An Introduction to Sparse Representations and Compressive Sensing

Part II

Paulo Gonçalves

CPE Lyon - 4ETI - Cours Semi-Optionnel

Méthodes Avancées pour le Traitement des Signaux

2014

Objectifs

Part I

- The motivation and the rationale of **sparse representations**
- Linear decompositions (Fourier, DCT, wavelets. . .)
- Sparsity and compression, estimation and other inverse problems
- (X-lets)

Part II

- **Compressive sensing** : The main idea
- Linear algebra formulation (an invertible ill-posed problem)
- Projection on Random Matrices
- Some striking examples

Bibliography

A wavelet tour of signal processing Stéphane Mallat. Academic Press, 1999

Ten Lectures on Wavelets Ingrid Daubechies. Siam, 1992

Compressive Sampling Emmanuel Candès. Int. Congress of Mathematics, 3, pp. 1433-1452, Madrid, Spain, 2006

Compressive sensing Richard Baraniuk. IEEE Signal Processing Magazine, 24(4), pp. 118-121, July 2007

Imaging via compressive sampling Justin Romberg. IEEE Signal Processing Magazine, 25(2), pp. 14 - 20, March 2008

Introduction to compressed sensing M. Davenport, M. Duarte, Y. Eldar, and G. Kutyniok. Chapter in Compressed Sensing : Theory and Applications, Cambridge University Press, 2012

Compressive sensing M. Fornasier and H. Rauhut. Chapter in Part 2 of the Handbook of Mathematical Methods in Imaging (O. Scherzer Ed.), Springer, 2011

Sparsity-Aware Learning and Compressed Sensing : An Overview S. Theodoridis, Y. Kopsinis, K. Slavakis, arXiv :1211.5231

http ://dsp.rice.edu/cs An updated list of publications related to compressive sensing

A survey of Compressive Sensing and Applications Lecture by Justin Romberg, Master 2, Computer Sc. Dept. ENS Lyon. 2012.

Signal processing trends

DSP: sample first, ask questions later

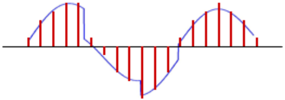
Explosion in sensor technology/ubiquity has caused two trends:

- 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...

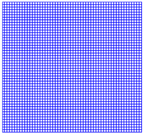
Compressive Sensing: sample smarter, not faster

Classical data acquisition

- *Shannon-Nyquist sampling theorem* (Fundamental Theorem of DSP): "if you sample at twice the bandwidth, you can perfectly reconstruct the data"



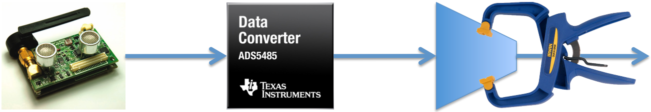
time



space

- Counterpart for "indirect imaging" (MRI, radar):
  Resolution is determined by bandwidth

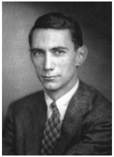
Sense, sample, process...



sensor “fast” ADC data compression

Compressive sensing (CS)

- Shannon/Nyquist theorem is *pessimistic*
  - 2×bandwidth is the worst-case sampling rate — holds uniformly for *any* bandlimited data
  - sparsity/compressibility is irrelevant
  - Shannon sampling based on a linear model, compression based on a nonlinear model

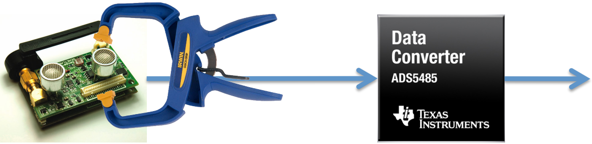


Shannon

- Compressive sensing
  - new sampling theory that *leverages compressibility*
  - key roles played by new *uncertainty principles* and *randomness*

Heisenberg

Compressive sensing



"compressive" sensor

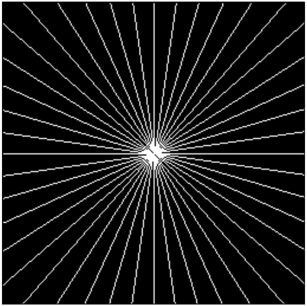
"slow" ADC

- Essential idea:
  "pre-coding" the signal in analog makes it "easier" to acquire
- Reduce power consumption, hardware complexity, acquisition time

A simple underdetermined inverse problem

Observe a subset Ω of the 2D discrete Fourier plane

phantom (hidden)



white star $=$ sample locations

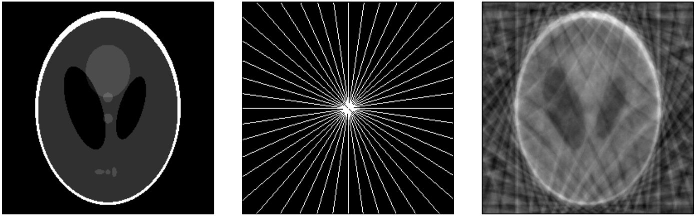
$N := 512^2 = 262,144$ pixel image
observations on 22 radial lines, $10,486$ samples, $\approx 4\%$ coverage

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

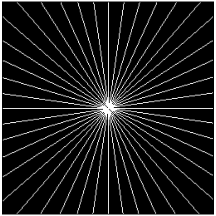
Find an image that

- 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

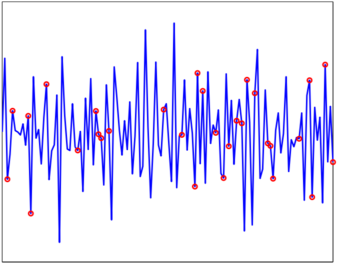
$g^* =$ original

perfect reconstruction

Sampling a superposition of sinusoids

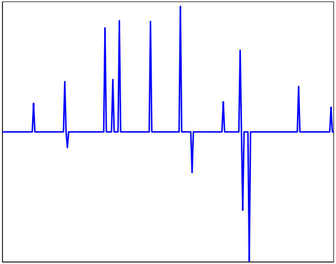
We take M samples of a superposition of S sinusoids:

Time domain $x_0(t)$



Measure M samples
(red circles = samples)

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

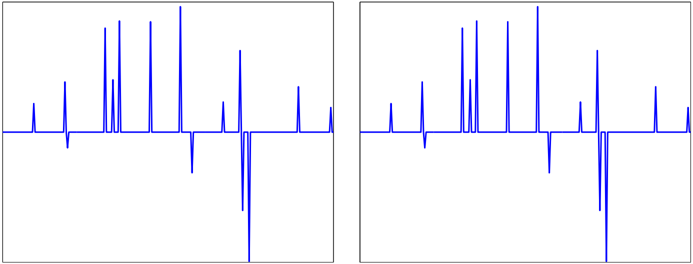


S nonzero components

Sampling a superposition of sinusoids

Reconstruct by solving

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



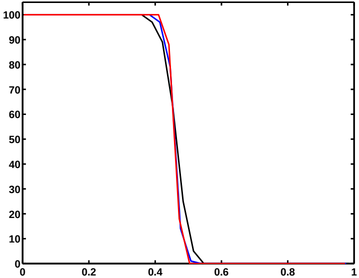
original $\hat{x}_0$, $S = 15$

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

% success



S/M

- 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

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

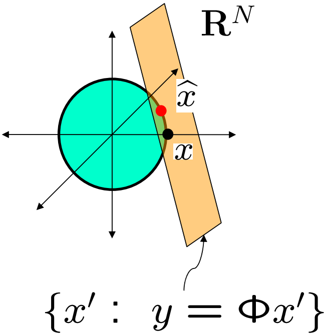
- Take time-domain samples (measurements) $y_m = x_0(t_m)$
- Solve

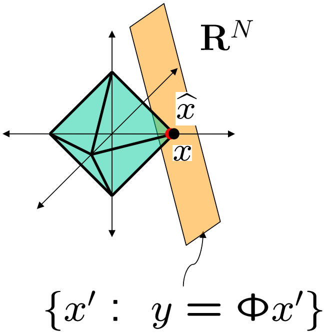
$$\min_x \|\hat{x}\|_{\ell_1} \quad \text{subject to} \quad x(t_m) = y_m, \quad m = 1, \ldots, M$$

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

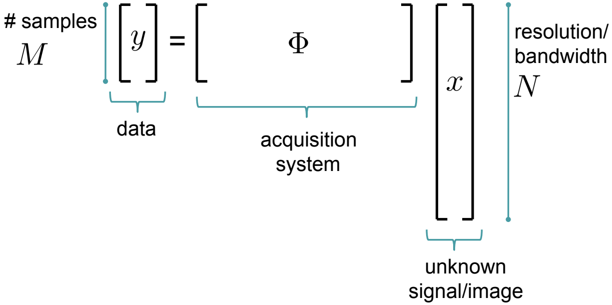
Graphical intuition for ℓ_1

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





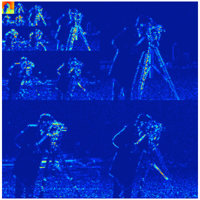
Acquisition as linear algebra



- 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"

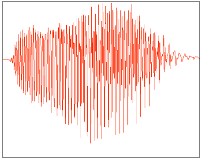
Sparsity/Compressibility

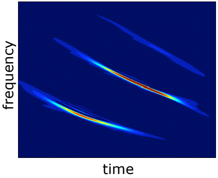
N
pixels



$S \ll N$
large
wavelet
coefficients

N
wideband
signal
samples





$S \ll N$
large
Gabor
coefficients

Wavelet approximation

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

original

approximated

rel. error = 0.031

Linear measurements

- Instead of samples, take *linear measurements* of signal/image x_0

$$y_1 = \langle x_0, \phi_1 \rangle, \quad y_2 = \langle x_0, \phi_2 \rangle, \quad \ldots, y_M = \langle x_0, \phi_K \rangle$$

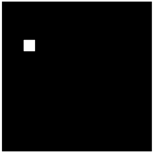
$$y = \Phi x_0$$

- Equivalent to transform-domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: pixels

$y_m = \Bigg\langle$

,



$\Bigg\rangle$

Linear measurements

- Instead of samples, take *linear measurements* of signal/image x_0

$$y_1 = \langle x_0, \phi_1 \rangle, \quad y_2 = \langle x_0, \phi_2 \rangle, \quad \ldots, y_M = \langle x_0, \phi_K \rangle$$

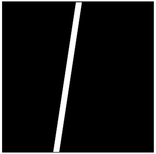
$$y = \Phi x_0$$

- Equivalent to transform-domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: line integrals (tomography)

$$y_m = \Bigg\langle$$

,



$$\Bigg\rangle$$

Linear measurements

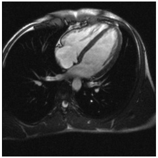
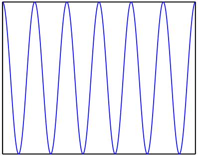
- Instead of samples, take *linear measurements* of signal/image x_0

$$y_1 = \langle x_0, \phi_1 \rangle, \quad y_2 = \langle x_0, \phi_2 \rangle, \quad \ldots, y_M = \langle x_0, \phi_K \rangle$$

$$y = \Phi x_0$$

- Equivalent to transform-domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: sinusoids (MRI)

$y_m = \Big\langle$  ,  $\Big\rangle$

Linear measurements

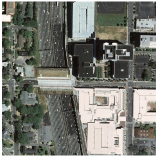
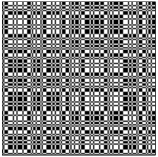
- Instead of samples, take *linear measurements* of signal/image x_0

$$y_1 = \langle x_0, \phi_1 \rangle, \quad y_2 = \langle x_0, \phi_2 \rangle, \quad \ldots, y_M = \langle x_0, \phi_K \rangle$$

$$y = \Phi x_0$$

- Equivalent to transform-domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: coded imaging

$y_m = \Big\langle$  ,  $\Big\rangle$

Linear measurements

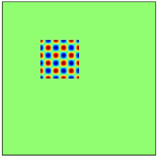
- Instead of samples, take *linear measurements* of signal/image x_0

$$y_1 = \langle x_0, \phi_1 \rangle, \quad y_2 = \langle x_0, \phi_2 \rangle, \quad \ldots, y_M = \langle x_0, \phi_K \rangle$$

$$y = \Phi x_0$$

- Equivalent to transform-domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: DCT ?

$y_m = \Big\langle$  ,  $\Big\rangle$

Linear measurements

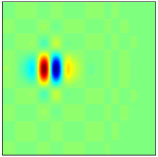
- Instead of samples, take *linear measurements* of signal/image x_0

$$y_1 = \langle x_0, \phi_1 \rangle, \quad y_2 = \langle x_0, \phi_2 \rangle, \quad \ldots, y_M = \langle x_0, \phi_K \rangle$$

$$y = \Phi x_0$$

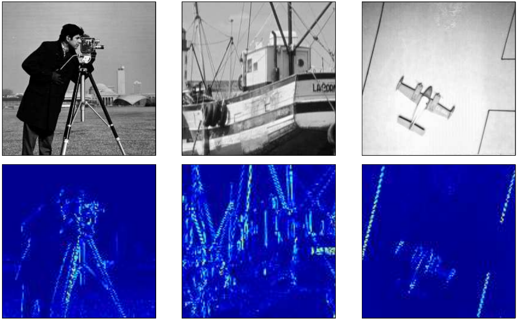
- Equivalent to transform-domain sampling,
  $\{\phi_m\}$ = basis functions

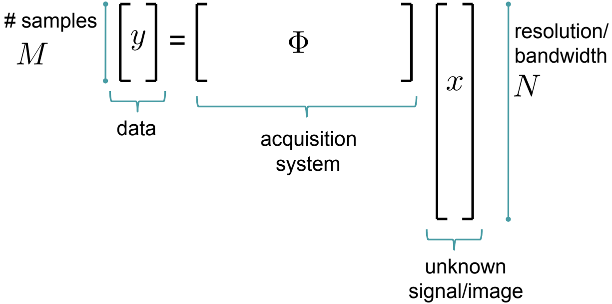
- Example: wavelets ?

$y_m = \Big\langle$  ,  $\Big\rangle$

Sparsity and Linear Measurements

- Since x_0 is sparse in Ψ, why don't we measure $\langle x_0, \psi_k \rangle$? Why not sample images in the wavelet domain?
- We'd love to sample wavelet coeffs, but which ones?





- 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

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

- 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

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

- 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*

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

  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)

Sparse recovery algorithms

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

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

Ψ = sparsifying transform, Φ = measurement system (need RIP for $\Phi\Psi^T$)

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*
- **Theorem** (Candès, R, Tao '06)

  If we observe $y = \Phi x_0 + e$, with $\|e\|_2 \leq \epsilon$, the solution $\hat{x}$ to

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

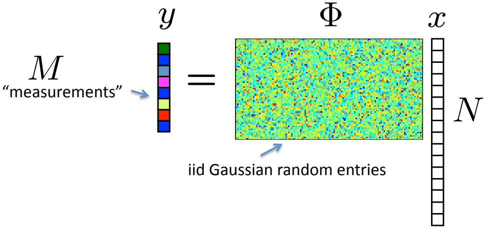
  will satisfy

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

  where
  - $x_{0,S}$ = S-term approximation of x_0
  - S is the largest value for which $\Phi\Psi^T$ 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!

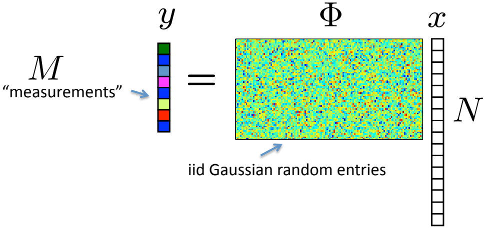


- For *any fixed* $x \in \mathbb{R}^N$, each measurement is

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

What kind of matrices are restricted isometries?

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



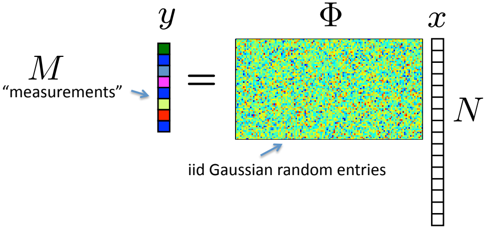
- For *any fixed* $x \in \mathbb{R}^N$, we have

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

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!

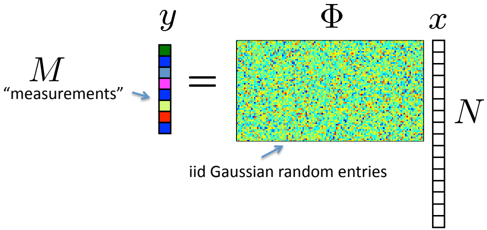


- For *any fixed* $x \in \mathbb{R}^N$, we have

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

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?

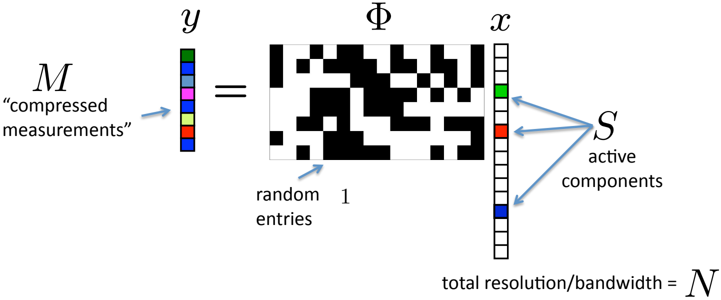
Four general frameworks:

- Random matrices (iid entries)
- Random subsampling
- Random convolution
- Randomly modulated integration

Note the role of randomness in all of these approaches

Slogan: *random projections keep sparse signal separated*

Random matrices (iid entries)

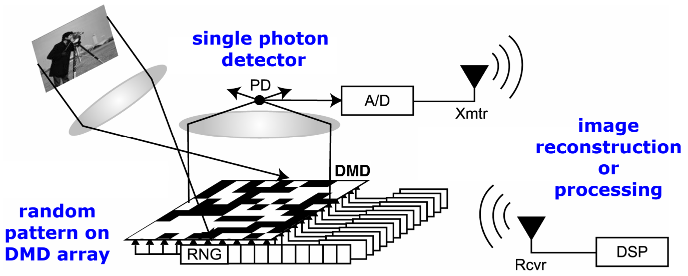


- *Random matrices* are provably efficient
- We can recover S-sparse x from

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

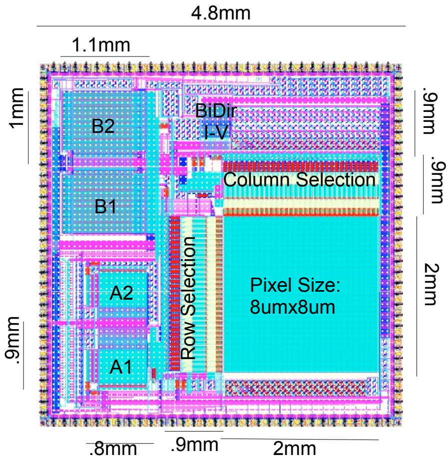
measurements

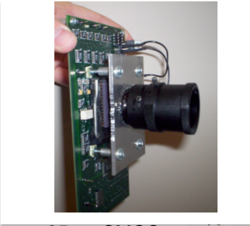
Rice single pixel camera



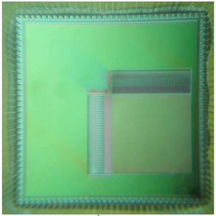
(Duarte, Davenport, Takhar, Laska, Sun, Kelly, Baraniuk '08)

Georgia Tech analog imager

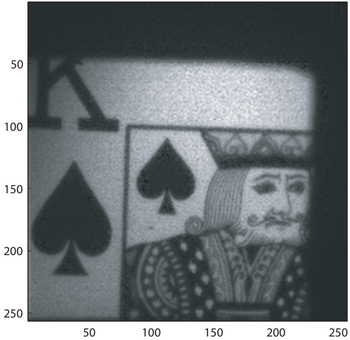




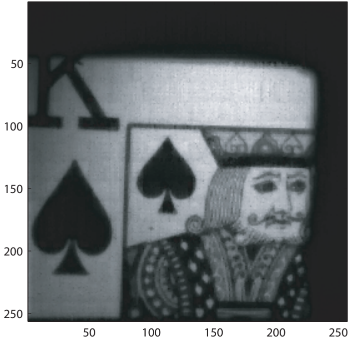
.35um CMOS process



Compressive sensing acquisition



10k DCT measurements

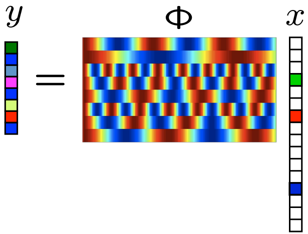


10k random measurements

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

Random matrices

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



Can recover S-sparse x from

(Rudelson and Vershynin '06, Candès and R '07)

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

measurements, where

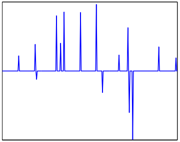
$$\mu = \sqrt{N} \max_{i,j} |(U^T \Psi)_{ij}|$$

is the *coherence*

Examples of incoherence

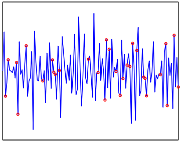
- Signal is sparse in time domain, sampled in Fourier domain

time domain $x(t)$



S nonzero components

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

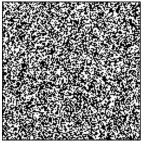


measure m samples

- Signal is sparse in wavelet domain, measured with noiselets

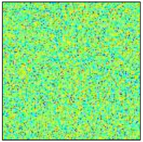
(Coifman et al '01)

example noiselet

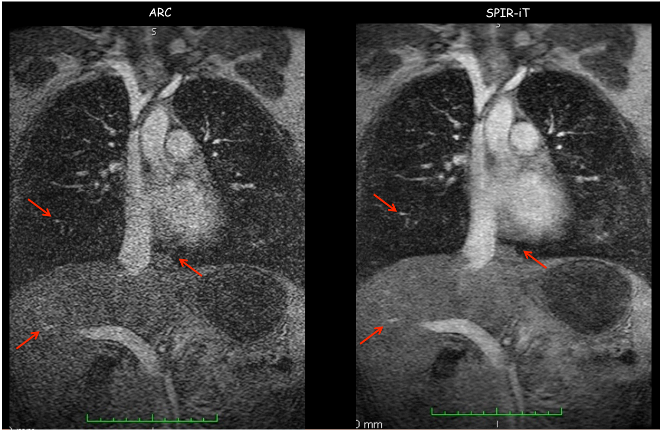


wavelet domain

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*

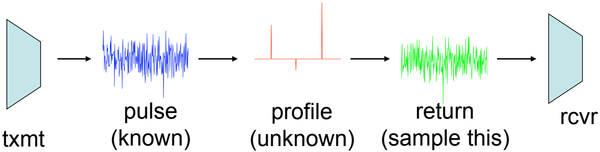
$$\sup_x |G(x)| = \sup_x |x^*\Phi^*\Phi x - x^*x| \;\;\leq\;\; \delta$$

  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

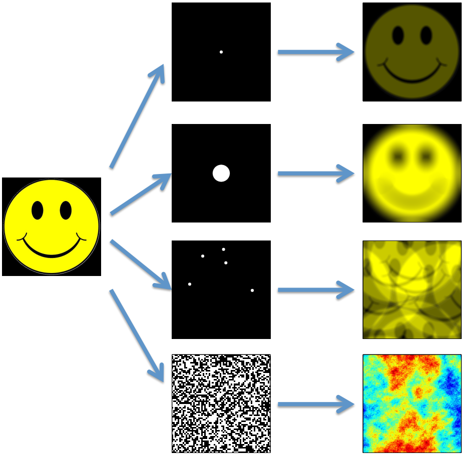
- Many *active imaging* systems measure a pulse convolved with a *reflectivity profile* (Green's function)



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

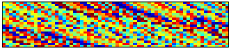
  (Tropp et al. '06, R '08, Herman et al. '08, Haupt et al. '09, Rauhut '09)

Coded aperture imaging



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



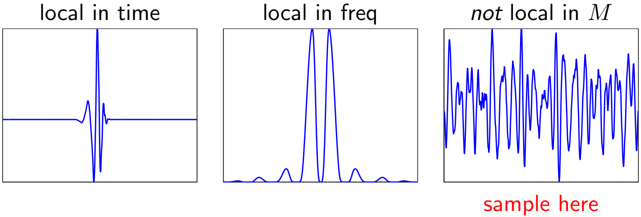
$$\Phi = R_\Omega \mathcal{F}^* \Sigma \mathcal{F}, \quad \Sigma = \mathrm{diag}(\{\sigma_\omega\})$$

- 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



Why is random convolution + subsampling universal?

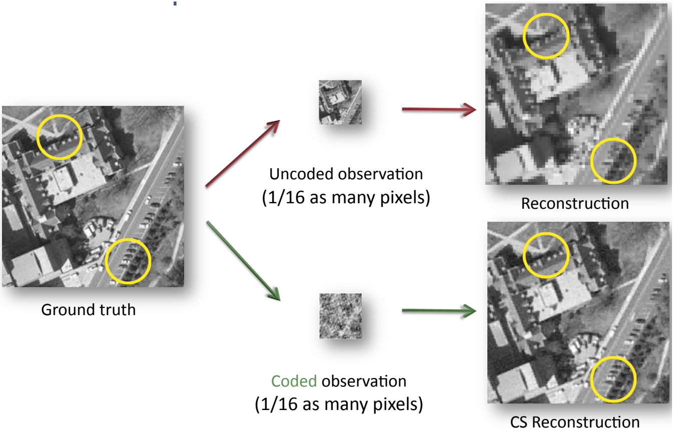
$$\begin{bmatrix} & & \\ & \mathcal{F} & \\ & & \\ & & \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} \begin{bmatrix} \hat{\psi}_1(\omega) & \hat{\psi}_2(\omega) & \cdots & \hat{\psi}_n(\omega) \end{bmatrix}$$

- One entry of $\Phi' = \Phi\hat{\Psi} = \mathcal{F}\Sigma\hat{\Psi}$:

$$\begin{aligned} \Phi'_{t,s} &= \sum_{\omega} e^{j2\pi\omega t}\sigma_\omega \hat{\psi}_s(\omega) \\ &= \sum_{\omega} \sigma'_\omega \hat{\psi}_s(\omega) \end{aligned}$$

- Size of each entry will be concentrated around $\|\hat{\psi}_s(\omega)\|_2 = 1$ *does not depend on the "shape" of* $\hat{\psi}_s(\omega)$

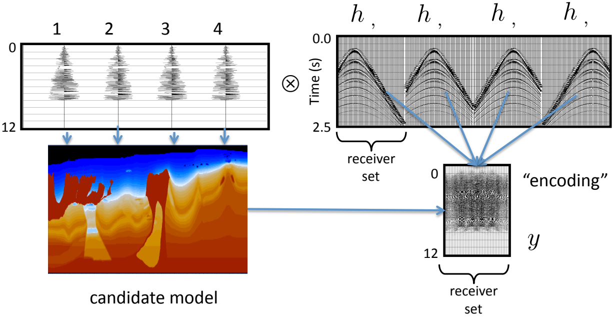
Super-resolved imaging



(Marcia and Willet '08)

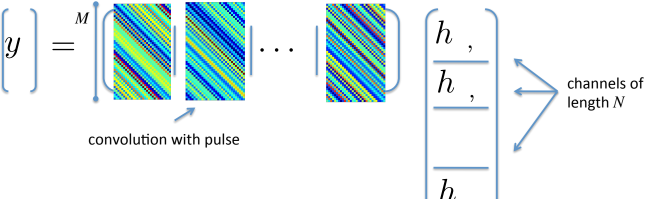
Seismic forward modeling

- 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

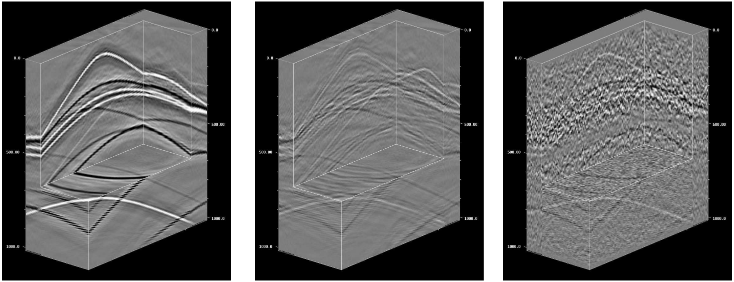
$$(1-\delta)\|h\|^2 \;\leq\; \|\Phi h\|_2^2 \;\leq\; (1+\delta)\|h\|_2^2 \quad \forall s\text{-sparse } h \in \mathbb{R}^{NC}$$

  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)

Seismic imaging simulation



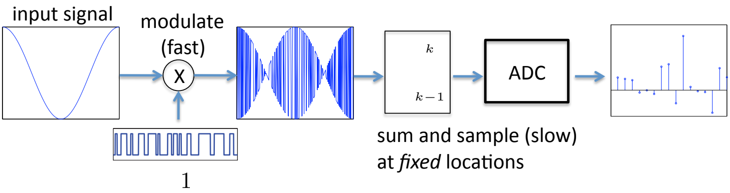
(a) Estimated (16x aster N = .6 d)

( ) Estimation error (igure 2 minus 5(a))

( ) Cross orrelation estimate

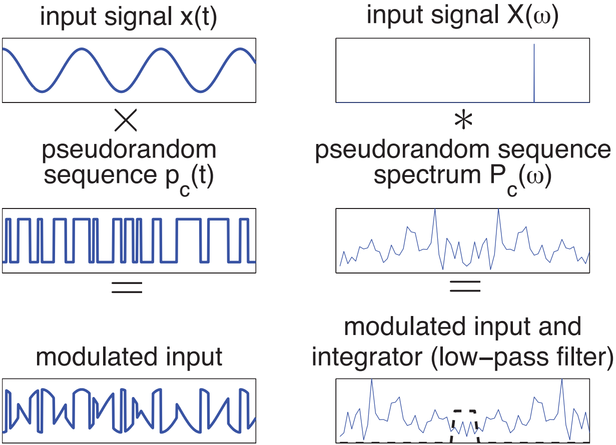
- Result produced with $16\times$ "compression" in the computations
- Can even take this example down to $32\times$

Randomly modulated integration

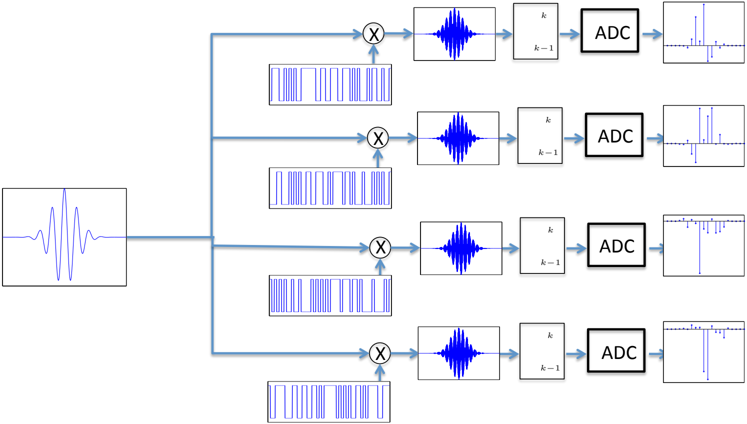


- Uses a standard "slow" ADC preceded by a "fast" binary mixing
- Mixing circuit much easier to build than a "fast" ADC
- In each sampling interval, the signal is summarized with a random sum
- Sample rate $\sim$ total *active* bandwidth

Random modulated integration in time and frequency

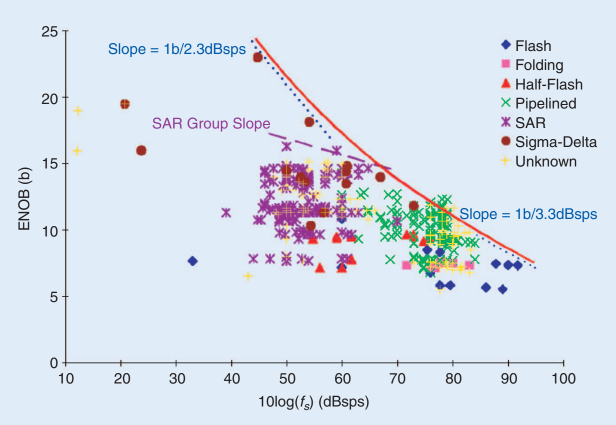


Multichannel modulated integration



This architecture is being implemented as part of DARPA's Analog-to-Information program

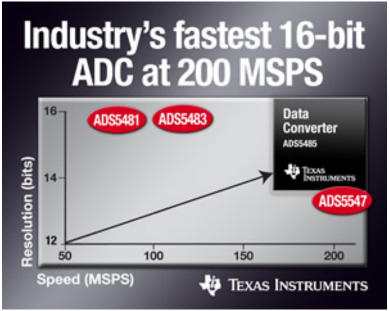
Analog-to-digital converter state-of-the-art



The bad news starts at 1 GHz (Le et al '05)

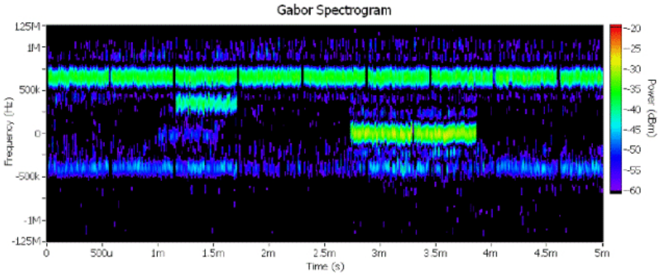
Analog-to-digital converter state-of-the-art

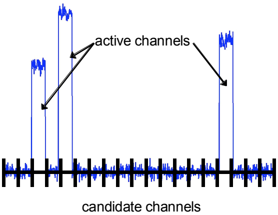
From 2008...



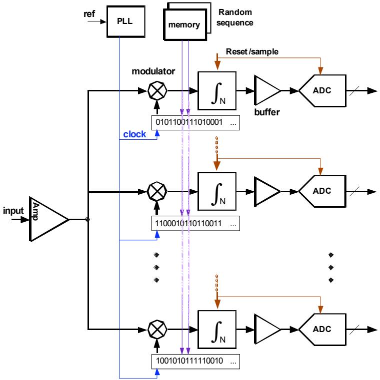
(Lots of RF signals have components in the 10s of gigahertz...)

Spectrally sparse RF signals



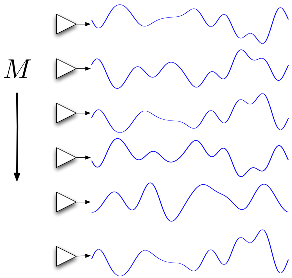


Randomly modulated integration receiver



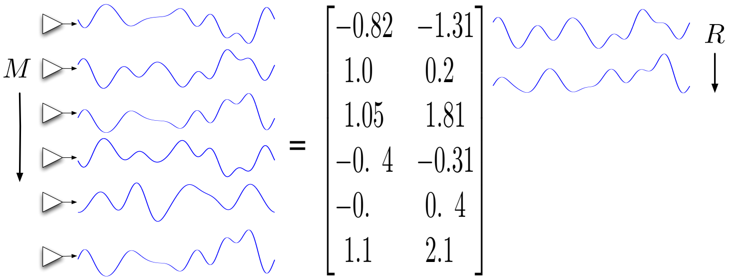
- Random demodulator being built at part of DARPA A2I program (Emami, Hoyos, Massoud)
- Multiple (8) channels, operating with different mixing sequences
- Effective BW/chan = 2.5 GHz
  Sample rate/chan = 50 MHz
- Applications: radar pulse detection, communications surveillance, geolocation

Sampling correlated signals



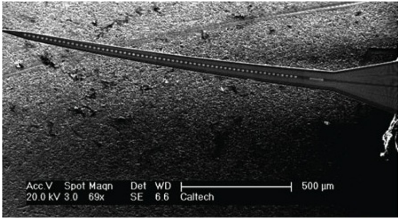
- Goal: acquire an *ensemble* of M signals
- Bandlimited to $W/2$
- "Correlated" $\rightarrow$ M signals are $\approx$ linear combinations of R signals

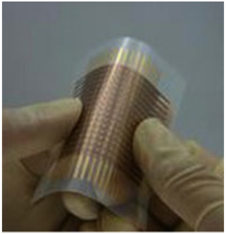
Sampling correlated signals

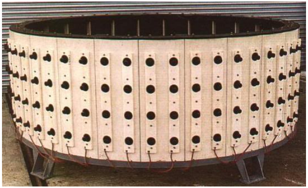


- Goal: acquire an *ensemble* of M signals
- Bandlimited to $W/2$
- "Correlated" $\rightarrow$ M signals are $\approx$ linear combinations of R signals

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:

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

  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

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

  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:

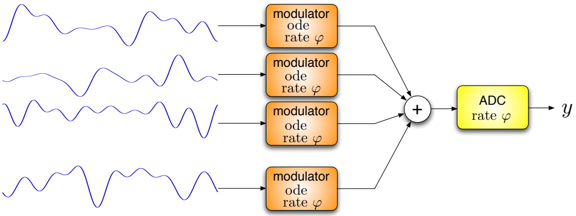
$$(1-\delta)\|\mathbf{X}\|_F^2 \ \le \ \|\mathcal{A}(\mathbf{X})\|_2^2 \ \le \ (1+\delta)\|\mathbf{X}\|_F^2 \quad \forall \text{ rank-}2R \ \mathbf{X},$$

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

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

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

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
  $\sim S \cdot \log N$ measurements

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

Compressive sensing tells us ...

Sensing...

- ... we can sample *smarter* not faster
- ... we can replace front-end acquisition complexity with back-end computing
- ... injecting randomness allows us to *super-resolve* high-frequency signals (or high-resolution images) from low-frequency (low-resolution) measurements
- ... the acquisition process can be *independent* of the types of signals we are interested in

Compressive sensing tells us ...

Sensing...

- ... we can sample *smarter* not faster
- ... we can replace front-end acquisition complexity with back-end computing
- ... injecting randomness allows us to *super-resolve* high-frequency signals (or high-resolution images) from low-frequency (low-resolution) measurements
- ... the acquisition process can be *independent* of the types of signals we are interested in

Mathematics...

- ... there are unique *sparse* solutions to underdetermined systems of equations
- ... random projections keep sparse signals separated
- ... a seemingly impossible optimization program (subset selection) can be solved using a tractable amount of computation