Compressive Parameter Estimation via Approximate Message Passing

Marco F. Duarte

UMASS AMHERST

Joint work with Shermin Hamzehei

IEEE ICASSP - April 22 2015

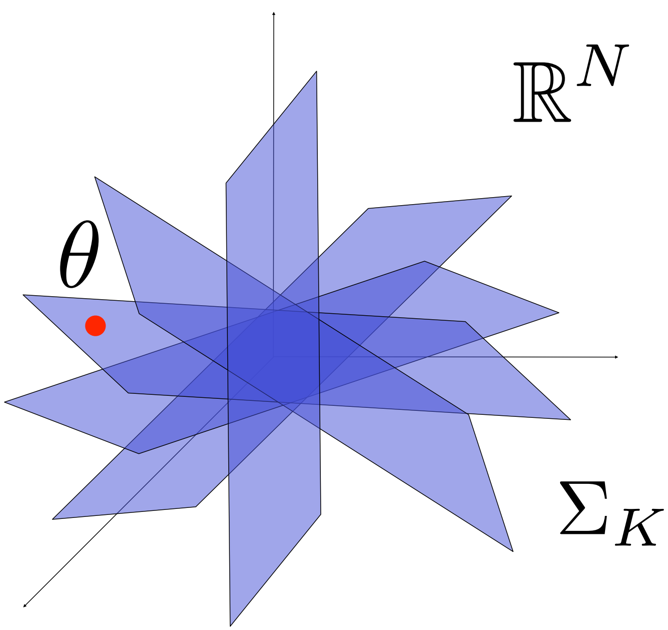
Concise Signal Structure

- **Sparse** signal: only K out of N coefficients are nonzero

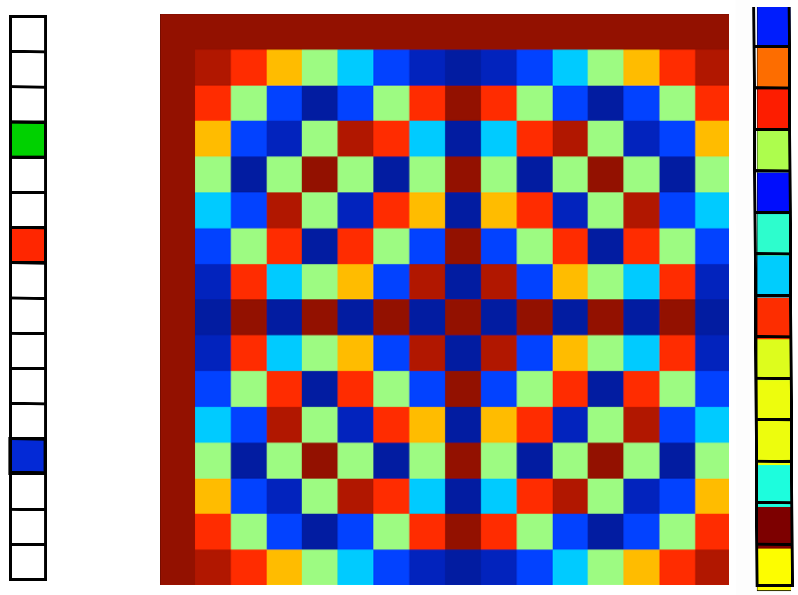
$$x = \sum_n \psi_n \theta_n$$

  - model: union of K-dimensional subspaces aligned w/ coordinate axes

$$x = \Psi\theta$$



$$\theta = \Psi^{-1} x$$



$N \times 1$

sparse signal

K

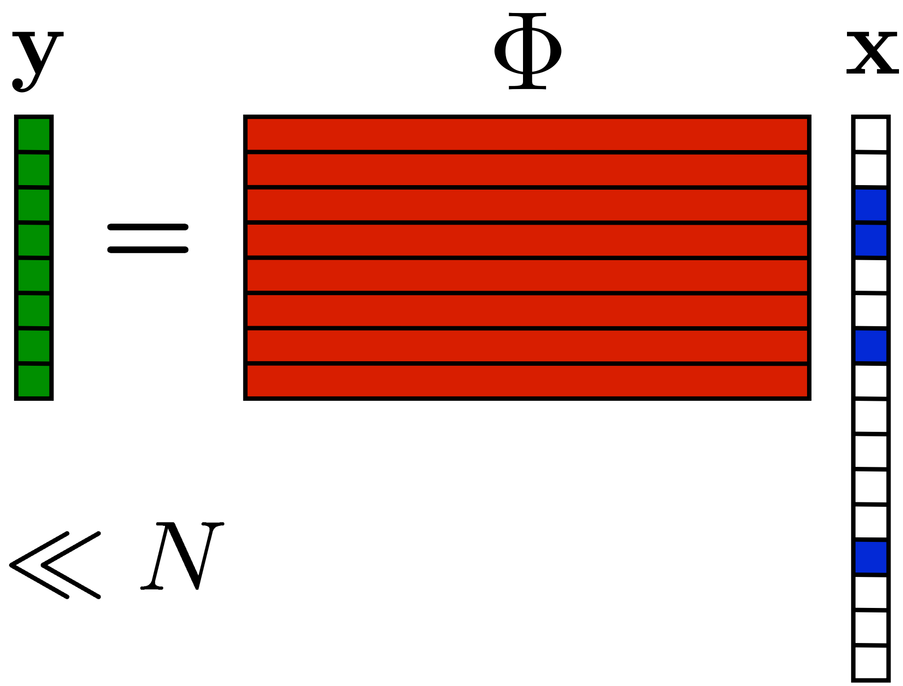
nonzero entries

From Samples to *Measurements*

- Replace **samples** by more general **encoder** based on a few linear projections (inner products)

$\mathbf{y} = \Phi\mathbf{x}$, $\mathbf{x}$ is sparse

$M \times 1$
measurements



$N \times 1$
sparse signal

K
information rate

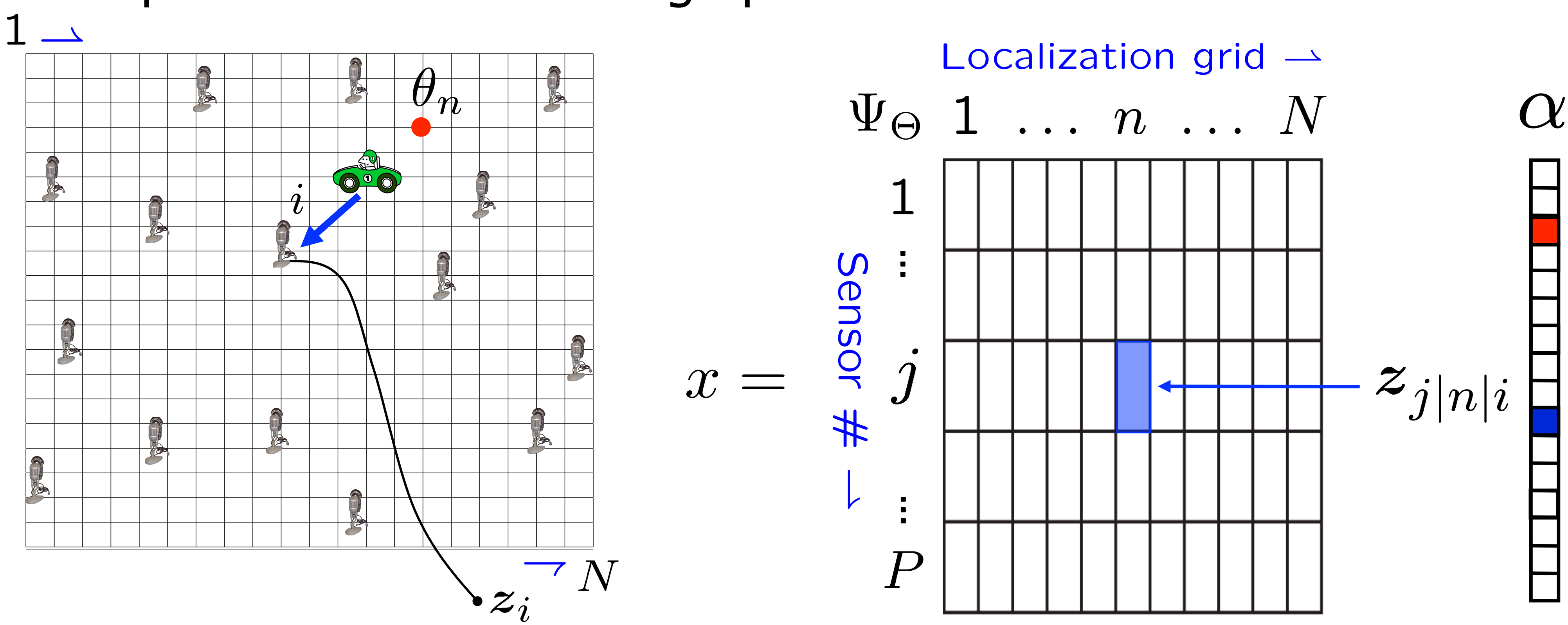
$$M \approx K \ll N$$

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)

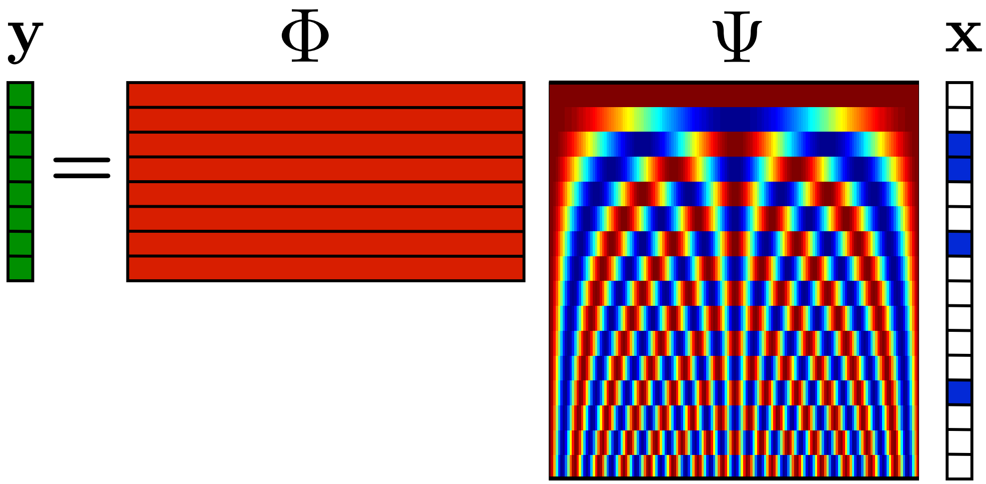
$$\Theta = \{\theta_1, \ldots, \theta_N\}$$

- 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][...]

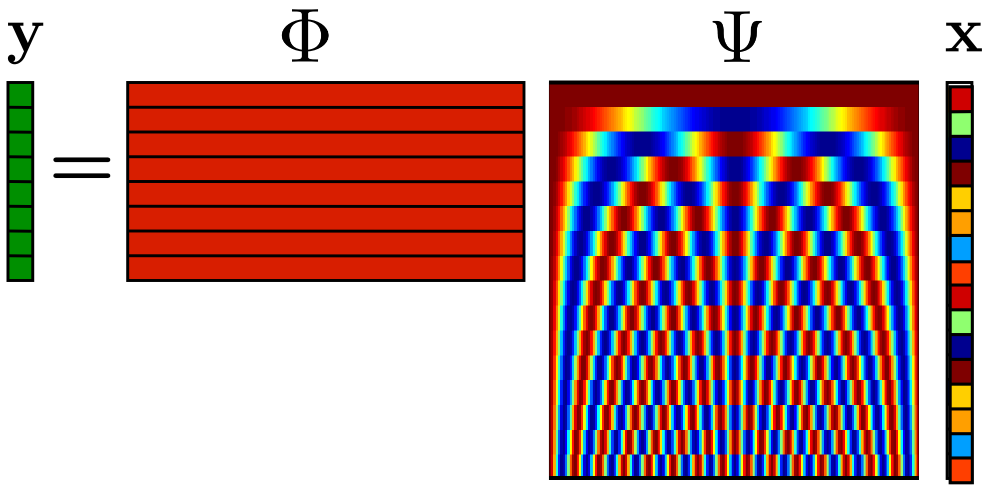
Parameter Estimation in Compressive Sensing



[Strohmer and Herman 2012]
[Chi, Pezeshki, Scharf, Calderbank 2012]
[Liao and Fannjiang 2012]
[Duarte and Baraniuk 2013]

- "Retrofitting" sparsity in CS for common parameter estimation problems (spectral estimation, localization, bearing estimation) results in several issues: dictionary coherence, basis/discretization mismatch, suboptimal sparsity…
- Need new signal models that rely on continuous parameter space and are widely applicable

Parameter Estimation in Compressive Sensing



[Strohmer and Herman 2012]
[Chi, Pezeshki, Scharf, Calderbank 2012]
[Liao and Fannjiang 2012]
[Duarte and Baraniuk 2013]

- “Retrofitting” sparsity in CS for common parameter estimation problems (spectral estimation, localization, bearing estimation) results in several issues: dictionary coherence, basis/discretization mismatch, suboptimal sparsity…
- Need new signal models that rely on continuous parameter space and are widely applicable

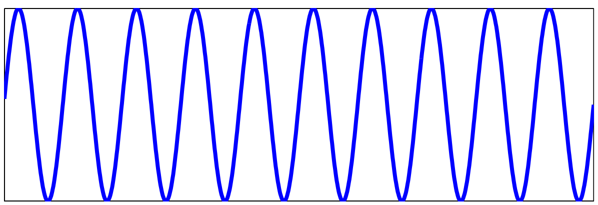
Parametric Signals and Basis Mismatch

$$x = \sum_{k=1}^{K} a_k e(f_k) \qquad X(f) = \sum_{k=1}^{K} a_k \delta(f - f_k) \quad \text{DTFT}$$

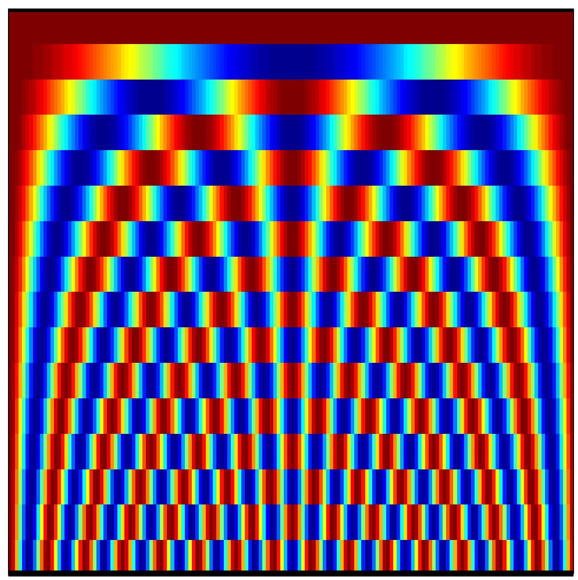
$$e(f) = \left[1 e^{j2\pi f/N} \;\; e^{j2\pi 2f/N} \;\; \ldots \;\; e^{j2\pi(N-1)f/N} \right]$$

$$\theta = \Psi^{-1} x$$

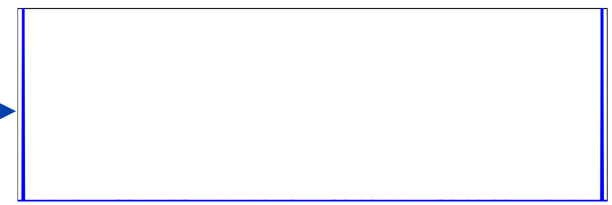
On DFT Grid



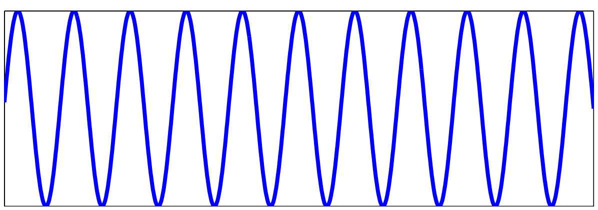
$$x[n] = \sin\left(\frac{2\pi n}{N} \times 10\right)$$



$N = 1024$

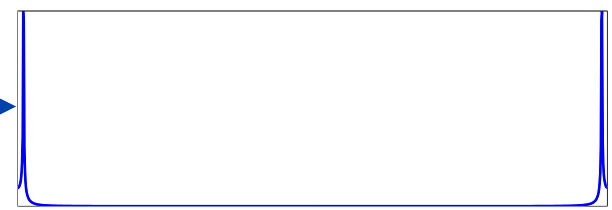


$$\|\theta\|_0 = 2, \|\theta - \theta_2\|_2 = 0$$



$$x[n] = \sin\left(\frac{2\pi n}{N} \times 10.5\right)$$

Off DFT Grid



$$\|\theta\|_0 = 1024, \quad \|\theta - \theta_2\|_2 = 0.76\|\theta\|_2$$

[Herman and Strohmer 2010][Chi, Scharf, Pezeshki, and Calderbank 2011]

Resolution in Frequency Domain

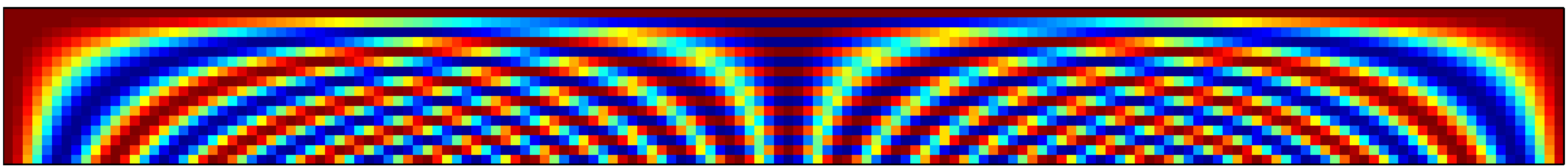
- Redundant Fourier Frame $\Psi(c)$

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

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

$N = 1024$

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



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

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 $\hat{s} = \hat{s}_i$

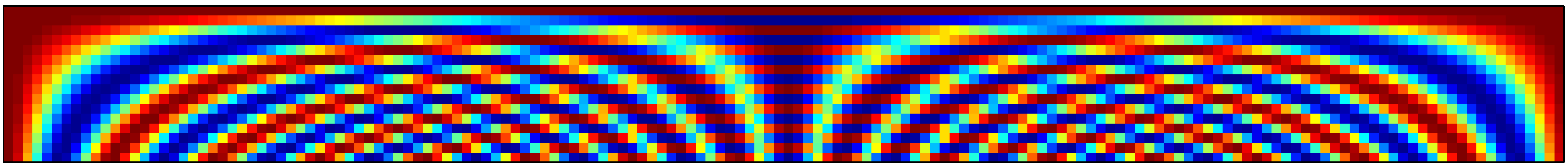
[Blumensath and Davies 2009]

Resolution in Frequency Domain

- Redundant Fourier Frame $\Psi(c)$

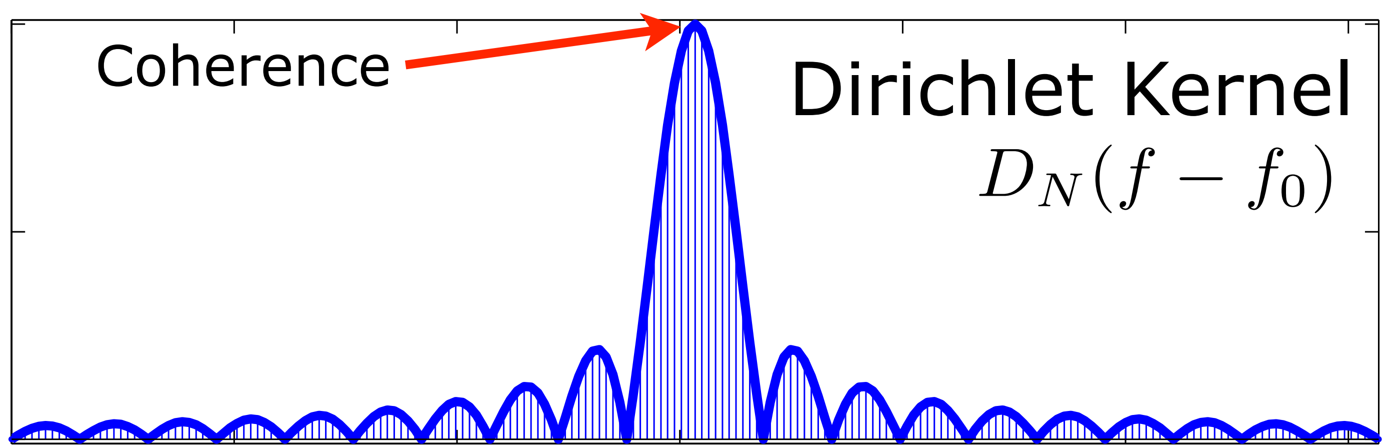
$$\Psi(c) = \left[e\left(\frac{1}{c}\right) \; e\left(\frac{2}{c}\right) \; \ldots \; e\left(\frac{N-1/c}{c}\right) \right] \quad N = 1024$$

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

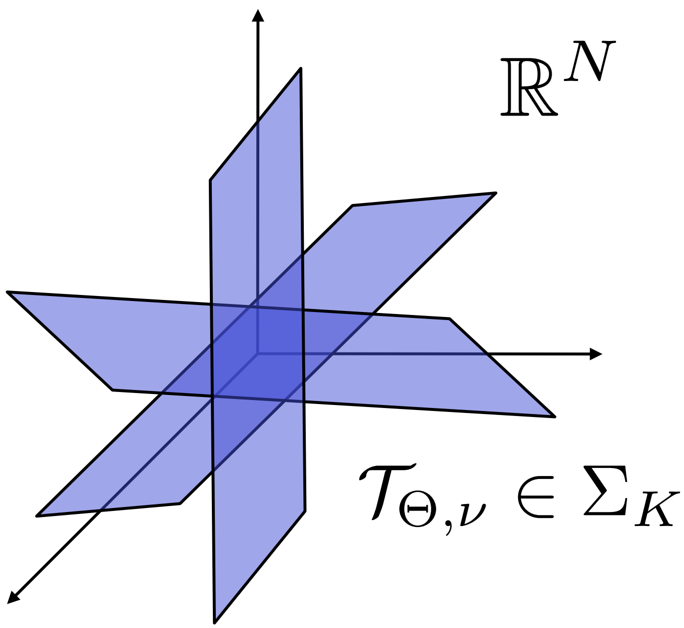


Sparse signal recovery resembles ***matched filtering***:

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



Structured Frequency-Sparse Signals

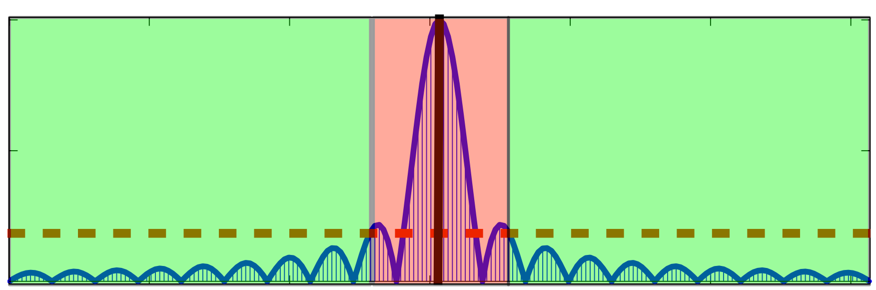


- A K**-structured PD-sparse** signal f consists of K PD elements that are mutually incoherent:

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

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

- 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***).



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

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]

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]

Standard Sparse Signal Recovery

Approximate Message Passing (AMP)

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,
  - $i \leftarrow i + 1$
  - $b \leftarrow \hat{s}_{i-1} + \Phi^T r_i$ *(estimate signal)*
  - $\hat{s}_i \leftarrow \mathcal{T}(b, K)$ *(obtain best sparse approx.)*
  - $r_i \leftarrow y - \Phi\hat{s}_i + \frac{1}{\delta} r_{i-1} \langle \mathcal{T}'(\Phi^T r_{i-1} + \hat{s}_i) \rangle$ (***Onsager correction term***)
- Return estimate $\hat{s} = \hat{s}_i$

Onsager correction term shapes b to resemble ***input signal*** x ***in Gaussian noise***

[Donoho, Maleki, and Montanari 2009]

The Power of Approximate Message Passing

- AMP: based on message passing algorithms
- ***Onsager correction term*** $O(x) = \langle \mathcal{T}'(x) \rangle$ "shapes" the distribution of the signal estimate b to resemble the original signal embedded in additive white Gaussian noise

[Donoho, Maleki, and Montanari 2009]

$\langle x \rangle = \frac{1}{N}\mathbf{1}^T x$

$\mathcal{T}'(x)$: Divergence of hard thresholding (sum of dimension-wise derivatives)

- Intuition: hard thresholding provides ***optimal denoising*** for sparse signals embedded in additive white Gaussian noise

[Metzler, Maleki, and Baraniuk 2014]

The Flexibility of AMP

- AMP algorithm can be extended to arbitrary signal models by using ***optimal denoising algorithm*** $\mathcal{D}(x)$ for signals in additive Gaussian noise
[Donoho, Johnstone, and Montanari 2013]
- Examples: hard thresholding for sparse signals; block thresholding for block-sparse signals; total variation denoisers for piecewise constant signals; image denoising algorithms
[Metzler, Maleki, and Baraniuk 2014]
[Tan, Ma, and Baron 2015]
- "Flexible" AMP requires formulation of ***new Onsager correction*** term specific to denoiser applied
- Can also estimate numerically via ***Monte Carlo*** iterations

$$\hat{O}(x) = \lim_{\epsilon \to 0} \mathbb{E}_b \left[\frac{b^T(\mathcal{D}(x + \epsilon b) - \mathcal{D}(x))}{\epsilon} \right]$$

with $b \sim \mathcal{N}(0, I)$; average over multiple draws of b with small ϵ to obtain numerical estimate of expected value.

[Metzler, Maleki, and Baraniuk 2014]

AMP with Parametric Denoisers

- ***Statistical parameter estimation*** algorithms can be paired with generative signal models to provide ***"parametric denoisers"***
- Rich literature in statistical parameter estimation for a multitude of problems (including line spectral estimation)
- Estimate Onsager correction term numerically:

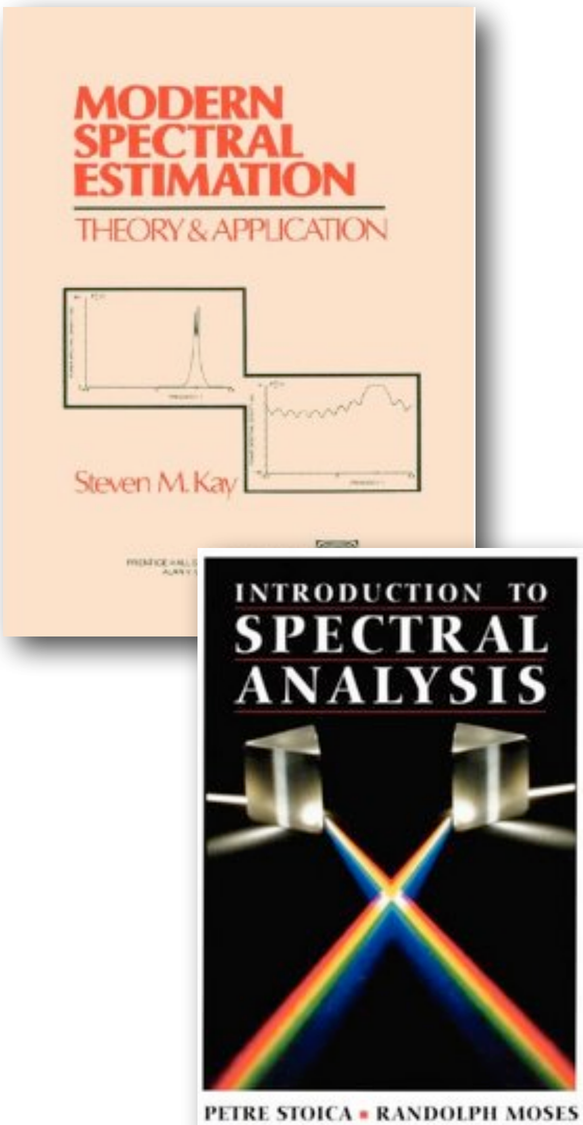
$$\hat{O}(x) = \lim_{\epsilon \to 0} \mathbb{E}_b \left[\frac{b^T (\mathcal{D}(x + \epsilon b) - \mathcal{D}(x))}{\epsilon} \right]$$

$b \sim \mathcal{N}(0, I)$; average over multiple realizations of b with a small value of ϵ (e.g., $\epsilon = \|x\|_\infty / 1000$) to obtain numerical estimate of expected value.

- In practice, 1-2 iterations often suffice.

[Metzler, Maleki, and Baraniuk 2014]

Parametric Denoising via Line Spectral Estimation



Line Spectral Estimation-Based Denoising Algorithm $\mathbb{T}_l(x, K)$

Inputs:

- Noisy observation x
- Target sparsity K

Output:

- Parameter estimates $\widehat{a}_1, \ldots, \widehat{a}_K$, $\widehat{f}_1, \ldots, \widehat{f}_K$
- Denoised signal $\widehat{x}$

x, K → MUSIC / Root MUSIC / ESPRIT / PHD / … → $\widehat{a}_1, \ldots, \widehat{a}_K$, $\widehat{f}_1, \ldots, \widehat{f}_K$

$$\mathbb{T}_l(x, K) \longrightarrow \widehat{x} = \sum_{k=1}^{K} \widehat{a}_k e(\widehat{f}_k)$$

Prior Use of Denoisers as Sparse Approximation Algorithms

IHT + Line Spectral Estimation

Inputs:

- Measurement vector y
- Measurement matrix Φ
- 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,
  - $i \leftarrow i + 1$
  - $b \leftarrow \hat{s}_{i-1} + \Phi^T r$ *(estimate signal)*
  - $\hat{s}_i \leftarrow \mathbb{T}_l(b, K)$ *(obtain* ***LSE parametric*** *sparse approx.)*
  - $r \leftarrow y - \Phi\Psi\hat{s}_i$ *(calculate residual)*
- Return estimate $\hat{s} = \hat{s}_i$

[Duarte and Baraniuk, 2012]

Denoising via Line Spectral Estimation

AMP + Line Spectral Estimation

Inputs:

- Measurement vector y
- Measurement matrix Φ
- Sparsity K

Output:

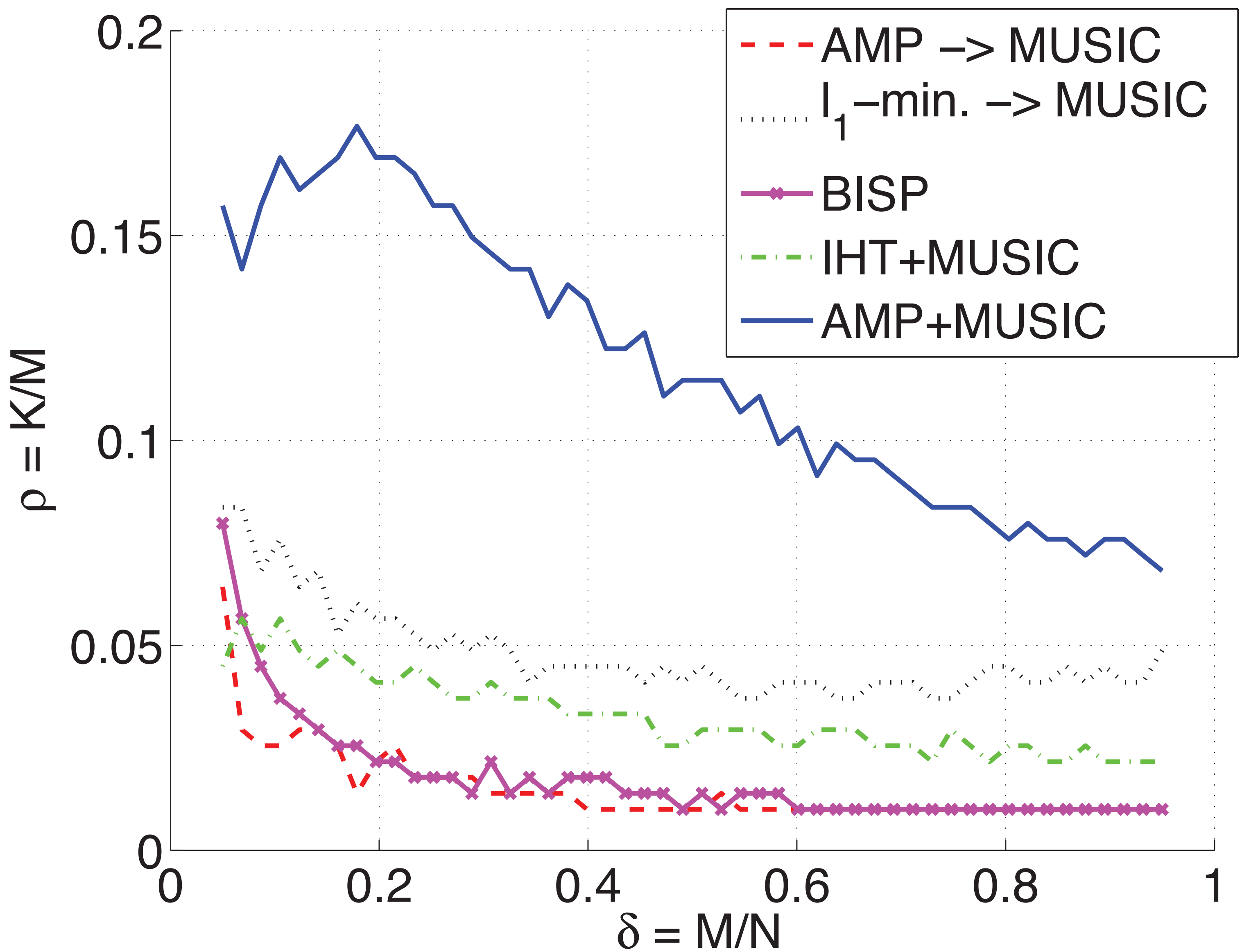
- PD coefficient estimate $\hat{s}$

- Initialize: $\hat{s}_0 = 0, r = y, i = 0$
- While halting criterion false,
  - $i \leftarrow i + 1$
  - $b \leftarrow \hat{s}_{i-1} + \Phi^T r_i$ *(estimate signal)*
  - $(\hat{s}_i, \{f_k\}_{k=1}^K) \leftarrow \mathbb{T}_l(b, K)$ *(obtain frequency estimates)*
  - $r_i \leftarrow y - \Phi \hat{s}_i + \frac{1}{\delta} r_{i-1} \langle \mathbb{T}'_l(\Phi^T r_{i-1} + \hat{s}_i) \rangle$

    (numerical estimation of Onsager correction term)
- Return estimate $\hat{s} = \hat{s}_i$

[Hamzehei and Duarte 2015]

Phase Transition Diagram for Compressive Line Spectral Estimation

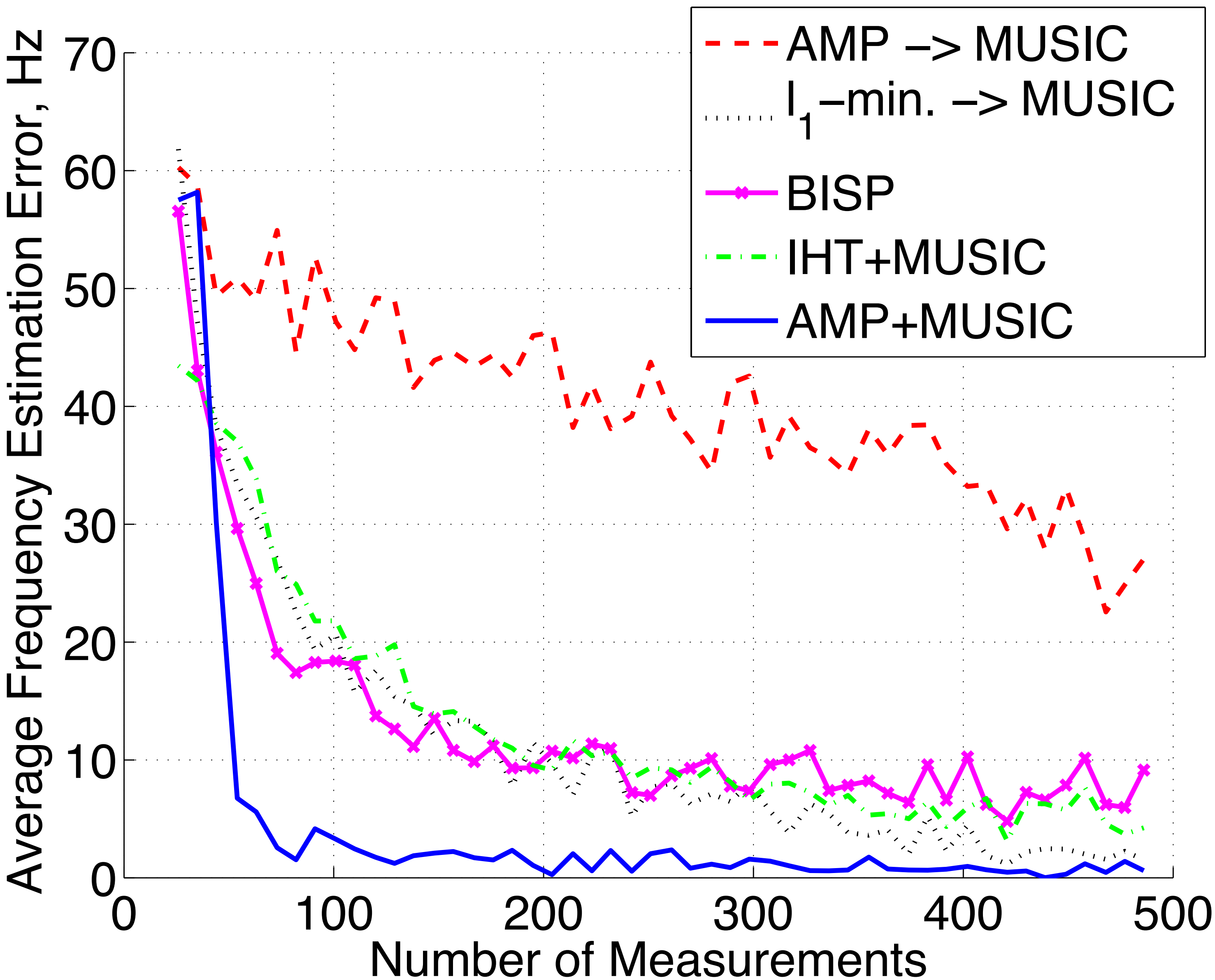


Success: Average frequency estimation error < 1 Hz over 100 trials

$$N = 512$$

BISP: Band-Excluding Interpolating Subspace Pursuit [Fyhn, Duarte, Jensen 2014]

Phase Transition Diagram for Compressive Line Spectral Estimation



$N = 512$
$K = 8$
100 trials

Conclusions

- Approximate Message Passing is ***flexible enough*** to be extended from compressive sensing (signal recovery) to compressive parameter estimation
- Leverage existing statistical estimation algorithms as ***"parametric denoisers"*** within AMP
- Sidestep discretization issues implicit in the use of parametric dictionaries and parameter tuning issues from structured sparsity models
- Additional computation in Onsager term estimation
- ***Future work***: theoretical analysis (state evolution, denoiser analysis, measurement bounds…)
- Many other example applications: bearing estimation, time delay estimation, localization…

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