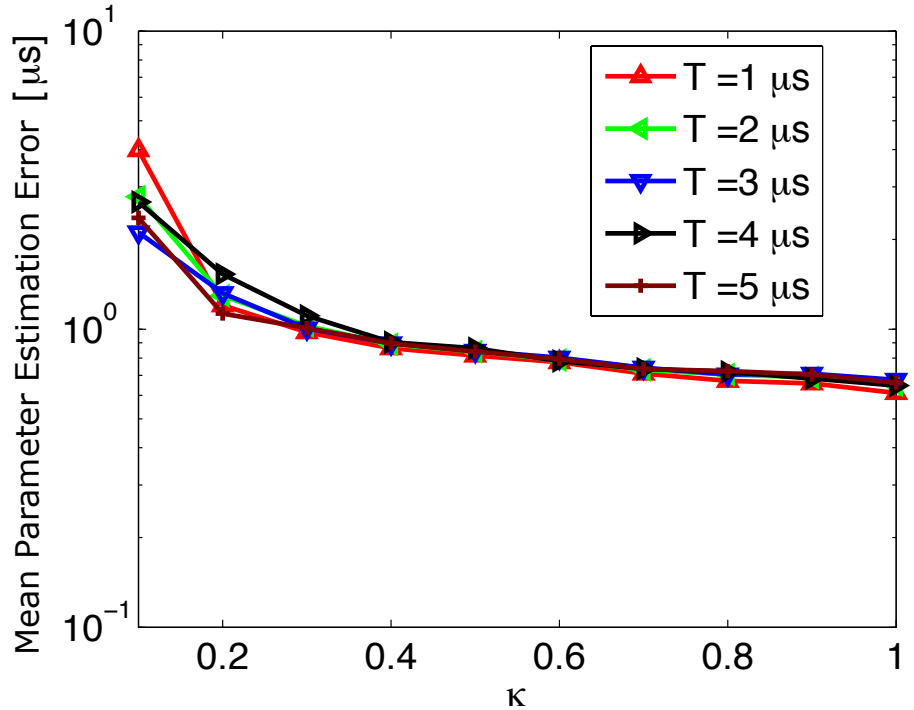
- Initialize: $\hat{s}_0 = 0, r = y, i = 0$
- While halting criterion false,
  - $i \leftarrow i + 1$
  - $b \leftarrow \hat{s}_{i-1} + \Psi^T \Phi^T r$ *(estimate signal)*
  - $\hat{s}_i = \arg\min_{\bar{b} \in \Sigma_K} \mathrm{EMD}(b, \bar{b})$ *(best sparse approx.* ***in EMD****)*
  - $r \leftarrow y - \Phi\Psi\hat{s}_i$ *(calculate residual)*
- Return estimate $\hat{s} = \hat{s}_i$

Can be applied to a variety of greedy algorithms (CoSaMP, OMP, Subspace Pursuit, etc.)

Numerical Results



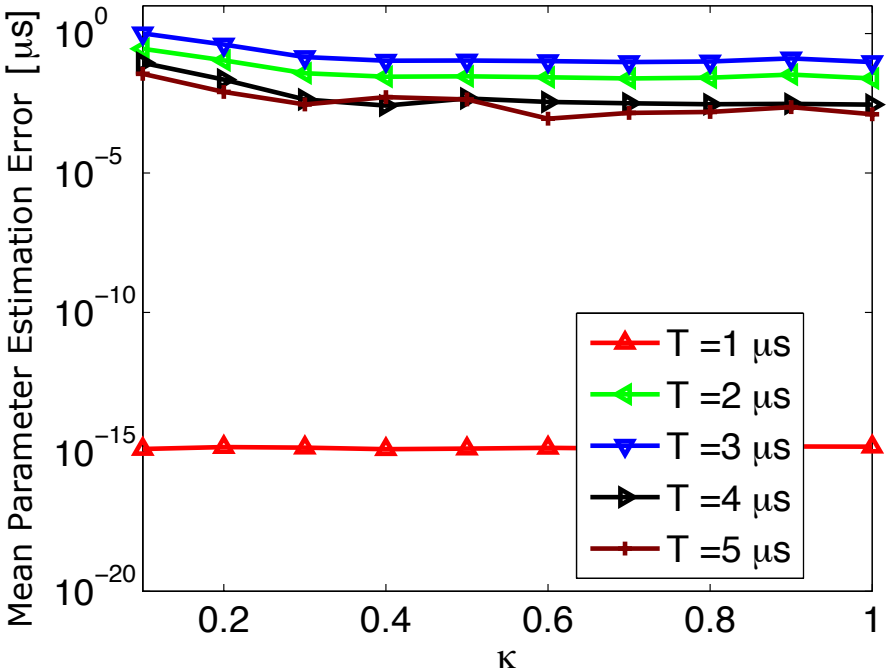
Band-Excluding Subspace Pursuit



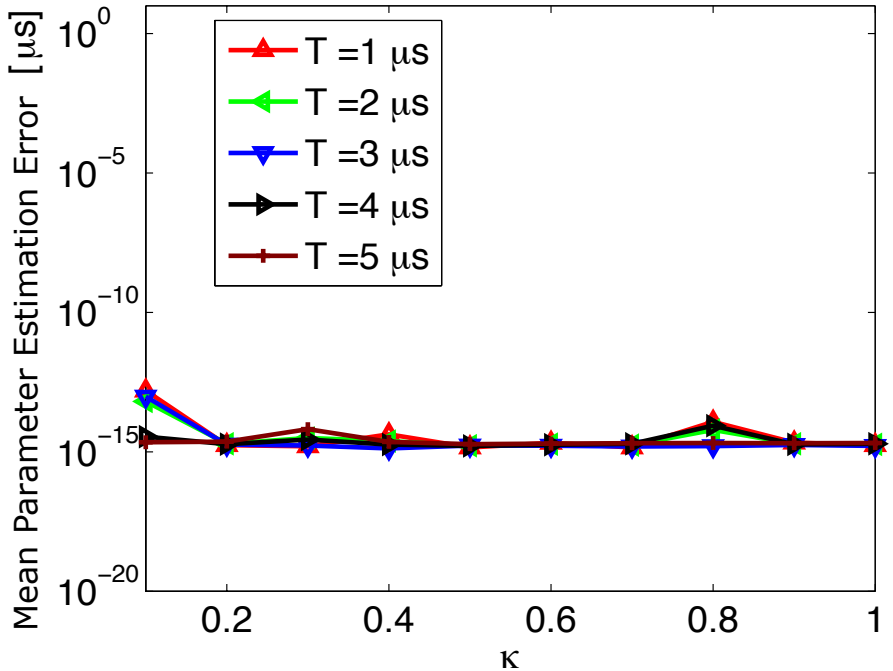
Clustered Subspace Pursuit

- Example: Compressive TDE w/ PD & random demodulator
- Performance depends on ***measurement ratio*** $\kappa = M/N$
- TDE performance ***varies widely*** as chirp length increases
- Consistent behavior for EMD-based signal recovery, but ***consistent bias observed***
- Bias partially due to ***parameter space discretization***

Numerical Results



Band-Excluding Subspace Pursuit



Clustered Subspace Pursuit

- Example: Compressive TDE w/ PD & random demodulator
- Performance depends on measurement ratio $\kappa = M/N$
- When integrated with ***polar interpolation***, performance of compressive TDE improves significantly
- Sensitivity of Band-Excluding SP becomes more ***severe***, while Clustered SP ***remains robust***

Conclusions

- In radar and other parameter estimation settings, retrofitting sparsity ***is not enough!***
  - PDs enable use of CS, but often are ***coherent***
  - band exclusion can help, but must be ***highly precise***
  - issues remain with guarantees (***Euclidean is not useful***)
  - PDs also ***discretize*** parameter space, limiting resolution
- Address discretization with tractable signal models
  - from PDs to manifolds via ***interpolation*** techniques
  - readily available models for time delay, frequency/doppler
- Earth Mover's Distance is a suitable metric
  - easily implementable by leveraging K***-median clustering***
  - EMD is suitable for dictionaries with well-behaved (compact) ***correlation functions***
- ***Ongoing work***: multidimensional extensions, sensitivity to noise, theoretical analysis of EMD...

References

- M. F. Duarte and R. G. Baraniuk, “Spectral Compressive Sensing,” Applied and Computational Harmonic Analysis, Vol. 35, No. 1, pp. 111-129, 2013.
- M. F. Duarte, “Localization and Bearing Estimation via Structured Sparsity Models,” IEEE Statistical Signal Processing Workshop (SSP), Aug. 2012, Ann Arbor, MI, pp. 333-336.
- K. Fyhn, H. Dadkhahi, and M. F. Duarte, “Spectral Compressive Sensing with Polar Interpolation,” IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 2013, Vancouver, Canada, pp. 6225-6229.
- K. Fyhn, M. F. Duarte, and S. Holdt Jensen, “Compressive Parameter Estimation for Sparse Translation-Invariant Signals Using Polar Interpolation,” arXiv:1305.3483, May 2013.
- D. Mo and M. F. Duarte, “Compressive Parameter Estimation with Earth Mover’s Distance via K-Means Clustering,” Wavelets and Sparsity XV, SPIE Optics and Electronics, Aug. 2013, San Diego, CA.

http://www.ecs.umass.edu/~mduarte mduarte@ecs.umass.edu