Compressive Sensing and Applications

Volkan Cevher
volkan.cevher@epfl.ch

Justin Romberg
jrom@ece.gatech.edu
Acknowledgements

- Rice DSP Group (Slides)
  - Richard Baraniuk
    - Mark Davenport,
    - Marco Duarte,
    - Chinmay Hegde,
    - Jason Laska,
    - Shri Sarvotham,
    - Mona Sheikh
    - Stephen Schnelle...
  - Mike Wakin, 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)
  – structured matrices and random convolution
  – signal recovery algorithms
  – structured sparsity

• CS in Action

• Summary
Sensing
Digital Revolution

12MP

25fps/1080p

4KHz

Multi touch
Digital Revolution

12MP
25fps/1080p
4KHz
1977 – 5 hours
<30 mins
Major Trends

higher resolution / denser sampling

12MP  25fps/1080p  4KHz

160MP  200,000fps  192,000Hz
Major Trends

higher resolution
faster sampling
large numbers of sensors
Major Trends

higher resolution / denser sampling
large numbers of sensors

increasing # of modalities / mobility
acoustic, RF, visual, IR, UV, x-ray, gamma ray, ...
Major Trends in Sensing

higher resolution / denser sampling

X

large numbers of sensors

X

increasing # of modalities / mobility

=  ?
Major Trends in Sensing
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

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

JPEG
JPEG2000
...

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

$N$ pixels

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

$N$ wideband signal samples

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

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

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

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

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

**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?**

\[ \mathbb{X} \overset{\text{sample}}{\rightarrow} \overset{N}{\text{compress}} \overset{K}{\text{transmit/store}} \]

- \(N \gg K\)
- **sparse / compressible**
- wavelet transform

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

linear processing
linear signal model
(bandlimited subspace)

nonlinear processing
nonlinear signal model
(union of subspaces)

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

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

sparse / compressible
wavelet
transform
Compressive Sensing

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

\[ K \approx M \ll N \]

\[ \mathcal{X} \xrightarrow{\text{compressive sensing}} M \xrightarrow{y} \text{transmit/store} \]

\[ \text{receive} \xrightarrow{M} \text{reconstruct} \xrightarrow{N} \hat{\mathcal{X}} \]
Compressive Sensing

Theory I

Geometrical Perspective
Sampling

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

$N \times 1$ sparse signal

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

• Samples

$N \times 1$ measurements

\[
\begin{align*}
\Phi &= I \\
N \times 1 &\quad \text{sparse signal} \\
K &\quad \text{nonzero entries}
\end{align*}
\]
Compressive Sampling

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

\[ y = \Phi x \]

\[ M \times 1 \text{ measurements} = \Phi \] 
\[ M \times N \]
\[ x \]

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

\[ K < M \ll N \]

\[ K \text{ nonzero entries} \]
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 $x$
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 $2K$
• **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 **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
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

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

\[ K < M \ll N \]

\[ K \] nonzero entries
Compressive Sensing Recovery

1. Sparse / compressible \( x \)

\textit{not sufficient alone}

\[
\begin{align*}
y & = \Phi x \\
M \times 1 & \quad M \times N \ (M < N) \\
N \times 1 &
\end{align*}
\]

2. Projection \( \Phi \)

\textit{information preserving}

\textit{(restricted isometry property - RIP)}

3. Decoding algorithms

\textit{tractable}
Compressive Sensing Recovery

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

- \( \ell_2 \) fast

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

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

pseudoinverse
Compressive Sensing Recovery

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

- \( \ell_2 \) fast, wrong

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

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

\( x \)

\( \hat{x} \) 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)
Compressive Sensing Recovery

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

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

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

"find sparsest \( x \) in translated nullspace"
Compressive Sensing Recovery

- Reconstruction/decoding: given \( y = \Phi x \) find \( x \)

- \( l_2 \) fast, wrong

- \( l_0 \) correct:
  - only \( M = 2K \) measurements required to reconstruct \( K \)-sparse signal

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

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

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

- \( l_2 \)
  fast, wrong

- \( l_0 \)
  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

slow: NP-hard algorithm
**Compressive Sensing Recovery**

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

- \( \ell_2 \) fast, wrong

- \( \ell_0 \) correct, slow

- \( \ell_1 \) **correct, efficient**
  - **mild oversampling**
  - [Candes, Romberg, Tao; Donoho]

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

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

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

**linear program**

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 $x_2$,

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

- There are more structured (but still random) compressed sensing matrices
- We can randomly sample in a domain whose basis vectors are \textit{incoherent} with the sparsity basis
- Example: sparse in time, sample in frequency
Structured Random Matrices

- Signal is sparse in the wavelet domain, measured with *noiselets* (Coifman et al. ‘01)

2D noiselet  wavelet domain  noiselet domain

- Stable recovery from

\[ M = \mathcal{O}(K \log^4 N) \]

measurements
Random Convolution

• A natural way to implement compressed sensing is through random convolution

• Applications include active imaging (radar, sonar,...)

• Many recent theoretical results

(R 08, Bajwa, Haupt et al 08, Rauhut 09)
Random Convolution Theory

- Convolution with a random pulse, then subsample
  \[ \Phi = R_\Omega F^* \Sigma F, \quad \Sigma = \text{diag}(\{\sigma_\omega\}) \]
  (each \( \sigma_\omega \) has unit magnitude and random phase)

- Stable recovery from
  \[ M = O(K \log^5 N) \]
  measurements
Coded Aperture Imaging

- Allows high-levels of light and high resolution

(Marcia and Willett 08, also Brady, Portnoy and others)
Super-resolved Imaging

Ground truth

Uncoded observation
(1/16 as many pixels)

Coded observation
(1/16 as many pixels)

Reconstruction

CS Reconstruction

(Marcia and Willet 08)
Stability

- Recovery is robust against noise and modeling error

- Suppose we observe

\[ y = \Phi x_0 + e, \quad \|e\|_2 \leq \epsilon \]

- Relax the recovery algorithm, solve

\[
\min_x \|x\|_{\ell_1} \quad \text{subject to} \quad \|y - \Phi x\|_2 \leq \epsilon
\]

- The recovery error obeys

\[
\|x^* - x_0\|_2 \lesssim \epsilon + \frac{\|x_{0,K} - x_0\|_{\ell_1}}{\sqrt{K}}
\]

*measurement error + approximation error*

\[ x_{0,K} = \text{best } K\text{-term approximation} \]
Consider and “$\ell_1$-descent vectors” $h$ for feasible $x$:

$$\|x_0 + h\|_{\ell_1} < \|x_0\|_{\ell_1}$$

$x_0$ is the solution if

$$\Phi h \neq 0$$

for all such descent vectors.
Geometrical Viewpoint, Noise

- Solution will be within $\epsilon$ of $H$
- Need that not too much of the $\ell_1$ ball near $x_0$ is feasible
Compressive Sensing
Recovery Algorithms
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
L1 with equality constraints = linear programming

The standard L1 recovery program

$$\min_x \|x\|_{\ell_1} \quad \text{s.t.} \quad y = \Phi x$$

is equivalent to the linear program

$$\min_{x, t} \sum_i t_i \quad \text{s.t.} \quad -t_i \leq x_i \leq t_i, \quad \Phi x = y$$

There has been a tremendous amount of progress in solving linear programs in the last 15 years.
SOCP

- Standard LP recovery
  \[ \min \|x\|_1 \text{ subject to } y = \Phi x \]

- Noisy measurements
  \[ y = \Phi x + n \]

- Second-Order Cone Program
  \[ \min \|x\|_1 \text{ subject to } \|y - \Phi x\|_2 \leq \epsilon \]

- Convex, quadratic program
Other Flavors of L1

- Quadratic relaxation (called LASSO in statistics)
  \[ \min_x \|x\|_{\ell_1} + \lambda \|y - \Phi x\|_2^2 \]

- Dantzig selector (residual correlation constraints)
  \[ \min_x \|x\|_{\ell_1} \quad \text{s.t.} \quad \|\Phi^T (y - \Phi x)\|_\infty \]

- L1 Analysis (\(\Psi\) is an overcomplete frame)
  \[ \min_x \|\Psi^T x\|_{\ell_1} \quad \text{s.t.} \quad \|y - \Phi x\|_2 \leq \epsilon \]
Solving L1

- “Classical” (mid-90s) interior point methods
  - main building blocks due to Nemirovski
  - second-order, series of local quadratic approximations
  - boils down to a series of linear systems of equations
  - formulation is very general (and hence adaptable)

- Modern progress (last 5 years) has been on “first order” methods
  - Main building blocks due to Nesterov (mid 80s)
  - iterative, require applications of $\Phi$ and $\Phi^T$ at each iteration
  - convergence in 10s-100s of iterations typically

- Many software packages available
  - Fixed-point continuation (Rice)
  - Bregman iteration-based methods (UCLA)
  - NESTA (Caltech)
  - GPSR (Wisconsin)
  - SPGL1 (UBC)....
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
CoSAMP

- Needell and Tropp, 2008
- Very simple greedy algorithm, provably effective

**Algorithm 2.1: CoSaMP Recovery Algorithm**

CoSaMP(Φ, u, s)  
**Input:** Sampling matrix Φ, noisy sample vector u, sparsity level s  
**Output:** An s-sparse approximation a of the target signal

\[
\begin{align*}
a^0 & \leftarrow 0 \quad \{ \text{Trivial initial approximation} \} \\
v & \leftarrow u \quad \{ \text{Current samples = input samples} \} \\
k & \leftarrow 0
\end{align*}
\]

repeat
\[
\begin{align*}
k & \leftarrow k + 1
\end{align*}
\]

\[
\begin{align*}
y & \leftarrow \Phi^* v \quad \{ \text{Form signal proxy} \} \\
\Omega & \leftarrow \text{supp}(y_{2s}) \quad \{ \text{Identify large components} \} \\
T & \leftarrow \Omega \cup \text{supp}(a^{k-1}) \quad \{ \text{Merge supports} \}
\end{align*}
\]

\[
\begin{align*}
b|_T & \leftarrow \Phi^*_T u \quad \{ \text{Signal estimation by least-squares} \} \\
b|_{T^c} & \leftarrow 0
\end{align*}
\]

\[
\begin{align*}
a^k & \leftarrow b_s \quad \{ \text{Prune to obtain next approximation} \} \\
v & \leftarrow u - \Phi a^k \quad \{ \text{Update current samples} \}
\end{align*}
\]

until halting criterion true
From Sparsity
to
Model-based (*structured*) Sparsity
Sparse Models

wavelets: natural images

Gabor atoms: chirps/tones

pixels: background subtracted images
Sparse Models

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

sparse image
Beyond Sparse Models

- Sparse/compressible 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
Signal Priors

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

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

- **Structured sparse** signal: reduced set of subspaces (or model-sparse)
  - model: a particular union of subspaces
    - ex: clustered or dispersed sparse patterns

\[
f(X_i) \leq 1
\]
Restricted Isometry Property (RIP)

- **Model:** $K$-sparse

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

- **RIP:** stable embedding

Random subGaussian (iid Gaussian, Bernoulli) matrix $\prec \succ$ RIP w.h.p.
Restricted Isometry Property (RIP)

- **Model:** $K$-sparse
  - significant coefficients lie on a rooted subtree (a known model for piecewise smooth signals)

- **Tree-RIP:** stable embedding

$$\delta_{\mathcal{M}_{2K}}: M = O(K) < O(K \log(N/K))$$
Restricted Isometry Property (RIP)

- **Model**: $K$-sparse

  Note the difference:

- **RIP**: stable embedding

\[
\delta_{2K} : M = O(K \log(N/K))
\]
Model-Sparse Signals

• Defn: A **K-sparse signal model** comprises a particular (*reduced*) set of $K$-dim canonical subspaces

• **Structured subspaces**

  <> *fewer measurements*

  <> *improved recovery perf.*

  <> *faster recovery*
CS Recovery

- **Iterative Hard Thresholding (IHT)
  [Nowak, Figueiredo; Kingsbury, Reeves; Daubechies, Defrise, De Mol; Blumensath, Davies; ...]

Given $y = \Phi x$, recover a sparse $x$

initialize: $\hat{x}_0 = 0$, $r = y$, $i = 0$

iteration:

  - $i \leftarrow i + 1$
  
  - $b \leftarrow \hat{x}_{i-1} + \Phi^T r$
    update signal estimate
  
  - $\hat{x}_i \leftarrow \text{thresh}(b, K)$
    prune signal estimate
    (best $K$-term approx)
  
  - $r \leftarrow y - \Phi \hat{x}_i$
    update residual

return: $\hat{x} \leftarrow \hat{x}_i$
Model-based CS Recovery

- **Iterative Model Thresholding**
  
  \[ \text{[VC, Duarte, Hegde, Baraniuk; Baraniuk, VC, Duarte, Hegde]} \]

  Given \( y = \Phi x \), recover a model sparse \( x \in M \)

  initialize: \( \hat{x}_0 = 0, r = y, i = 0 \)

  iteration:
  
  \( i \leftarrow i + 1 \)

  \( b \leftarrow \hat{x}_{i-1} + \Phi^T r \)  \hspace{1cm} \text{update signal estimate}

  \( \hat{x}_i \leftarrow M(b, K) \)  \hspace{1cm} \text{prune signal estimate (best } K\text{-term model approx)}

  \( r \leftarrow y - \Phi \hat{x}_i \)  \hspace{1cm} \text{update residual}

  return: \( \hat{x} \leftarrow \hat{x}_i \)
Tree-Sparse Signal Recovery

\[ N = 1024 \]
\[ M = 80 \]

Target signal

CoSaMP, (MSE=1.12)

L1-minimization
(MSE=0.751)

Tree-sparse CoSaMP
(MSE=0.037)
Tree-Sparse Signal Recovery

- Number samples for correct recovery with noise

- Piecewise cubic signals + wavelets

- Plot the number of samples to reach the noise level

\[
O(K \log N)
\]

\[
O(K)
\]
Clustered Sparsity

- **\((K,C)\) sparse signals** (1-D)
  - \(K\)-sparse within at most \(C\) clusters

- For stable recovery

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

- Model approximation using **dynamic programming**

- Includes **block sparsity** as a special case

[VC, Indyk, Hedge, Baraniuk]
Clustered Sparsity

- Model clustering of significant pixels in space domain using **graphical model** (MRF)

- Ising model approximation via **graph cuts**

  [VC, Duarte, Hedge, Baraniuk]
Clustered Sparsity

20% Compression
No performance loss in tracking
Neuronal Spike Trains

- Model the firing process of a single neuron via 1D Poisson process with spike trains
  - stable recovery

\[ M = O \left( K \log\left( \frac{N}{K} - \Delta \right) \right) \]

- Model approximation solution:
  - integer program
    - efficient & provable solution due to total unimodularity of linear constraint
  - dynamic program

[Hegde, Duarte, VC] – best paper award
Performance of Recovery

- Using model-based IHT and CoSaMP

\[ M = \mathcal{O} \left( \log |\mathcal{M}_K| \right) \]

|\mathcal{M}_K| : \# of subspaces

- Model-sparse signals

[Baraniuk, VC, Duarte, Hegde]

\[ \|x - \hat{x}\|_2 \leq C_1 \frac{\|x - x\mathcal{M}_K\|_1}{K^{1/2}} + C_2 \|n\|_2 \]

- Model-compressible signals w/restricted amplification property

\[ \|x - \hat{x}\|_2 \leq C_1 \log \left( \frac{N}{K} \right) \frac{\|x - x\mathcal{M}_K\|_1}{K^{1/2}} + C_2 \|n\|_2 \]
Compressive Sensing

In Action

Cameras
“Single-Pixel” CS Camera

scene

random pattern on DMD array

single photon detector

DMD

image reconstruction or processing

w/ Kevin Kelly
“Single-Pixel” CS Camera

- **Random pattern on DMD array**
- **Single photon detector**
- **Scene**
- **Image reconstruction or processing**

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

Object
LED (light source)
Photodiode circuit
Lens 2
Lens 1
DMD+ALP Board
First Image Acquisition

- Target: 65536 pixels
- 11,000 measurements (16%)
- 1,300 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
Georgia Tech Analog Imager

- Robucci and Hasler 07
- Transforms image *in analog*, reads out transform coefficients
Compressive Sensing Acquisition

10k DCT coefficients

10k random measurements

(Robucci et al 09)
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, ...
The bad news starts at 1 GHz...
ADC State of the Art
From 2008...

Industry’s fastest 16-bit ADC at 200 MSPS

Data Converter
ADS5485

Texas Instruments
ADS5547

Resolution (bits)

16
14
12

50 100 150 200

Speed (MSPS)

Texas Instruments
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

spectrogram

20x sub-Nyquist sampling

sparsogram
Multichannel Random Demodulation

- 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
Compressive Sensing

*In Action*

Data Processing
Information Scalability

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

  detection < classification < estimation < reconstruction

  fairly computationally intense
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
Low-dimensional signal models

$N$ pixels

$K \ll N$

large wavelet coefficients

(blue = 0)

sparse signals

structured sparse signals

parameter manifolds
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

- Target templates from generative model or training data (points)
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
CS for Manifolds

• **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

![Graphs showing the relationship between number of measurements and average shift estimate error, and between number of measurements and classification rate. The graphs demonstrate how more noise affects the performance of the Smashed Filter.]
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)