An Introduction to Compressive Sensing and its Applications

Justin Romberg

Georgia Tech, School of ECE

Dutch-Flemish Numerical Analysis Conference

October 5, 2011
Woudschoten, Zeist, Netherlands

Acquisition as linear algebra

samples M

$$\underbrace{\begin{bmatrix} y \end{bmatrix}}_{\text{data}} = \underbrace{\begin{bmatrix} \quad \Phi \quad \end{bmatrix}}_{\text{acquisition system}} \underbrace{\begin{bmatrix} x \end{bmatrix}}_{\text{unknown signal/image}}$$

resolution/ bandwidth N

- 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"

Signal processing trends

DSP: sample first, ask questions later

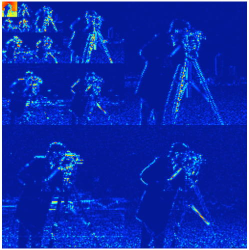
Explosion in sensor technology/ubiquity has caused two trends:

- Physical capabilities of hardware are being stressed, increasing speed/resolution becoming *expensive*
  - gigahertz+ analog-to-digital conversion
  - accelerated MRI
  - industrial imaging
- Deluge of data
  - camera arrays and networks, multi-view target databases, streaming video...

Compressive Sensing: sample smarter, not faster

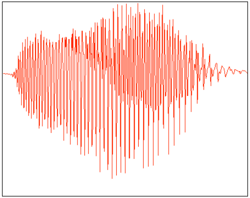
Sparsity/Compressibility

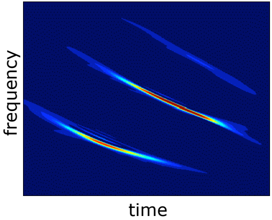
N
pixels



$S \ll N$
large
wavelet
coefficients

N
wideband
signal
samples





$S \ll N$
large
Gabor
coefficients

Wavelet approximation

Take 1% of *largest* coefficients, set the rest to zero (adaptive)

original

approximated

rel. error $= 0.031$

samples M

$$\underbrace{\begin{bmatrix} y \end{bmatrix}}_{\text{data}} = \underbrace{\begin{bmatrix} \quad & \Phi & \quad \end{bmatrix}}_{\text{acquisition system}} \underbrace{\begin{bmatrix} x \end{bmatrix}}_{\text{unknown signal/image}}$$

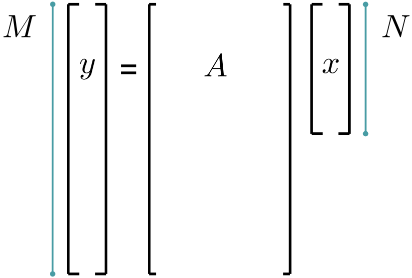
resolution/ bandwidth N

- If x is *sparse* and Φ is *diverse*, then these systems can be "inverted"

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)*

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

- A relaxed (convex) program

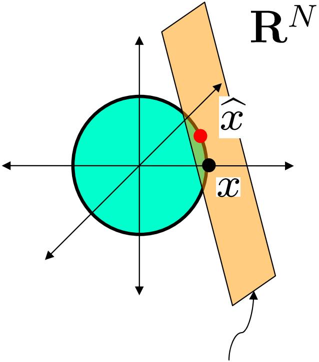
$$\min_x \; \|x\|_1 \quad \text{subject to} \quad \Phi x \approx y$$

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

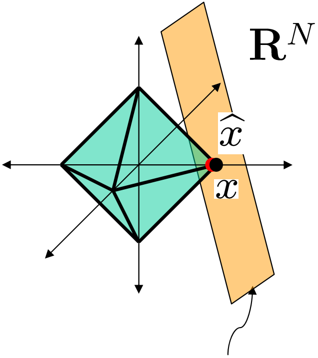
- This program is very tractable (linear program)

Graphical intuition for ℓ_1

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



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



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

Sparse recovery algorithms

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

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

  $\Psi =$ sparsifying transform, $\Phi =$ measurement system
  (need RIP for $\Phi\Psi$)

  Convex (linear) program, can relax for robustness to noise

  Performance has theoretical guarantees
- Other recovery methods include greedy algorithms and iterative thresholding schemes

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

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

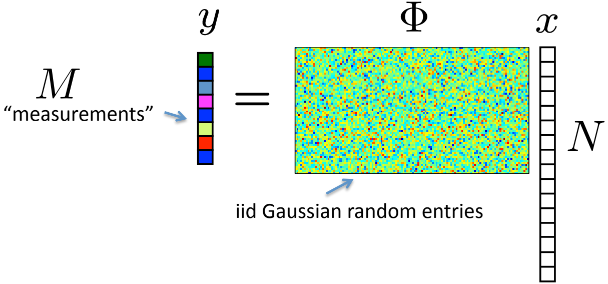
  will satisfy

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

  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 kind of matrices are restricted isometries?

- They are very hard to design, but they exist everywhere!

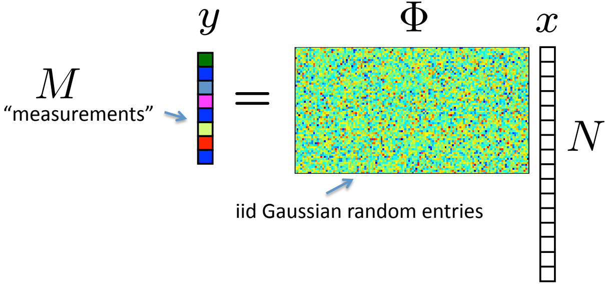


- For *any fixed* $x \in \mathbb{R}^N$, each measurement is

$$y_k \sim \mathrm{Normal}(0, \|x\|_2^2/M)$$

What kind of matrices are restricted isometries?

- They are very hard to design, but they exist everywhere!



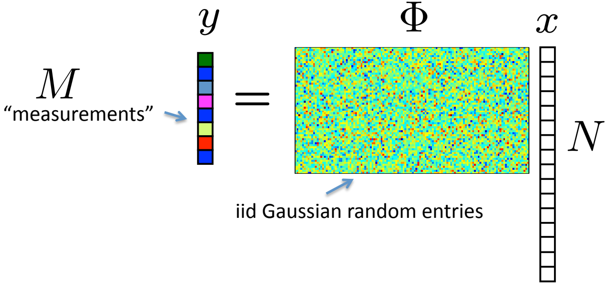
- For *any fixed* $x \in \mathbb{R}^N$, we have

$$\mathrm{E}[\|\Phi x\|_2^2] = \|x\|_2^2$$

the mean of the measurement energy is exactly $\|x\|_2^2$

What kind of matrices are restricted isometries?

- They are very hard to design, but they exist everywhere!

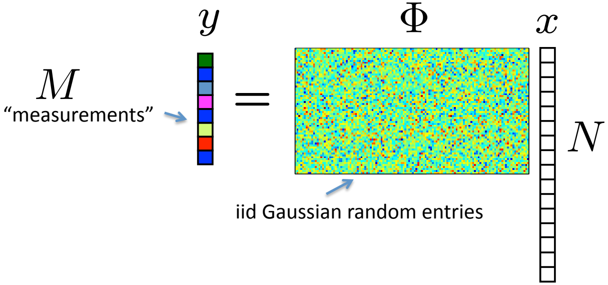


- For *any fixed* $x \in \mathbb{R}^N$, we have

$$\mathrm{P}\left\{\left|\|\Phi x\|_2^2 - \|x\|_2^2\right| < \delta\|x\|_2^2\right\} \;\geq\; 1 - e^{-M\delta^2/4}$$

What kind of matrices are restricted isometries?

- They are very hard to design, but they exist everywhere!



- For *all $2S$-sparse* $x \in \mathbb{R}^N$, we have

$$\mathrm{P}\left\{\max_x \left|\|\Phi x\|_2^2 - \|x\|_2^2\right| < \delta\|x\|_2^2\right\} \;\geq\; 1 - e^{c\cdot S\log(N/S)}e^{-M\delta^2/4}$$

So we can make this probability close to 1 by taking

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

What other types of matrices are restricted isometries?

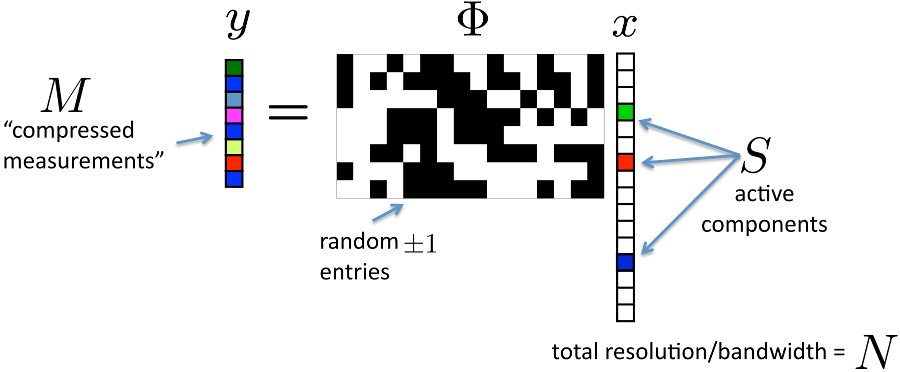
Four general frameworks:

- Random matrices (iid entries)
- Random subsampling
- Random convolution
- (Randomly modulated integration — we'll skip this today)

Note the role of randomness in all of these approaches

Slogan: *random projections keep sparse signal separated*

Random matrices (iid entries)

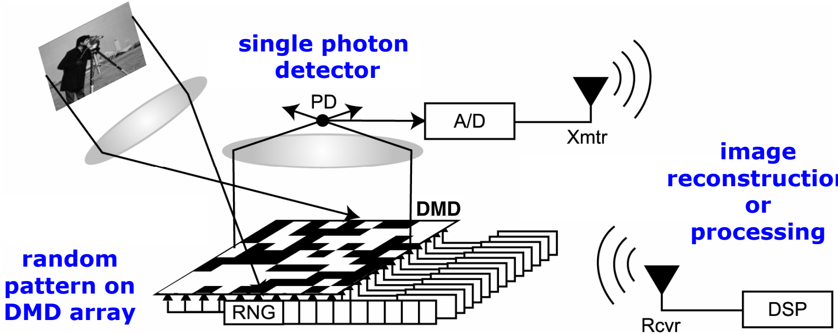


- *Random matrices* are provably efficient
- We can recover S-sparse x from

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

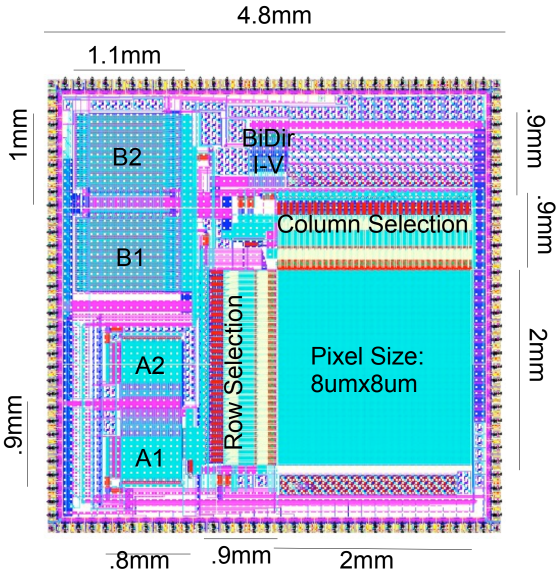
measurements

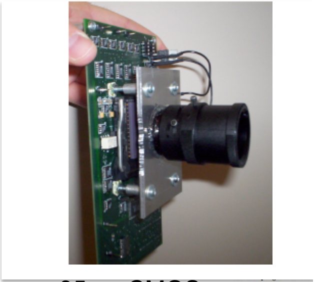
Rice single pixel camera



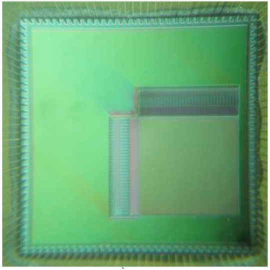
(Duarte, Davenport, Takhar, Laska, Sun, Kelly, Baraniuk '08)

Georgia Tech analog imager

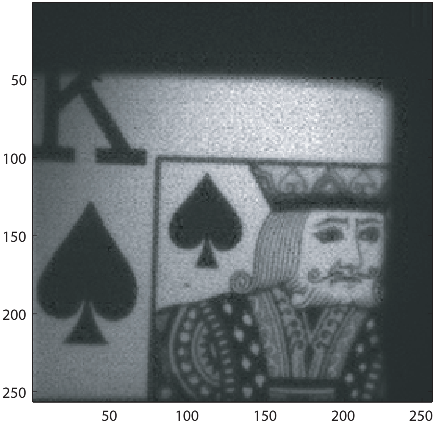




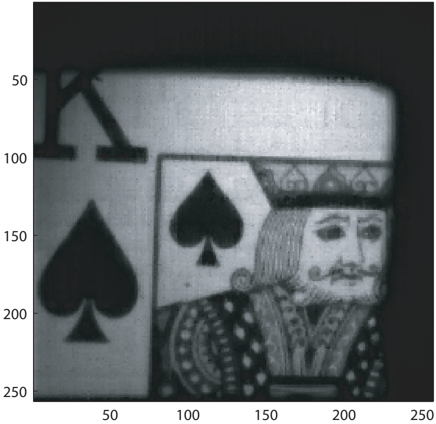
.35um CMOS process



Compressive sensing acquisition



10k DCT measurements

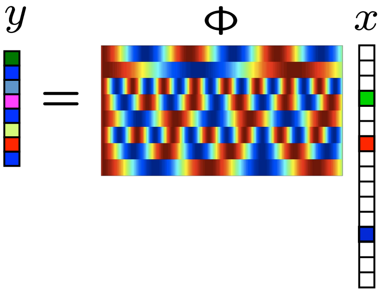


10k random measurements

(Robucci, Chiu, Gray, R, Hasler '09)

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

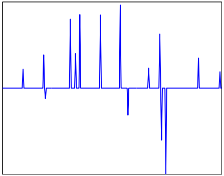
$$\mu \;=\; \sqrt{N} \max_{i,j} |(U^T \Psi)_{ij}|$$

is the *coherence*

Examples of incoherence

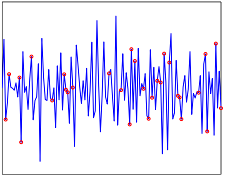
- Signal is sparse in time domain, sampled in Fourier domain

time domain $x(t)$



S nonzero components

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

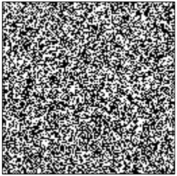


measure m samples

- Signal is sparse in wavelet domain, measured with noiselets

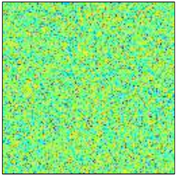
(Coifman et al '01)

example noiselet

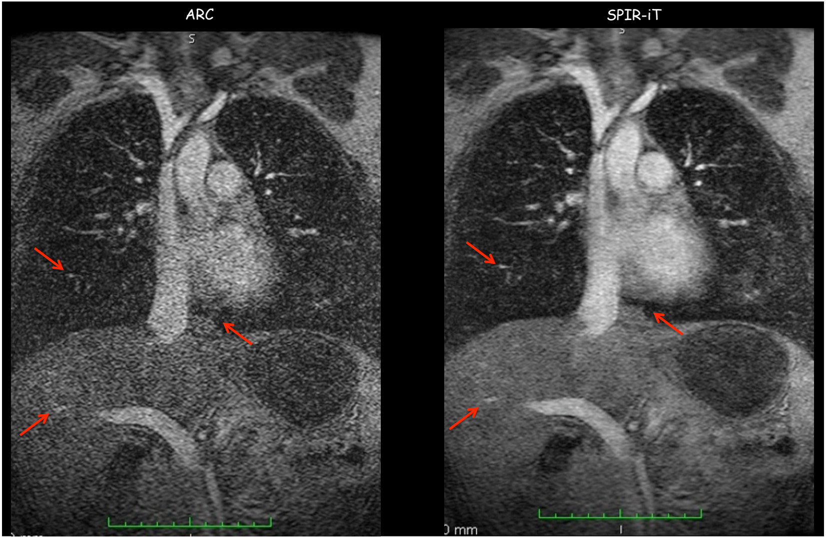


wavelet domain

noiselet domain



Accelerated MRI



(Lustig et al. '08)

Empirical processes and structured random matrices

- For matrices with this type of *structured randomness*, we simply do not have enough concentration to establish

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

  "the easy way"

- Re-write the RIP as a the *supremum of a random process*

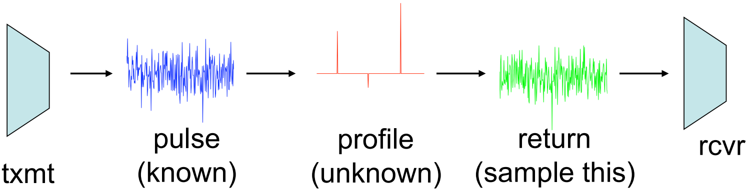
$$\sup_x |G(x)| = \sup_x |x^*\Phi^*\Phi x - x^*x| \;\leq\; \delta$$

  where the sup is taken over all $2S$-sparse signals

- Estimate this sup using tools from probability theory (e.g. the Dudley inequality) — approach pioneered by Rudelson and Vershynin

Random convolution

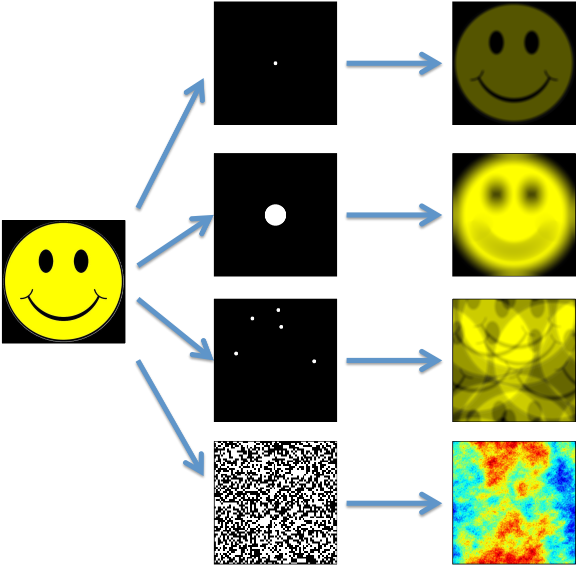
- Many *active imaging* systems measure a pulse convolved with a *reflectivity profile* (Green's function)



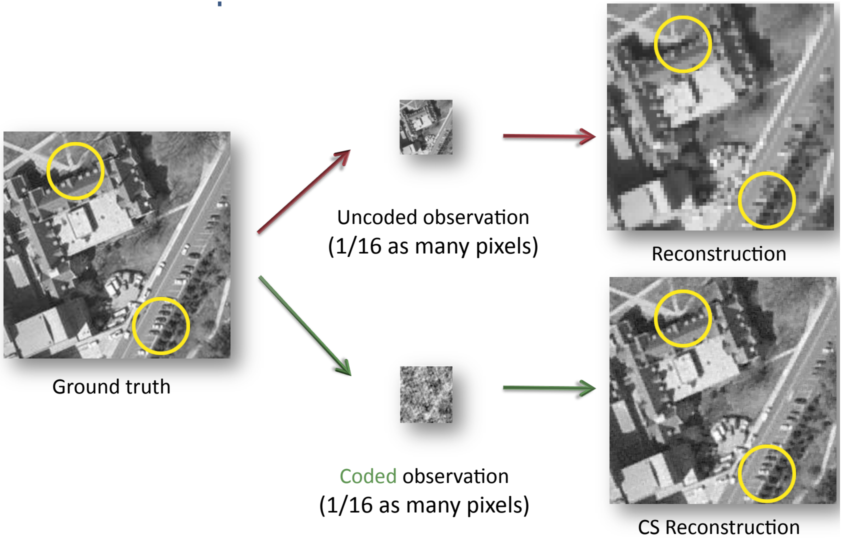
- Applications include:
  - radar imaging
  - sonar imaging
  - seismic exploration
  - channel estimation for communications
  - super-resolved imaging
- Using a *random pulse* = compressive sampling

  (Tropp et al. '06, R '08, Herman et al. '08, Haupt et al. '09, Rauhut '09)

Coded aperture imaging



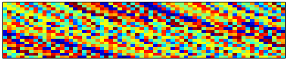
Super-resolved imaging



(Marcia and Willet '08)

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)

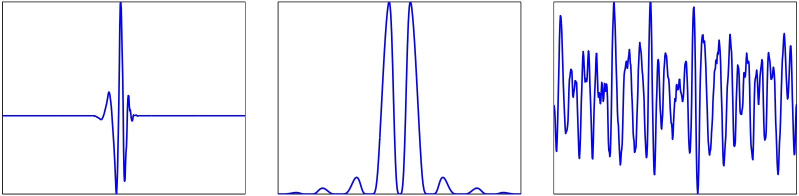
$$M \;\gtrsim\; S \log^5 N$$

  Note that this result is *universal*

- Both the random sampling and the flat Fourier transform are needed for universality

Randomizing the phase

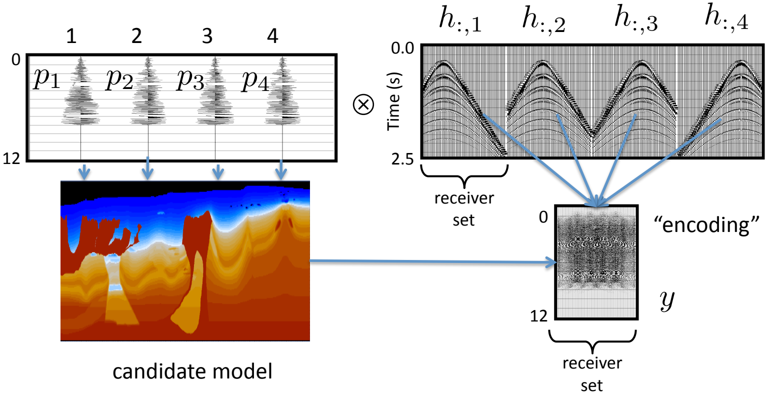
local in time　　local in freq　　*not* local in M



sample here

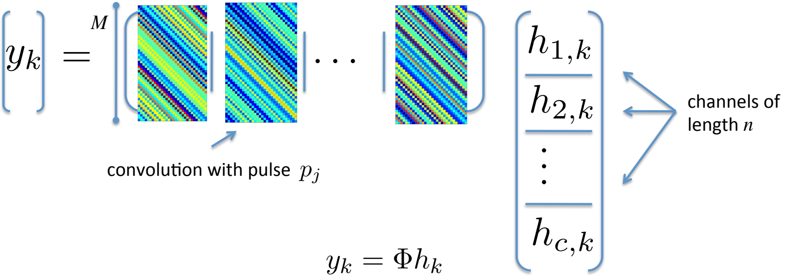
Seismic forward modeling

- Run a single simulation with all of the sources activated simultaneously with random waveforms
- The channel responses interfere with one another, but the randomness "codes" them in such a way that they can be separated later



Related work: Herrmann et. al '09

Restricted isometries for multichannel systems



$$y_k = \Phi h_k$$

- With each of the pulses as iid Gaussian sequences, Φ obeys

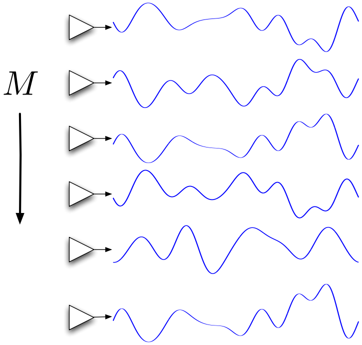
$$(1-\delta)\|h\|^2 \;\le\; \|\Phi h\|_2^2 \;\le\; (1+\delta)\|h\|_2^2 \quad \forall 2S\text{-sparse } h \in \mathbb{R}^{nc}$$

  when (R and Neelamani '09)

$$M \;\gtrsim\; S \cdot \log^5(nc) \;+\; n$$

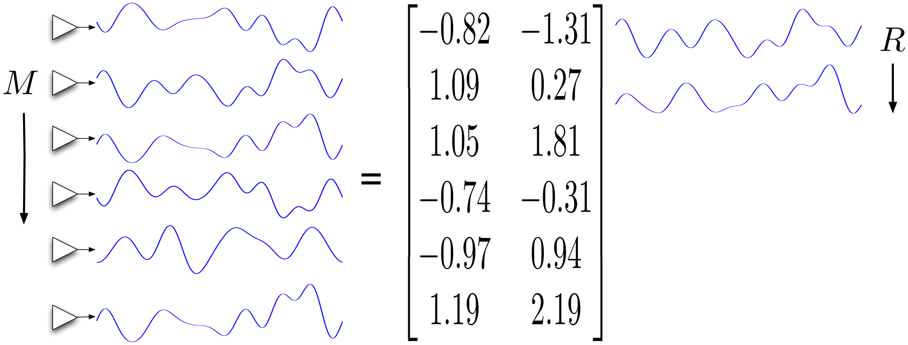
- **Consequence:** we can separate the channels using short random pulses (using ℓ_1 min or other sparse recovery algorithms)

Sampling correlated signals



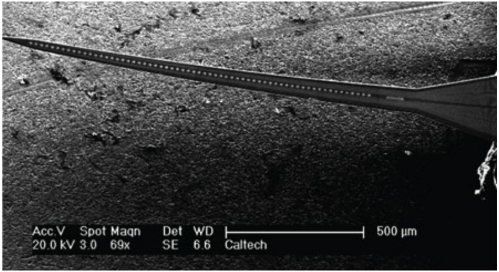
- Goal: acquire an *ensemble* of M signals
- Bandlimited to $W/2$
- "Correlated" $\rightarrow$ M signals are $\approx$ linear combinations of R signals

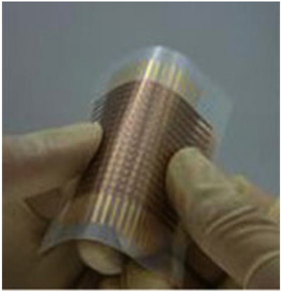
Sampling correlated signals

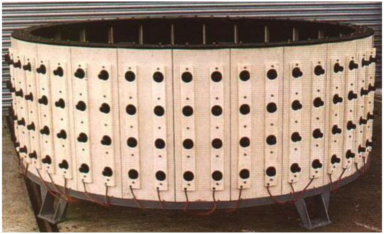


- Goal: acquire an *ensemble* of M signals
- Bandlimited to $W/2$
- "Correlated" $\rightarrow$ M signals are $\approx$ linear combinations of R signals

Sensor arrays







Low-rank matrix recovery

- Given P *linear samples* of a matrix,

$$y = \mathcal{A}(\mathbf{X}_0), \quad y \in \mathbb{R}^P, \quad \mathbf{X}_0 \in \mathbb{R}^{M \times W}$$

we solve

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

where $\|\mathbf{X}\|_*$ is the nuclear norm: the sum of the singular values of $\mathbf{X}$.

Low-rank matrix recovery

- Given P *linear samples* of a matrix,

$$y = \mathcal{A}(\mathbf{X}_0), \quad y \in \mathbb{R}^P, \quad \mathbf{X}_0 \in \mathbb{R}^{M \times W}$$

  we solve

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

  where $\|\mathbf{X}\|_*$ is the nuclear norm: the sum of the singular values of $\mathbf{X}$.

- If $\mathbf{X}_0$ is rank-R and $\mathcal{A}$ obeys the mRIP:

$$(1-\delta)\|\mathbf{X}\|_F^2 \ \le \ \|\mathcal{A}(\mathbf{X})\|_2^2 \ \le \ (1+\delta)\|\mathbf{X}\|_F^2 \quad \forall \text{ rank-}2R \ \mathbf{X},$$

  then we can stably recover $\mathbf{X}_0$ from y. (Recht et. al '07)

Low-rank matrix recovery

- Given P *linear samples* of a matrix,

$$y = \mathcal{A}(\mathbf{X}_0), \quad y \in \mathbb{R}^P, \quad \mathbf{X}_0 \in \mathbb{R}^{M \times W}$$

  we solve

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

  where $\|\mathbf{X}\|_*$ is the nuclear norm: the sum of the singular values of $\mathbf{X}$.

- If $\mathbf{X}_0$ is rank-R and $\mathcal{A}$ obeys the mRIP:

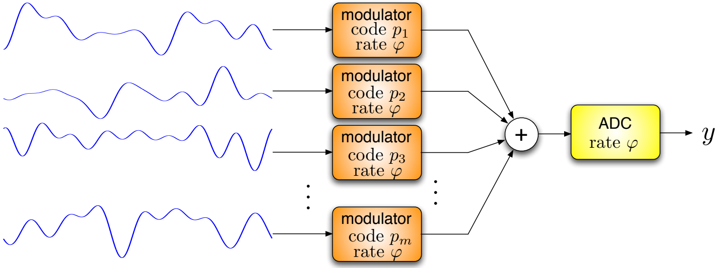
$$(1-\delta)\|\mathbf{X}\|_F^2 \ \le \ \|\mathcal{A}(\mathbf{X})\|_2^2 \ \le \ (1+\delta)\|\mathbf{X}\|_F^2 \quad \forall \text{ rank-}2R \ \mathbf{X},$$

  then we can stably recover $\mathbf{X}_0$ from y. (Recht et. al '07)

- An 'generic' (iid random) sampler $\mathcal{A}$ (stably) recovers $\mathbf{X}_0$ from y when

$$\begin{aligned} \#\text{samples} &\gtrsim \ R \cdot \max(M, W) \\ &\gtrsim RW \quad \text{(in our case)} \end{aligned}$$

CS for correlated signals: modulated multiplexing



- If the signals are spread out uniformly in time, then the ADC and modulators can run at rate

$$\varphi \gtrsim RW \log^{3/2}(MW)$$

- Requires signals to be (mildly) spread out in time

Summary

- Main message of CS:

  We can recover an S-sparse signal in $\mathbb{R}^N$ from
  $$\sim S \cdot \log N \text{ measurements}$$

  We can recover a rank-R matrix in $\mathbb{R}^{M \times W}$ from
  $$\sim R \cdot \max(M, W) \text{ measurements}$$

- Random matrices (iid entries)
  - easy to analyze, optimal bounds
  - universal
  - hard to implement and compute with
- Structured random matrices (random sampling, random convolution)
  - structured, and so computationally efficient
  - physical
  - much harder to analyze, bound with extra log-factors

Agenda

Friday: Mathematical proof (analysis!)

We will prove two fundamental results in compressive sensing.

Not much background is required:

basic probability (Gaussian random variables, Markov inequality,
union bound/Boole inequality,...)

basic linear algebra (operator norm, singular value,...)

basic geometry (triangle inequality, covering a set, ...)

Agenda

On Friday, we will prove from top to bottom two things:

- That an $M \times N$ iid Gaussian random matrix satisfies

$$(1-\delta)\|x\|_2^2 \;\leq\; \|\Phi x\|_2^2 \;\leq\; (1+\delta)\|x\|_2^2 \quad \forall\ 3S\text{-sparse } x \tag{1}$$

  with (extraordinarily) high probability when

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

- Suppose an $M \times N$ matrix Φ obeys (1). Let x_0 be an S-sparse vector, and suppose we observe $y = \Phi x_0$. Given y, the solution to

$$\min_x \;\|x\|_{\ell_1} \quad \text{such that} \quad \Phi x_0 = y$$

  is *exactly* x_0.