Compressive Sensing

A New Framework for Sparse Signal Acquisition and Processing

Richard Baraniuk

Rice University
Better, Stronger, Faster
Accelerating Data Deluge

- **1250 billion gigabytes** generated in 2010
  - # digital bits > # stars in the universe
  - growing by a factor of 10 every 5 years

- Total data **generated** > total storage

- Increases in **generation rate** >> increases in **transmission rate**
Case in Point: DARPA ARGUS-IS

- 1.8 Gpixel image sensor
  - video rate output: 770 Gbits/s
  - data rate input: 274 Mbits/s
  - factor of 2800x way out of reach of existing compression technology

- **Reconnaissance without conscience**
  - too much data to transmit to a ground station
  - too much data to make effective real-time decisions
Accelerating Data Deluge
Today’s Menu

• What’s wrong with today’s sensor systems?
  *why go to all the work to acquire massive amounts of multimedia data only to throw much/most of it away?*

• One way out: dimensionality reduction (compressive sensing)
  *enables the design of radically new sensors and systems*

• Theory: mathematics of sparsity
  *new nonlinear signal models and recovery algorithms*

• Practice: compressive sensing in action
  *new cameras, imagers, ADCs, …*
Sense by Sampling
Sense by Sampling

$\mathcal{X} \xrightarrow{\text{sample}} N \xrightarrow{\text{too much data!}}$
Sense then *Compress*

\[ x \rightarrow \text{sample} \rightarrow N \gg K \rightarrow K \]

JPEG
JPEG2000
...

\[ K \rightarrow \text{decompress} \rightarrow N \rightarrow \hat{x} \]
Sparsity

\[ N \text{ pixels} \]

\[ K \ll N \text{ large wavelet coefficients} \]
(blue = 0)

\[ N \text{ wideband signal samples} \]

\[ K \ll N \text{ large Gabor (TF) coefficients} \]
Concise Signal Structure

- **Sparse** signal: only $K$ out of $N$ coordinates nonzero
  - model: union of $K$-dimensional subspaces

- **Compressible** signal: sorted coordinates decay rapidly with power-law
Concise Signal Structure

- **Sparse** signal: only $K$ out of $N$ coordinates nonzero
  - model: union of $K$-dimensional subspaces

- **Compressible** signal: sorted coordinates decay rapidly with power-law
  - model: $\ell_p$ ball: $\|x\|_p^p = \sum_i |x_i|^p \leq 1$, $p \leq 1$
What’s Wrong with this Picture?

- Why go to all the work to acquire \( N \) samples only to discard all but \( K \) pieces of data?
What’s Wrong with this Picture?

linear processing
linear signal model
(bandlimited subspace)

nonlinear processing
nonlinear signal model
(union of subspaces)

\( \mathcal{X} \rightarrow \text{sample} \)

\( \overset{N}{\longrightarrow} \text{compress} \)

\( \overset{K}{\longrightarrow} \)

\( \overset{K}{\longrightarrow} \text{decompress} \)

\( \overset{N}{\longrightarrow} \hat{\mathcal{X}} \)
Compressive Sensing

- Directly acquire “compressed” data via dimensionality reduction
- Replace samples by more general “measurements”

\[ K \approx M \ll N \]

\[ \begin{align*}
\mathbf{x} &\rightarrow \text{compressive sensing} \\
&M \rightarrow \mathbf{y}
\end{align*} \]
Sampling

• Signal $x$ is $K$-sparse in basis/dictionary $\Psi$
  - WLOG assume sparse in space domain $\Psi = I$

• Sampling

$$N \times 1$$
measurements

$y$

$\Phi = I$

$x$

$N \times 1$
sparse signal

$K$
nonzero entries
Compressive Sampling

- When data is sparse/compressible, can directly acquire a **condensed representation** with no/little information loss through linear **dimensionality reduction**

\[ y = \Phi x \]

- \( M \times 1 \) measurements
- \( \Phi \) is a \( M \times N \) matrix
- \( x \) is a \( N \times 1 \) sparse signal
- \( K < M \ll N \) where \( K \) is the number of nonzero entries

\[
\begin{align*}
M & \times 1 \\
\text{measurements} & \\
\Phi & \\
\begin{array}{c}
M \times N \\
\text{nonzero entries}
\end{array}
\end{align*}
\]
How Can It Work?

• Projection $\Phi$ 
  *not full rank*...

\[ M < N \]

... and so

**loses information** in general

• Ex: Infinitely many $x$’s map to the same $y$ (null space)
How Can It Work?

- Projection $\Phi$ not full rank...
  
  $M < N$
  
  ... and so loses information in general

- But we are only interested in **sparse** vectors
How Can It Work?

- Projection $\Phi$ not full rank...

$$M < N$$

... and so loses information in general

- But we are only interested in sparse vectors

- $\Phi$ is effectively $MxK$
How Can It Work?

• Projection $\Phi$ not full rank...

$$M < N$$

... and so loses information in general

• But we are only interested in \textit{sparse} vectors

• \textbf{Design $\Phi$} so that each of its $MxK$ submatrices are full rank (ideally close to orthobasis)
  - \textit{Restricted Isometry Property (RIP)}
RIP = Stable Embedding

- An information preserving projection $\Phi$ preserves the **geometry** of the set of sparse signals

- RIP ensures that $\|x_1 - x_2\|_2 \approx \|\Phi x_1 - \Phi x_2\|_2$
How Can It Work?

• Projection $\Phi$ not full rank...

\[ M < N \]

... and so loses information in general

• Design $\Phi$ so that each of its $MxK$ submatrices are full rank (RIP)

• Unfortunately, a combinatorial, NP-Hard design problem
Insight from the 70’s [Kashin, Gluskin]

- Draw $\Phi$ at random
  - iid Gaussian
  - iid Bernoulli $\pm 1$

- Then $\Phi$ has the RIP with high probability provided

$$M = O(K \log(N/K)) \ll N$$
Randomized Sensing

• Measurements \( y = \text{random linear combinations} \) of the entries of \( x \)

• **No information loss** for sparse vectors \( x \) whp

\[
M \times 1 \begin{array}{c}	ext{measurements} \\
\end{array} = \Phi \begin{array}{c}	ext{\( M \times N \)} \\
\end{array} x
\]

\[M = O(K \log(N/K))\]
CS Signal Recovery

- **Goal**: Recover signal \( x \) from measurements \( y \)
- **Problem**: Random projection \( \Phi \) not full rank (ill-posed inverse problem)
- **Solution**: Exploit the sparse/compressible *geometry* of acquired signal \( x \)
CS Signal Recovery

• Random projection $\Phi$ not full rank

• Recovery problem: given $y = \Phi x$ find $x$

• Null space

• Search in null space for the “best” $x$ according to some criterion
  – ex: least squares

\[ y = \Phi x \]

$(N-M)$-dim hyperplane at random angle
\[ l_2 \text{ Signal Recovery} \]

- **Recovery:**
  (ill-posed inverse problem)
  \[
  \text{given } y = \Phi x \\
  \text{find } \hat{x} \text{ (sparse)}
  \]

- **Optimization:**
  \[
  \hat{x} = \arg \min_{\Phi x = y} \|x\|_2
  \]

- **Closed-form solution:**
  \[
  \hat{x} = (\Phi^T \Phi)^{-1} \Phi^T y
  \]

- **Wrong answer!**
$\ell_2$ Signal Recovery

- Recovery: given $y = \Phi x$
  find $x$ (sparse)

- Optimization:

- Closed-form solution:

  $\hat{x} = \arg\min_{y = \Phi x} \|x\|_2$

  $\hat{x} = (\Phi^T\Phi)^{-1}\Phi^Ty$

- Wrong answer!

![Diagram of signal recovery process]
\( l_0 \) Signal Recovery

- Recovery:
  (ill-posed inverse problem)

- Optimization:

  - **Correct!**

  \[
  \hat{x} = \arg \min_{\|x\|_0} \| y - \Phi x \|_0
  \]
  
  “find sparsest vector in translated nullspace”

- But **NP-Complete** alg

\[ y = \Phi x \]
\[ \| \_1 \text{ Signal Recovery} \]

- **Recovery:**
  (ill-posed inverse problem)
  \[
  \text{given } y = \Phi x \\
  \text{find } x \text{ (sparse)}
  \]

- **Optimization:**
  \[
  \hat{x} = \arg \min_{y = \Phi x} \| x \|_1
  \]
  
  **Convexify** the \( \ell_0 \) optimization

Candes  Romberg  Tao  Donoho
\( \ell_1 \) Signal Recovery

- **Recovery:**
  \( (\text{ill-posed inverse problem}) \)
  \( \text{given} \quad y = \Phi x \)
  \( \text{find} \quad x \quad \text{(sparse)} \)

- **Optimization:**
  \( \hat{x} = \arg \min_{y=\Phi x} \|x\|_1 \)

- **Convexify** the \( \ell_0 \) optimization

- **Correct!**

- **Polynomial time** alg
  (linear programming)
CS Hallmarks

- **Stable**
  - acquisition/recovery process is numerically stable

- **Asymmetrical** (most processing at decoder)
  - conventional: smart encoder, dumb decoder
  - CS: dumb encoder, smart decoder

- **Democratic**
  - each measurement carries the same amount of information
  - robust to measurement loss and quantization
  - “digital fountain” property

- Random measurements *encrypted*

- **Universal**
  - same random projections / hardware can be used for *any* sparse signal class *(generic)*
Universality

- Random measurements can be used for signals sparse in any basis

\[ x = \Psi \alpha \]
Universality

- Random measurements can be used for signals sparse in *any* basis

\[ y = \Phi x = \Phi \Psi \alpha \]
Universality

- Random measurements can be used for signals sparse in any basis

\[ y = \Phi x = \Phi \Psi \alpha = \Phi' \alpha \]
Compressive Sensing

*In Action*

Cameras
“Single-Pixel” CS Camera

scene

random pattern on DMD array

single photon detector

image reconstruction or processing

w/ Kevin Kelly
“Single-Pixel” CS Camera

- Flip mirror array $M$ times to acquire $M$ measurements
- Sparsity-based (linear programming) recovery
First Image Acquisition

- Target: 65536 pixels
- 11000 measurements (16%)
- 1300 measurements (2%)
Utility?

Fairchild 100Mpixel CCD

Single photon detector
true color low-light imaging
256 x 256 image with 10:1 compression
[Nature Photonics, April 2007]
CS Infrared Imager

IR photodiode

raster scan IR

CS IR
CS Hyperspectral Imager

Hyperspectral data cube
450-850nm
$N=1M$ space x wavelength voxels
$M=200k$ random measurements
Compressive Sensing

*In Action*

Video Acquisition
From Image to Video Sensing

• Nontrivial extension of CS image acquisition
  – immoral to treat time as 3rd spatial dimension

• **Ephemeral** temporal events
  – should measure temporal events at their “information rate”
  – fleeting events hard to predict and capture

• Computational **complexity** involved in recovering billions of video voxels
Simple LDS Model

• **Linear dynamical system** model
  – image sequence lies along a curve on a linear subspace

• Reasonable model for certain physical phenomena
  – flows, waves, ...

• Leverage modern *state space techniques* to estimate image sequence from compressive measurements
Flame Video

(a) Ground truth

(b) $f_s = 256$ Hz, $\tilde{M} = 30$, $\tilde{M} = 170$, Meas. rate = 5%, SNR = 13.73 dB.

(c) $f_s = 512$ Hz, $\tilde{M} = 30$, $\tilde{M} = 70$, Meas. rate = 2.44%, SNR = 13.73 dB.

(d) $f_s = 1024$ Hz, $\tilde{M} = 30$, $\tilde{M} = 20$, Meas. rate = 1.22%, SNR = 12.63 dB.
Traffic Video

ground truth

CS video recovery

measurement rate = 4%
Compressive Sensing

In Action

A/D Converters
Analog-to-Digital Conversion

- Nyquist rate limits reach of today’s ADCs

- “Moore’s Law” for ADCs:
  - technology Figure of Merit incorporating sampling rate and dynamic range doubles every 6-8 years

- Analog-to-Information (A2I) converter
  - wideband signals have high Nyquist rate but are often sparse/compressible
  - develop new ADC technologies to exploit
  - new tradeoffs among Nyquist rate, sampling rate, dynamic range, ...
Streaming Measurements

- Streaming applications: cannot fit entire signal into a processing buffer at one time

\[ y = \Phi x \]

streaming requires special \( \Phi \)
Streaming Measurements

- Streaming applications: cannot fit entire signal into a processing buffer at one time

\[ y = \Phi x \]

streaming requires special \( \Phi \)

\[ M \text{ measurements} = \Phi \]

\[ x \]
Streaming Measurements

- Many applications: Signal sparse in **frequency** (Fourier transform)

\[ y = \Phi x \]

streaming requires special \( \Phi \)

\[
\begin{array}{cccc}
\text{y} & \Phi & \psi & \alpha \\
\text{\_} & \text{\_} & \text{\_} & \text{\_}
\end{array}
\]
Random Demodulator

\[ x(t) \times p_c(t) \]

\[ \int_{t}^{t - \frac{1}{M}} \]

\[ y[n] \]

\[ t = \frac{n}{M} \]

input signal \( x(t) \)

\[ \times \]

pseudorandom sequence \( p_c(t) \)

modulated input

input signal \( X(\omega) \)

\[ \ast \]

pseudorandom sequence spectrum \( P_c(\omega) \)

modulated input and integrator (low-pass filter)
Random Demodulator

$x(t) \times p_c(t) \rightarrow \int_{t-\frac{1}{M}}^t \rightarrow y[n]$

Seed → Pseudorandom Number Generator

\[ y = \Phi \]

\[ x \]

\[ \Psi \]

\[ \alpha \]
Random Demodulator

\[ x(t) \times p_c(t) \xrightarrow{\int_{t-\frac{1}{M}}^{t}} y[n] \]

Seed → Pseudorandom Number Generator

\[ y = \Phi \psi \alpha \]

\[ \omega \]
Sampling Rate

- **Goal:** Sample near signal’s (low) “information rate” rather than its (high) Nyquist rate

\[ M = O(K \log(N/K)) \]

- A2I sampling rate
- number of tones / window
- Nyquist bandwidth
Sampling Rate

• **Theorem** [Tropp, B, et al 2007]

If the sampling rate satisfies

\[ M > cK \log^2(N/\delta), \quad 0 < \delta < 1 \]

then locally Fourier $K$-sparse signals can be recovered exactly with probability

\[ 1 - \delta \]
Empirical Results

\[ M \leq CK \log\left(\frac{N}{K} + 1\right) \]

\[ C \sim 1.7 \]
Example: Frequency Hopper

Nyquist rate sampling

spectrogram

20x sub-Nyquist sampling

sparsogram
Example: Frequency Hopper

Nyquist rate sampling

- spectrogram
- conventional ADC
- 20MHz sampling rate

20x sub-Nyquist sampling

- sparsogram
- CS-based AIC
- 1MHz sampling rate
Dynamic Range

- **Key result:** Random measurements don’t affect dynamic range

Theoretical slope = $1/3 \, \text{b/dB}$

Actual Slope = $1/2.3 \, \text{b/dB}$

High resolution (many bits per sample) only at low sampling rate
Application: Frequency Tracking

- Compressive **Phase Locked Loop (PLL)**
  - key idea: phase detector in PLL computes inner product between signal and oscillator output
  - RIP ensures we can compute this inner product between corresponding low-rate CS measurements
Summary: CS

- **Compressive sensing**
  - randomized dimensionality reduction
  - exploits signal **sparsity** information
  - integrates sensing, compression, processing

- Why it works: with high probability, random projections preserve information in signals with concise geometric structures

- Enables new sensing architectures
  - ADCs, radios, cameras, ...

- Can process signals/images directly from their compressive measurements
Open Research Issues

• **Links with information theory**
  – new encoding matrix design via codes (LDPC, fountains)
  – new decoding algorithms (BP, etc.)
  – quantization and rate distortion theory

• **Links with machine learning**
  – Johnson-Lindenstrauss, manifold embedding, RIP

• **Processing/inference** on random projections
  – filtering, tracking, interference cancellation, ...

• **Multi-signal CS**
  – array processing, localization, sensor networks, ...

• **CS hardware**
  – ADCs, receivers, cameras, imagers, radars, ...