Compressive Sensing
and Applications

Volkan Cevher
volkan@rice.edu
Rice University
Acknowledgements

- Rice DSP Group (Slides)
  - Richard Baraniuk
    - Mark Davenport,
    - Marco Duarte,
    - Chinmay Hegde,
    - Jason Laska,
    - Shri Sarvotham,
    - Mona Sheikh
    - Stephen Schnelle...
  - Mike Wakin, Justin Romberg, Petros Boufounos, Dror Baron
Outline

• Introduction to Compressive Sensing (CS)
  – motivation
  – basic concepts

• CS Theoretical Foundation
  – geometry of sparse and compressible signals
  – coded acquisition
  – restricted isometry property (RIP)
  – signal recovery

• CS in Action

• Summary
Sensing
Digital Revolution
Pressure is on Digital Sensors

- Success of digital data acquisition is placing increasing pressure on signal/image processing hardware and software to support
  
  **higher resolution / denser sampling**
  - ADCs, cameras, imaging systems, microarrays, ...

  **large numbers of sensors**
  - image data bases, camera arrays, distributed wireless sensor network, ...

  **increasing numbers of modalities**
  - acoustic, RF, visual, IR, UV, x-ray, gamma ray, ...
Pressure is on Digital Sensors

- Success of digital data acquisition is placing increasing pressure on signal/image processing hardware and software to support

  **higher resolution / denser sampling**
  - ADCs, cameras, imaging systems, microarrays, ...

  **large numbers of sensors**
  - Image data bases, camera arrays, distributed wireless sensor networks, ...

  **increasing numbers of modalities**
  - Acoustic, RF, visual, IR, UV

  **deluge of data**
  - How to acquire, store, fuse, process efficiently?
Digital Data Acquisition

• Foundation: *Shannon/Nyquist sampling theorem*
  
  "if you sample densely enough (at the Nyquist rate), you can perfectly reconstruct the original analog data"
Sensing by *Sampling*

- Long-established paradigm for digital data acquisition
  - uniformly *sample* data at Nyquist rate (2x Fourier bandwidth)
Sensing by *Sampling*

- Long-established paradigm for digital data acquisition
  - uniformly *sample* data at Nyquist rate (2x Fourier bandwidth)
Sensing by *Sampling*

- Long-established paradigm for digital data acquisition
  - uniformly **sample** data at Nyquist rate (2x Fourier bandwidth)
  - **compress** data

\[ \mathcal{X} \xrightarrow{\text{sample}} N \xrightarrow{N \gg K} \text{compress} \xrightarrow{K} \text{transmit/store} \]

\[ \mathcal{X} \xrightarrow{\text{receive}} K \xrightarrow{K} \text{decompress} \xrightarrow{N} \hat{\mathcal{X}} \]
Sparsity / Compressibility

$N$ pixels

$K \ll N$
large wavelet coefficients
(blue = 0)

$N$ wideband
signal samples

$K \ll N$
large Gabor (TF)
coefficients

$N$ pixels

$K \ll N$
large wavelet coefficients
(blue = 0)

$N$ wideband
signal samples

$K \ll N$
large Gabor (TF)
coefficients
Sample / Compress

- Long-established paradigm for digital data acquisition
  - uniformly **sample** data at Nyquist rate
  - **compress** data

\[ \mathcal{X} \xrightarrow{\text{sample}} \text{compress} \xrightarrow{N \gg K} \text{transmit/store} \]

\[ \text{receive} \xrightarrow{K} \text{decompress} \xrightarrow{N} \hat{\mathcal{X}} \]

**sparse / compressible** wavelet transform
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?**

\[ \mathcal{X} \rightarrow \text{sample} \xrightarrow{N} \text{compress} \xrightarrow{K} \text{transmit/store} \]

\( N \gg K \)

**Sparse / compressible** 

wavelet transform

\[ \text{receive} \xrightarrow{K} \text{decompress} \xrightarrow{N} \hat{\mathcal{X}} \]
What’s Wrong with this Picture?

**Linear** processing

**Linear** signal model
(bandlimited subspace)

**Nonlinear** processing

**Nonlinear** signal model
(union of subspaces)

\( \mathcal{X} \xrightarrow{N} \text{compress} \xrightarrow{K} \text{transmit/store} \)

\( \mathcal{X} \xrightarrow{K} \text{decompress} \xrightarrow{N} \widehat{\mathcal{X}} \)

**Sparse / compressible**
wavelet transform
Compressive Sensing

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

\[ K \approx M \ll N \]

\[ x \rightarrow \text{compressive sensing} \rightarrow \frac{M}{y} \rightarrow \text{transmit/store} \]

\[ \text{receive} \rightarrow \frac{M}{y} \rightarrow \text{reconstruct} \rightarrow \frac{N}{\hat{x}} \]
Compressive Sensing

Theory I
Geometrical Perspective
Sampling

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

\[ x \]

\[ N \times 1 \]
spare signal

\[ K \]
nonzero entries
Sampling

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

• Samples

\[
\begin{align*}
N \times 1 & \quad \text{measurements} \\
\begin{array}{c}
\vdots \\
\text{green} \\
\vdots \\
\end{array} & \quad = & \quad \begin{array}{c}
\vdots \\
\text{red} \\
\vdots \\
\end{array} \\
\begin{array}{c}
\vdots \\
\text{green} \\
\vdots \\
\end{array} & \quad \Phi = I & \quad \begin{array}{c}
\vdots \\
\text{green} \\
\vdots \\
\end{array}
\end{align*}
\]

\[
\begin{align*}
N \times 1 & \quad \text{sparse signal} \\
\begin{array}{c}
\vdots \\
\text{green} \\
\vdots \\
\end{array} & \quad = & \quad \begin{array}{c}
\vdots \\
\text{red} \\
\vdots \\
\end{array} \\
\begin{array}{c}
\vdots \\
\text{green} \\
\vdots \\
\end{array} & \quad \Phi = I & \quad \begin{array}{c}
\vdots \\
\text{green} \\
\vdots \\
\end{array}
\end{align*}
\]

\[
\begin{align*}
K & \quad \text{nonzero entries}
\end{align*}
\]
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
- \( M \times N \)
- \( K < M \ll N \)
- \( N \times 1 \) sparse signal
- \( K \) nonzero entries
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 $y$
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
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 $M \times K$
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 $M \times K$ submatrices are full rank
How Can It Work?

- **Goal:** Design $\Phi$ so that its $M \times 2K$ submatrices are full rank

  - difference $x_1 - x_2$ between two $K$-sparse vectors is $2K$ sparse in general
  - preserve information in $K$-sparse signals
  - **Restricted Isometry Property** (RIP) of order $2K$
Unfortunately...

- **Goal:** Design $\Phi$ so that its $M \times 2K$ submatrices are full rank (Restricted Isometry Property – RIP)

- Unfortunately, a combinatorial, **NP-complete design problem**
Insight from the 80’s [Kashin, Gluskin]

- Draw $\Phi$ at \textbf{random}
  - iid Gaussian
  - iid Bernoulli $\pm 1$
  - ...

- Then $\Phi$ has the RIP with high probability as long as $M = O(K \log(N/K)) \ll N$
  - $Mx2K$ submatrices are full rank
  - stable embedding for sparse signals
  - extends to compressible signals in $\ell_p$ balls
Compressive Data Acquisition

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

• WHP does not distort structure of sparse signals
  – no information loss

\[
\begin{align*}
M \times 1 & \text{ measurements} & \Phi & \text{ nonzero entries} \\
M \times N & \text{ } & N \times 1 & \text{ sparse signal} \\
K < M & \ll N & K & \text{ nonzero entries}
\end{align*}
\]
CS Signal Recovery

- **Goal**: Recover signal $x$ from measurements $y$

- **Challenge**: Random projection $\Phi$ not full rank (ill-posed inverse problem)

- **Solution**: Exploit the sparse/compressible *geometry* of acquired signal $x$
Concise Signal Structure

- **Sparse** signal: only $K$ out of $N$ coordinates nonzero
Concise Signal Structure

- **Sparse** signal: only $K$ out of $N$ coordinates nonzero
  - model: union of $K$-dimensional subspaces aligned w/ coordinate axes

\[ |x_i| \]

\[ K \quad \text{sorted index} \quad N \]
Concise Signal Structure

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

- **Compressible** signal: sorted coordinates decay rapidly to zero

\[ |x_i| \]

Sorted index \( K \), power-law decay, \( N \)
Concise Signal Structure

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

- **Compressible** signal: sorted coordinates decay rapidly to zero
  - model: $\ell_p$ ball: $\|x\|_p^p = \sum_i |x_i|^p \leq 1$, $p \leq 1$
CS Signal Recovery

- Random projection $\Phi$ not full rank

- Recovery problem: given $y = \Phi x$, find $x$

- Null space

- So search in null space for the "best" $x$ according to some criterion
  - ex: least squares
CS Signal Recovery

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

- $\ell_2$ fast

\[ \hat{x} = \text{arg min}_{y=\Phi x} \|x\|_2 \]

\[ \hat{x} = (\Phi^T \Phi)^{-1} \Phi^T y \]
	pseudoinverse
CS Signal Recovery

- Recovery: given $y = \Phi x$ find $x$ (sparse)
  
- $\ell_2$ fast, wrong
  
\[ \hat{x} = \arg\min_{y=\Phi x} \|x\|_2 \]

![Signal Recovery Diagram]

\[ x \]

\[ \hat{x} = (\Phi^T \Phi)^{-1} \Phi^T y \]

pseudoinverse
Why $\ell_2$ Doesn’t Work

for signals sparse in the **space/time domain**

\[
\hat{x} = \arg \min_{y = \Phi x'} \|x'\|_2
\]

least squares, minimum $\ell_2$ solution is almost **never sparse**

null space of $\Phi$
translated to $x$
(random angle)
CS Signal Recovery

- Reconstruction/decoding: given \( y = \Phi x \)
  (ill-posed inverse problem) find \( x \)

- \( \ell_2 \) fast, wrong
  \[ \hat{x} = \underset{y=\Phi x}{\text{arg min}} \|x\|_2 \]

- \( \ell_0 \)
  \[ \hat{x} = \underset{y=\Phi x}{\text{arg min}} \|x\|_0 \]

  "find sparsest \( x \) in translated nullspace"

number of nonzero entries
CS Signal Recovery

- Reconstruction/decoding: given \( y = \Phi x \) (ill-posed inverse problem) find \( x \)

- \( \ell_2 \) fast, wrong

- \( \ell_0 \) correct: only \( M = 2K \) measurements required to reconstruct \( K \)-sparse signal

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

\[
\hat{x} = \arg \min_{y=\Phi x} \|x\|_0
\]

number of nonzero entries
CS Signal Recovery

- Reconstruction/decoding: given $y = \Phi x$ (ill-posed inverse problem) find $x$
- $\ell_2$ fast, wrong
- $\ell_0$ \textbf{correct:}
  - only $M=2K$ measurements required to reconstruct $K$-sparse signal
  - \textbf{slow:} NP-complete algorithm

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

\[
\hat{x} = \arg \min_{y=\Phi x} \|x\|_0
\]

\textit{number of nonzero entries}
CS Signal Recovery

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

- **\( \ell_2 \)** fast, wrong
  \[ \hat{x} = \arg \min_{y=\Phi x} \|x\|_2 \]

- **\( \ell_0 \)** correct, slow
  \[ \hat{x} = \arg \min_{y=\Phi x} \|x\|_0 \]

- **\( \ell_1 \)** correct, efficient, mild oversampling
  \[ \hat{x} = \arg \min_{y=\Phi x} \|x\|_1 \]

  [Candes, Romberg, Tao; Donoho]

  number of measurements required
  \[ M = O(K \log(N/K)) \ll N \]
Why $\ell_1$ Works

for signals sparse in the space/time domain

\[
\hat{x} = \arg \min_{y = \Phi x'} \| x' \|_1
\]

minimum $\ell_1$ solution
= sparsest solution
(with high probability) if

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

- Directly acquire "compressed" data
- Replace $N$ samples by $M$ random projections

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

Theory II
Stable Embedding
Johnson-Lindenstrauss Lemma

- JL Lemma: random projection stably embeds a cloud of $Q$ points whp provided $M = O(\log Q)$

Proved via concentration inequality

Same techniques link JLL to RIP

Connecting JL to RIP

Consider effect of random JL $\Phi$ on each K-plane
- construct covering of points Q on unit sphere
- JL: isometry for each point with high probability
- union bound $\Rightarrow$ isometry for all points q in Q
- extend to isometry for all x in K-plane
Connecting JL to RIP

Consider effect of random JL $\Phi$ on each K-plane
- construct covering of points $Q$ on unit sphere
- JL: isometry for each point with high probability
- union bound $\Rightarrow$ isometry for all points $q$ in $Q$
- extend to isometry for all $x$ in K-plane
- union bound $\Rightarrow$ isometry for all K-planes
Favorable JL Distributions

- **Gaussian**
  \[ \phi_{i,j} \sim \mathcal{N} \left( 0, \frac{1}{M} \right) \]

- **Bernoulli/Rademacher** [Achlioptas]
  \[ \phi_{i,j} := \begin{cases} 
  +\frac{1}{\sqrt{M}} & \text{with probability } \frac{1}{2}, \\
  -\frac{1}{\sqrt{M}} & \text{with probability } \frac{1}{2}
 \end{cases} \]

- **“Database-friendly”** [Achlioptas]
  \[ \phi_{i,j} := \begin{cases} 
  +\sqrt{\frac{3}{M}} & \text{with probability } \frac{1}{6}, \\
  0 & \text{with probability } \frac{2}{3}, \\
  -\sqrt{\frac{3}{M}} & \text{with probability } \frac{1}{6}
 \end{cases} \]

- **Random Orthoprojection to \( \mathbb{R}^M \)** [Gupta, Dasgupta]
RIP as a “Stable” Embedding

• RIP of order $2K$ implies: for all $K$-sparse $x_1$ and $x_2$,

$$(1 - \delta_{2K}) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2^2}{\|x_1 - x_2\|_2^2} \leq (1 + \delta_{2K})$$
CS Recovery Algorithms

• Convex optimization:
  – noise-free signals
    ▪ Linear programming (Basis pursuit)
    ▪ FPC
    ▪ Bregman iteration, ...
  – noisy signals
    ▪ Basis Pursuit De-Noising (BPDN)
    ▪ Second-Order Cone Programming (SOCP)
    ▪ Dantzig selector
    ▪ GPSR, ...

• Iterative greedy algorithms
  – Matching Pursuit (MP)
  – Orthogonal Matching Pursuit (OMP)
  – StOMP
  – CoSaMP
  – Iterative Hard Thresholding (IHT), ...

software @
dsp.rice.edu/cs
SOCP

• Standard LP recovery

\[ \min \|x\|_1 \quad \text{subject to} \quad y = \Phi x \]

• Noisy measurements

\[ y = \Phi x + n \]

• Second-Order Cone Program

\[ \min \|x\|_1 \quad \text{subject to} \quad \|y - \Phi x\|_2 \leq \epsilon \]

• Convex, \textit{quadratic program}
BPDN

- Standard LP recovery
  \[ \min \|x\|_1 \quad \text{subject to} \quad y = \Phi x \]
- Noisy measurements
  \[ y = \Phi x + n \]
- Basis Pursuit De-Noising
  \[ \min \frac{1}{2} \| y - \Phi x \|_2^2 + \lambda \|x\|_1 \]
- Convex, \textit{quadratic program}
Matching Pursuit

- Greedy algorithm

**Key ideas:**

1. Measurements $y$ composed of sum of $K$ columns of $\Phi$

2. Identify which $K$ columns sequentially according to size of contribution to $y$
Matching Pursuit

- For each column $\phi_i$
  compute
  \[ \hat{x}_i = \langle y, \phi_i \rangle \]

- Choose largest $|\hat{x}_i|$ (greedy)

- Update estimate $\hat{x}$ by adding in $\hat{x}_i$

- Form residual measurement and iterate until convergence
  \[ y' = y - x_i \phi_i \]
Orthogonal Matching Pursuit

• Same procedure as Matching Pursuit

• Except at each iteration:
  – remove selected column $\phi_i$
  – re-orthogonalize the remaining columns of $\Phi$

• Converges in $K$ iterations
Compressive Sensing

*In Action*

Cameras
“Single-Pixel” CS Camera

Scene

Single photon detector

Random pattern on DMD array

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
Single Pixel Camera
Single Pixel Camera

Object

LED (light source)

Lens 1

Lens 2

Photodiode circuit

DMD+ALP Board
Single Pixel Camera

Object

LED (light source)

Lens 1

Lens 2

Photodiode circuit

DMD+ALP Board
Single Pixel Camera
First Image Acquisition

- Target: 65536 pixels
- 11000 measurements (16%)
- 1300 measurements (2%)
Second Image Acquisition

4096 pixels  500 random measurements
CS Low-Light Imaging with PMT

true color low-light imaging
256 x 256 image with 10:1 compression
[Nature Photonics, April 2007]
Hyperspectral Imaging

spectrometer

blue

red

near IR
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

• DARPA Analog-to-Information (A2I) program
  – 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, ...
Analog-to-Information Conversion

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

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

- Practical hardware: randomized demodulator (CDMA receiver)
Example: Frequency Hopper

Nyquist rate sampling

20x sub-Nyquist sampling

spectrogram

sparsogram
Compressive Sensing

In Action

Data Processing
Information Scalability

- Many applications involve signal *inference* and not *reconstruction*

  detection $<$ classification $<$ estimation $<$ reconstruction
Information Scalability

• Many applications involve signal *inference* and not *reconstruction*

  detection < classification < estimation < reconstruction

• **Good news:** CS supports efficient learning, inference, processing directly on compressive measurements

• **Random projections ~ sufficient statistics** for signals with concise geometrical structure
Matched Filter

- Detection/classification with $K$ unknown articulation parameters
  - Ex: position and pose of a vehicle in an image
  - Ex: time delay of a radar signal return

- **Matched filter**: joint parameter estimation and detection/classification
  - compute sufficient statistic for each potential target and articulation
  - compare “best” statistics to detect/classify
Matched Filter Geometry

- Detection/classification with $K$ unknown articulation parameters

- Images are points in $\mathbb{R}^N$

- **Classify** by finding closest target template to data for each class (AWG noise)
  - distance or inner product
Matched Filter Geometry

- Detection/classification with $K$ unknown articulation parameters

- Images are points in $\mathbb{R}^N$

- Classify by finding closest target template to data

- As template articulation parameter changes, points map out a $K$-dim **nonlinear manifold**

- Matched filter classification = closest manifold search
Theorem: \( M = O(K \log N) \)
random measurements stably embed manifold whp

[Baraniuk, Wakin, *FOCM* ’08]
related work:
[Indyk and Naor, Agarwal et al., Dasgupta and Freund]

- Stable embedding
- Proved via concentration inequality arguments (JLL/CS relation)
CS for Manifolds

- **Theorem:**
  \[ M = O(K \log N) \]
  random measurements stably embed manifold whp

- Enables parameter estimation and MF detection/classification directly on compressive measurements
  - \( K \) very small in many applications (# articulations)
Example: Matched Filter

- Detection/classification with $K=3$ unknown articulation parameters
  1. horizontal translation
  2. vertical translation
  3. rotation
Smashed Filter

- Detection/classification with $K=3$ unknown articulation parameters (manifold structure)

- Dimensionally reduced matched filter directly on compressive measurements

$$M = O(K \log N)$$
Smashed Filter

- Random shift and rotation ($K=3$ dim. manifold)
- Noise added to measurements
- Goal: identify most likely position for each image class; identify most likely class using nearest-neighbor test
Compressive Sensing

Summary
CS Hallmarks

- CS changes the rules of the data acquisition game
  - exploits a priori signal *sparsity* information

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

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

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

- Random projections weakly *encrypted*
CS Hallmarks

- **Democratic**
  - each measurement carries the same amount of information
  - robust to measurement loss and quantization simple encoding

- Ex: wireless streaming application with data loss

  - conventional: complicated (unequal) error protection of compressed data
    - DCT/wavelet low frequency coefficients

  - CS: merely stream additional measurements and reconstruct using those that arrive safely (fountain-like)
After the Break

Beyond Sparsity with structured sparsity models.
volkan@rice.edu
dsp.rice.edu/cs