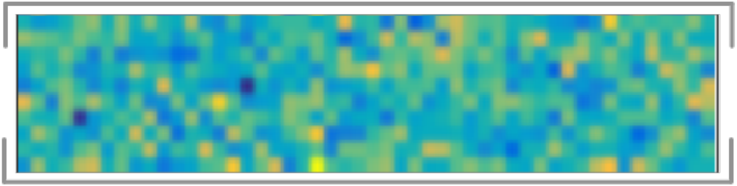


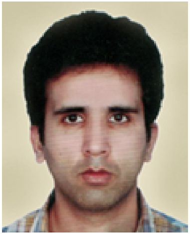
Low-rank Recovery Problems in Signal Processing

Justin Romberg, Georgia Tech ECE

NMI, IISc, Bangalore, India

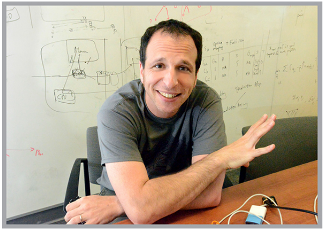
February 21, 2015

Collaborators

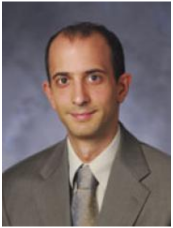


Ali Ahmed

Sohail Bahmani



Ben Recht



Karim Sabra

Ning Tian

Recovering a matrix from limited observations

Suppose we are interested in recovering the values of a **matrix** $\boldsymbol{X}$

$$\boldsymbol{X} = \begin{bmatrix} X_{1,1} & X_{1,2} & X_{1,3} & X_{1,4} & X_{1,5} \\ X_{2,1} & X_{2,2} & X_{2,3} & X_{2,4} & X_{2,5} \\ X_{3,1} & X_{3,2} & X_{3,3} & X_{3,4} & X_{3,5} \\ X_{4,1} & X_{4,2} & X_{4,3} & X_{4,4} & X_{4,5} \\ X_{5,1} & X_{5,2} & X_{5,3} & X_{5,4} & X_{5,5} \end{bmatrix}$$

We are given a series of different *linear combinations* of the entries

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X})$$

Example: matrix completion

Suppose we do not see all the entries in a matrix ...

$$\boldsymbol{X} = \begin{bmatrix} X_{1,1} & - & X_{1,3} & - & X_{1,5} \\ - & X_{2,2} & - & X_{2,4} & - \\ - & X_{3,2} & X_{3,3} & - & - \\ X_{4,1} & - & - & X_{4,4} & X_{4,5} \\ - & - & - & X_{5,4} & X_{5,5} \end{bmatrix}$$

... can we "fill in the blanks"?

Low rank structure

$$\underbrace{\left[\ \boldsymbol{X}\ \right]}_{K \times N} = \underbrace{\left[\ \boldsymbol{L}\ \right]}_{K \times R}\ \underbrace{\left[\ \boldsymbol{R}^T\ \right]}_{R \times N}$$

Agenda

- Many applications of low-rank recovery in machine learning: recommendation systems, covariance estimation, etc.

- This talk: how this theory relates to fundamental problems in signal processing

Topics include:

- sampling large ensembles of correlated signals
- blind deconvolution
- source separation
- super-resolution with unknown spreading function

Low rank recovery from linear measurements

- We have an underdetermined linear operator $\mathcal{A}$

$$\mathcal{A} : \mathbb{R}^{K\times N} \to \mathbb{R}^M, \quad M \ll KN, \quad \mathcal{A}(\boldsymbol{X}) = \{\langle \boldsymbol{X}, \boldsymbol{A}_m \rangle\}_{m=1}^M$$

  and observe

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}_0) + \text{noise}$$

  where $\boldsymbol{X}_0$ has rank R

- One recovery technique: nuclear norm relaxation

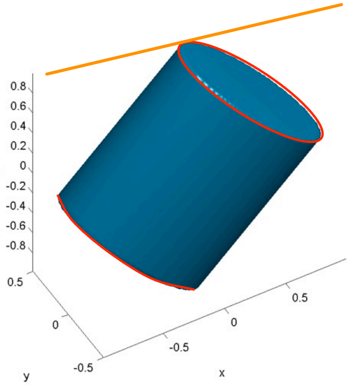
$$\min_{\boldsymbol{X}} \ \|\boldsymbol{X}\|_* \quad \text{subject to} \quad \mathcal{A}(\boldsymbol{X}) = \boldsymbol{y}$$

  where $\|\boldsymbol{X}\|_* =$ sum of the singular values of $\boldsymbol{X}$

Recovering low rank matrices

Given $\boldsymbol{y}$, we solve the (convex) optimization program

$$\text{minimize} \ \ \|\boldsymbol{X}\|_* = \sum_i \sigma_i(\boldsymbol{X}) \quad \text{subject to} \quad \mathcal{A}(\boldsymbol{X}) = \boldsymbol{y}$$



When $\mathcal{A}$ is distance preserving, this is provably effective.

When can we recover a low rank matrix?

Two main approaches for establishing effectiveness:

- Uniform embeddings: Show $\mathcal{A}$ keeps rank-R matrices separated,

$$\|\mathcal{A}(\boldsymbol{X}_1 - \boldsymbol{X}_2)\|_2^2 \;\approx\; \|\boldsymbol{X}_1 - \boldsymbol{X}_2\|_F^2 \quad \text{for all rank-}R\ \boldsymbol{X}_1, \boldsymbol{X}_2$$

  Very powerful, hard to establish
  Works for subgaussian projections, "fast JLT" projections

When can we recover a low rank matrix?

Two main approaches for establishing effectiveness:

- Uniform embeddings: Show $\mathcal{A}$ keeps rank-R matrices separated,

  $$\|\mathcal{A}(\boldsymbol{X}_1 - \boldsymbol{X}_2)\|_2^2 \approx \|\boldsymbol{X}_1 - \boldsymbol{X}_2\|_F^2 \quad \text{for all rank-}R\ \boldsymbol{X}_1, \boldsymbol{X}_2$$

  Very powerful, hard to establish
  Works for subgaussian projections, "fast JLT" projections

- Duality theory: show you can construct a dual certificate for

  $$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_* \quad \text{subject to} \quad \mathcal{A}(\boldsymbol{X}) = \mathcal{A}(\boldsymbol{X}_0)$$

  for a particular $\boldsymbol{X}_0$.
  Result holds for a particular $\boldsymbol{X}_0$, strong stability harder to establish
  Works for many $\mathcal{A}$ with *structured randomness*

Duality for low rank recovery

The matrix $\boldsymbol{X}_0$ is a solution to

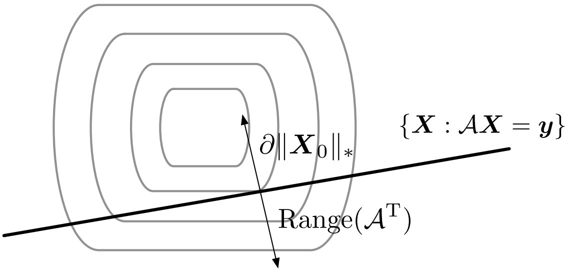
$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_* \quad \text{subject to} \quad y_m = \langle \boldsymbol{X}, \boldsymbol{A}_m \rangle, \quad m = 1, \ldots, M$$

when

$$\mathcal{A}(\boldsymbol{X}_0) = \boldsymbol{y}, \quad \text{and there is a } \boldsymbol{z} \text{ s.t.} \quad \mathcal{A}^{\mathrm{T}}(\boldsymbol{z}) = \sum_{m=1}^{M} z_m \boldsymbol{A}_m \in \partial \|\boldsymbol{X}_0\|_*$$

where with $\boldsymbol{X}_0 = \boldsymbol{U}_0 \boldsymbol{\Sigma}_0 \boldsymbol{V}_0^{\mathrm{T}}$,

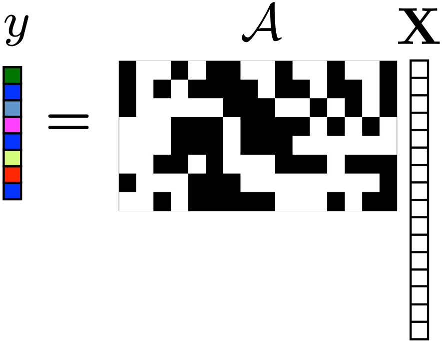
$$\partial \|\boldsymbol{X}_0\|_* = \{\boldsymbol{U}_0 \boldsymbol{V}_0^{\mathrm{T}} + \boldsymbol{W}, \quad \|\boldsymbol{W}\| \leq 1\}$$



Matrix Measurements

Matrix Recovery: random measurements

Take vectorize $\boldsymbol{X}$, stack up vectorized $\boldsymbol{A}_m$ as rows of a matrix

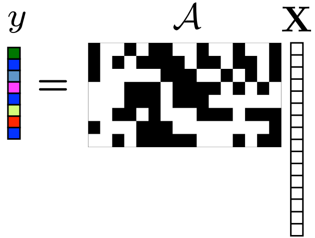


Independent subgaussian entries in the $\boldsymbol{A}_m$ embeds rank-R matrices when

$$M \gtrsim R(K+N)$$

(Recht, Fazel, Parillo, Candes, Plan, ...)

Matrix Recovery: random measurements



Embedding established in a similar manner as yesterday:

- **Concentration:** For a fixed $\boldsymbol{X}$,

$$\mathrm{P}\left(\left|\|\mathcal{A}(\boldsymbol{X})\|_2^2 - \|\boldsymbol{X}\|_F^2\right| > \delta\right) \;\leq\; C \cdot e^{-c\delta^2 M}$$

- **Covering:** Rank-R matrices come from an infinite union of subspaces, standard covering bounds allow a net of size $\sim e^{R(K+N)}$ for same order δ

Matrix Recovery: structured randomness

Krahmer, Ward'10: If $M \times N$ $\mathbf{\Phi}$ obeys RIP for S sparse, then

$$\mathbf{\Phi}' = \mathbf{\Phi}\boldsymbol{D}, \quad \boldsymbol{D} \text{ diagonal, random}$$

obeys the concentration inequality

$$\mathrm{P}\left(\left|\|\mathbf{\Phi}'\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| > \delta\right) < C \cdot e^{-cS}$$

If $\mathbf{\Phi}$ is an "efficient" CS matrix, we have for all $\boldsymbol{X}_1, \boldsymbol{X}_2$ rank-R,

$$\|\mathcal{A}(\boldsymbol{X}_1)\|_2^2 \approx \|\boldsymbol{X}_1 - \boldsymbol{X}_2\|_F^2 \quad \text{when} \quad M \gtrsim R(K+N)\log^q(KN)$$

Matrix Recovery: structured randomness

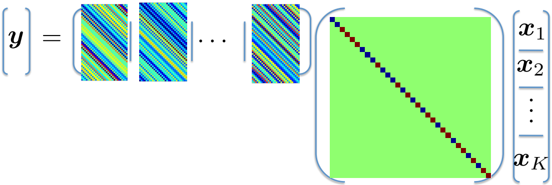
Krahmer, Ward'10: If $M \times N$ $\mathbf{\Phi}$ obeys RIP for S sparse, then

$$\mathbf{\Phi}' = \mathbf{\Phi}\boldsymbol{D}, \quad \boldsymbol{D} \text{ diagonal, random}$$

obeys the concentration inequality

$$\mathrm{P}\left(\left|\|\mathbf{\Phi}'\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| > \delta\right) < C \cdot e^{-cS}$$

Example: Modulate each column, then convolve each with random pulse, then add



Matrix Recovery: structured randomness

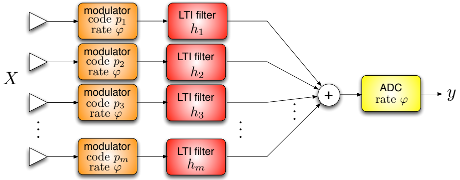
Krahmer, Ward'10: If $M \times N$ $\boldsymbol{\Phi}$ obeys RIP for S sparse, then

$$\boldsymbol{\Phi}' = \boldsymbol{\Phi}\boldsymbol{D}, \quad \boldsymbol{D} \text{ diagonal, random}$$

obeys the concentration inequality

$$\mathrm{P}\left(\left|\|\boldsymbol{\Phi}'\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| > \delta\right) < C \cdot e^{-cS}$$

Example: Modulate each column, then convolve each with random pulse, then add



Example: matrix completion

Suppose we do not see all the entries in a matrix ...

$$\boldsymbol{X} = \begin{bmatrix} X_{1,1} & - & X_{1,3} & - & X_{1,5} \\ - & X_{2,2} & - & X_{2,4} & - \\ - & X_{3,2} & X_{3,3} & - & - \\ X_{4,1} & - & - & X_{4,4} & X_{4,5} \\ - & - & - & X_{5,4} & X_{5,5} \end{bmatrix}$$

... we can fill them in from

$$M \;\gtrsim\; R(K+N)\cdot\log^2(KN)$$

randomly chosen samples if $\boldsymbol{X}$ is *diffuse*.

(Recht, Gross, Candes, Tao, Montenari, Oh, ...)

Rank 1 inner products

Measurements of the form

$$y_m = \boldsymbol{\nu}_m^{\mathrm{T}} \boldsymbol{X} \boldsymbol{\phi}_m = \langle \boldsymbol{X}, \boldsymbol{\nu}_m \boldsymbol{\phi}_m^{\mathrm{T}} \rangle$$

inner products with rank 1 matrices

With $\boldsymbol{\nu} = \boldsymbol{\phi}_k$, and $\boldsymbol{X} = \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ itself rank 1 and symmetric, this is the "phase retrieval" problem

$$y_m = \boldsymbol{\phi}_m^{\mathrm{T}} \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} \boldsymbol{\phi}_m = |\langle \boldsymbol{u}, \boldsymbol{\phi}_m \rangle|^2$$

- Recovery for $M \gtrsim N$, based on weak embeddings for rank-1, $\boldsymbol{\phi}$ random (Candes, Strohmer, Voroninski '12)
- ℓ_1/ℓ_2 embeddings for rank-R for $M \gtrsim RN$ (Chen, Chi, Goldsmith '13)

Rank 1 inner products

Measurements of the form

$$y_m = \boldsymbol{\nu}_m^{\mathrm{T}} \boldsymbol{X} \boldsymbol{\phi}_m = \langle \boldsymbol{X}, \boldsymbol{\nu}_m \boldsymbol{\phi}_m^{\mathrm{T}} \rangle$$

inner products with rank 1 matrices

With $\boldsymbol{\nu}, \boldsymbol{\phi}$ different and $\boldsymbol{X} = \boldsymbol{r}\boldsymbol{c}^{\mathrm{T}}$ rank 1, this is equialent to "blind deconvolution"

$$y_m = \boldsymbol{\nu}_m^{\mathrm{T}} \boldsymbol{r} \boldsymbol{c}^{\mathrm{T}} \boldsymbol{\phi}_m = \langle \boldsymbol{r}, \boldsymbol{\nu}_m \rangle \cdot \langle \boldsymbol{c}, \boldsymbol{\phi}_m \rangle$$

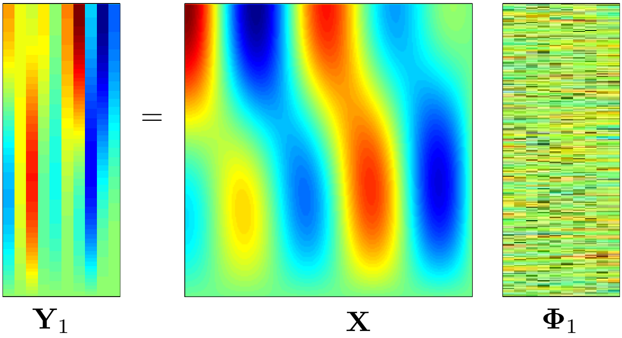
- $\boldsymbol{\phi}_m$ random, $\boldsymbol{\nu}$ incoherent in Fourier domain,
  Recovery for rank-1 $\boldsymbol{X}$ for $M \gtrsim (N+K)\log^3(NK)$
  (Ahmed, R, Recht '12)
  Recovery for rank-R $\boldsymbol{X}$ for $M \gtrsim R(N+K)\log^3(NK)$
  (Ahmed, R '13)
- $\boldsymbol{\nu}_m, \boldsymbol{\phi}_m$ both random,
  strong embedding (RIP) for $M \gtrsim R(N+K)\log(NK)$
  (Ahmed, Krahmer, R '15)

Randomized linear algebra

Given an $N \times K$ matrix $\boldsymbol{X}$ with rank R, we can recover the column space from

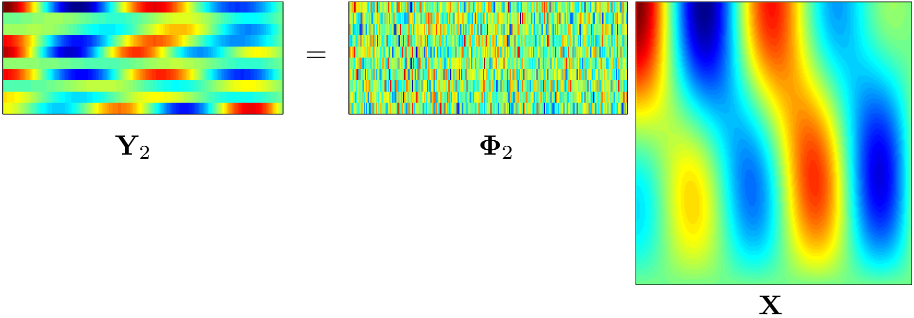
$$\boldsymbol{X}\boldsymbol{\phi}_1,\ \boldsymbol{X}\boldsymbol{\phi}_2, \ldots,\ \boldsymbol{X}\boldsymbol{\phi}_p$$

for $p \approx R$, where the $\boldsymbol{\phi}_i$ are random vectors



Factor $\boldsymbol{Y}_1 = \boldsymbol{Q}_c\boldsymbol{R}_c$ to get an orthobasis for the column space

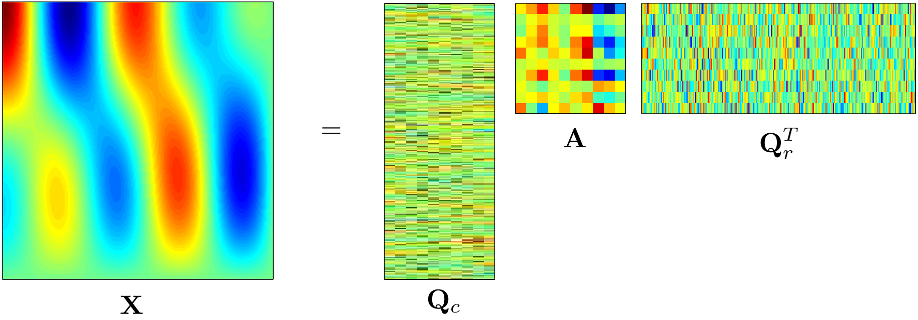
Given an $K \times N$ matrix $\boldsymbol{X}$ with rank R, multiplying by a $q \times N$ random projection $\boldsymbol{U}^{\mathrm{T}}$ preserves the row space for $q \approx R$



Randomized linear algebra

With orthobases for the column space Q_c and the row space Q_r identified, we can recover $\boldsymbol{m}X$ from the two sets of measurements using a *least-squares* algorithm

$$\min_{\boldsymbol{A}} \|\boldsymbol{Y}_1 - \boldsymbol{Q}_c \boldsymbol{A} \boldsymbol{Q}_r^{\mathrm{T}} \boldsymbol{\Phi}_1\|_F^2 + \|\boldsymbol{Y}_2 - \boldsymbol{\Phi}_2 \boldsymbol{Q}_c \boldsymbol{A} \boldsymbol{Q}_r^{\mathrm{T}}\|_F^2$$



Randomized Linear algebra and Rank 1

- Measurements $\boldsymbol{Y}_1 = \boldsymbol{X}\boldsymbol{\Phi}1, \boldsymbol{Y}_2 = \boldsymbol{\Phi}_2^{\mathrm{T}}\boldsymbol{X}$ can be written as

$$y_m = \langle \boldsymbol{X}, \boldsymbol{e}_i \boldsymbol{\phi}_k^{\mathrm{T}} \rangle, \quad \text{or} \quad \langle \boldsymbol{X}, \boldsymbol{\phi}_k \boldsymbol{e}_{i'}^{\mathrm{T}} \rangle$$

  where $\boldsymbol{\phi}_k$ is random and $\boldsymbol{e}_i$ are standard basis vectors.

- Recovery us least-squares for

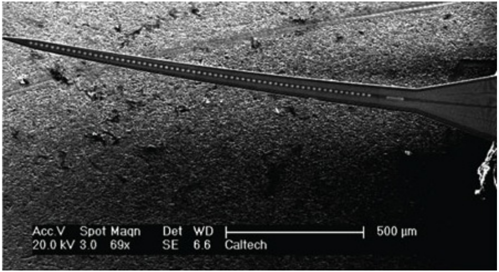
$$M \gtrsim R(K + N)$$

  measurements

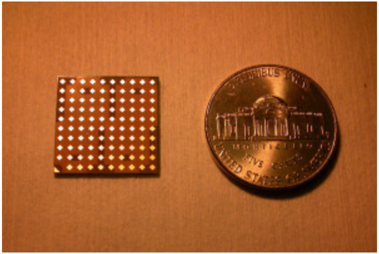
- Stability with noise added?

Sampling ensembles of correlated signals

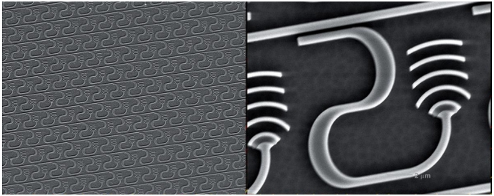
Sensor arrays



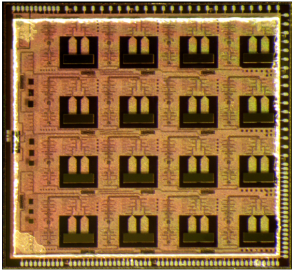
Caltech multielectrode



IBM phased array

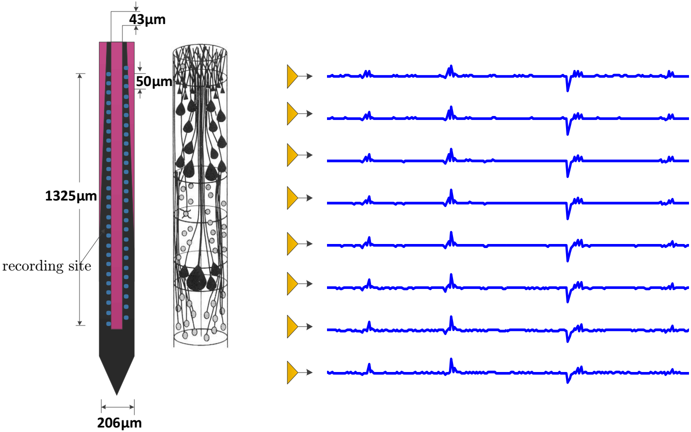


MIT nanophotonic array



UCSD phased

Neural probes

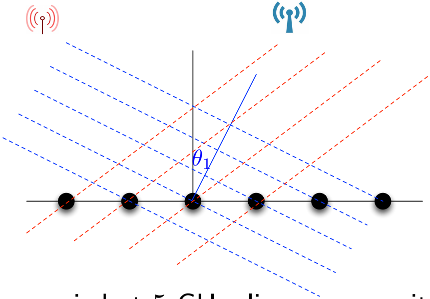


Up to 100s of channels sampled at $\sim$ 100 kHz

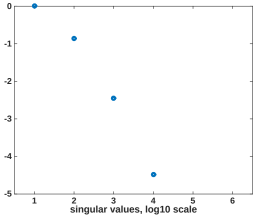
10s of millions of samples/second

Near Future: 1 million channels, terabits per second

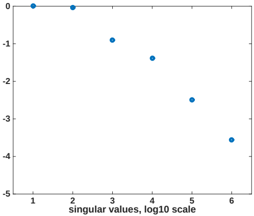
Array processing of narrowband signals



100 MHz bandwidth, carried at 5 GHz, linear array with 100 elements

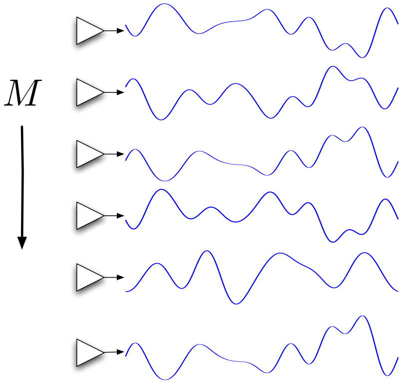


one source



two sources

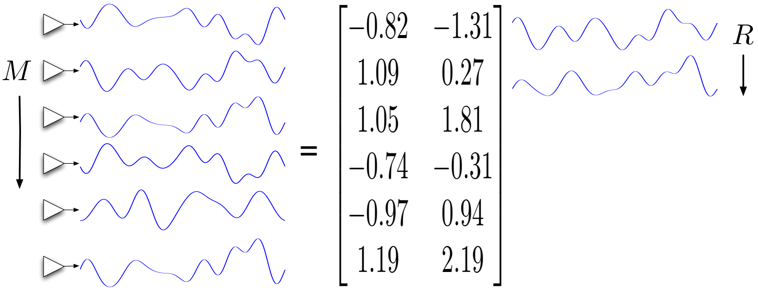
Correlated signals



Nyquist acquisition:

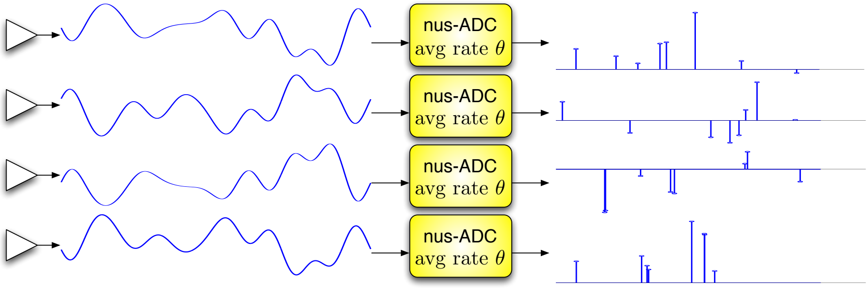
$$\text{sampling rate} \approx (\text{number of signals}) \times (\text{bandwith}) \\ = M \cdot W$$

Correlated signals



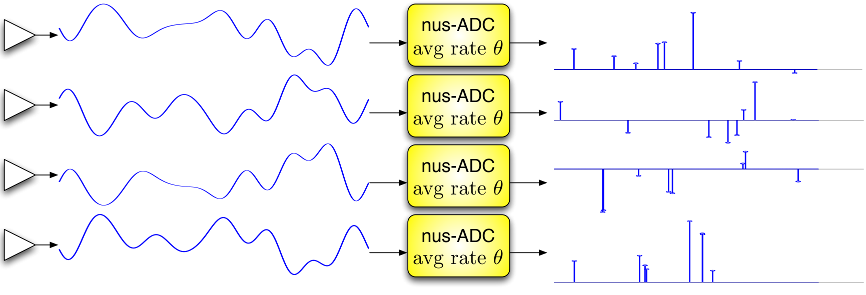
Can we exploit the *latent* correlation structure to reduce the sampling rate?

One non-uniform ADC per channel



- M individual nonuniform-ADCs with average rate θ
- Same as choosing $M\theta$ random samples from $M \times W$ matrix

One non-uniform ADC per channel



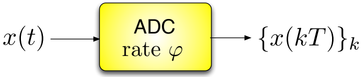
- Direct application of matrix completion results:
  we can recover "incoherent" ensembles when

$$\text{sampling rate} = \theta \ \gtrsim \ \frac{R}{M} W \cdot \log^2(W)$$

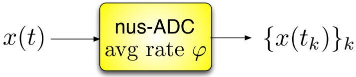
- Incoherent $\Rightarrow$
  signal energy is spread out evenly across time and channels

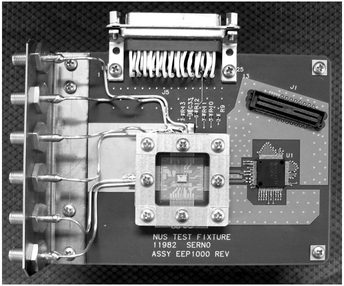
Analog to digital converters

Uniform



Non-uniform



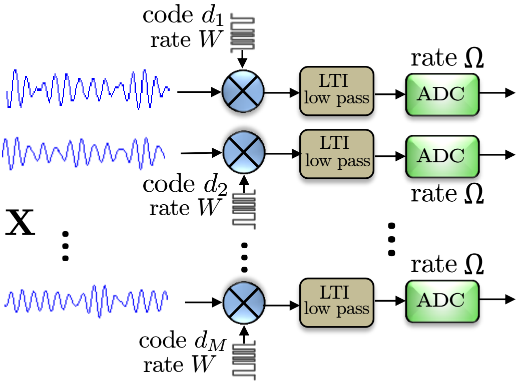


225 Msps, 2 GHz bandwidth

Many architectures for compressive sampling of spectrally sparse signals based on non-uniform sampling

Bresler, Feng, Candes, R, Tao, ...

Sampling using the random demodulator

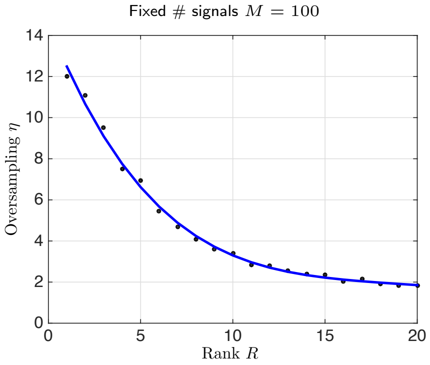


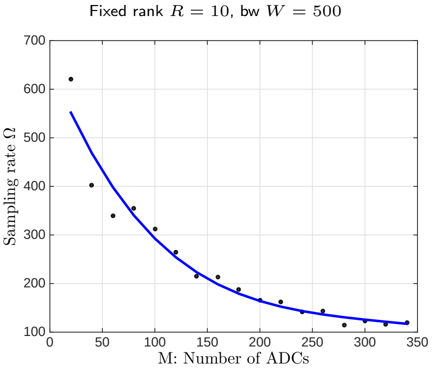
- Instead of running each ADC at rate $\Omega \geq W$, we can take

$$\Omega \gtrsim \frac{R}{M} W \cdot \log^3(W)$$

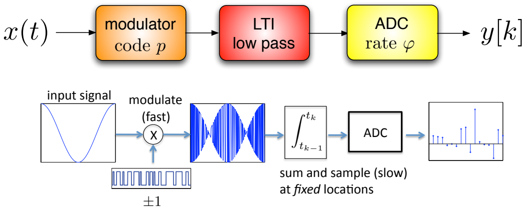
subject to (weaker) incoherence conditions

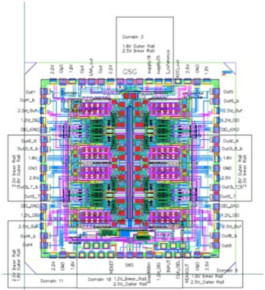
Correlated sampling: numerical results





Random demodulation





(Architecture of Yoo and Emami)

- Architectures for (compressive) sampling of spectrally sparse signals

  Tropp, Duarte, Laska, R, Baraniuk '08

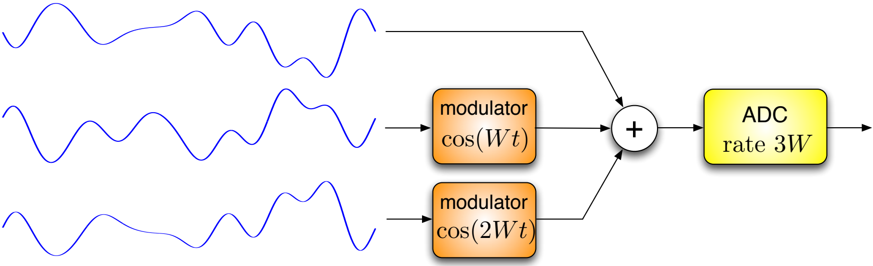
  Mishali, Eldar '09

- Hardware implementations with 10s of channels at 5 GHz

Multiplexing onto one channel

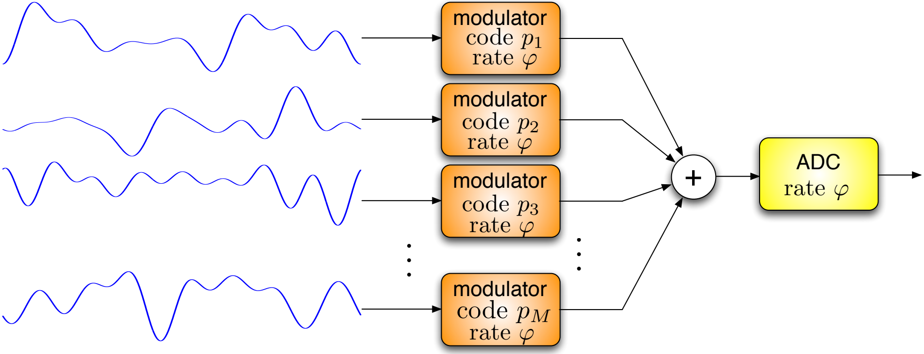
- We can always combine M channels into 1 by *multiplexing* in either time or frequency

  Frequency multiplexer:



- Replace M ADCs running at rate W with 1 ADC at rate MW

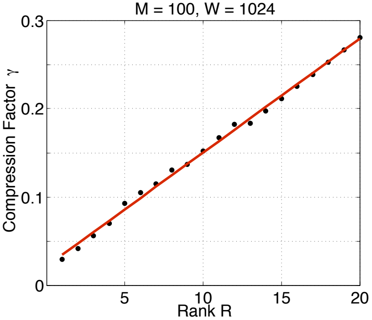
Coded multiplexing



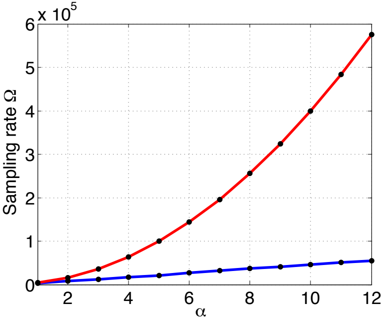
Architecture that achieves

$$\text{sampling rate} \approx (\text{independent signals}) \times (\text{bandwidth})$$
$$\gtrsim \; RW \cdot \log^{3/2} W$$

Coded multiplexing: numerical results



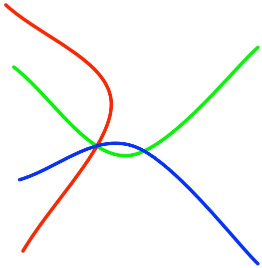
$M = 20\alpha$, $W = 100\alpha$



red=Nyquist, blue=cmux

Blind deconvolution and source separation

Quadratic and bilinear equations



Second-order equations contain unknown terms multiplied by one another

$$v_1^2 + 3v_1v_2 - 6v_1v_3 + v_2^2 = 7$$

$$u_1v_1 + 5u_1v_2 + 7u_2v_3 = -12 \qquad \text{both } u, v \text{ unknown}$$

$$u_3v_1 - 9u_2v_2 + 4u_3v_2 = 2$$

Their nonlinearity makes them trickier to solve, and the computational framework is nowhere nearly as strong as for linear equations

Quadratic and bilinear equations

Simple (but only recently appreciated) observation:
Systems of bilinear equations, e. g.

$$u_1 v_1 + 5u_1 v_2 + 7u_2 v_3 = -12$$
$$u_3 v_1 - 9u_2 v_2 + 4u_3 v_2 = 2$$

can be recast as *linear system of equations on a matrix that has rank 1*:

$$uv^T = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 & \cdots & u_1v_N \\ u_2v_1 & u_2v_2 & u_2v_3 & \cdots & u_2v_N \\ u_3v_1 & u_3v_2 & u_3v_3 & \cdots & u_3v_N \\ \vdots & \vdots & & \ddots & \\ u_Kv_1 & u_Kv_2 & u_Kv_3 & \cdots & u_Kv_N \end{bmatrix}$$

Quadratic and bilinear equations

Simple (but only recently appreciated) observation:
Systems of bilinear equations, e. g.

$$u_1v_1 + 5u_1v_2 + 7u_2v_3 = -12$$
$$u_3v_1 - 9u_2v_2 + 4u_3v_2 = 2$$

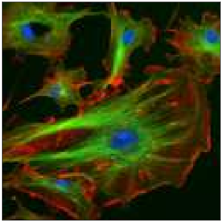
can be recast as *linear system of equations on a matrix that has rank 1*:

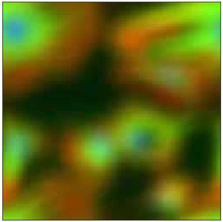
$$uv^T = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 & \cdots & u_1v_N \\ u_2v_1 & u_2v_2 & u_2v_3 & \cdots & u_2v_N \\ u_3v_1 & u_3v_2 & u_3v_3 & \cdots & u_3v_N \\ \vdots & \vdots & & \ddots & \\ u_Kv_1 & u_Kv_2 & u_Kv_3 & \cdots & u_Kv_N \end{bmatrix}$$

Compressive (low rank) recovery $\Rightarrow$
"Generic" quadratic/bilinear systems with cN equations and N unknowns can be solved using nuclear norm minimization

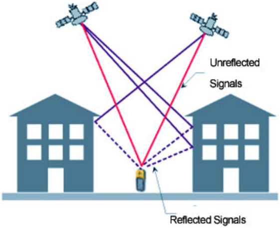
Blind deconvolution

image deblurring





multipath in wireless comm



(image from EngineeringsALL)

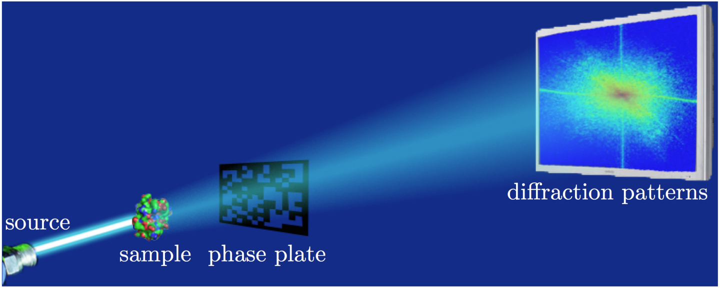
We observe

$$y[\ell] = \sum_n s[n]\, h[\ell - n]$$

and want to "untangle" $\boldsymbol{s}$ and $\boldsymbol{h}$.

(Recent identifiability results by Choudhary, Mitra)

Phase retrieval



(image courtesy of M. Soltanolkotabi)

Observe the *magnitude* of the Fourier transform $|\hat{x}(\omega)|^2$

$\hat{x}(\omega)$ is complex, and its phase carries important information

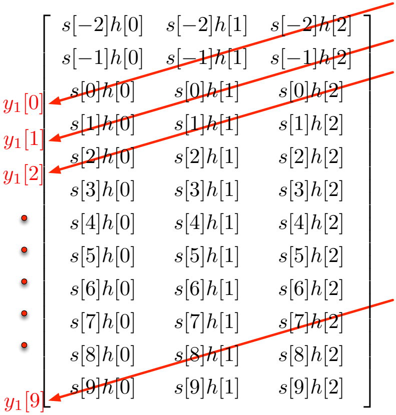
(Recently analyzed by Candes, Li, Soltanolkotabi, Strohmer, and Voroninski)

Blind deconvolution as low rank recovery

Each sample of $\boldsymbol{y} = \boldsymbol{s} * \boldsymbol{h}$ is a bilinear combination of the unknowns,

$$y[\ell] = \sum_n s[n]h[\ell - n]$$

and is a *linear* combination of $\boldsymbol{s}\boldsymbol{h}^{\mathrm{T}}$:



Blind deconvolution as low rank recovery

Given $\boldsymbol{y} = \boldsymbol{s} * \boldsymbol{h}$, it is impossible to untangle $\boldsymbol{s}$ and $\boldsymbol{h}$ unless we make some *structural assumptions*

Structure: $\boldsymbol{s}$ and $\boldsymbol{h}$ live in known *subspaces* of $\mathbb{R}^L$; we can write

$$\boldsymbol{s} = \boldsymbol{B}\boldsymbol{u}, \qquad \boldsymbol{h} = \boldsymbol{C}\boldsymbol{v}, \quad B : L \times K, \;\; C : L \times N$$

where $\boldsymbol{B}$ and $\boldsymbol{C}$ are matrices whose columns form bases for these spaces

We can now write blind deconvolution as a *linear inverse problem with a rank contraint*:

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}_0), \quad \text{where} \;\; \boldsymbol{X}_0 = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \;\; \text{has rank=1}$$

The action of $\mathcal{A}(\cdot)$ can be broken down into three linear steps:

$$\boldsymbol{X}_0 \;\rightarrow\; \boldsymbol{B}\boldsymbol{X}_0 \;\rightarrow\; \boldsymbol{B}\boldsymbol{X}_0\boldsymbol{C}^{\mathrm{T}} \;\rightarrow\; \text{take skew-diagonal sums}$$

Blind deconvolution theoretical results

We observe

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{s} * \boldsymbol{h}, \qquad \boldsymbol{h} = \boldsymbol{B}\boldsymbol{u}, \quad \boldsymbol{s} = \boldsymbol{C}\boldsymbol{v} \\ &= \mathcal{A}(\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}), \qquad \boldsymbol{u} \in \mathbb{R}^K, \quad \boldsymbol{v} \in \mathbb{R}^N, \end{aligned}$$

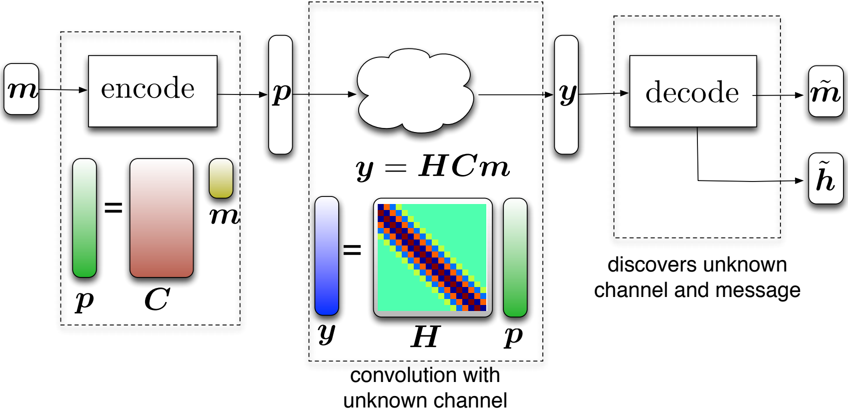
and then solve

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{X}\|_* \quad \text{subject to} \quad \mathcal{A}(\boldsymbol{X}) = \boldsymbol{y}.$$

Ahmed, Recht, R, '13:
If $\boldsymbol{B}$ is "incoherent" in the Fourier domain, and $\boldsymbol{C}$ is randomly chosen, then we will recover $\boldsymbol{X}_0 = \boldsymbol{s}\boldsymbol{h}^{\mathrm{T}}$ exactly (with high probability) when

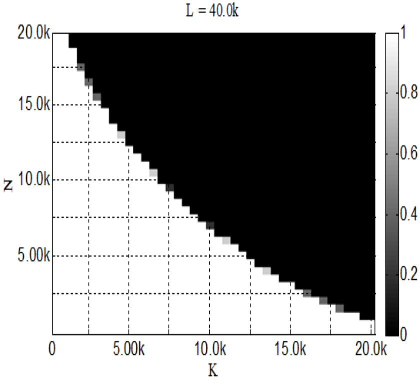
$$L \ \geq \ \mathrm{Const} \cdot (K + N) \cdot \log^3(KN)$$

Multipath protection

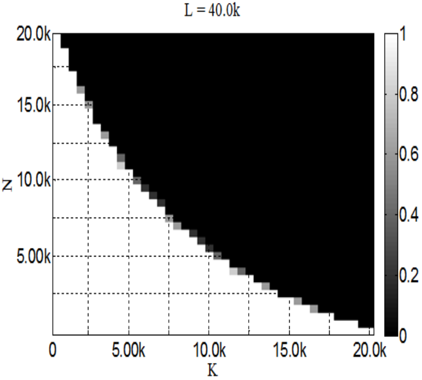


Numerical results

white = 100% success, black = 0% success



$\boldsymbol{h}$ sparse, $\boldsymbol{s}$ randomly coded



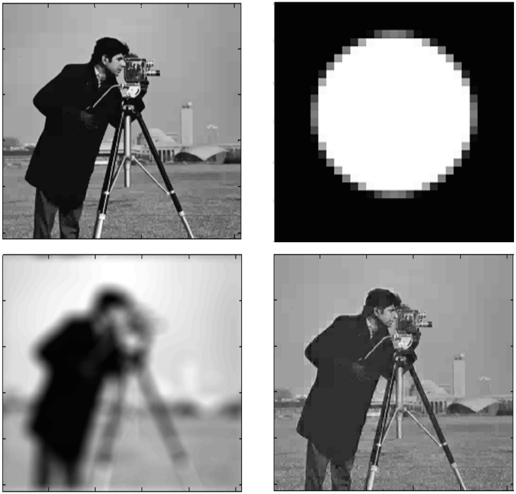
$\boldsymbol{h}$ short, $\boldsymbol{s}$ randomly coded

In the cases above, we can take

$$(N + K) \lesssim L/3$$

Numerical results

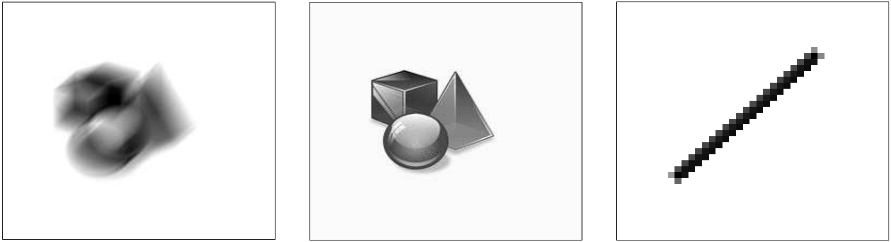
Unknown image with known support in the wavelet domain,
Unknown blurring kernel with known support in spatial domain



observed　　　　recovered

Numerical results

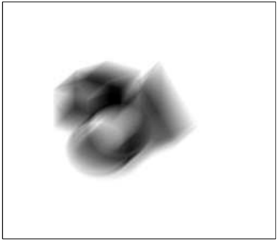
Oracle recovery



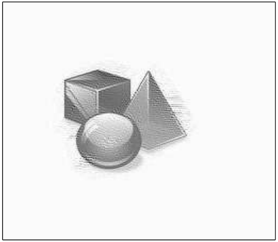
observed | recovered image | recovered kernel

Numerical results

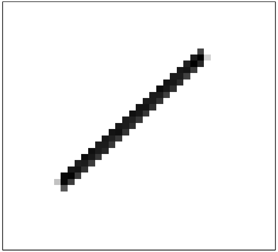
Adaptive recovery



observed

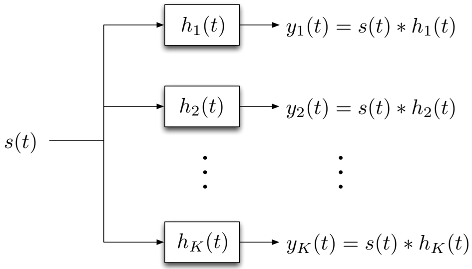


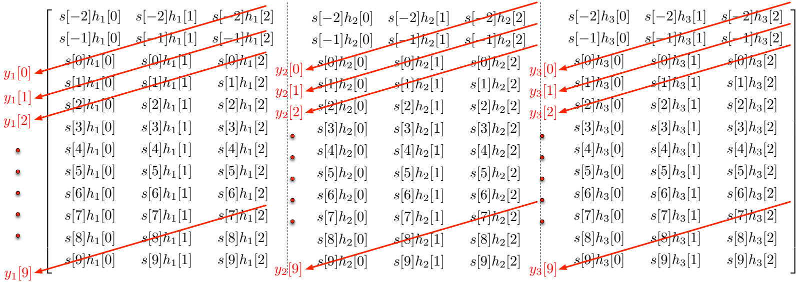
recovered image



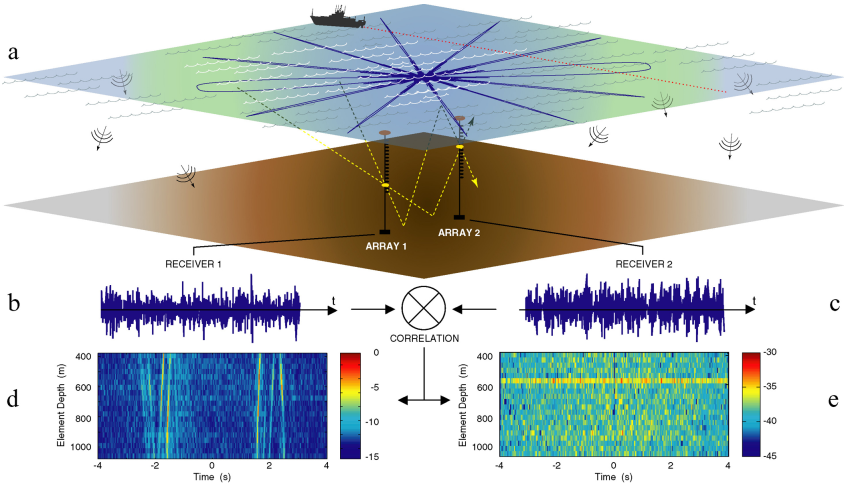
recovered kernel

Passive estimation of multiple channels





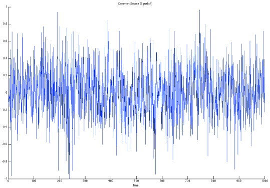
Passive imaging of the ocean

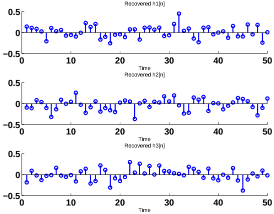


Recovery results

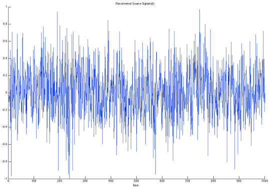
Source / output length: 1000
Number of channels: 100
Channel impulse response length: 50

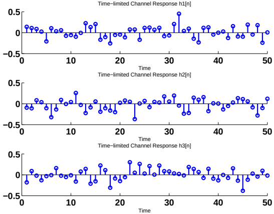
Original:



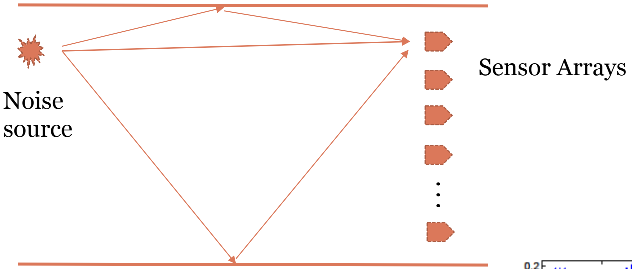


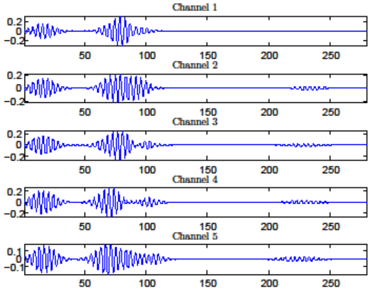
Recovered:





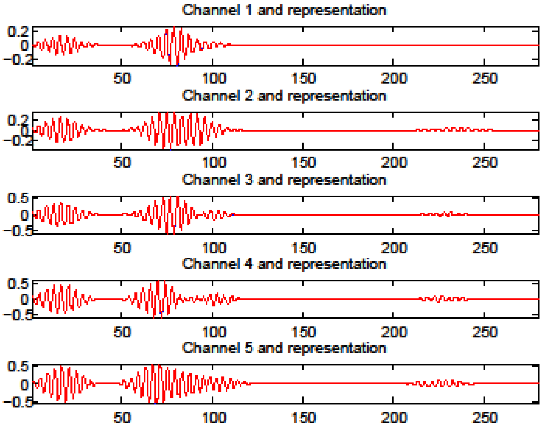
Realistic (simulated) ocean channels





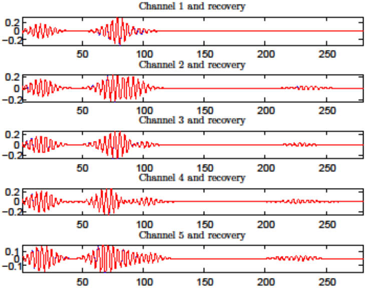
- Noise signal is in the broad band 400~600Hz
- The distance between the noise source and sensor arrays is approximate 1km

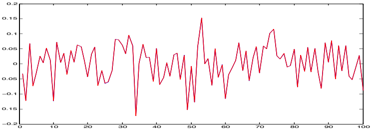
Realistic (simulated) ocean channels



Build a subspace model using bandwidth and approximate arrival times (about 20 dimensions per channel)

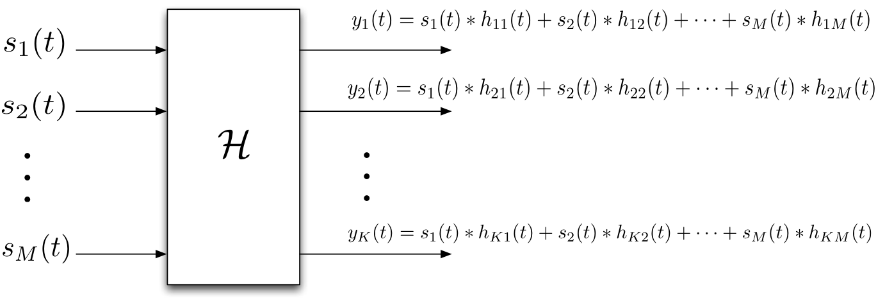
Simulated recovery





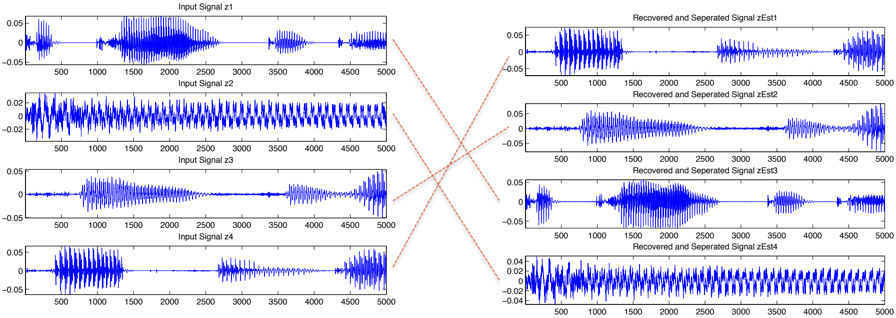
~ 100 channels total, ~ 2000 samples per channel,
Normalized error $\sim 10^{-4}$ (no noise), robust with noise

Multiple sources



- Memoryless: structured matrix factorization (SMF) problem
  ICA, NNMF, dictionary learning, etc.
- Use matrix recovery to make convolutional channels "memoryless":
  recover rank M matrix, run SMF on column space

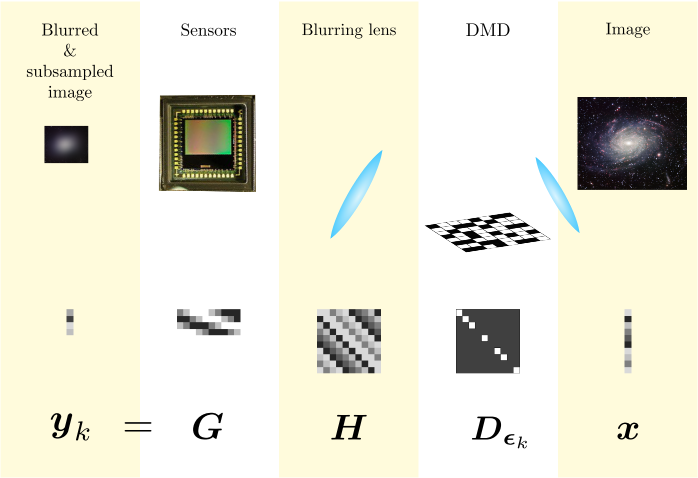
Low-rank recovery + ICA on broadband voice



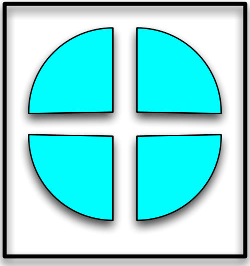
4 sources
30 channels (microphones)
2000 time samples
10 taps per channel

“Blind super-resolution” in coded imaging

Imaging architecture



- Small number of sensors with gaps between them
- Blurring introduced to “fill in” these gaps
- Uncalibrated: blur kernel is unknown



Masked imaging linear algebra

$$\boldsymbol{y}_k = \left[\boldsymbol{GH} \right] \left[\boldsymbol{D}_k \right] \left[\boldsymbol{x} \right]$$

$\boldsymbol{GH}$: unknown

linear constraints

$$\boldsymbol{GH} = \sum_i a[i]\, \boldsymbol{B}_i$$

$\boldsymbol{D}_k$: known

$\boldsymbol{x}$: unknown image,

unconstrained

- Operator coefficients $\boldsymbol{a}$, image $\boldsymbol{x}$ unknown
- Observations: $\mathcal{A}(\boldsymbol{a}\boldsymbol{x}^{\mathrm{T}})$
- Alternative interpretation: *structured matrix factorization*

$$\boldsymbol{Y} = (\boldsymbol{GH})\,\mathrm{diag}(\boldsymbol{X})\boldsymbol{\Phi}^{\mathrm{T}}$$

Masked imaging: theoretical results

$$\boldsymbol{y}_k = \begin{bmatrix} \boldsymbol{GH} \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \end{bmatrix}$$

known

unknown

linear constraints

$$\boldsymbol{GH} = \sum_i a[i]\, \boldsymbol{B}_i$$

unknown image,

unconstrained

L pixels, N sensors, K codes

Theorem (Bahimani, R '14):

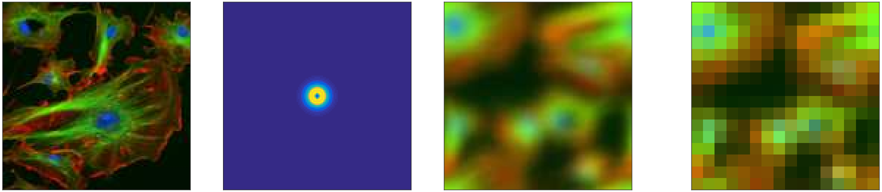
We can jointly recover the blur $\boldsymbol{H}$ and the image $\boldsymbol{X}$ for a number of codes:

$$K \gtrsim \mu^2 \frac{L}{N} \cdot \log^3(L) \log\log N$$

$\mu^2 \geq 1$ measures how spread out blur is in frequency

(Related work by Tang and Recht '14)

Masked imaging: numerical results

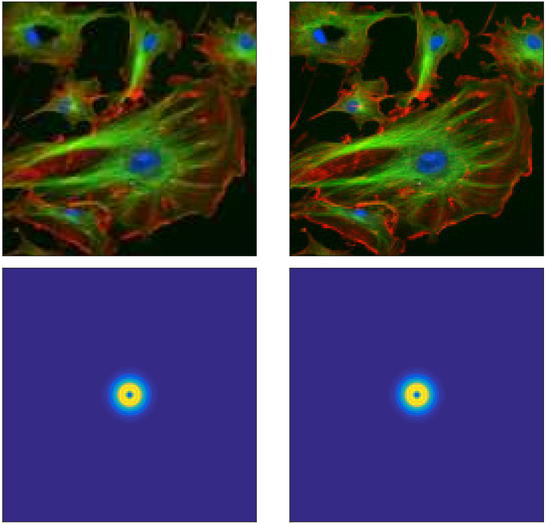


original blur blurred image blurred, subsampled

- No structural model for the image
- Blur model: build basis from psfs over a range of focal lengths (EPFL PSF Generator, Born and Wolf model)

Masked imaging: numerical results

Recovery results: 16k pixels, 64 sensors, 200 codes



originals recovery

Notes on computation and extentions

Computational concerns

Calculating

$$\min_{\boldsymbol{X} \in \mathbb{R}^{K \times N}} \|\boldsymbol{X}\|_* \quad \text{subject to} \quad \mathcal{A}(\boldsymbol{X}) = \boldsymbol{y}$$

is expensive — it's an optimization program in KN variables.

The solution is low rank, so we would like to keep iterates low-rank as well.

Replace with

$$\min_{\boldsymbol{L},\boldsymbol{R}} \|\boldsymbol{L}\|_F^2 + \|\boldsymbol{R}\|_F^2 \quad \text{subject to} \quad \mathcal{A}(\boldsymbol{L}\boldsymbol{R}^{\mathrm{T}}) = \boldsymbol{y}$$

with $\boldsymbol{R} : K \times R'$ and $\boldsymbol{L} : N \times R'$.

Nonconvex heuristic

$$\min_{\boldsymbol{L},\boldsymbol{R}} \|\boldsymbol{L}\|_F^2 + \|\boldsymbol{R}\|_F^2 \quad \text{subject to} \quad \mathcal{A}(\boldsymbol{L}\boldsymbol{R}^{\mathrm{T}}) = \boldsymbol{y}$$

with $\boldsymbol{R} : K \times R'$ and $\boldsymbol{L} : N \times R'$.

- Requires $\sim R'(K+N)$ storage, as opposed to $\sim KN$
- Nonconvex
- Same solution as nuclear norm when $R' \geq \mathrm{rank}(\boldsymbol{X}_0)$
- For small enough rank, local minima correspond to global minima (need many measurements for convergence guarantees, though)

Convexification for rank 1

Given rank-1 measurements of a rank-1 matrix,

$$y_m = \langle \boldsymbol{w}_0 \boldsymbol{z}_0^{\mathrm{T}}, \boldsymbol{\nu}_m \boldsymbol{\phi}_m^{\mathrm{T}} \rangle$$

it is natural to solve

$$\min_{\boldsymbol{w},\boldsymbol{z}} \|\boldsymbol{w}\|_2^2 + \|\boldsymbol{z}\|_2^2 \quad \text{subject to} \quad \langle \boldsymbol{w}, \boldsymbol{\nu}_m \rangle \langle \boldsymbol{z}, \boldsymbol{\phi}_m \rangle = y_m, \quad m = 1, \ldots, M$$

Convexification for rank 1

Given rank-1 measurements of a rank-1 matrix,

$$y_m = \langle \boldsymbol{w}_0 \boldsymbol{z}_0^{\mathrm{T}}, \boldsymbol{\nu}_m \boldsymbol{\phi}_m^{\mathrm{T}} \rangle$$

it is natural to solve

$$\min_{\boldsymbol{w},\boldsymbol{z}} \|\boldsymbol{w}\|_2^2 + \|\boldsymbol{z}\|_2^2 \quad \text{subject to} \quad \langle \boldsymbol{w}, \boldsymbol{\nu}_m \rangle \langle \boldsymbol{z}, \boldsymbol{\phi}_m \rangle = y_m, \quad m = 1, \ldots, M$$

Dual is an SDP:

$$\min_{\boldsymbol{\lambda}} \langle \boldsymbol{\lambda}, \boldsymbol{y} \rangle \quad \text{subject to} \quad \begin{bmatrix} \mathbf{I} & \sum_m \lambda_m \boldsymbol{\nu}_m \boldsymbol{\phi}_m^{\mathrm{T}} \\ \sum_m \lambda_m \boldsymbol{\nu}_m \boldsymbol{\phi}_m^{\mathrm{T}} & \mathbf{I} \end{bmatrix} \succeq \boldsymbol{m}0$$

Convexification for rank 1

Given rank-1 measurements of a rank-1 matrix,

$$y_m = \langle \boldsymbol{w}_0 \boldsymbol{z}_0^{\mathrm{T}}, \boldsymbol{\nu}_m \boldsymbol{\phi}_m^{\mathrm{T}} \rangle$$

it is natural to solve

$$\min_{\boldsymbol{w},\boldsymbol{z}} \|\boldsymbol{w}\|_2^2 + \|\boldsymbol{z}\|_2^2 \quad \text{subject to} \quad \langle \boldsymbol{w}, \boldsymbol{\nu}_m \rangle \langle \boldsymbol{z}, \boldsymbol{\phi}_m \rangle = y_m, \quad m = 1, \ldots, M$$

Dual-of-the-dual is equivalent to

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{X}\|_* \quad \text{subject to} \quad \mathcal{A}(\boldsymbol{X}) = \boldsymbol{y}$$

Nuclear norm minimization is the natural relaxation

Alternating minimization

A "classical" way to solve a bilinear problem: find $\boldsymbol{U}^\star, \boldsymbol{V}^\star$ such that

$$\mathcal{A}(\boldsymbol{U}^\star \boldsymbol{V}^{\star\mathrm{T}}) = \boldsymbol{y}$$

by choosing initial $\boldsymbol{U}_0$, then iterating

$$\begin{aligned} \boldsymbol{V}_k &= \arg\min_{\boldsymbol{V}} \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{U}_{k-1}\boldsymbol{V}^{\mathrm{T}})\|_2^2 \\ \boldsymbol{U}_k &= \arg\min_{\boldsymbol{U}} \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{U}\boldsymbol{V}_k^{\mathrm{T}})\|_2^2 \end{aligned}$$

Both of these are linear least-squares problems

Recently, Jain, Netrapalli, and Sanghavi have analyzed the initialization and convergence of this for many types of measurements (random projections, matrix completion, phase retrieval)

Powerful results for rank 1 recovery for phase retrieval in Candes, Li, Soltanolkotabi

Simultaneous structure

What if the target $\boldsymbol{X}_0$ is *simultaneously* sparse and low rank?

There are multiple negative results for convex relaxation; for example

$$\min_{\boldsymbol{X}} \frac{1}{2}\|\boldsymbol{y} - \mathcal{A}(\boldsymbol{X})\|_F^2 + \tau_1\|\boldsymbol{X}\|_* + \tau_2\|\boldsymbol{X}\|_1$$

where $\mathcal{A}$ is a random projection, is not fundamentally better than using rank alone (Oymak et al. '12)

For phase retrieval, the number of measurements required for convex relation is large ($\sim N^2$). (Li et al '13)

Simultaneous structure

For alternating minimization, there seems to be more hope:
Example: $\boldsymbol{X}_0$ is $K \times N$ row S-sparse ($S \ll K$) and low rank ($R < K$) then iterating

$$\boldsymbol{U}_k = \mathrm{SparseApprox}_{\boldsymbol{U}}(\boldsymbol{y},\ \mathcal{A}'), \quad \mathcal{A}'_k(\boldsymbol{U}) = \mathcal{A}(\boldsymbol{U}\boldsymbol{V}_{k-1})$$

$$\boldsymbol{V}_k = \arg\min_{\boldsymbol{U}} \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{U}_k \boldsymbol{V}^{\mathrm{T}})\|_2$$

from a known starting point is effective when $\mathcal{A}$ obeys "SSLR-RIP"

(Lee, Wu, Bresler '13)

Random rank-1 measurements $y_m = \langle \boldsymbol{X}, \boldsymbol{\nu}_m \phi_m^{\mathrm{T}} \rangle$ obey SSLR-RIP for

$$M \gtrsim (K+N)\log^5(K+N)$$

(Ahmed, Krahmer, R '15)

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. Ahmed, B. Recht, and J. Romberg, "Blind Deconvolution using Convex Programming," *IEEE Transactions on Information Theory*, March 2014.
- S. Bahmani and J. Romberg, "Compressive deconvolution in random mask imaging," submitted to *IEEE Transactions on Computational Imaging*, December 2014.
- S. Bahmani and J. Romberg, "Blind deconvolution and super-resolution using random masks," manuscript in preparation, to be submitted to *SIAM J. Imaging Science*, December 2014.

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