Subspaces and sparsity on the continuum

Mark A. Davenport

School of Electrical and Computer Engineering

Georgia Tech

Santhosh Karnik

Rakshith Sharma

Justin Romberg

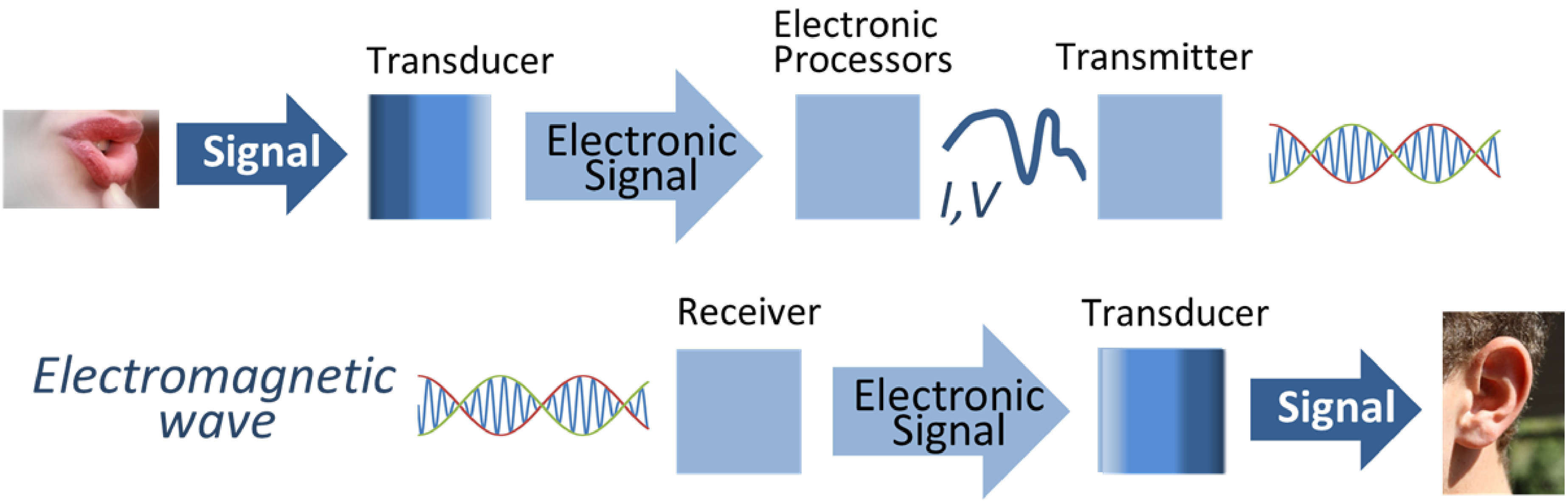
Georgia Tech

Zhihui Zhu

Michael Wakin

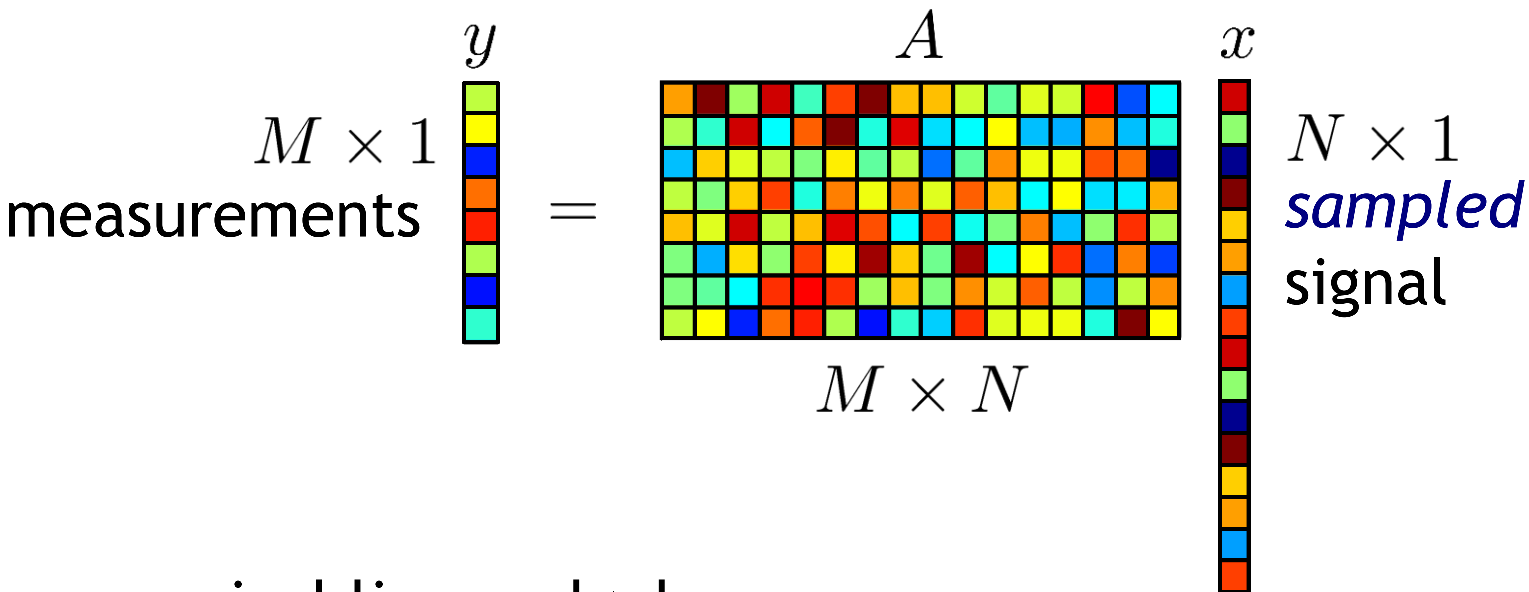
COLORADO SCHOOL OF MINES

Traditional view of signal processing



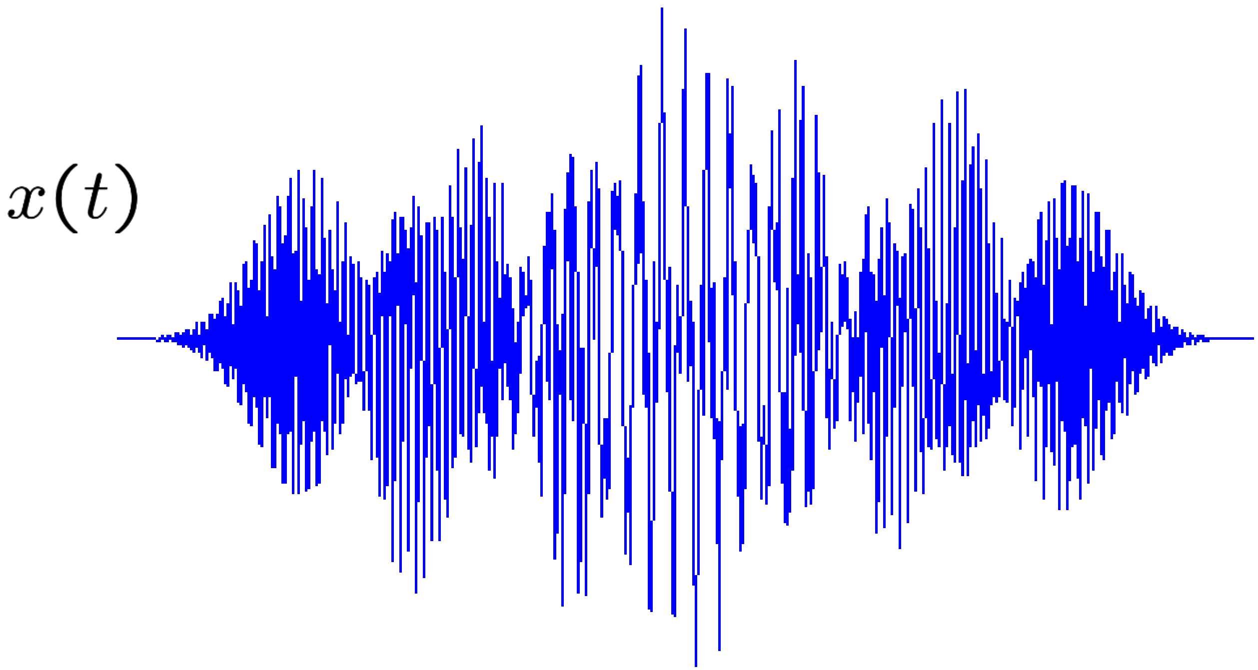
[Wikipedia]

Modern signal processing



- numerical linear algebra
- optimization
- subspaces
- sparsity

Modeling on the continuum



In many applications, the most natural signal models are inherently ***continuous***

Translating this to a discrete, finite setting can be subtle

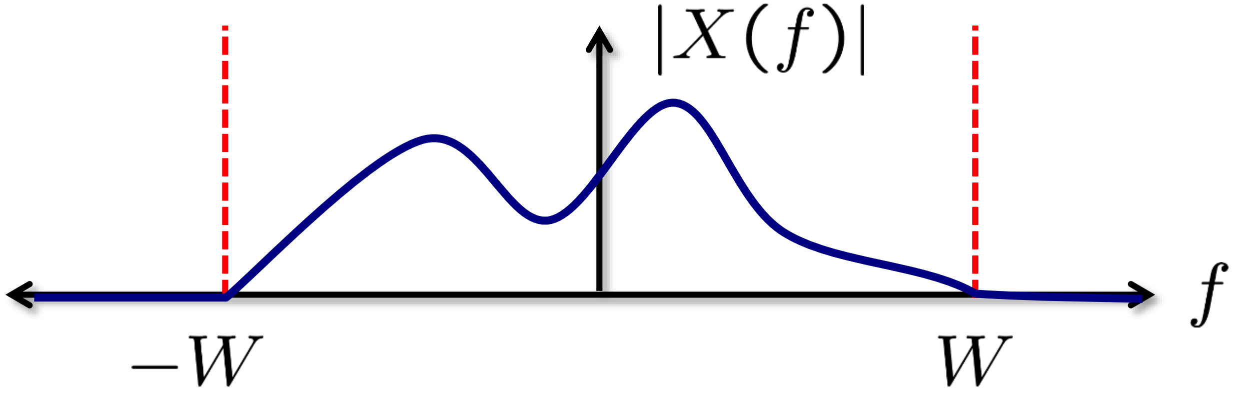
Bandlimited functions

Perhaps the most basic model is that $x(t)$ is ***bandlimited***

The ***continuous-time Fourier transform*** of a function $x(t)$ is given by

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi f t}\, dt, \;\; f \in \mathbb{R}$$

We say that $x(t)$ is ***bandlimited*** (with bandlimit W) if $X(f) = 0$ for $|f| > W$

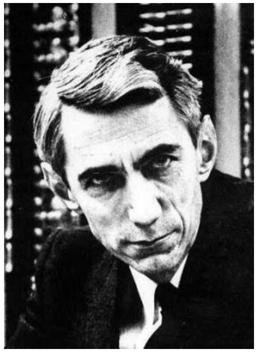


Sampling bandlimited functions

"If we sample a signal at twice its highest frequency, then we can recover it exactly."

Whittaker-Nyquist-Kotelnikov-Shannon



More specifically, let T_S denote the sampling period and let $x[n] = x(nT_s)$ denote the sequence of samples we obtain

The sampling theorem shows us that no information is lost provided $W \leq \frac{1}{2T_S}$

Windows of samples

To simplify our notation, we will assume without loss of generality that $T_s = 1$ so that

$$\boldsymbol{x}[n] = x(n), \;\; n = 0, 1, \ldots, N-1$$

$W = \frac{1}{2}$: sampling at the Nyquist rate

→ N degrees of freedom

$W < \frac{1}{2}$: sampling faster than the Nyquist rate

→ $< N$ degrees of freedom?

Models for bandlimited signals

If $W \ll \frac{1}{2}$, we expect that $\boldsymbol{x}$ has $\ll N$ degrees of freedom

How can we represent this mathematically?

From bandlimitedness we have

$$x[n] = \int_{-W}^{W} X(f) e^{-i2\pi f n} \, df$$

The discrete Fourier transform (DFT) gives a representation of the form

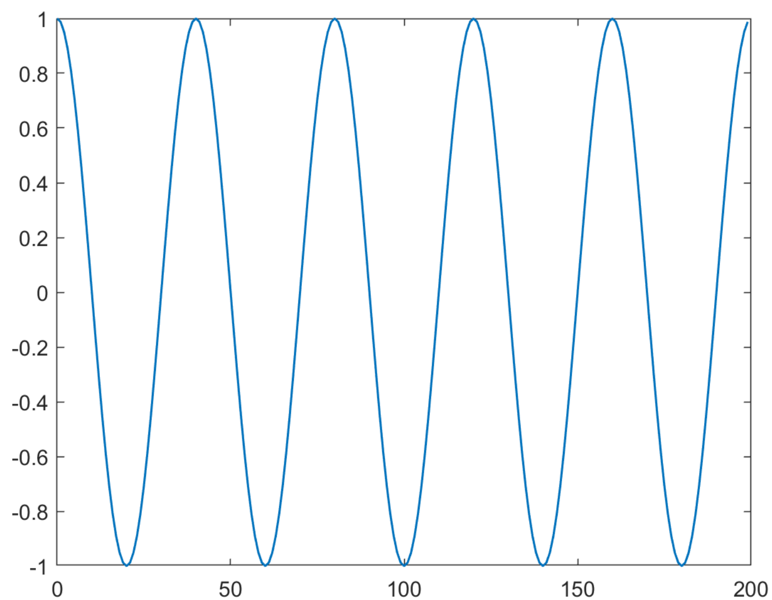
$$\boldsymbol{x}[n] = \frac{1}{N} \sum_{k=0}^{N-1} \boldsymbol{X}[k] e^{-i2\pi(k/N)n}$$

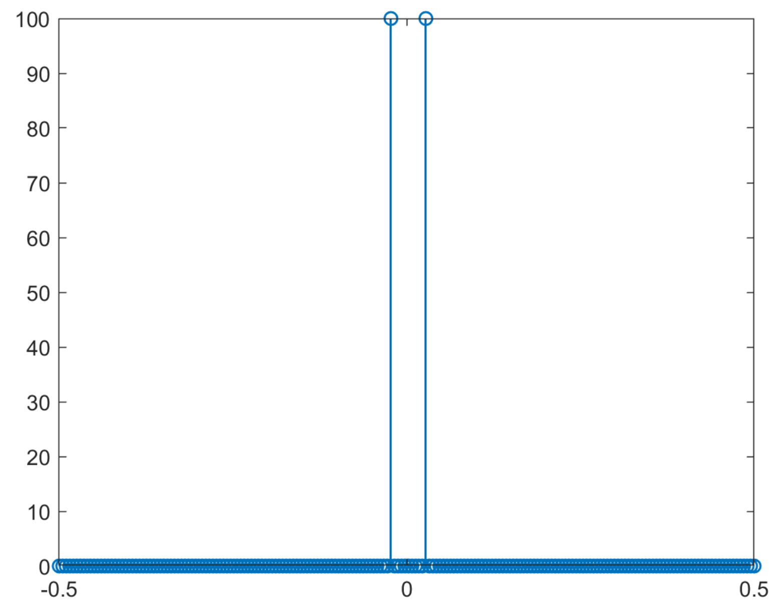
Models for bandlimited signals

If $W \ll \frac{1}{2}$, we expect that x has $\ll N$ degrees of freedom

How can we represent this mathematically?

The DFT should be ***sparse***



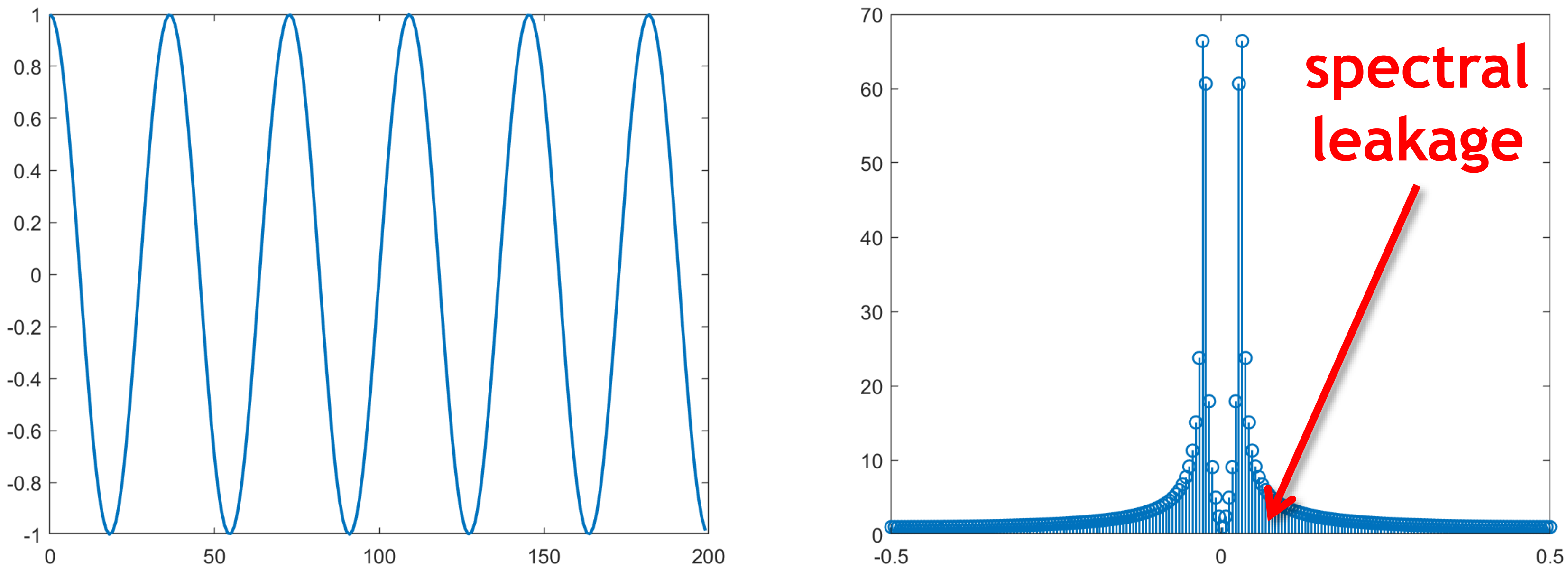


Models for bandlimited signals

If $W \ll \frac{1}{2}$, we expect that x has $\ll N$ degrees of freedom

How can we represent this mathematically?

The DFT should be ***sparse*** - **but it usually isn't…**

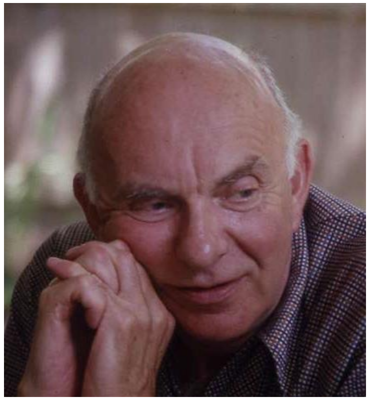


A better model

The DFT is simply the wrong basis for compactly representing this structure

A much better choice: *discrete prolate spheroidal sequences*

Slepian basis. Defined by the vectors that satisfy the eigenvalue equation



$$\mathcal{T}_N(\mathcal{B}_W(\boldsymbol{s}_\ell)) = \lambda_{N,W}^{(\ell)} \boldsymbol{s}_\ell$$

The first $\approx 2NW$ eigenvalues ≈ 1.
The remaining eigenvalues ≈ 0.

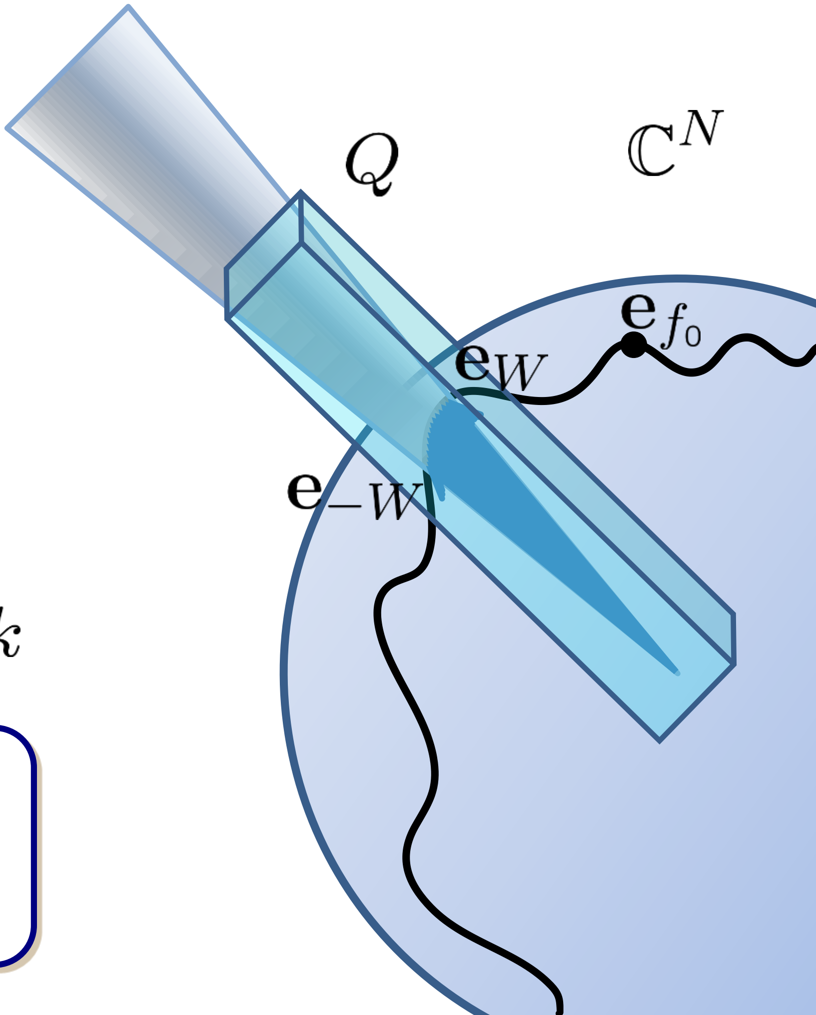
Another perspective: Subspace fitting

$$\mathbf{e}_f := \begin{bmatrix} e^{i2\pi f 0} \\ e^{i2\pi f} \\ \vdots \\ e^{i2\pi f(N-1)} \end{bmatrix}$$

Suppose that we wish to minimize

$$\int_{-W}^{W} \|\mathbf{e}_f - P_Q \mathbf{e}_f\|_2^2 \, df$$

over all subspaces Q of dimension k



Optimal subspace is spanned by the first k Slepian basis elements

The prolate matrix

From either perspective, it is not hard to show that the Slepian basis elements are the eigenvectors of the ***prolate matrix*** $\boldsymbol{B}_{N,W}$

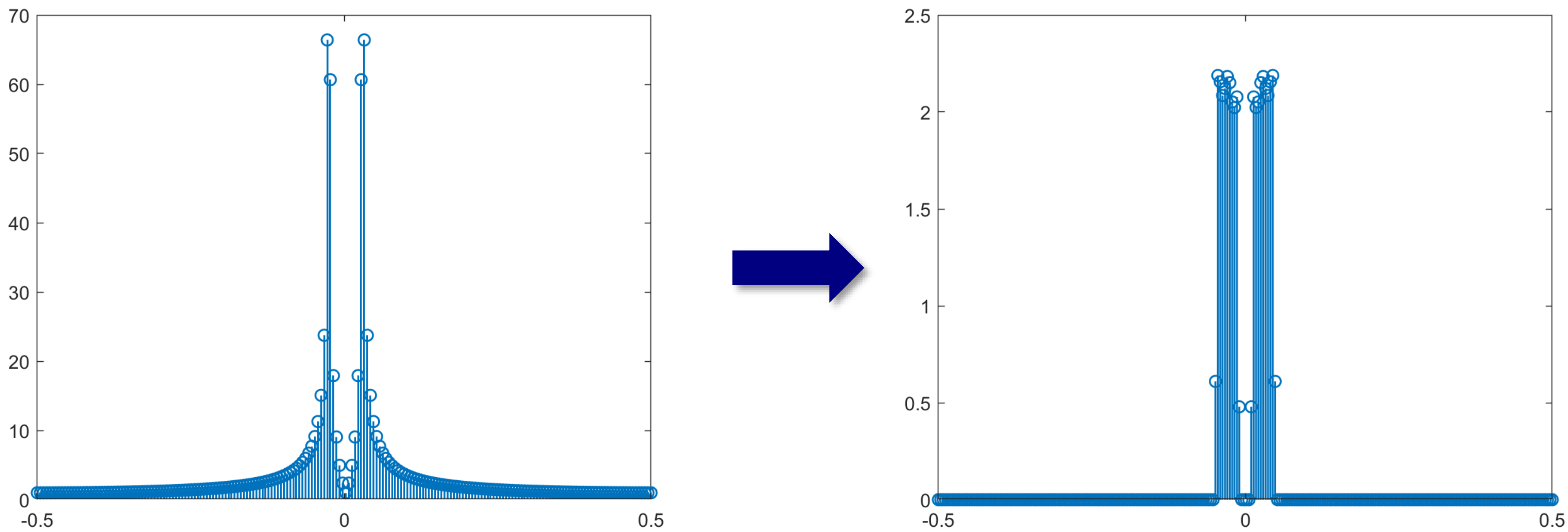
$$\boldsymbol{B}_{N,W}[m,n] = \begin{cases} \frac{\sin(2\pi W(m-n))}{\pi(m-n)} & \text{if } m \neq n \\ 2W & \text{if } m = n \end{cases}$$

$$\boldsymbol{B}_{N,W} = \boldsymbol{S}_{N,W}\boldsymbol{\Lambda}_{N,W}\boldsymbol{S}^*_{N,W}$$

Bottom line

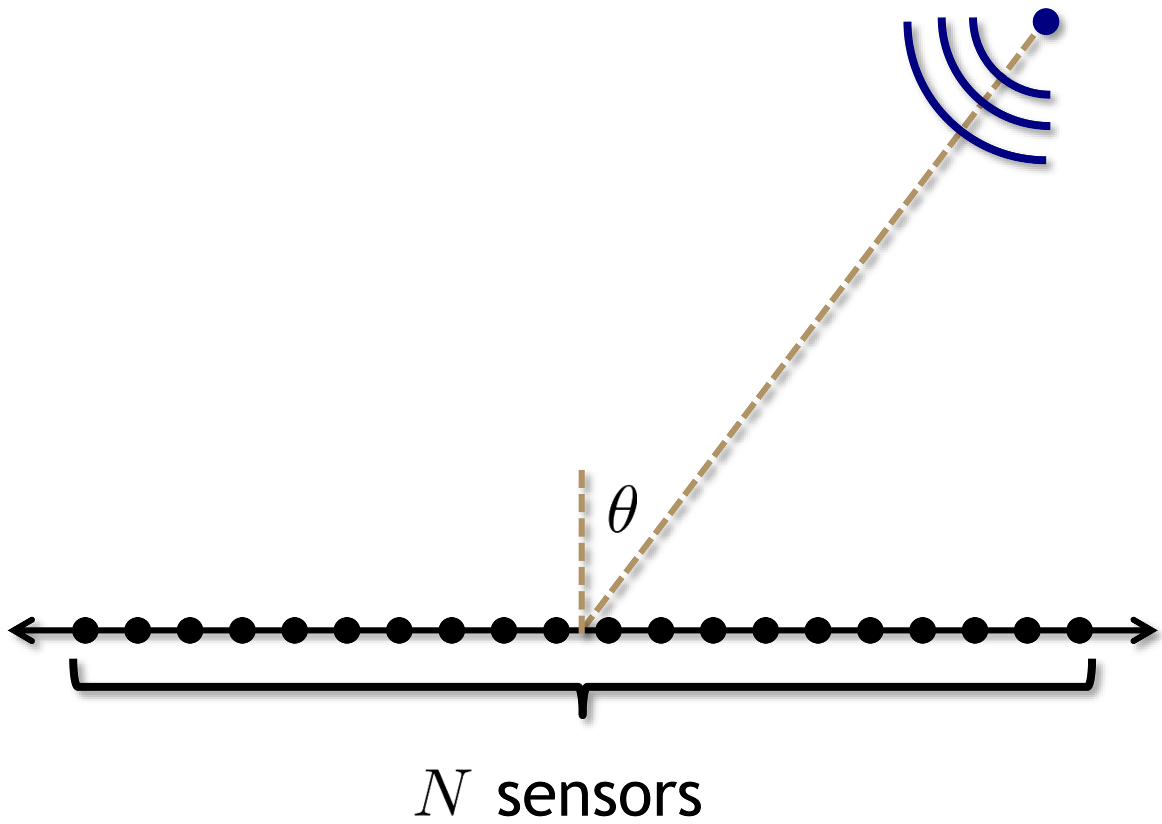
Windowed and sampled bandlimited signals live in a subspace with an effective dimension of $\approx 2NW$

Other frequency bands handled by simply modulating the Slepian basis elements to different center frequencies



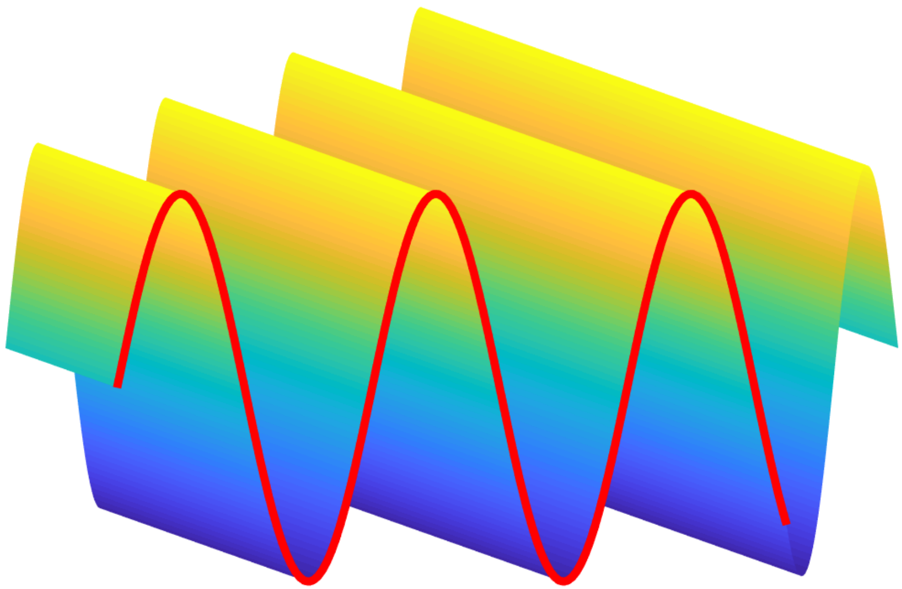
Narrowband DOA

Consider the problem of estimating the direction-of-arrival (DOA) of a narrowband source using a linear array of sensors



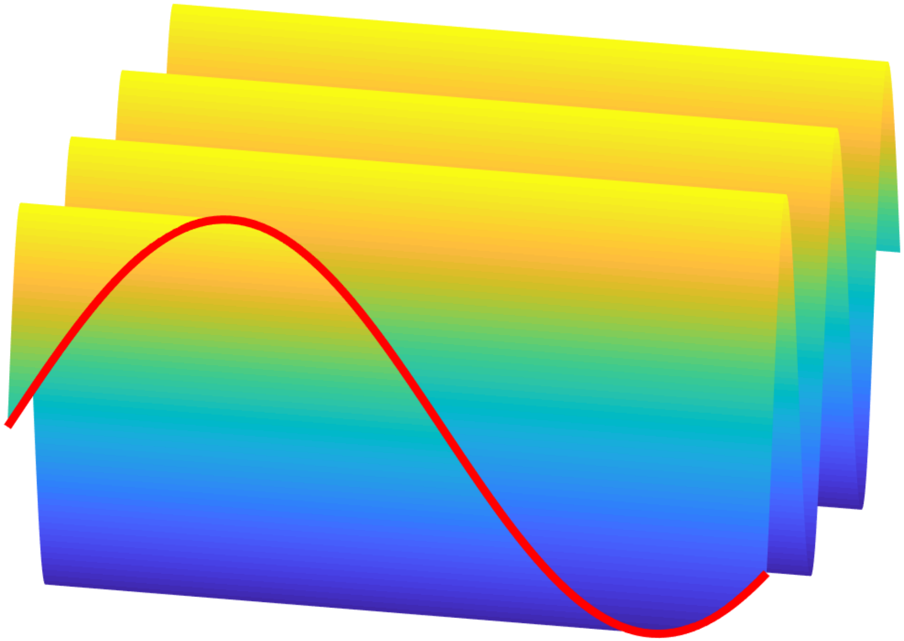
Narrowband DOA – Far field

Assume we can approximate the source as a plane wave



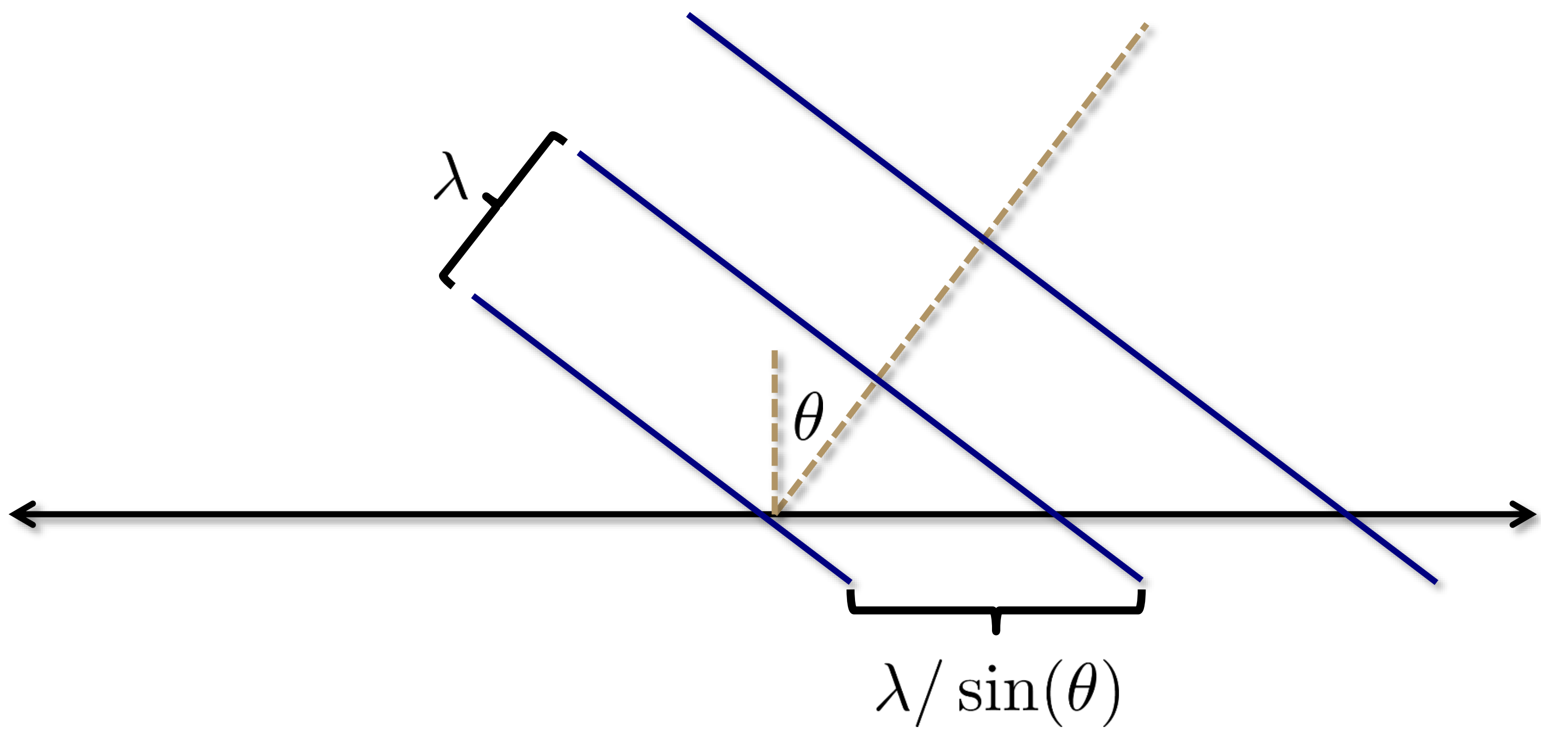
Narrowband DOA – Far field

Assume we can approximate the source as a plane wave



Narrowband DOA – Far field

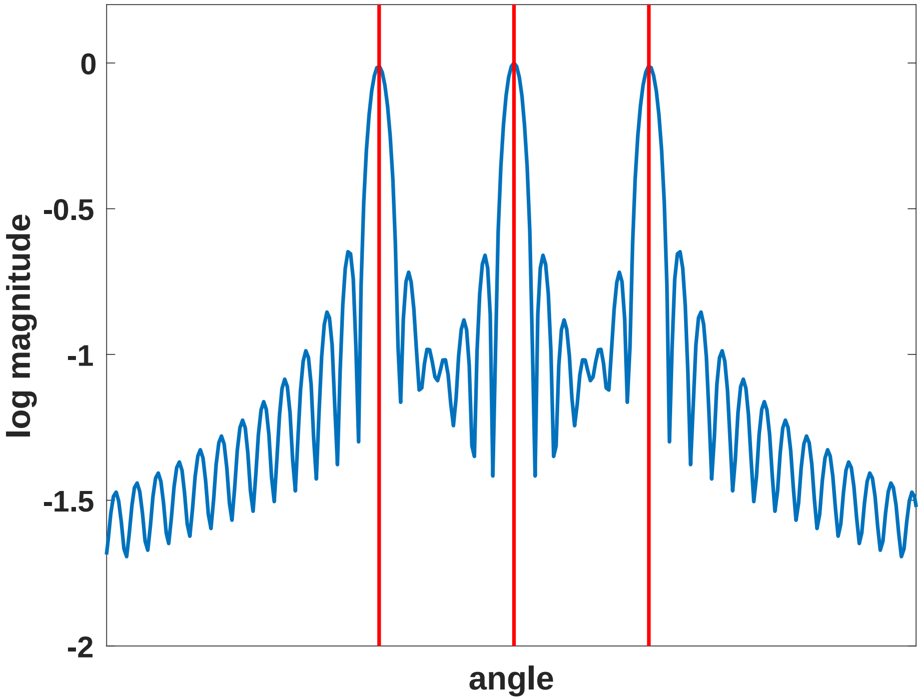
Assume we can approximate the source as a plane wave



Sinusoid at frequency f ➡ Sinusoid at frequency $f \sin(\theta)$

DOA as spectral estimation

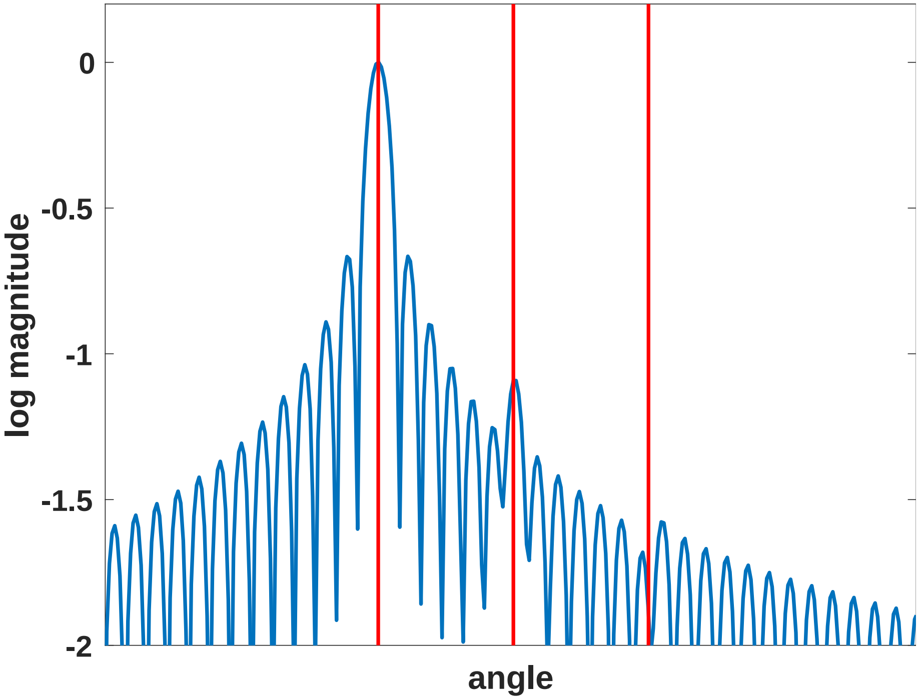
What happens when we have multiple sources?



Three sources, equal magnitude

DOA as spectral estimation

What happens when we have multiple sources?



Magnitudes 1.0, 0.05, 0.01

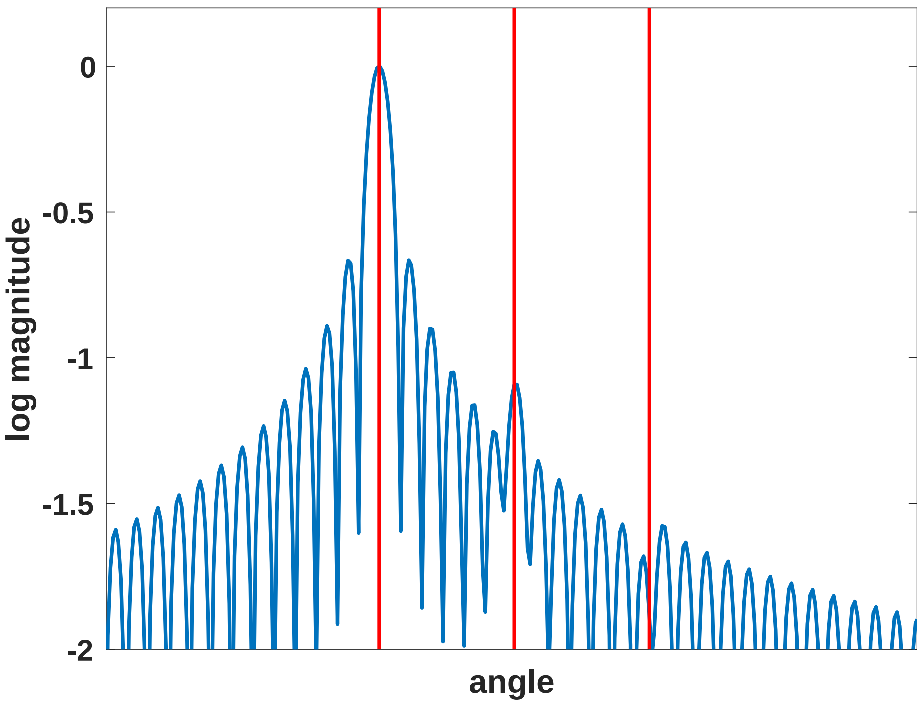
Iterative source localization

Essential procedure

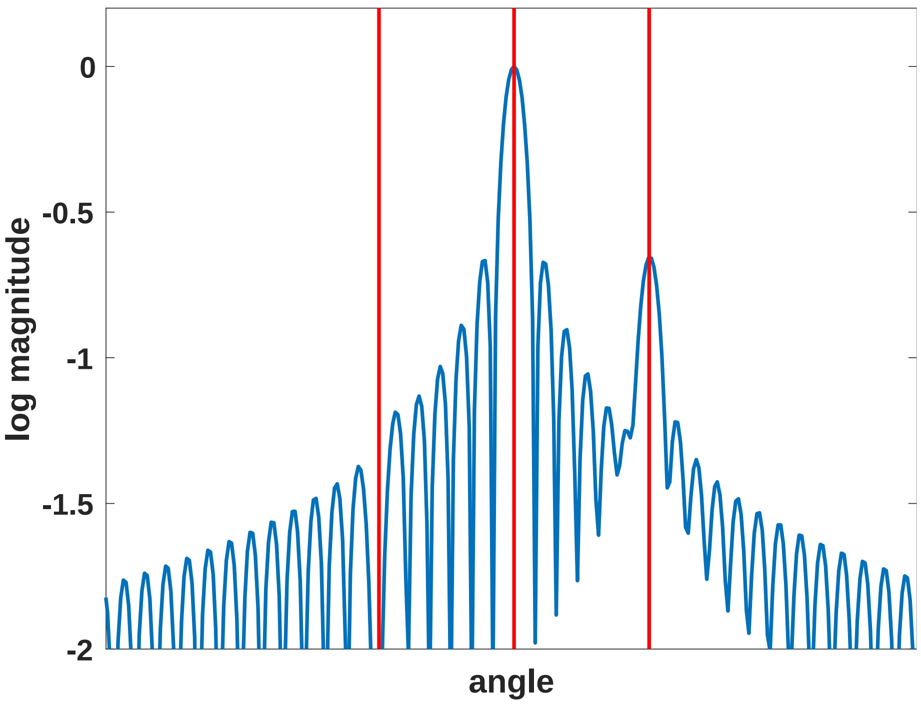
1. Find a source
2. Null it out
   (remove its effect from the measurements)
3. Repeat until no more sources

Nulling must be performed with care…

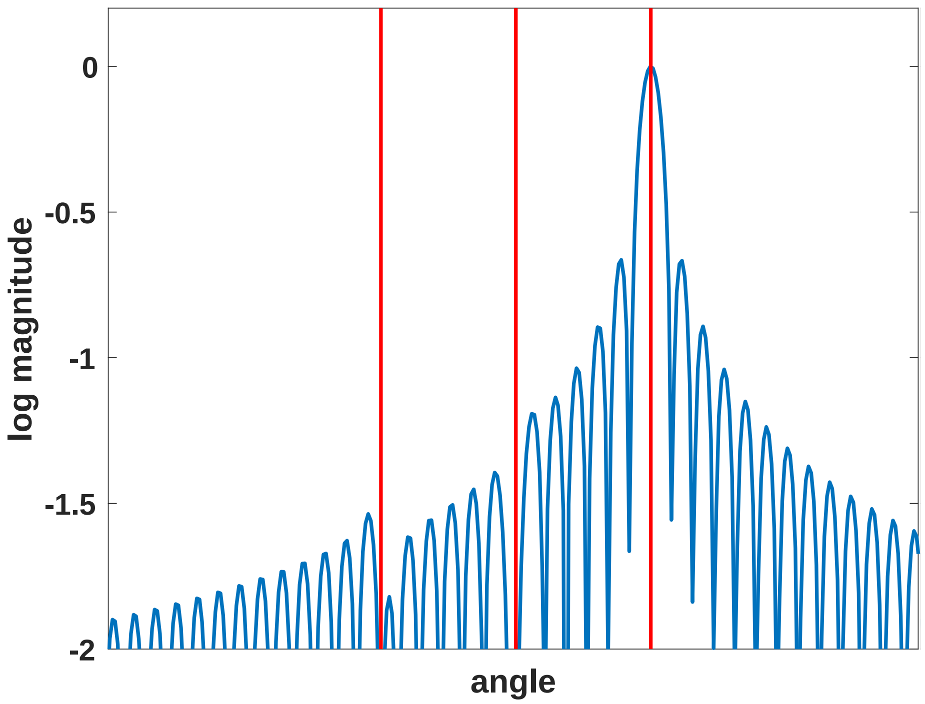
Iterative source localization



Iterative source localization



Iterative source localization



"Nulling" procedure

Given several sources, we observe their sum $\mathbf{x}_1 + \mathbf{x}_2 + \dots$

We do not expect to estimate the angle θ_j (or equivalently, the frequency f_j) corresponding to source $\mathbf{x}_j$ ***exactly***

Suppose that we have an estimate $\widehat{f}_j \in [f_j - \epsilon, f_j + \epsilon]$

Then $\mathbf{x}_j$ lives in a subspace of dimension $\approx 2N\epsilon$, spanned by the Slepian basis elements modulated to $\widehat{f}_j$

We can null by projecting onto the orthogonal complement of this subspace

- choose slightly ***more*** than $2N\epsilon$ basis elements to null nearly all of the energy
- can choose ϵ to account for non-point sources

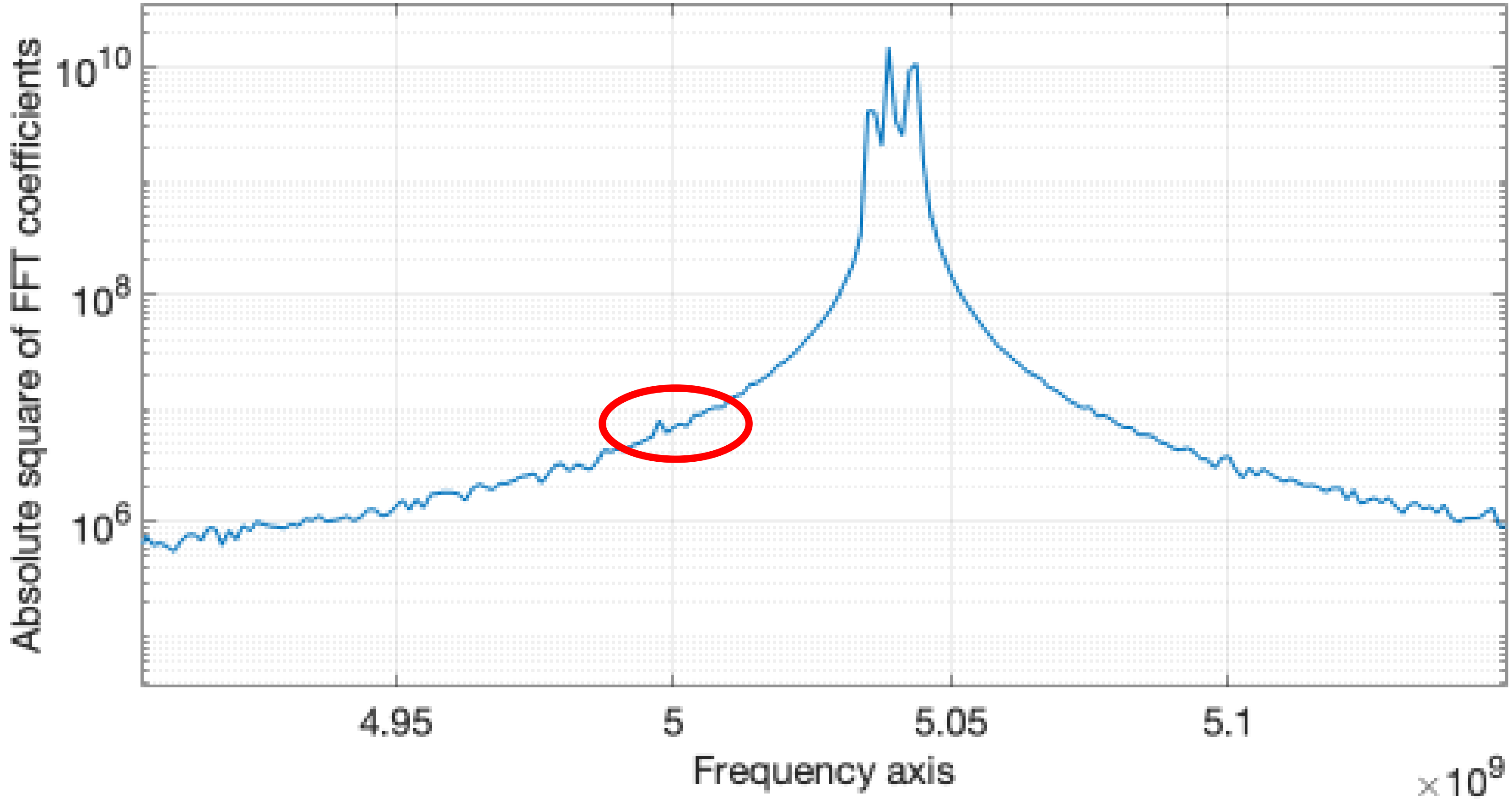
“Rolling Slepian” spectral estimation

We may also wish to use a Slepian-inspired approach to estimating the source angles as follows:

1. Project onto the “Slepian subspace” corresponding to an interval of bandwidth $[f - \epsilon, f + \epsilon]$
2. Compute the energy in this projection
3. Sweep over all frequencies f

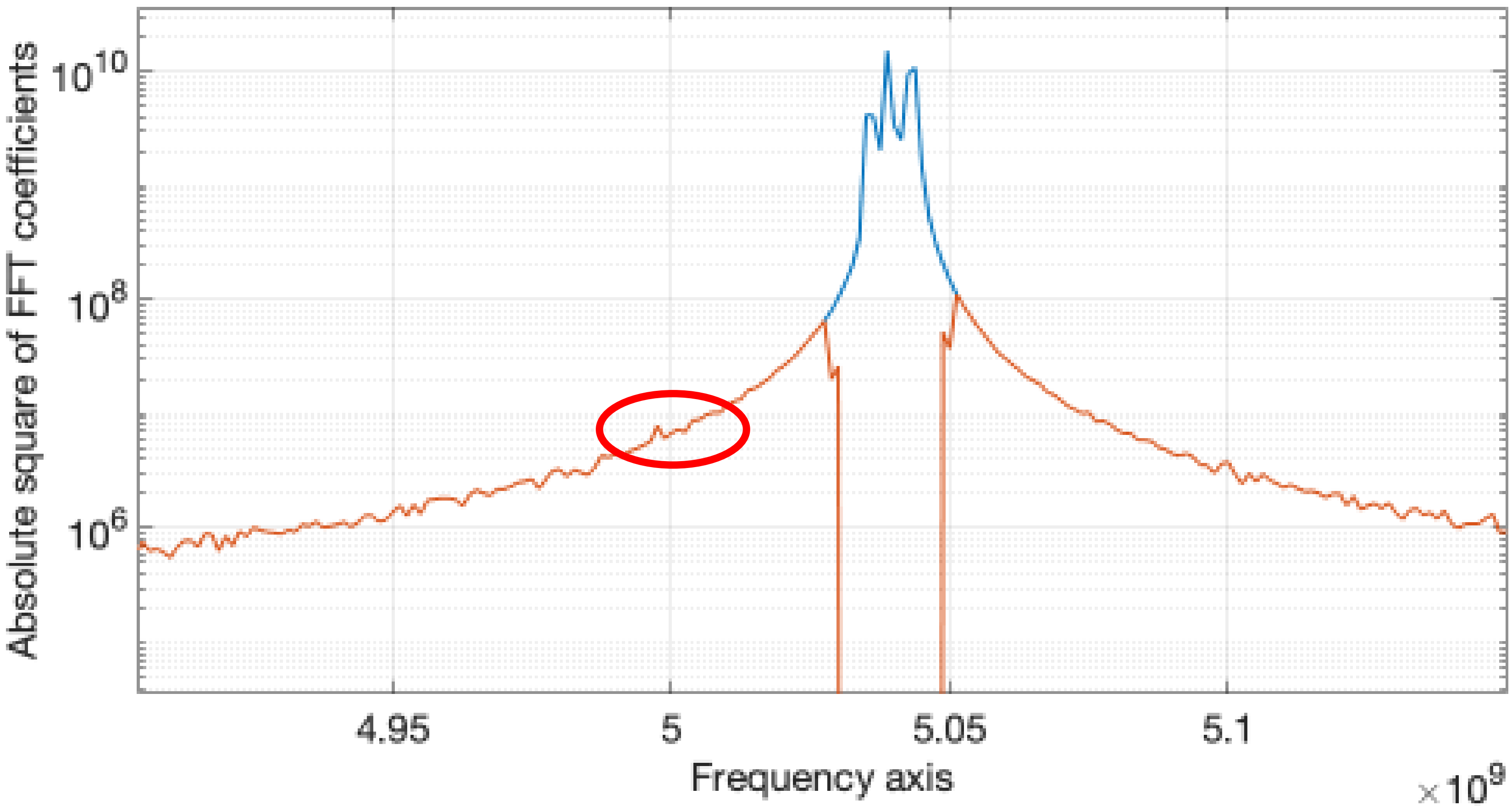
By choosing slightly ***fewer*** than $2N\epsilon$ basis elements, we can nearly eliminate spectral leakage

Example



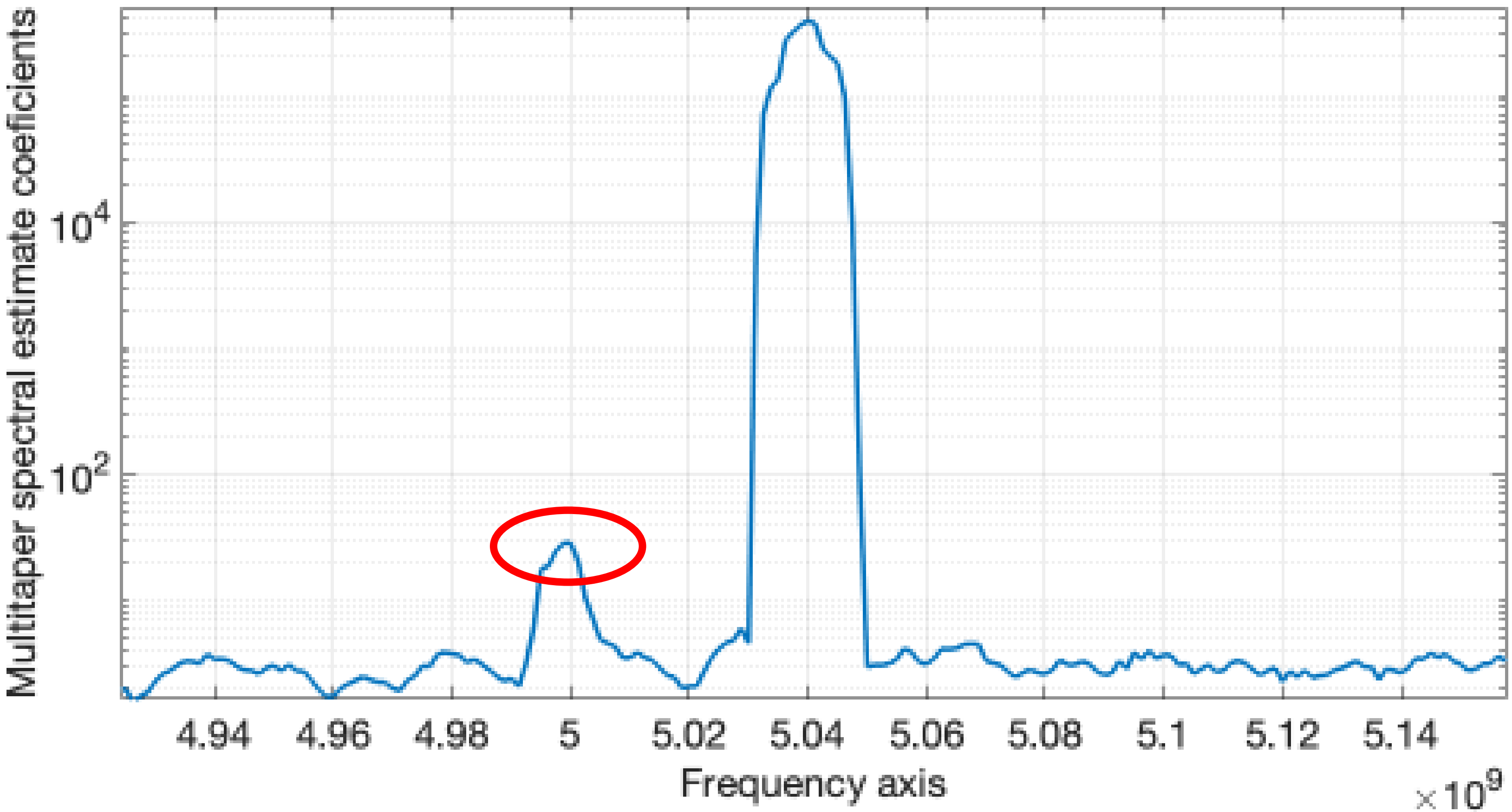
Two 10 MHz bandwidth sources:
One centered at 5.04 GHz ….. one 100x lower at 5.0 GHz

Example



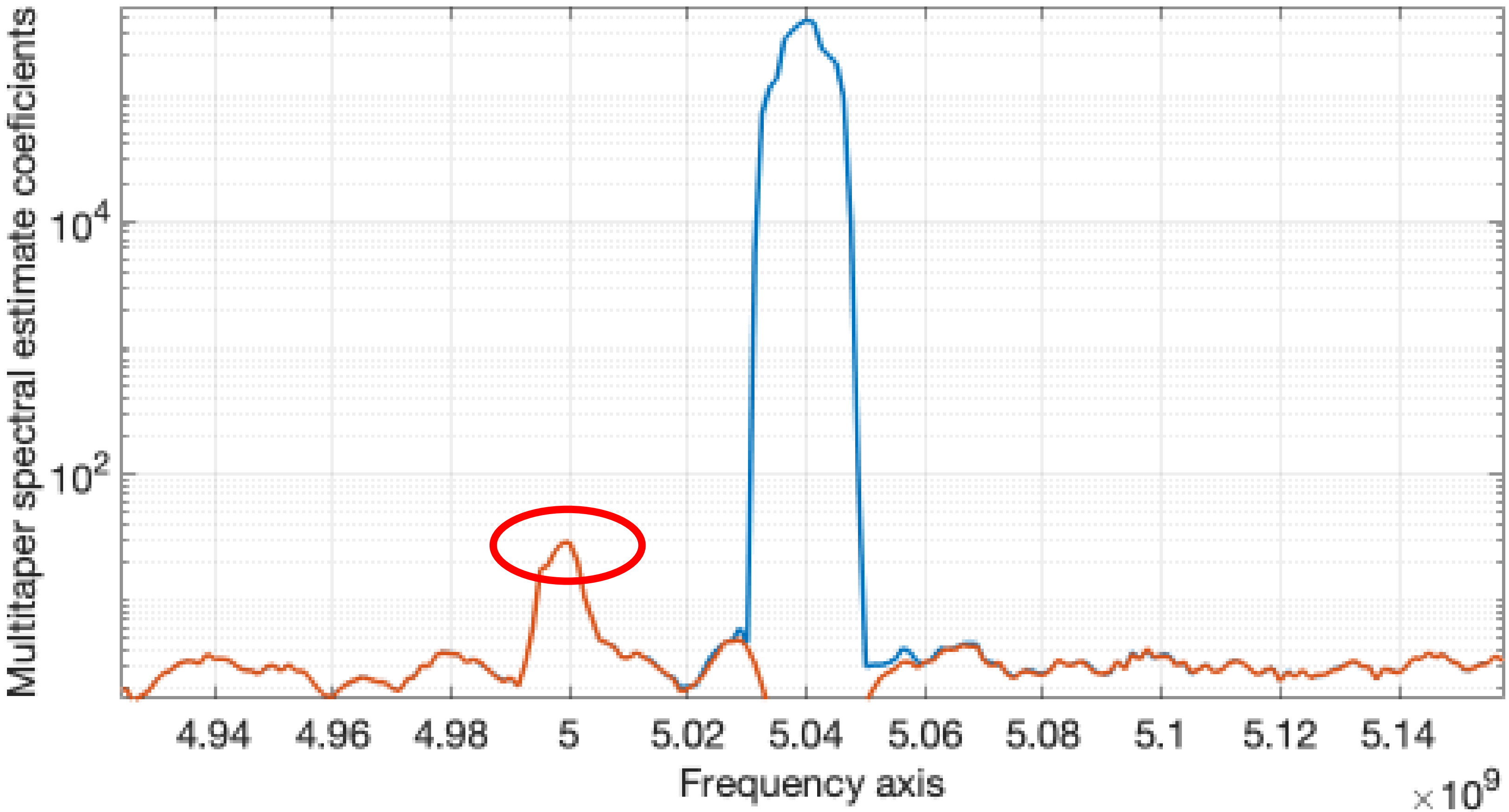
Residual when we "null" using DFT vectors

Example



"Rolling Slepian" spectral estimate

Example



Slepian projection cleanly reveals smaller source

Thomson's multitaper method

Is the "Rolling Slepian" spectral estimate related to Thomson's multitaper method?

They are equivalent!

$$\widehat{S}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \left| \sum_{n=0}^{N-1} \underbrace{\mathbf{x}[n]\mathbf{s}_k[n]}_{\text{windowing}} e^{-i2\pi f n} \right|^2$$

windowing

DFT energy

Average over windows

Thomson's multitaper method

Is the "Rolling Slepian" spectral estimate related to Thomson's multitaper method?

They are equivalent!

$$\widehat{S}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \left| \sum_{n=0}^{N-1} \mathbf{x}[n] \underbrace{\mathbf{s}_k[n] e^{-i2\pi f n}}_{\text{Slepian basis element}} \right|^2$$

Slepian basis element

projection energy

Sum across all basis elements

Summary so far

The Slepian basis provides a natural tool for working with (finite windows) of sampled bandlimited signals

- Subspace modeling
  projecting a vector in $\mathbb{C}^N$ onto span of the first $\approx 2NW$ Slepian basis elements to enforce/exploit bandlimited model

- Applications in DOA estimation

- Applications in spectral estimation (Thomson's method)

How can we do this at a speed comparable with the FFT?

Motivating example

Motivating example

- Radar array whose goal is to track every object of size 10 cm or larger in low-Earth orbit
- First radar site expected to go online this year in Marshall Islands, another planned for Australia
- Each radar site has a digital phased array consisting of ~100,000 (S-band) receivers
- TREMENDOUS data volume, need for scalable algorithms

Towards fast Slepian computations

Recall that the Slepian basis can be computed via an eigendecomposition of the so-called ***prolate matrix***

$$\boldsymbol{B}_{N,W} = \boldsymbol{S}_{N,W} \boldsymbol{\Lambda}_{N,W} \boldsymbol{S}_{N,W}^*$$

Let $\boldsymbol{F}_{N,W}$ be the matrix whose columns correspond to the $2NW$ lowest frequency length-N DFT vectors

One can show that $\boldsymbol{B}_{N,W} = \boldsymbol{F}_{N,W} \boldsymbol{F}_{N,W}^* + \boldsymbol{L} + \boldsymbol{E}$, where

$$\operatorname{rank}(\boldsymbol{L}) \lesssim \log(N) \log\left(\tfrac{1}{\epsilon}\right) \qquad \|\boldsymbol{E}\| \leq \epsilon$$

Number of eigenvalues in transition region

$$\boldsymbol{B}_{N,W} = \boldsymbol{S}_{N,W} \boldsymbol{\Lambda}_{N,W} \boldsymbol{S}^*_{N,W} \approx \boldsymbol{F}_{N,W} \boldsymbol{F}^*_{N,W} + \boldsymbol{L}$$

The rank of $\boldsymbol{L}$ gives as a nonasymptotic bound on the number of eigenvalues of $\boldsymbol{B}_{N,W}$ in the "transition region"

Specifically,

$$\#\{\ell : \epsilon < \lambda^{(\ell)}_{N,W} < 1 - \epsilon\} \lesssim \log(N) \log\left(\frac{1}{\epsilon}\right)$$

Improves on previous asymptotic bounds by Slepian and nonasymptotic bounds by Zhu and Wakin

Fast Slepian Projection

Let $\boldsymbol{S}_K$ be the matrix formed by the first K columns of $\boldsymbol{S}_{N,W}$

Theorem

For any $W \in (0, \frac{1}{2})$, $\epsilon \in (0, \frac{1}{2})$, and K such that $\epsilon < \lambda_{N,W}^{(K-1)} < 1 - \epsilon$, there exist matrices $\boldsymbol{L}$ and $\boldsymbol{E}$ such that

$$\boldsymbol{S}_K \boldsymbol{S}_K^* = \boldsymbol{B}_{N,W} + \boldsymbol{L} + \boldsymbol{E}$$

where

$$\mathrm{rank}(\boldsymbol{L}) \lesssim \log(N) \log\left(\frac{1}{\epsilon}\right) \qquad \|\boldsymbol{E}\| \leq \epsilon$$

Fast Slepian Projection

$$\boldsymbol{S}_K \boldsymbol{S}_K^* = \boldsymbol{B}_{N,W} + \boldsymbol{L} + \boldsymbol{E}$$

Toeplitz

low rank

small

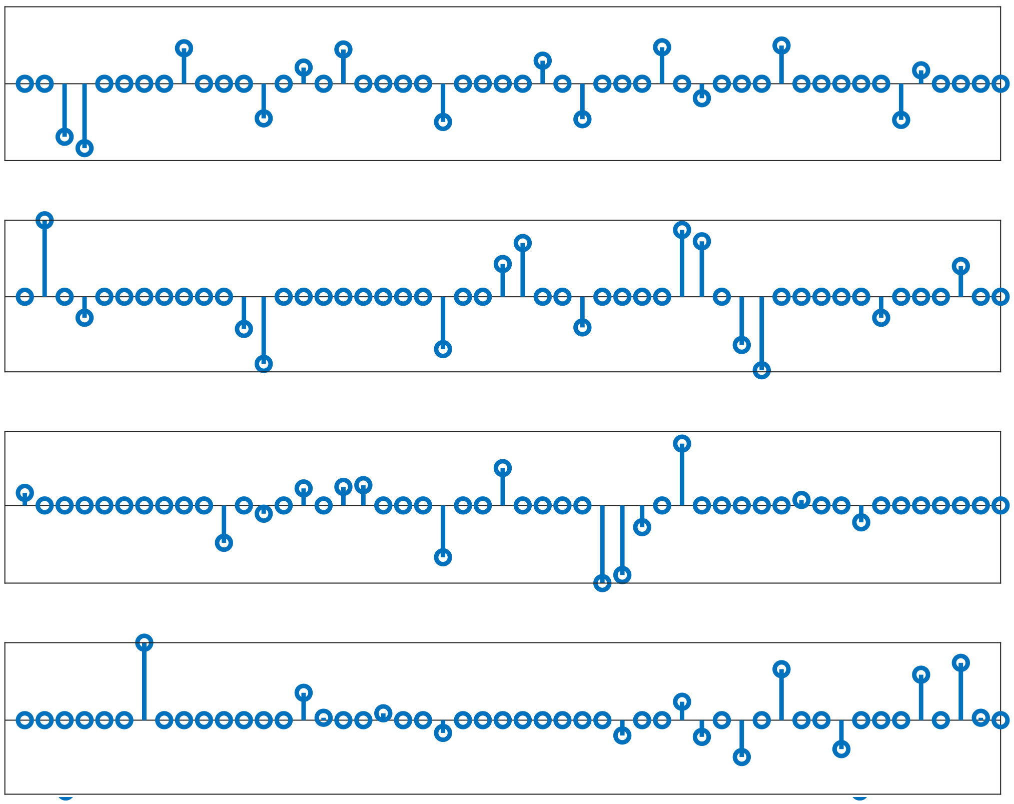
We can apply the approximation $\boldsymbol{B}_{N,W} + \boldsymbol{L}$ to a vector in $O(N \log N \log \frac{1}{\epsilon})$ operations

Similar fast algorithms can be developed for Thomson's method as well as solving related problems

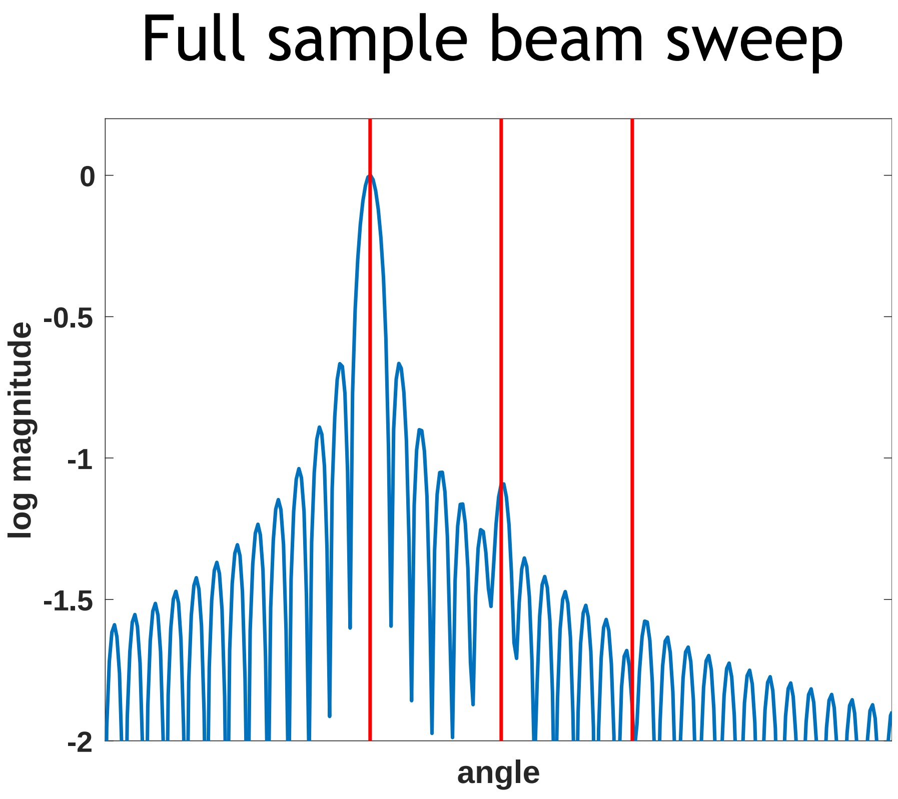
Extensions

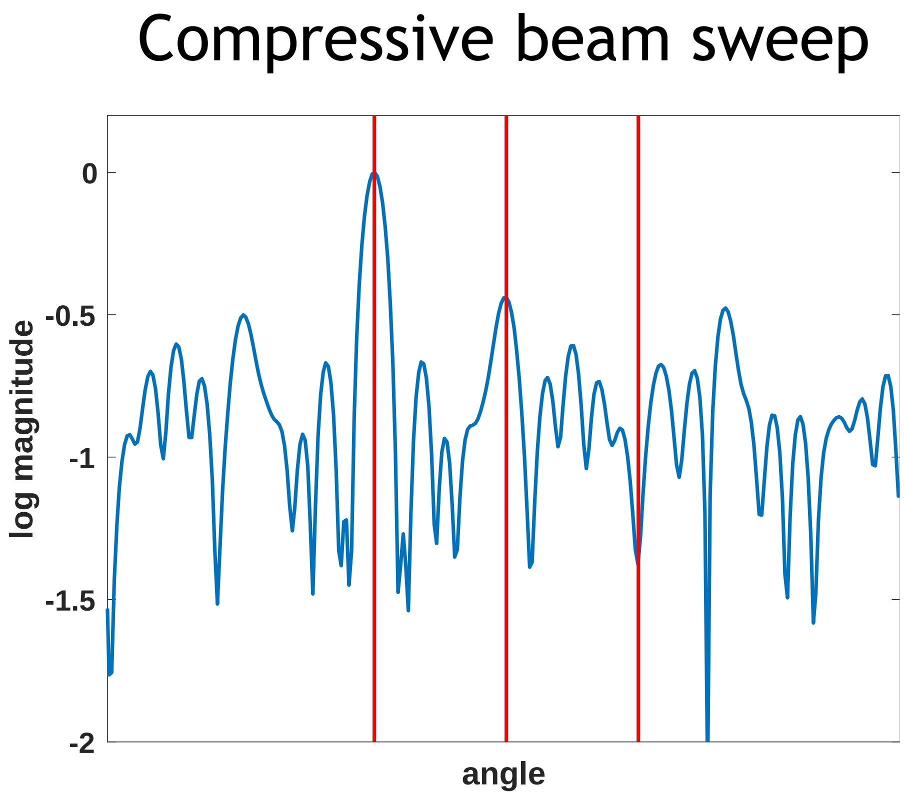
- Higher dimensions
  - array geometry and source environment can be two- or three-dimensional
- DOA with unknown frequencies
  - given a sequence of samples in time, we can consider a joint search over both angle-of-arrival and frequency
- Compressive acquisition
  - subsample array elements in time
  - subsample using spatial coded aperture techniques

Simple compression: Subsampling

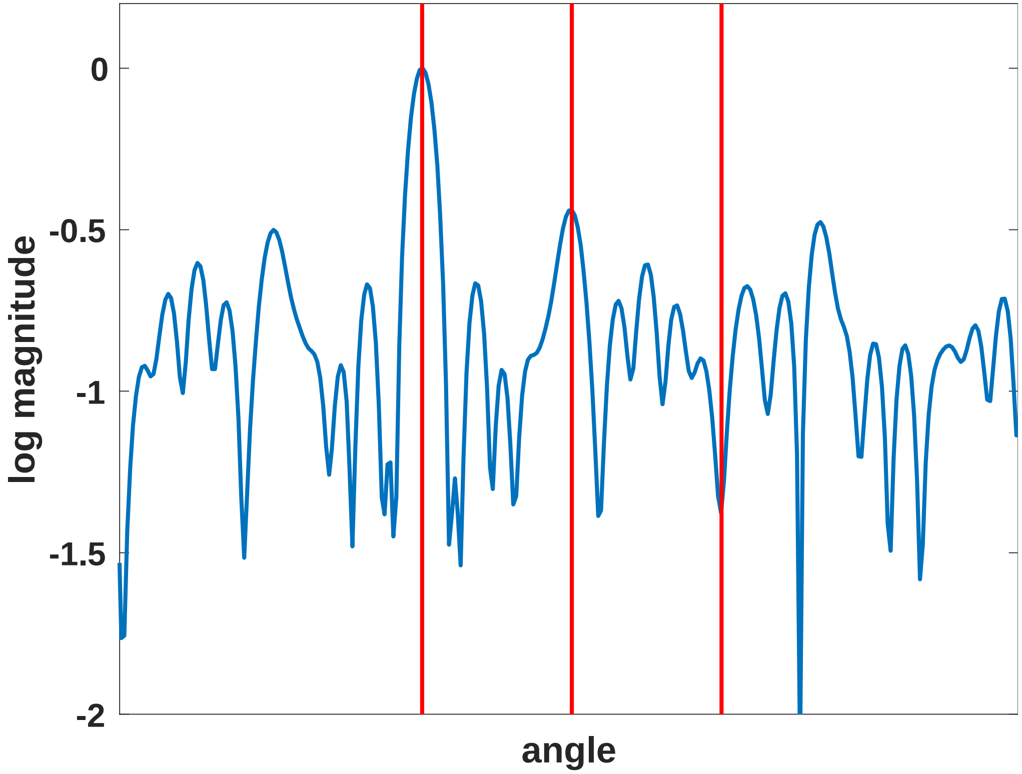


Compressive beamforming

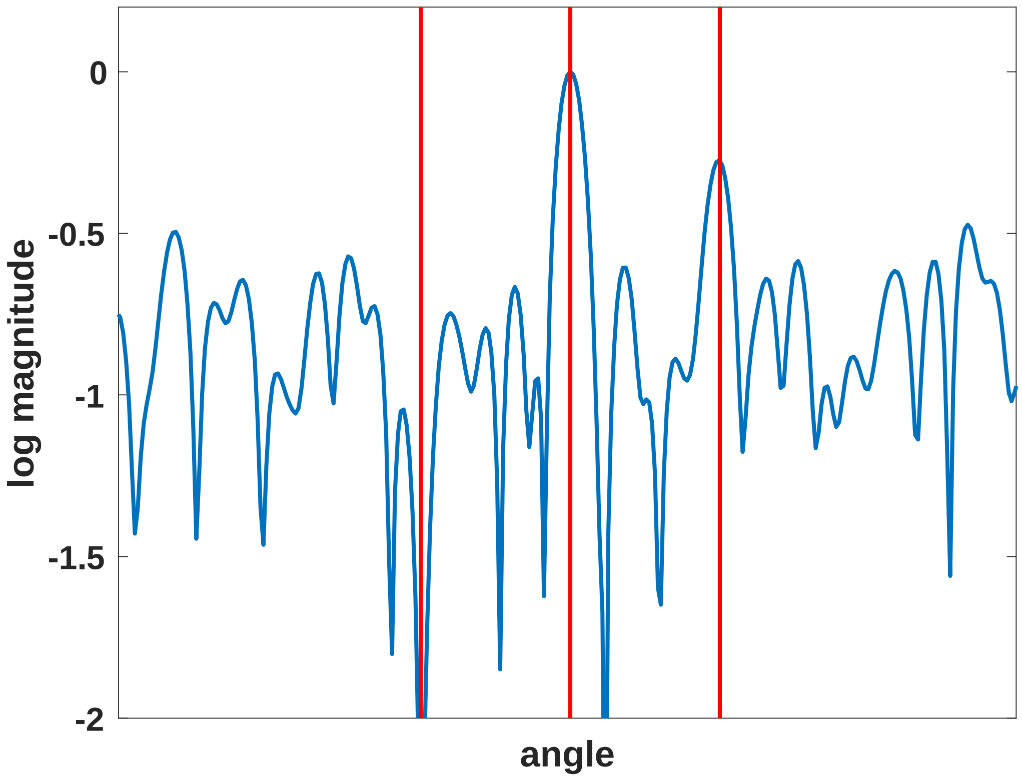




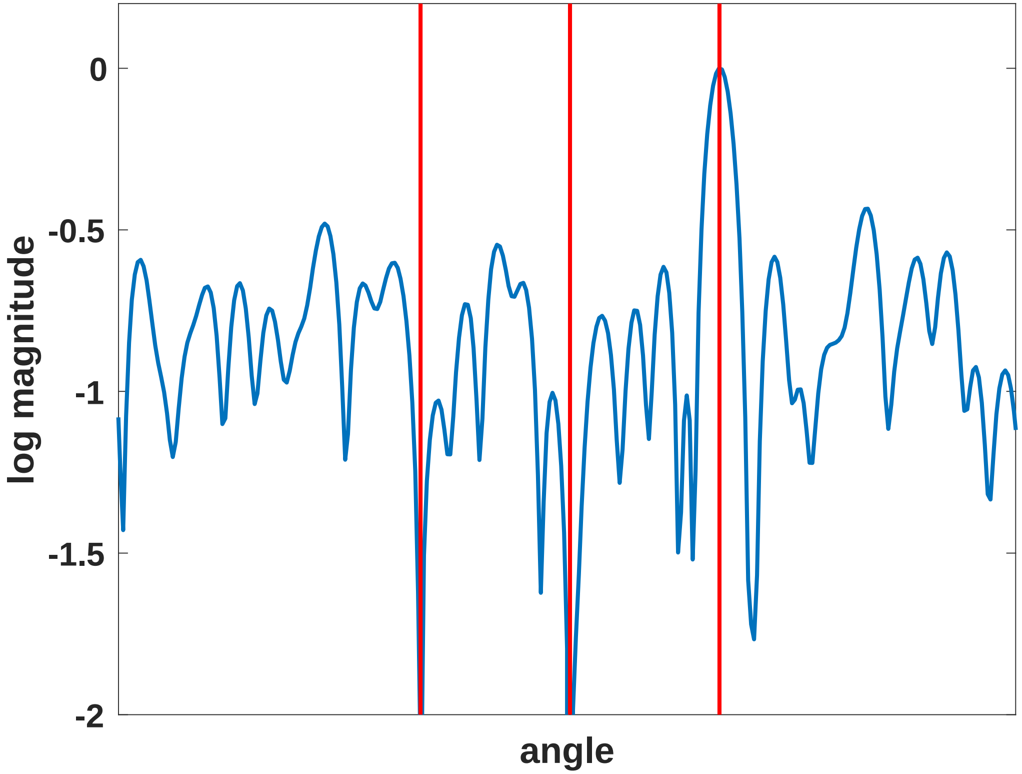
Iterative (compressive) source localization



Iterative (compressive) source localization

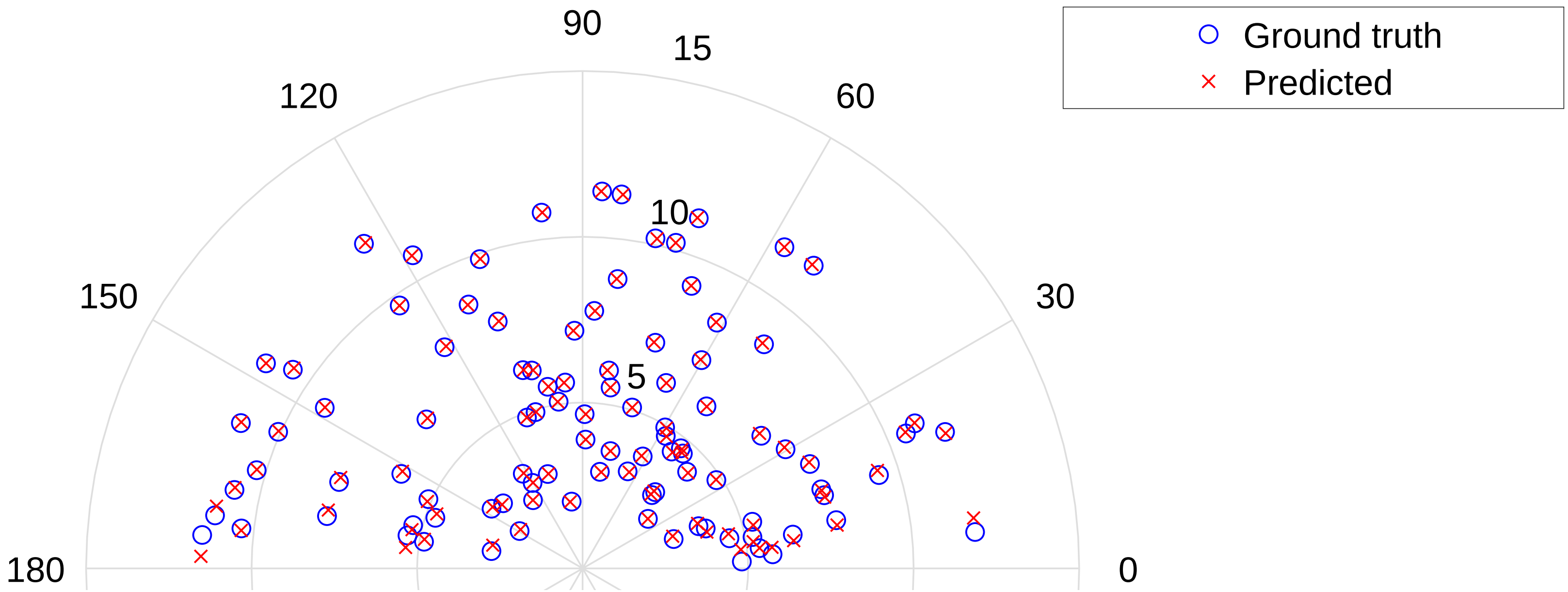


Iterative (compressive) source localization



Large scale simulation

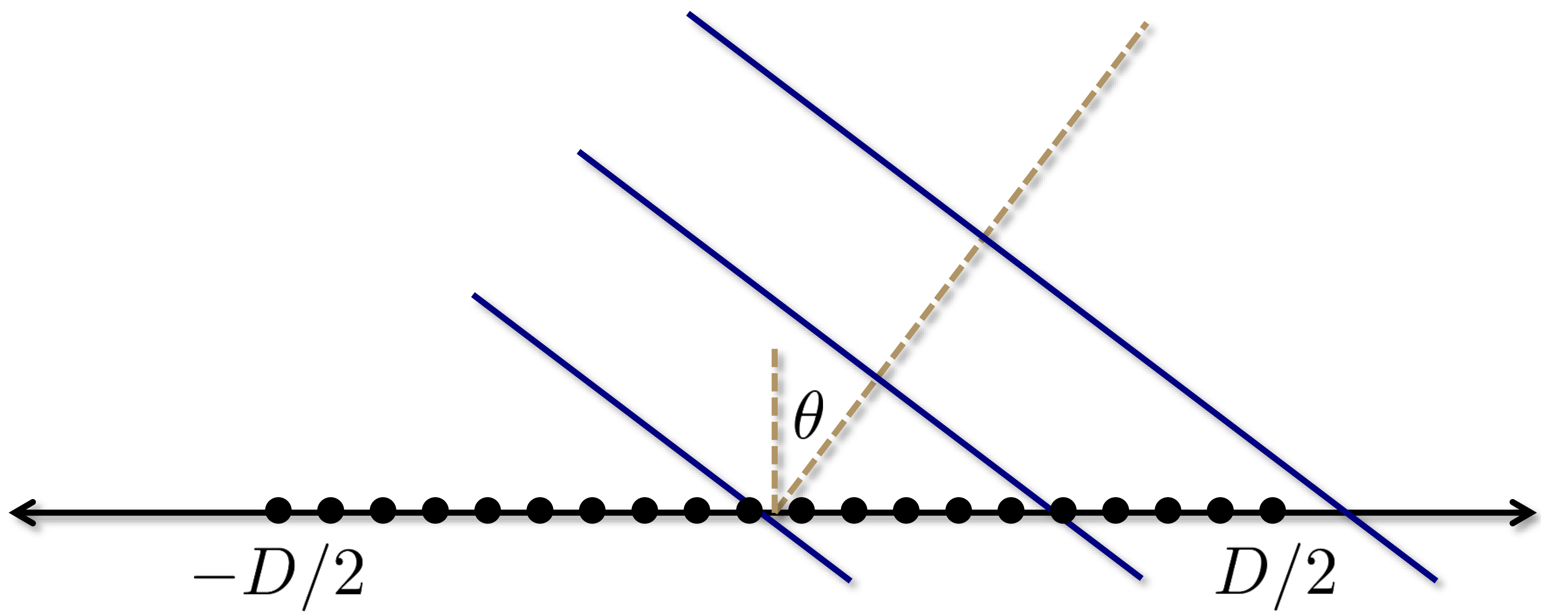
- 64 x 64 antenna array
- 100 sources, each 10 MHz, located in [2 GHz, 12 GHz]
- 0 dB SNR, keep only 12% of samples



These results simply would not be possible without (fast) Slepian projections

Exploiting bandwidth in active sensing

Recall the linear array setup:



Sinusoid at frequency f_0 ➡ Sinusoid at frequency $f_0 \sin(\theta)$

In general, for a target profile $x(\theta)$ at a constant range, our linear array observes (a warped) version of the Fourier transform of $x(\theta)$ over the range $[-f_0 D, f_0 D]$

A linear model

If we discretize $x(\theta)$ we can write our observations as

$$\mathbf{y}_{f_0} = \mathbf{A}_{f_0}\mathbf{x}$$

where the columns of $\mathbf{A}_{f_0}$ are uniformly spaced complex exponentials over the range $[-f_0 D, f_0 D]$

What happens if we repeat this for many different frequencies?

$$\begin{bmatrix} \mathbf{y}_{f_0} \\ \mathbf{y}_{f_1} \\ \vdots \\ \mathbf{y}_{f_{K-1}} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{f_0} \\ \mathbf{A}_{f_1} \\ \vdots \\ \mathbf{A}_{f_{K-1}} \end{bmatrix} \mathbf{x}$$

Does bandwidth buy us anything?

If $f_0 > f_1 > \cdots > f_{K-1}$, we seemingly do not gain any new information beyond what is contained in $\mathbf{y}_{f_0} = \mathbf{A}_{f_0}\mathbf{x}$

Effectively, $\mathcal{R}(\mathbf{A}_j) \subset \mathcal{R}(\mathbf{A}_0)$ for all j

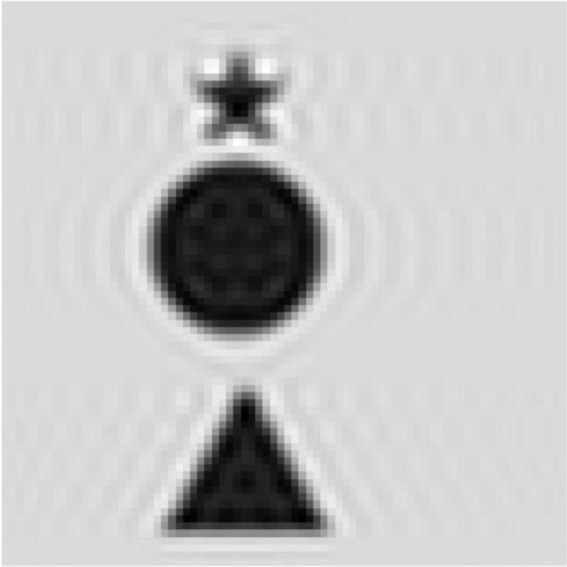
But what if we subsample?

$$\begin{bmatrix} \mathbf{\Phi}\mathbf{y}_{f_0} \\ \mathbf{\Phi}\mathbf{y}_{f_1} \\ \vdots \\ \mathbf{\Phi}\mathbf{y}_{f_{K-1}} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}\mathbf{A}_{f_0} \\ \mathbf{\Phi}\mathbf{A}_{f_1} \\ \vdots \\ \mathbf{\Phi}\mathbf{A}_{f_{K-1}} \end{bmatrix} \mathbf{x}$$

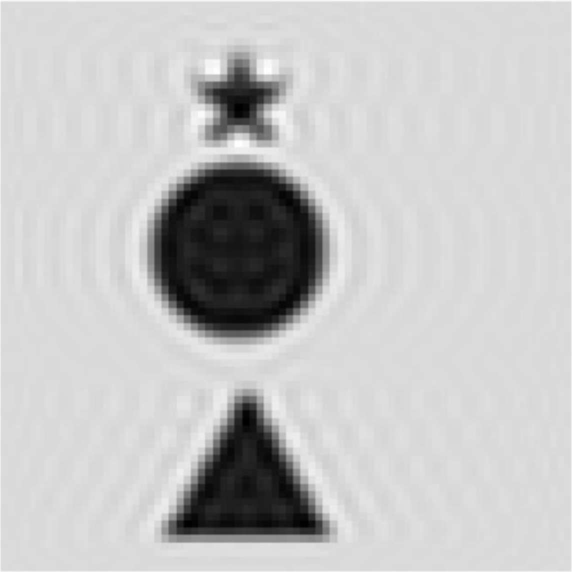
One can show that $\mathbf{\Phi}$ can be highly dimensionality-reducing without compromising our ability to estimate $\mathbf{x}$

Simulated results

- 40x40 sensor array, sensors placed 3.75cm apart
- Traditional imaging using excitation wavelength of 7.5cm would require ~1100 beams
- By exploiting bandwidth (lower frequencies) we can dramatically reduce the number of “beams”



1100 beams



80 “generic” beams

Summary

- The Slepian basis is a natural choice in many applications
  - any time you are working with finite windows of samples of bandlimited/narrowband/multiband functions
- We now have fast (approximate) algorithms for working with the Slepian basis, with complexity that scales comparably to the FFT
- Can play an important role in large-scale problems, especially in the context of compressive acquisition

References

D and Wakin. Compressive sensing of analog signals using discrete prolate spheroidal sequences. *Applied and Computational Harmonic Analysis* (2012)

D, Needell, and Wakin. Signal space CoSaMP for sparse recovery with redundant dictionaries. *IEEE Trans. on Information Theory* (2013)

Zhu, Karnik, Wakin, D, Romberg, ROAST: Rapid Orthogonal Approximate Slepian Transform. *IEEE Trans. on Signal Processing* (2018)

Karnik, Zhu, Wakin, Romberg, D. The Fast Slepian Transform. *Applied and Computational Harmonic Analysis* (2019)

Karnik, Romberg, D. Fast multitaper spectral estimation. *SAMPTA* (2019)

Srinivasa, D, Romberg, Trading beams for bandwidth: Imaging with randomized beamforming. http://arxiv.org/abs/1902.00075 (2019)