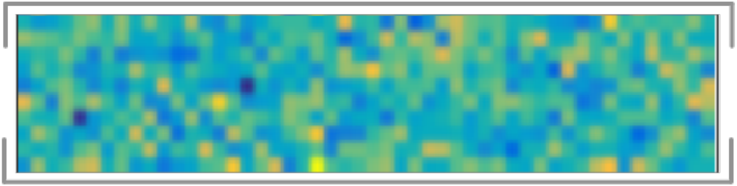
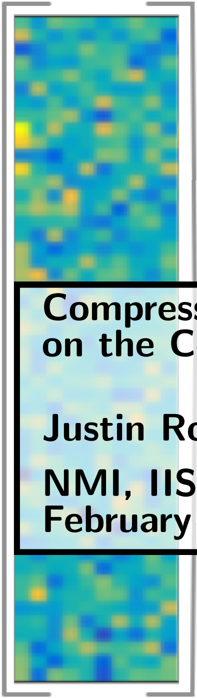


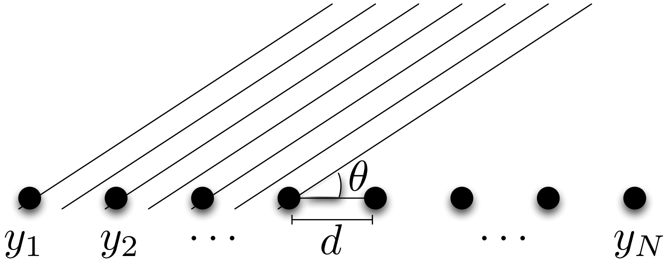


Compressed Subspace Matching on the Continuum

Justin Romberg, Georgia Tech ECE

NMI, IISc, Bangalore, India
February 22, 2015

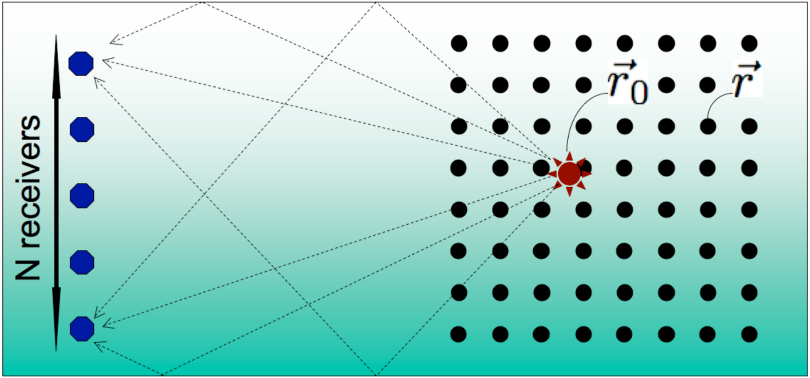
Direction of arrival estimation



Sinusoidal source (wavelength λ, complex amplitude A) at angle θ induces

$$\boldsymbol{y} = A \begin{bmatrix} 1 \\ e^{-j\frac{2\pi}{\lambda} d\cos\theta} \\ e^{-j\frac{2\pi}{\lambda} 2d\cos\theta} \\ \cdots \\ e^{-j\frac{2\pi}{\lambda} (N-1)d\cos\theta} \end{bmatrix}$$

Source localization



We observe a narrowband source emitting from (unknown) location $\vec{r_0}$:

$$\boldsymbol{y} = \alpha \boldsymbol{G}(\vec{r_0}) + \text{noise}, \quad \boldsymbol{y} \in \mathbb{C}^N$$

The dependence of $\boldsymbol{G}(\vec{r})$ on $\vec{r}$ might be complicated, “implicit”

Subspace matching

Collection of subspaces $\{\mathcal{S}_\theta : \theta \in \Theta\}$

- each $\mathcal{S}_\theta$ is K-dimensional
- θ is a continuous (multi) parameter

Subspace matching

Collection of subspaces $\{\mathcal{S}_\theta : \theta \in \Theta\}$

- each $\mathcal{S}_\theta$ is K-dimensional
- θ is a continuous (multi) parameter

Subspace matching: Given $\boldsymbol{h}_0 \in \mathbb{R}^N$, find closest subspace

$$\bar{\theta} = \arg\min_{\theta \in \Theta} \|\boldsymbol{h}_0 - \boldsymbol{P}_\theta \boldsymbol{h}_0\|_2^2$$

Subspace matching

Collection of subspaces $\{\mathcal{S}_\theta : \theta \in \Theta\}$

- each $\mathcal{S}_\theta$ is K-dimensional
- θ is a continuous (multi) parameter

Subspace matching: Given $\boldsymbol{h}_0 \in \mathbb{R}^N$, find closest subspace

$$\bar{\theta} = \arg\min_{\theta \in \Theta} \|\boldsymbol{h}_0 - \boldsymbol{P}_\theta \boldsymbol{h}_0\|_2^2$$

Compressed subspace matching: Given $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{h}_0$, where $\boldsymbol{\Phi}$ is $M \times N$, random, solve

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \min_{\boldsymbol{g} \in \mathcal{S}_\theta} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{g}\|_2^2 = \arg\min_{\theta \in \Theta} \|\boldsymbol{y} - \tilde{\boldsymbol{P}}_\theta \boldsymbol{y}\|_2^2,$$

$\tilde{\boldsymbol{P}}_\theta =$ orthorpojector onto range of $\boldsymbol{\Phi}\boldsymbol{P}_\theta$

Subspace matching

Collection of subspaces $\{\mathcal{S}_\theta : \theta \in \Theta\}$

- each $\mathcal{S}_\theta$ is K-dimensional
- θ is a continuous (multi) parameter

Subspace matching: Given $\boldsymbol{h}_0 \in \mathbb{R}^N$, find closest subspace

$$\bar{\theta} = \arg\min_{\theta \in \Theta} \|\boldsymbol{h}_0 - \boldsymbol{P}_\theta \boldsymbol{h}_0\|_2^2$$

Compressed subspace matching: Given $\boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{h}_0$, where $\boldsymbol{\Phi}$ is $M \times N$, random, solve

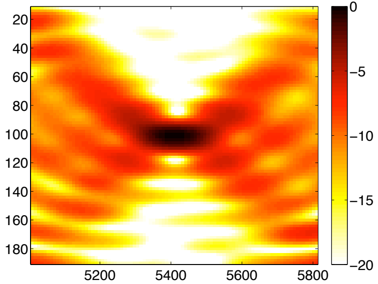
$$\hat{\theta} = \arg\min_{\theta \in \Theta} \min_{\boldsymbol{g} \in \mathcal{S}_\theta} \|\boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{g}\|_2^2 = \arg\min_{\theta \in \Theta} \|\boldsymbol{y} - \tilde{\boldsymbol{P}}_\theta \boldsymbol{y}\|_2^2,$$

$\tilde{\boldsymbol{P}}_\theta =$ orthorpojector onto range of $\boldsymbol{\Phi} \boldsymbol{P}_\theta$

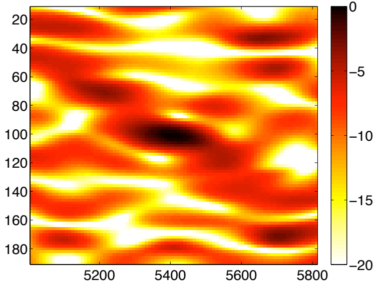
When is $\hat{\theta}$ as good as $\bar{\theta}$?

Compressive ambiguity functions

ambiguity function $\|\boldsymbol{P}_\theta \boldsymbol{h}\|_2^2$



compressed amb func $\|\tilde{\boldsymbol{P}}_\theta \boldsymbol{y}\|_2^2$



$M = 10$ (compare to $N = 37$)

- The compressed ambiguity function is a *random process* whose mean is the true ambiguity function
- For very modest M, these two functions peak in the same place

Performance gap

$\bar{\theta}$ = best subspace from direct observation,
$\hat{\theta}$ = best subspace from compressed observation

Performance gap (assume $\|\boldsymbol{h}_0\|_2 = 1$):

$$\hat{E}^2 - \bar{E}^2 = \|\boldsymbol{P}_{\bar{\theta}}\boldsymbol{h}_0\|_2^2 - \|\boldsymbol{P}_{\hat{\theta}}\boldsymbol{h}_0\|_2^2$$

Performance gap

$\bar{\theta}$ = best subspace from direct observation,
$\hat{\theta}$ = best subspace from compressed observation

Performance gap (assume $\|\boldsymbol{h}_0\|_2 = 1$):

$$\hat{E}^2 - \bar{E}^2 = \|\boldsymbol{P}_{\bar{\theta}}\boldsymbol{h}_0\|_2^2 - \|\boldsymbol{P}_{\hat{\theta}}\boldsymbol{h}_0\|_2^2$$

Theorem: For fixed $\boldsymbol{h}_0$, if we have

$$\sup_{\theta\in\Theta} \|\boldsymbol{P}_\theta - \boldsymbol{P}_\theta \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{P}_\theta\| \;\leq\; \delta_1, \qquad \sup_{\theta\in\Theta} \|\boldsymbol{P}_\theta \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{P}_\theta^\perp \boldsymbol{h}_0\|_2 \;\leq\; \delta_2$$

then

$$\hat{E}^2 - \bar{E}^2 \;\leq\; F(\delta_1, \delta_2) \approx C(\delta_1 + \delta_2)$$

Keeping subspaces separated

With $\mathbf{\Phi}$ random (entries iid, Gaussian),

$$\sup_{\theta \in \Theta} \|\boldsymbol{P}_\theta - \boldsymbol{P}_\theta \mathbf{\Phi}^{\mathrm{T}} \mathbf{\Phi} \boldsymbol{P}_\theta\|$$

is the suprema of a (matrix-valued) random process

We are essentially asking $\mathbf{\Phi}$ to stably embed every subset in the collection $\{\mathcal{S}_\theta : \theta \in \Theta\}$

$$\sup_{\theta \in \Theta} \sup_{\substack{\boldsymbol{x} \in \mathcal{S}_\theta \\ \|\boldsymbol{x}\|_2 \leq 1}} \left| \|\mathbf{\Phi}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right|$$

There is a lot of context for this type of problem ...

Embedding a subspace of $\mathbb{R}^N$

Let $\mathcal{S}$ be a K dimensional subspace of $\mathbb{R}^N$. For $\mathbf{\Phi}$ random, when do we have

$$(1-\delta)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 \;\leq\; \|\mathbf{\Phi}\boldsymbol{x}_1 - \mathbf{\Phi}\boldsymbol{x}_2\|_2^2 \;\leq\; (1+\delta)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2,$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{S}$ with appropriately high probability?

δ is directly related to the *singular values* of $\mathbf{\Phi}$, and

$$\delta \;\lesssim\; \sqrt{\frac{K}{M}}.$$

This is a "classical" result by Marchenko, Pastur (1960s), and later Szarek (1990s).

Embedding a finite collection of subspaces of $\mathbb{R}^N$

Let $\{\mathcal{S}_\theta : \theta \in \Theta\}$ be a *finite collection of subspaces* of dimension K. For $\mathbf{\Phi}$ random, when do we have

$$(1-\delta)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 \;\le\; \|\mathbf{\Phi}\boldsymbol{x}_1 - \mathbf{\Phi}\boldsymbol{x}_2\|_2^2 \;\le\; (1+\delta)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2,$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{S}_\theta$ with appropriately high probability?

A simple union bound yields

$$\delta \;\lesssim\; \sqrt{\frac{K + \log|\Theta|}{M}}$$

(RIP: Candes,Tao; Rudelson, Vershynin; Davenport et al., mid-2000s)

Embedding an infinite collection of subspaces of $\mathbb{R}^N$

Let $\{\mathcal{S}_\theta : \theta \in \Theta\}$ be an *infinite collection of subspaces* of dimension K.
For $\mathbf{\Phi}$ random, when do we have

$$(1-\delta)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 \;\leq\; \|\mathbf{\Phi}\boldsymbol{x}_1 - \mathbf{\Phi}\boldsymbol{x}_2\|_2^2 \;\leq\; (1+\delta)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2,$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{S}_\theta$ with appropriately high probability?

A chaining argument (between subspaces) yields

$$\delta \;\lesssim\; \sqrt{\frac{K(\Delta + \log K)}{M}}$$

where Δ is a measure of *geometrical complexity* of Θ.

(Mantzel and R. '13)

In typical cases of interest, $\Delta \sim \log\left(\max\{K, \text{“effective dimension” of } \Theta\}\right)$

Embedding an infinite collection of subspaces of $\mathbb{R}^N$

Let $\{\mathcal{S}_\theta : \theta \in \Theta\}$ be an *infinite collection of subspaces* of dimension K.
For $\boldsymbol{\Phi}$ random, when do we have

$$(1-\delta)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 \;\leq\; \|\boldsymbol{\Phi}\boldsymbol{x}_1 - \boldsymbol{\Phi}\boldsymbol{x}_2\|_2^2 \;\leq\; (1+\delta)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2,$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{S}_\theta$ with appropriately high probability?

A (more subtle) chaining argument between subspaces yields

$$\delta \;\lesssim\; \sqrt{\frac{K+\Delta}{M}}$$

where Δ is a measure of *geometrical complexity* of Θ.

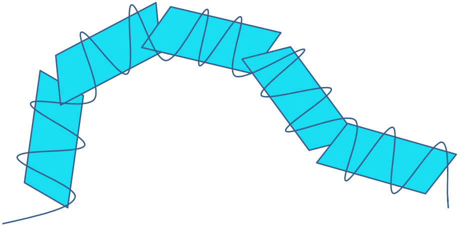
(Dirksen '14)

In typical cases of interest, $\Delta \sim \log\left(\max\{K,\right.$ "effective dimension" of $\left.\Theta\}\right)$

Geometrical complexity

$N(\{\mathcal{S}_\theta\}, d, \epsilon)$ = size of smallest cover with

$$d(\mathcal{S}_{\theta_1}, \mathcal{S}_{\theta_2}) = \|\boldsymbol{P}_{\theta_1} - \boldsymbol{P}_{\theta_2}\|$$



Δ captures fast the cover grows as $\epsilon \to 0$,
With N_0, α such that

$$N(\{\mathcal{S}_\theta\}, d, \epsilon) \;\leq\; N_0 \left(\frac{1}{\epsilon}\right)^\alpha$$

we can take

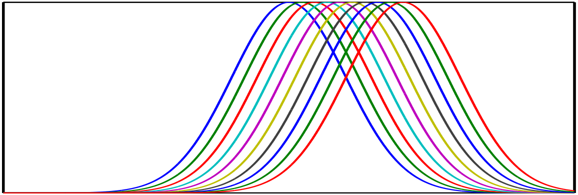
$$\Delta = \alpha \log(8) + 2 \log N_0$$

Typical: $\alpha = 1$ or 2, $N_0 = \mathrm{poly}(K)$.

Shiftable subspaces

Say $\{\mathcal{S}_\theta\}$ is generated by continuum of shifts of a known pulse $(K = 1)$

$$\{v(t-\theta),\ \theta \in [0,T]\}$$



$v(t)$ **smooth**:

with $\|v(t)\|_{W_1}$ = Sobolev norm,

$$\Delta \le 2\log(\|v\|_{W_1} T) + 4.08$$

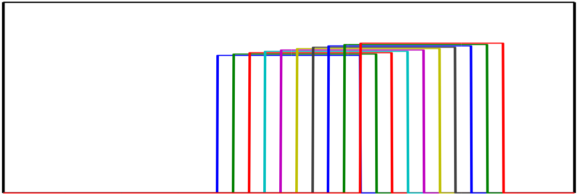
Example: Gaussian with width σ:

$$\Delta \le \log(T/\sigma) + 4.08$$

Shiftable subspaces

Say $\{\mathcal{S}_\theta\}$ is generated by continuum of shifts of a known pulse $(K = 1)$

$$\{v(t-\theta),\ \theta \in [0, T]\}$$



$v(t)$ **not smooth**:

with $\|v(t)\|_{TV} =$ total variation,

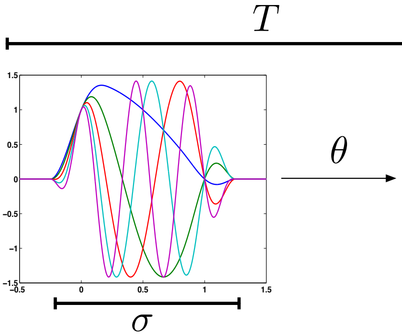
$$\Delta \leq 4 \log(\|v\|_{TV}\, T) + 7.55$$

Example: Square with width σ:

$$\Delta \leq 2 \log(T/\sigma) + 8.94$$

Shiftable bands

Smooth window, modulated by K different cosines (LOT).
Width of functions $= \sigma$
Shift over interval of length T



In this case, we have

$$\Delta \sim \log(K) + \log(T/\sigma)$$

Recall: Performance gap

$\bar{\theta}$ = best subspace from direct observation,
$\hat{\theta}$ = best subspace from compressed observation

Performance gap (assume $\|\boldsymbol{h}_0\|_2 = 1$):

$$\hat{E}^2 - \bar{E}^2 = \|\boldsymbol{P}_{\bar{\theta}}\boldsymbol{h}_0\|_2^2 - \|\boldsymbol{P}_{\hat{\theta}}\boldsymbol{h}_0\|_2^2$$

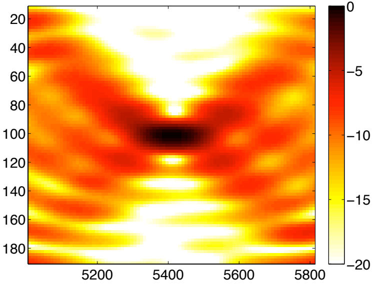
Theorem: For fixed $\boldsymbol{h}_0$, we have

$$\hat{E}^2 - \bar{E}^2 \approx \sqrt{\frac{K + \Delta}{M}}$$

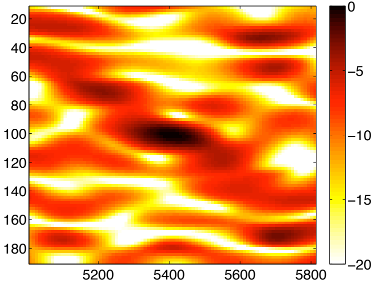
(Mantzel, R '13, Dirksen '14)

From approximation gap to parameter estimate

$\|\boldsymbol{P}_\theta \boldsymbol{h}_0\|_2^2$



$\|\tilde{\boldsymbol{P}}_\theta \boldsymbol{y}\|_2^2$ (compressed)



$M = 10$ (compare to 37 receivers)

We actually establish a uniform result:

$$\sup_{\theta \in \Theta} \left| \|\tilde{\boldsymbol{P}}_\theta \boldsymbol{y}\|_2^2 - \|\boldsymbol{P}_\theta \boldsymbol{h}_0\|_2^2 \right| \;\leq\; \delta$$

Separation of the max from the "sidelobes"

$\Rightarrow$ we have an accurate parameter estimate as well

References

- A. Ahmed and J. Romberg, "Compressive multiplexers for correlated signals," accepted to *IEEE Transactions on Information Theory*, September 2014.

- A. Ahmed and J. Romberg, "Compressive sampling of ensembles of correlated signals," submitted to *Applied and Computational Harmonic Analysis*, December 2014.

- A. Eftekhari, J. Romberg, and M. Wakin, "Matched filtering from limited frequency samples," *IEEE Trans. Inform. Theory*, June 2013.

- W. Mantzel, K. Sabra, J. Romberg, "Compressive matched-field processing," *J. Acoustic Soc. Amer.*, July 2012.

- W. Mantzel, J. Romberg, and K. Sabra, "Round-robin multiple-source localization," *J. Acoustic Soc. Amer.*, January 2014.

- W. Mantzel and J. Romberg, "Compressed subspace matching on the continuum," to appear in *Information and Inference*, 2015.

`http://users.ece.gatech.edu/~justin/Publications.html`