Randomized Dimensionality Reduction

with Applications to Signal Processing and Communications

Richard Baraniuk
Rice University
The Digital Universe

- **Size:** 281 billion gigabytes generated in 2007
  - digital bits > stars in the universe
  - growing by a factor of 10 every 5 years
  - > Avogadro’s number (6.02x10^{23}) in 15 years

- Growth fueled by **multimedia data**
  - audio, images, video, surveillance cameras, sensor nets, ...

- In 2007 digital data **generated > total storage**
  - by 2011, ½ of digital universe will have no home

[Source: IDC Whitepaper “The Diverse and Exploding Digital Universe” March 2008]
Act I  What’s wrong with today’s multimedia sensor systems?
why go to all the work to acquire massive amounts of multimedia data only to throw much/most of it away?

Act II  A way out: dimensionality reduction (compressive sensing)
enables the design of radically new sensors and systems

Act III  Compressive sensing in action
new cameras, imagers, ADCs, ...

Act IV  Intersections with info theory
new problems, new solutions
Sense by **Sampling**

\[ \mathcal{X} \xrightarrow{\text{sample}} \overset{N}{\longrightarrow} \]

**too much data!**
Sense then *Compress*

\[ \mathcal{X} \xrightarrow{\text{sample}} N \xrightarrow{N \gg K} K \xrightarrow{K} \hat{\mathcal{X}} \]

*JPEG*

*JPEG2000*

*...*
Sparsity

$N$ pixels

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

$N$ wideband signal samples

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

frequency
time
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)
Compressive Sensing

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

\[ K \approx M \ll N \]

[Diagram showing data flow from \( \mathcal{X} \) to compressive sensing, then recovering \( \hat{x} \)]
Compressive Sensing Theory

A Geometrical Perspective
Sampling

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

• Sampling

\[ N \times 1 \text{ measurements} \]

\[ y = \Phi = I \]

\[ N \times 1 \text{ sparse signal} \]

\[ K \text{ nonzero entries} \]
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
- \( N \times 1 \) sparse signal
- \( K \) nonzero entries

\( K < M \ll N \)
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 \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 (ideally orthobasis)
Geometrical Situation

- **Sparse** signal: All but \( K \) coordinates are zero

- **Q:** Structure of the set of all \( K \)-sparse vectors?
Geometrical Situation

- **Sparse** signal: All but $K$ coordinates are zero

- **Model:** union of $K$-dimensional subspaces aligned w/ coordinate axes (highly nonlinear!)

\[
N = 3 \\
K = 2
\]
Stable Embedding

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

- How to do this? Ensure $\|x_1 - x_2\|_2 \approx \|\Phi x_1 - \Phi x_2\|_2$
Restricted Isometry Property

- RIP of order $2K$ implies $\Phi$ preserves distances between any two $K$-sparse vectors $x_1, x_2$

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

- $x_1, x_2$ each $K$-sparse; $x_1 - x_2$ is $2K$-sparse

- RIP: any selection of $2K$ columns is close to an orthobasis
Restricted Isometry Property

- RIP of order $2K$ implies $\Phi$ preserves distances between any two $K$-sparse vectors $x_1, x_2$

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

- $x_1, x_2$ each $K$-sparse; $x_1 - x_2$ is $2K$-sparse

- RIP: any selection of $2K$ columns is close to an orthobasis

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

- Draw $\Phi$ at random
  - iid Gaussian
  - iid Bernoulli $\pm 1$
- Then $\Phi$ has the RIP with high probability provided

$$M = O(K \log(N/K)) \ll N$$
One Proof Technique

1. Consider one $K$-dim subspace
2. Cover unit sphere in subspace with a point cloud
3. Apply Johnson-Lindenstrauss Lemma to show isometry for all points in cloud
   \[ \log(Q) \] random projections preserve inter-point distances in a cloud of $Q$ points in $\mathbb{R}^N$ whp
4. Extend to entire $K$-dim subspace
5. Extend via union bound to all $\binom{N}{K}$ subspaces
Randomized Dim. Reduction

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

- **No information loss** for sparse vectors $x$ whp

\[ M \times 1 \text{ measurements} \quad \Phi \quad N \times 1 \text{ sparse signal} \]

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

\[
\begin{align*}
\Phi y &= x \\
\end{align*}
\]
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

$(N-M)$-dim hyperplane at random angle
\( \ell_2 \) Signal Recovery

- **Recovery:**
  (ill-posed inverse problem)
  given \( y = \Phi x \)
  find \( x \) (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!**
\( \ell_0 \) Signal Recovery

- **Recovery:**
  (ill-posed inverse problem)

- **Optimization:**

- **Correct!**

- But **NP-Complete** alg

given \( y = \Phi x \)
find \( x \) (sparse)

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

“find sparsest vector in translated nullspace”
$\ell_1$ Signal Recovery

- Recovery: given $y = \Phi x$ find $x$ (sparse)
  - (ill-posed inverse problem)

- Optimization:
  \[ \hat{x} = \arg \min_{y=\Phi x} \|x\|_1 \]

- **Convexify** the $\ell_0$ optimization

Candes  Romberg  Tao  Donoho
\[ \ell_1 \text{ Signal Recovery} \]

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

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

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

- **Correct!**

- **Polynomial time** alg
Summary: CS

- **Encoding:** \( y = \) random linear combinations of the entries of \( x \)

\[
M \times 1 \quad \text{measurements} \quad \begin{bmatrix} y \end{bmatrix} = \Phi x
\]

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

- **Decoding:** Recover \( x \) from \( y \) via optimization
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*
Compressive Sensing

*In Action*

Cameras
"Single-Pixel" CS Camera

scene

$x$

random pattern on DMD array

single photon detector

$\Phi$

DMD

PD

A/D

$y$

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?

single photon detector

Fairchild 100Mpixel CCD
CS Low-Light Imaging with PMT

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

**spectrometer**

**hyperspectral data cube**

450-850nm
1M space x wavelength voxels
200k random sums
CS In Action

• CS makes sense when measurements are expensive

• **Ultrawideband A/D converters**
  [DARPA “Analog to Information” program, Dennis Healy]

• **Medical imaging**
  – faster imaging/lower dosage [Lustig et al]

• **Camera networks**
  – sensing/compression/fusion

• **Radar, sonar, array processing**
  – exploit spatial sparsity of targets

• **DNA microarrays**
  – smaller, more agile arrays for bio-sensing [Milenkovic, Orlitsky, B]
Intersections with Information Theory

Structured Sparsity
Beyond Sparse Models

- Sparse signal model captures simplistic primary structure

wavelets: natural images  
Gabor atoms: chirps/tones  
pixels: background subtracted images
Beyond Sparse Models

- Sparse signal model captures **simplistic primary structure**

- Modern compression/processing algorithms capture **richer secondary coefficient structure**

wavelets: natural images

Gabor atoms: chirps/tones

pixels: background subtracted images
Sparse Signals

- **K-sparse signals** comprise a particular set of $K$-dim subspaces
Structured-Sparse Signals

- A \textit{K}-sparse signal model comprises a particular (\textit{reduced}) set of \textit{K}-dim subspaces

[Blumensath and Davies]

- Fewer subspaces
  \(<>\) relaxed RIP
  \(<>\) stable recovery using fewer measurements \(M\)
Wavelet Sparse

- Typical of wavelet transforms of natural signals and images (piecewise smooth)
Tree-Sparse

- **Model:** $K$-sparse coefficients + significant coefficients lie on a **rooted subtree**

- Typical of wavelet transforms of natural signals and images (piecewise smooth)
Wavelet Sparse

- **RIP**: stable embedding

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

- **Model:** $K$-sparse coefficients + significant coefficients lie on a rooted subtree

- **Tree-RIP:** stable embedding

\[ M = O(K) < O(K \log(N/K)) \]
Tree-Sparse Signal Recovery

$N=1024$
$M=80$

target signal

CoSaMP,
(RMSE=1.12)

Tree-sparse CoSaMP
(RMSE=0.037)

L1-minimization
(RMSE=0.751)

[B, Cevher, Duarte, Hegde’08]
Structured Sparsity

- **Clustered coefficients** via Ising graphical model
  [Cevher, Duarte, Hegde, B’08]

- **Dispersed coefficients** via block sparsity
  [Bresler et al; B, Cevher, Duarte, Hegde’08; Eldar et al; ...]

**Images:**
- Target
- Ising-model recovery
- CoSaMP recovery
- LP (FPC) recovery
Intersections with Information Theory

Detection/Classification
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
Classification

• Simple object classification problem
  – AWGN: nearest neighbor classifier

• Common issue:
  – $L$ unknown articulation parameters

• Common solution: matched filter
  – find nearest neighbor under all articulations
Matched Filter Geometry

- Classification with $L$ 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 $L$ 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 $L$-dim **nonlinear manifold**

- Matched filter classification = closest manifold search
CS for Manifolds

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

[B, Wakin, *FOCM ’08]

- \( \Phi \) preserves distances between points on the manifold

- Extension of CS RIP
CS for Manifolds

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

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

- Naïve approach
  - take $M$ CS measurements, $M = O(K \log N)$
  - recover $N$-pixel image from CS measurements (expensive)
  - conventional matched filter
Smashed Filter

- Worldly approach
  - take $M$ CS measurements, $M = O(L \log N)$
  - matched filter directly on CS measurements (inexpensive)
Smashed Filter

- Random shift and rotation ($L=3$ dim. manifold)
- WG noise added to measurements
- Goals: identify most likely shift/rotation parameters
  identify most likely class

![Graphs showing average shift estimate error and classification rate vs. number of measurements $M$ with different noise levels $\sigma$.]
Intersections with Information Theory

Fast Encoding/Decoding
“Dense” Encoding

- **Dense, unstructured** $\Phi$ leads to expensive encoding and decoding
  
  (think Shannon’s Gaussian codebook)
Sparse Encoding

- Consider **sparse** $\Phi$ w/ $J$ nonzeros/column
- Supports **fast encoding**
- Unfortunately, **not info preserving** in the RIP-$\ell_2$ sense

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

unless $M = O(K^2)$
Sparse Encoding

- Consider **sparse** $\Phi$
  w/ $J$ nonzeros/column

- Supports **fast encoding**

- However, **info preserving** in the RIP-$\ell_1$ sense

  
  $$J(1 - \delta) \leq \frac{\|\Phi x_1 - \Phi x_2\|_1}{\|x_1 - x_2\|_1} \leq J(1 + \delta)$$

  sufficient condition: underlying graph of $\Phi$
  is a $(K, J(1 - \delta/2))$ **expander**

  [Berinde-Gilbert-Indyk-Karloff-Strauss’08]
  [Xu-Hassibi’07, Indyk’08, Jafarpour-Xu-Hassibi-Calderbank’08]
Sparse Encoding

- Consider **sparse** $\Phi$ w/ $J$ nonzeros/column
- Supports **fast encoding**
- However, **info preserving**

in the RIP- $\ell_1$ sense

$$J(1 - \delta) \leq \frac{\|\Phi x_1 - \Phi x_2\|_1}{\|x_1 - x_2\|_1} \leq J(1 + \delta)$$

sufficient condition: underlying graph of $\Phi$ is a $(K, J(1 - \delta/2))$ **expander**

[Berinde-Gilbert-Indyk-Karloff-Strauss’08] [Xu-Hassibi’07, Indyk’08, Jafarpour-Xu-Hassibi-Calderbank’08]
Fast Decoding via BP

• Bayesian graphical model formulation

• Signal $x$ generated from a sparse prior (ex: mixture model)

• Measurements $y$ linked to $x$ via graph $\Phi$ (plus AWGN)

• MAP estimation via message passing (belief propagation – BP)

• Fast encoding $O(N \log(N))$
• Fast decoding $O(N \log^2(N))$

• Related to LDPC decoding, multiuser detection, ...
Many More Intersections with IT

CS measurement and recovery

Finite rate of innovation sampling [Vetterli et al]

Finite rate of innovation Reed-Solomon codes (Vandermonde) [Tarokh et al]

Multi-signal CS [Baron, Duarte, Wakin, Sarvotham, B]

Multi-user detection

Greedy recovery algs [MP, OMP, CoSaMP, SP,...]

Slepian-Wolf coding

Democracy [Laska, Davenport, Boufounos, B]

Fountain codes
Important Ideas Not Covered

- CS for “compressible” signals
  (approximate sparsity)

- **Finite Rate of Innovation Sampling**
  [Vetterli et al]

- “Random” / Co-set Sampling
  [Bresler et al; Gilbert and Strauss; Eldar et al]

- **Greedy recovery algorithms**
  [MP, OMP, CoSaMP, SP, ...]
Other Open Research Issues

- **Quantizing** CS measurements

- **Applications in** machine learning

  - **target**
    - 4096
    - 8-bit pixels

  - **recovery**
    - 4096 1-bit msnts

  - **recovery**
    - 512 1-bit msnts

  - 32768 bits
  - 4096 bits
  - 512 bits

  - 1 bit/pixel
  - 1/8 bit/pixel

[Petros Boufounos, B’07]

[Davenport, B; Siva, Calderbank]
Summary

• **Compressive sensing**
  – Randomized dimensionality reduction
  – integrates sensing, compression, processing
  – exploits signal sparsity information
  – enables new sensing modalities, architectures, systems

• **Why CS works:** preserves information in signals with concise geometric structure
  sparse signals | compressible signals | manifolds

• **Information scalability** for compressive inference
  – compressive measurements ~ sufficient statistics

• Many intersections with **information theory**
dsp.rice.edu/cs
Connexions (cnx.org)

- **non-profit open publishing project**
- **goal:** make high-quality educational content available to anyone, anywhere, anytime for free on the web and at very low cost in print
- open-licensed repository of **Lego**-block modules for authors, instructors, and learners to **create, rip, mix, burn**
- **global reach:** >1M users monthly from 200 countries
- **collaborators:** [IEEE](https://ieeecnx.org), Govt. Vietnam, TI, NI, ...