A Survey of Compressed Sensing and Applications to Medical Imaging

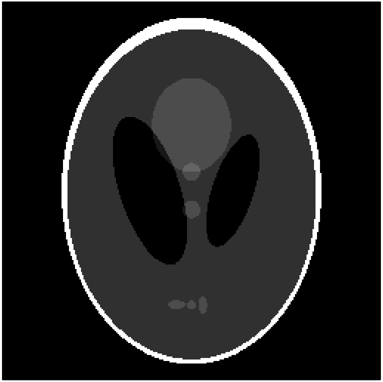
Justin Romberg

Georgia Tech, School of ECE

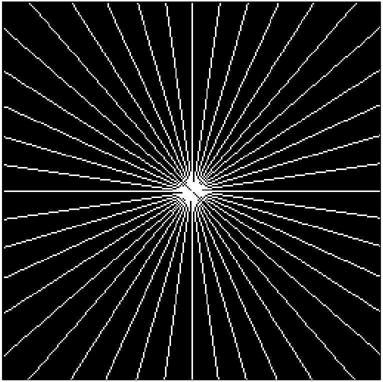
Fields-MITACS Conference on Mathematics of Medical Imaging
June 21, 2011
Toronto, Ontario

A simple underdetermined inverse problem

Observe a subset Ω of the 2D discrete Fourier plane



phantom (hidden)



white star = sample locations

$N := 512^2 = 262,144$ pixel image

observations on 22 radial lines, $10,486$ samples, $\approx 4\%$ coverage

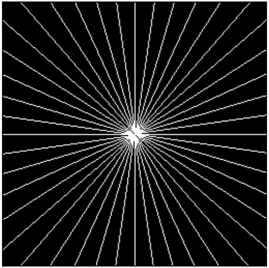
Minimum energy reconstruction

Reconstruct g^* with

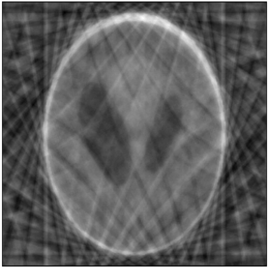
$$\hat{g}^*(\omega_1, \omega_2) = \begin{cases} \hat{f}(\omega_1, \omega_2) & (\omega_1, \omega_2) \in \Omega \\ 0 & (\omega_1, \omega_2) \notin \Omega \end{cases}$$

Set unknown Fourier coeffs to zero, and inverse transform

original



Fourier samples



g^*

Total-variation reconstruction

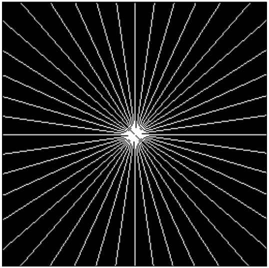
Find an image that

- Fourier domain: *matches observations*
- Spatial domain: has a *minimal amount of oscillation*

Reconstruct g^* by solving:

$$\min_{g} \sum_{i,j} |(\nabla g)_{i,j}| \quad \text{s.t.} \quad \hat{g}(\omega_1, \omega_2) = \hat{f}(\omega_1, \omega_2), \quad (\omega_1, \omega_2) \in \Omega$$

original



Fourier samples

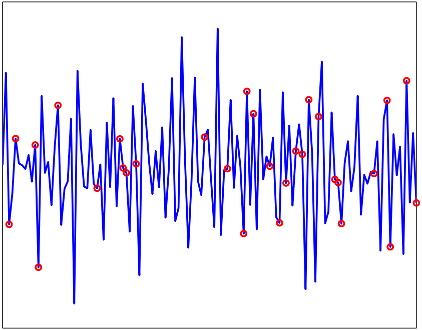
$g^* =$ original

perfect reconstruction

Sampling a superposition of sinusoids

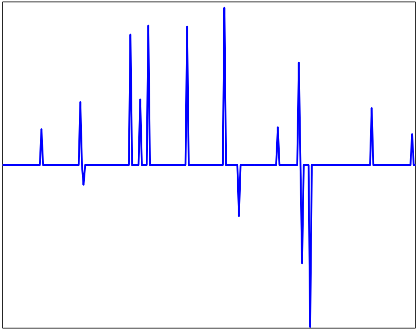
We take M samples of a superposition of S sinusoids:

Time domain $x_0(t)$



Measure M samples
(red circles = samples)

Frequency domain $\hat{x}_0(\omega)$

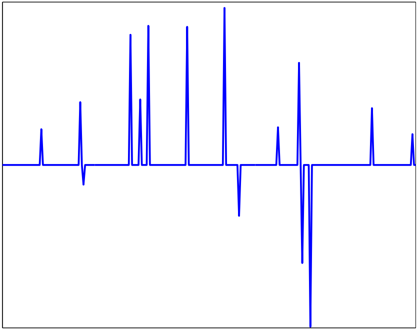


S nonzero components

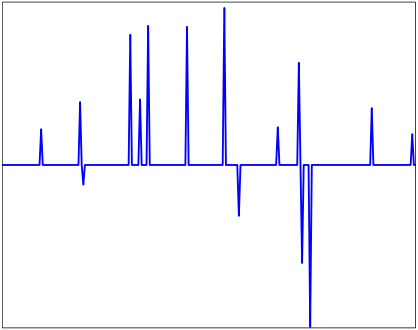
Sampling a superposition of sinusoids

Reconstruct by solving

$$\min_{x} \; \|\hat{x}\|_{\ell_1} \quad \text{subject to} \quad x(t_m) = x_0(t_m), \quad m = 1, \ldots, M$$



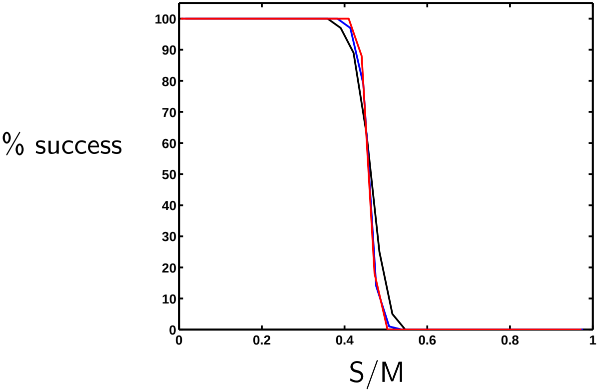
original $\hat{x}_0$, $S = 15$



perfect recovery from 30 samples

Numerical recovery curves

- Resolutions $N = 256, 512, 1024$ (black, blue, red)
- Signal composed of S randomly selected sinusoids
- Sample at M randomly selected locations



- In practice, perfect recovery occurs when $M \approx 2S$ for $N \approx 1000$

A nonlinear sampling theorem

Exact Recovery Theorem (Candès, R, Tao, 2004):

- Unknown $\hat{x}_0$ is supported on set of size S
- Select M sample locations $\{t_m\}$ "at random" with

$$M \;\geq\; \mathrm{Const} \cdot S \log N$$

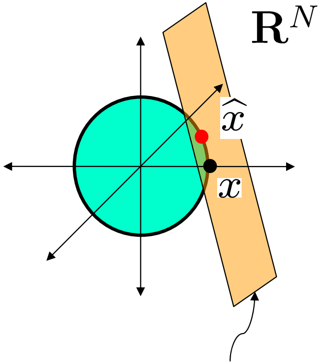
- Take time-domain samples (measurements) $y_m = x_0(t_m)$
- Solve

$$\min_x \|\hat{x}\|_{\ell_1} \quad \text{subject to} \quad x(t_m) = y_m, \quad m = 1, \ldots, M$$

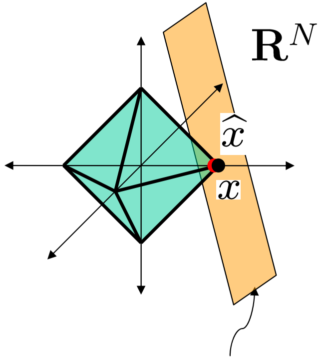
- Solution is *exactly* f with extremely high probability
- In total-variation/phantom example, S=number of jumps

Graphical intuition for ℓ_1

$\min_x \|x\|_2$ s.t. $\Phi x = y$ $\qquad$ $\min_x \|x\|_1$ s.t. $\Phi x = y$

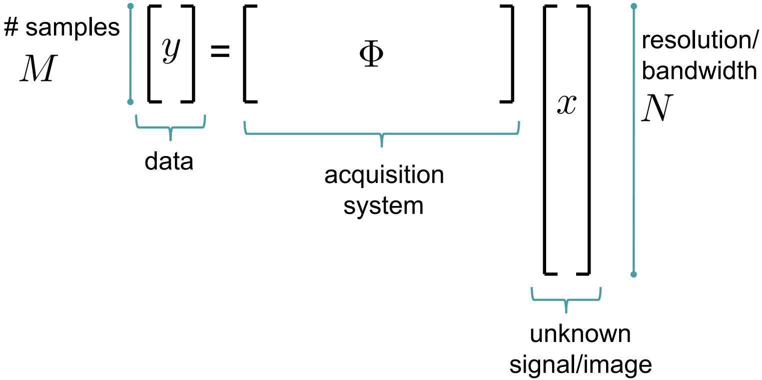


$\{x' : \; y = \Phi x'\}$



$\{x' : \; y = \Phi x'\}$

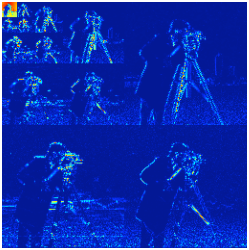
Acquisition as linear algebra



- Small number of samples $=$ underdetermined system
  Impossible to solve in general
- If x is *sparse* and Φ is *diverse*, then these systems can be "inverted"

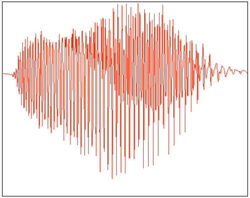
Sparsity/Compressibility

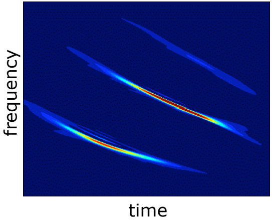
N
pixels



$S \ll N$
large
wavelet
coefficients

N
wideband
signal
samples





$S \ll N$
large
Gabor
coefficients

Classical: When can we stably “invert” a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

- Q: When is this recovery stable? That is, when is

$$\|\hat{x} - x_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

- Q: When is this recovery stable? That is, when is

$$\|\hat{x} - x_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

- A: When the matrix A is an *approximate isometry*...

$$\|Ax\|_2^2 \approx \|x\|_2^2 \quad \text{for all } x \in \mathbb{R}^N$$

  i.e. A preserves *lengths*

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

- Q: When is this recovery stable? That is, when is

$$\|\hat{x} - x_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

- A: When the matrix A is an *approximate isometry*...

$$\|A(x_1 - x_2)\|_2^2 \approx \|x_1 - x_2\|_2^2 \quad \text{for all } x_1, x_2 \in \mathbb{R}^N$$

  i.e. A preserves *distances*

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

- Q: When is this recovery stable? That is, when is

$$\|\hat{x} - x_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

- A: When the matrix A is an *approximate isometry*...

$$(1 - \delta) \leq \sigma_{\min}^2(A) \ \leq \ \sigma_{\max}^2(A) \leq (1 + \delta)$$

  i.e. A has *clustered singular values*

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

- Q: When is this recovery stable? That is, when is

$$\|\hat{x} - x_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

- A: When the matrix A is an *approximate isometry*...

$$(1-\delta)\|x\|_2^2 \;\leq\; \|Ax\|_2^2 \;\leq\; (1+\delta)\|x\|_2^2$$

for some $0 < \delta < 1$

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system Φ (FEWER measurements than unknowns), and observe

$$y = \Phi x_0 + \text{noise}$$

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system Φ (FEWER measurements than unknowns), and observe

$$y = \Phi x_0 + \text{noise}$$

- We can recover x_0 when Φ is a *keeps sparse signals separated*

$$(1-\delta)\|x_1 - x_2\|_2^2 \;\leq\; \|\Phi(x_1 - x_2)\|_2^2 \;\leq\; (1+\delta)\|x_1 - x_2\|_2^2$$

for all S-sparse x_1, x_2

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system Φ (FEWER measurements than unknowns), and observe

$$y = \Phi x_0 + \text{noise}$$

- We can recover x_0 when Φ is a *restricted isometry (RIP)*

$$(1-\delta)\|x\|_2^2 \;\leq\; \|\Phi x\|_2^2 \;\leq\; (1+\delta)\|x\|_2^2 \quad \text{for all } 2S\text{-sparse } x$$

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system Φ (FEWER measurements than unknowns), and observe

$$y = \Phi x_0 + \text{noise}$$

- We can recover x_0 when Φ is a *restricted isometry (RIP)*

$$(1-\delta)\|x\|_2^2 \;\leq\; \|\Phi x\|_2^2 \;\leq\; (1+\delta)\|x\|_2^2 \quad \text{for all } 2S\text{-sparse } x$$

- To recover x_0, we solve

$$\min_x \; \|x\|_0 \quad \text{subject to} \quad \Phi x \approx y$$

$\|x\|_0 =$ number of nonzero terms in x

- This program is intractable

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system Φ (FEWER measurements than unknowns), and observe

$$y = \Phi x_0 + \text{noise}$$

- We can recover x_0 when Φ is a *restricted isometry (RIP)*

$$(1-\delta)\|x\|_2^2 \ \le \ \|\Phi x\|_2^2 \ \le \ (1+\delta)\|x\|_2^2 \quad \text{for all } 2S\text{-sparse } x$$

- A relaxed (convex) program

$$\min_x \ \|x\|_1 \quad \text{subject to} \quad \Phi x \approx y$$

$\|x\|_1 = \sum_k |x_k|$

- This program is very tractable (linear program)

Sparse recovery algorithms

- Given y, look for a sparse signal which is consistent.
- One method: ℓ_1 minimization (or *Basis Pursuit*)

$$\min_x \ \|\Psi^T x\|_1 \quad \text{s.t.} \quad \Phi x = y$$

Ψ = sparsifying transform, Φ = measurement system

- $2S$-RIP for $\Phi\Psi \;\Rightarrow\;$ perfect recovery of S-sparse signals
- Convex (linear) program, can relax for robustness to noise...

Stable recovery

- Despite its nonlinearity, sparse recovery is stable in the presence of
  - *modeling mismatch* (approximate sparsity), and
  - *measurement error*
- If we observe $y = \Phi x_0 + e$, with $\|e\|_2 \leq \epsilon$, the solution $\hat{x}$ to

$$\min_x \|\Psi^T x\|_1 \quad \text{s.t.} \quad \|y - \Phi x\|_2 \leq \epsilon$$

  will satisfy

$$\|\hat{x} - x_0\|_2 \;\leq\; \mathrm{Const} \cdot \left(\epsilon + \frac{\|x_0 - x_{0,S}\|_1}{\sqrt{S}} \right)$$

  where
  - $x_{0,S} = S$-term approximation of x_0
  - S is the largest value for which $\Phi\Psi$ satisfies the RIP
- Similar guarantees exist for other recovery algorithms
  - greedy (Needell and Tropp '08)
  - iterative thresholding (Blumensath and Davies '08)

What types of matrices are restricted isometries?

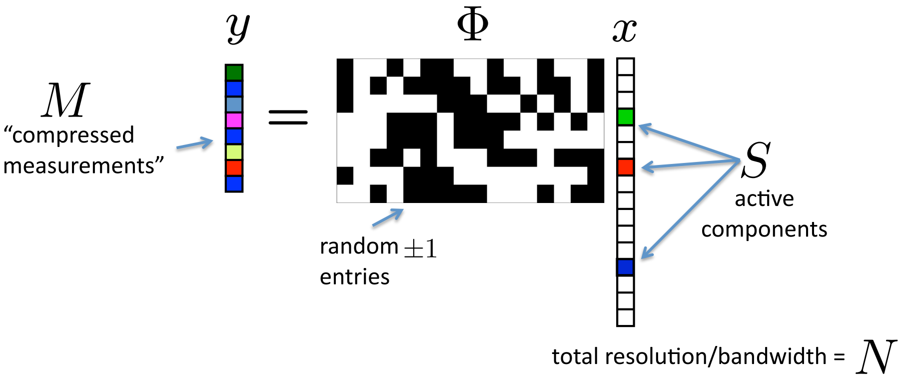
Three examples:

- Random matrices (iid entries)
- Random subsampling
- Random convolution

Note the role of randomness in all of these approaches

Slogan: *random projections keep sparse signal separated*

Random matrices (iid entries)

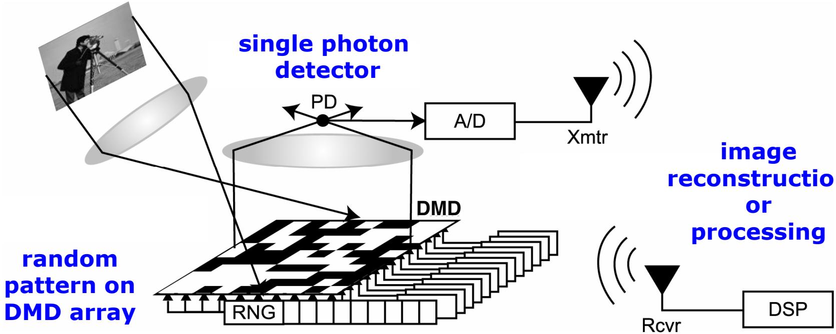


- *Random matrices* are provably efficient
- We can recover S-sparse x from

$$M \gtrsim S \cdot \log(N/S)$$

measurements

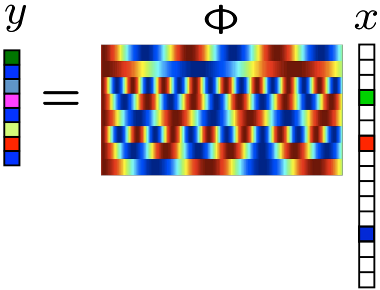
Rice single pixel camera



(Duarte, Davenport, Takhar, Laska, Sun, Kelly, Baraniuk '08)

Random matrices

Example: Φ consists of *random rows* from an *orthobasis* U



Can recover S-sparse x from

$$M \gtrsim \mu^2 \; S \cdot \log^4 N$$

measurements, where

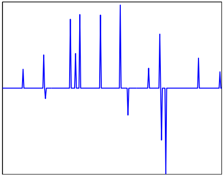
$$\mu = \sqrt{N} \max_{i,j} |(U^T \Psi)_{ij}|$$

is the *coherence*

Examples of incoherence

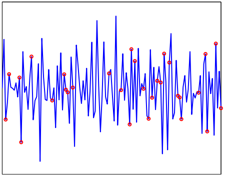
- Signal is sparse in time domain, sampled in Fourier domain

time domain $x(t)$



S nonzero components

freq domain $\hat{x}(\omega)$

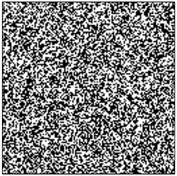


measure m samples

- Signal is sparse in wavelet domain, measured with noiselets

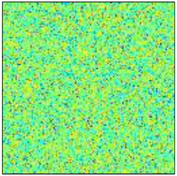
(Coifman et al '01)

example noiselet

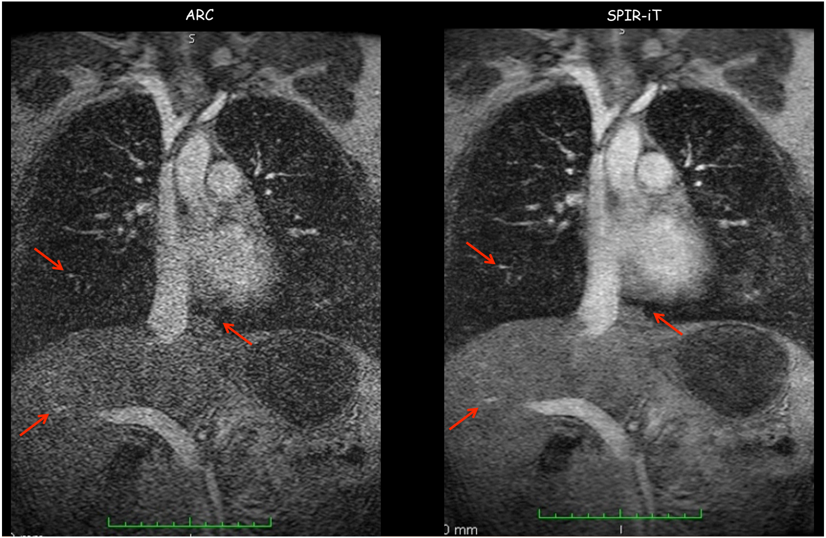


wavelet domain

noiselet domain



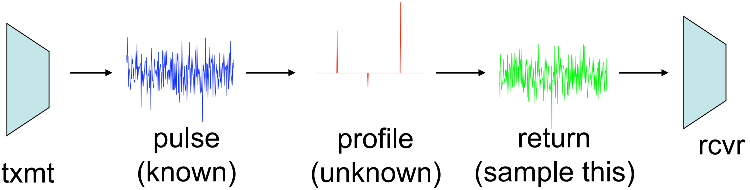
Accelerated MRI



(Lustig et al. '08)

Random convolution

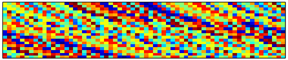
- Many *active imaging* systems measure a pulse convolved with a *reflectivity profile* (Green's function)



- Applications include:
  - radar imaging
  - sonar imaging
  - seismic exploration
  - channel estimation for communications
  - super-resolved imaging
- Using a *random pulse* = compressive sampling
  (Tropp et al. '06, R '08, Herman et al. '08, Haupt et al. '09, Rauhut '09)

Random convolution for CS, theory

- Signal model: sparsity in *any orthobasis* Ψ
- Acquisition model:
  generate a "pulse" whose FFT is a sequence of random phases (unit magnitude),
  convolve with signal,
  sample result at m *random* locations Ω



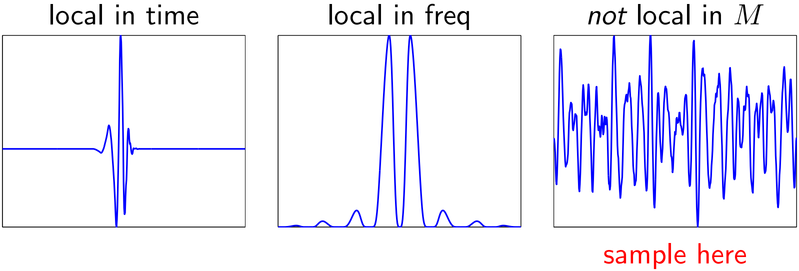
$$\Phi = R_\Omega \mathcal{F}^* \Sigma \mathcal{F}, \quad \Sigma = \mathrm{diag}(\{\sigma_\omega\})$$

- The RIP holds for (R '08)

$$M \;\gtrsim\; S \log^5 N$$

  Note that this result is *universal*
- Both the random sampling and the flat Fourier transform are needed for universality

Randomizing the phase



Dynamic Sparse Recovery

Streaming sparse recovery

- Solving an optimization program like

$$\min_x \ \tau \|x\|_{\ell_1} + \frac{1}{2}\|\Phi x - y\|_2^2$$

  can be costly

- We want to *update* the solution when the underlying signal changes slightly

Time-varying sparse signals

- Initial measurements. Observe

$$y_0 = \Phi x_0 + e_0$$

- Initial reconstruction. Solve

$$\min_x \ \tau\|x\|_{\ell_1} + \frac{1}{2}\|\Phi x - y_0\|_2^2$$

- A new set of measurements arrives:

$$y_1 = \Phi x_1 + e_1$$

- Reconstruct again using ℓ_1-min:

$$\min_x \ \tau\|x\|_{\ell_1} + \frac{1}{2}\|\Phi x - y_1\|_2^2$$

- We can gradually move from the first solution to the second solution using *homotopy*

$$\min \ \tau\|x\|_{\ell_1} + (1-\epsilon)\frac{1}{2}\|\Phi x - y_0\|_2^2 + \epsilon\frac{1}{2}\|\Phi x - y_1\|_2^2$$

Take ϵ from $0 \to 1$

Update direction

$$\min \ \tau\|x\|_{\ell_1} + \frac{1-\epsilon}{2}\|\Phi x - y_{\text{old}}\|_2^2 + \frac{\epsilon}{2}\|\Phi x - y_{\text{new}}\|_2^2$$

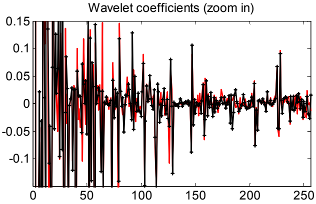
- Path from old solution to new solution is *piecewise linear*
- Optimality conditions for fixed ϵ:

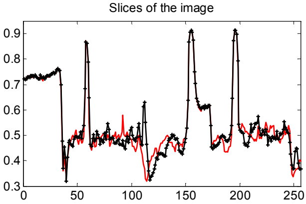
$$\begin{aligned} \Phi_\Gamma^T(\Phi x - (1-\epsilon)y_{\text{old}} - \epsilon y_{\text{new}}) &= -\tau \operatorname{sign} x_\Gamma \\ \|\Phi_{\Gamma^c}^T(\Phi x - (1-\epsilon)y_{\text{old}} - \epsilon y_{\text{new}})\|_\infty &< \tau \end{aligned}$$

Γ = active support

- Update direction:

$$\partial x = \begin{cases} -(\Phi_\Gamma^T\Phi_\Gamma)^{-1}(y_{\text{old}} - y_{\text{new}}) & \text{on } \Gamma \\ 0 & \text{off } \Gamma \end{cases}$$

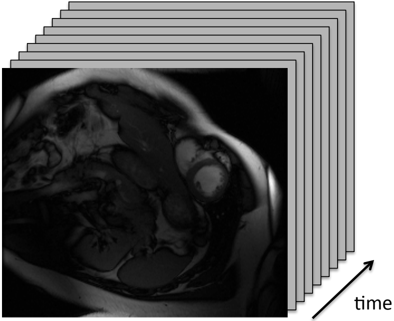




Average number of applications of Φ or Φ^T:
DynamicX 26.2, GPSR-BB: 92.24, FPC_AS: 90.9

Reconstructing time-varying images

We want to acquire a “data cube” X_0 (a time series of 2D images)



The structure across time is different than that across space...

Motion-compensated reconstruction

- Regularize in space using sparsity, regularize in time using a motion model, $X_{k+1} \approx M_k(X_k)$

$$\min_X \sum_k \left(\|\Phi_k X_k - Y_k\|_F^2 + \tau \, \mathrm{Sparsity}(X_k) + \eta \|M_k X_k - X_{k+1}\|_2^2 \right)$$

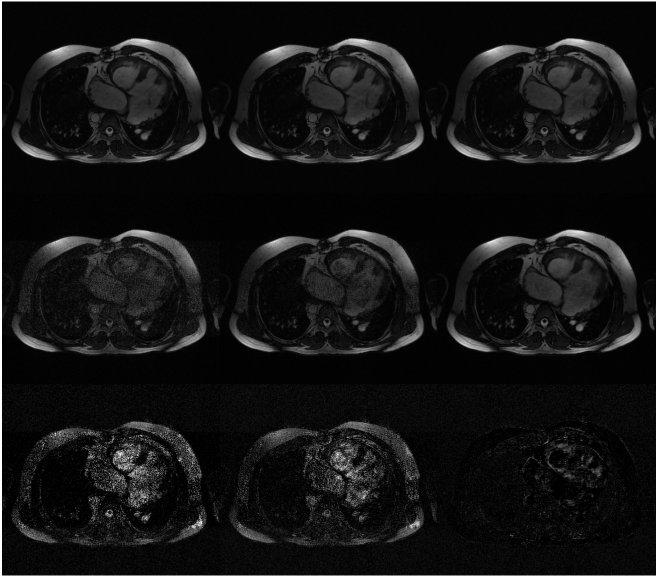
- Given the image sequence $\{X_k\}$, we can use standard techniques from video coding (block matching, local phase etc.) to estimate the motion operators M_k
- Strategy:
  - Reconstruct a "smoothed" version of $\{X_k\}$
  - Estimate the motion from this smoothed version
  - Reconstruct a more accurate version using motion compensation
  - Repeat (if desired) ...

Single frame of reconstruction

original

reconstruct

error



no motion high-freq penalty motion comp.

Questions?

jrom@ece.gatech.edu