AN INTRODUCTION TO COMPRESSIVE SENSING

Rodrigo B. Platte

Arizona State University
School of Mathematical and Statistical Sciences

APM/EEE598 Reverse Engineering of Complex Dynamical Networks
INTRODUCTION

INCOHERENCE

RIP

POLYNOMIAL MATRICES

DYNAMICAL SYSTEMS

OUTLINE

1 INTRODUCTION

2 INCOHERENCE

3 RIP

4 POLYNOMIAL MATRICES

5 DYNAMICAL SYSTEMS
The Rice DSP website

Resources for papers, codes, and more ....

http://www.dsp.ece.rice.edu/cs/

References:

m-files and some links are available in the course page
VIDEO LECTURES

Some well known CS people:

- Emmanuel Candès (Stanford University)
  Sequence of papers with Terence Tao and Justin Romberg in 2004.
- David Donoho (Stanford University)
- Richard Baraniuk (Rice University)
- Ronald A. DeVore (Texas A&M)
- Anna C. Gilbert (Univ. of Michigan)
- Jared Tanner (University of Edinburgh)
- ...

A good way to learn the basics of CS is to watch these IMA video lectures:

http://www.ima.umn.edu/videos/

→ IMA New Directions short courses → Compressive Sampling and Frontiers in Signal Processing (two weeks long)
**Underdetermined Systems**

Solve

\[ Ax = b, \]

where \( A \) is \( m \times N \) and \( m < N \).
**Underdetermined systems**

In CS we want to obtain sparse solutions, i.e., \( x_j \approx 0 \), for several \( j \)'s.

Solve

\[
Ax = b,
\]

where \( A \) is \( m \times N \) and \( m < N \).
SOLUTION

DESIGNING A STABLE MEASUREMENT MATRIX

The measurement matrix $\Phi$ plays a crucial role in recovering sparse and compressible signals. Its structure and properties, such as the Restricted Isometry Property (RIP), are essential for the successful recovery of signals with as few as $K$ nonzero entries, where $K < N$. The RIP states that for any $K$-sparse vector $\mathbf{x}$, the following condition holds:

$$
\|A\mathbf{x}\|_2 \approx \|\mathbf{x}\|_2
$$

where $A$ is the measurement matrix, $\|\cdot\|_2$ denotes the Euclidean norm, and $\mathbf{x}$ is the $K$-sparse vector.

To achieve a stable solution, one must ensure that the RIP is satisfied for all possible $K$-sparse vectors. This condition is necessary and sufficient for exact recovery of the signal.

**Example:**

Consider a $3 \times 3$ measurement matrix $\Phi$ with entries $\{0, 1\}$.

\[
\begin{bmatrix}
1 & 1 & 0 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
\end{bmatrix}
\]

This matrix is not RIP-preserving for $K = 2$ since it cannot recover a $2$-sparse vector $\mathbf{x}$ with $\|A\mathbf{x}\|_2 > \|\mathbf{x}\|_2$.

**Remark:**

The location of the nonzero entries in $\mathbf{x}$ is not known in advance, which makes the recovery problem challenging. The RIP condition provides a framework for dealing with this uncertainty.

**One option:**

Minimize $\|x\|_1$ subject to $Ax = b$.

**Why $p = 1$?**

This choice of $p$ is motivated by the fact that $\|x\|_1$ is a measure of sparsity, as it sums the absolute values of the entries in $x$. However, minimizing $\|x\|_1$ is an NP-hard problem due to its combinatorial nature.

**Remark:**

The location of the nonzero entries in $\mathbf{x}$ is not known in advance, which makes the recovery problem challenging. The RIP condition provides a framework for dealing with this uncertainty.
**Why \( \ell_1 \)?**

Unit ball:

\[
\ell_0, \ell_{1/2}, \ell_1, \ell_2, \ell_4, \ell_\infty
\]

\[
\|x\|_{\ell_p} = (|x_0|^p + \cdots + |x_N|^p)^{1/p}
\]

or, for \( 0 \leq p < 1 \),

\[
\|x\|_{\ell_p} = (|x_0|^p + \cdots + |x_N|^p)
\]

- \( \|x\|_{\ell_0} = \) # of nonzero entries in \( x \)
  ideal (?) but leads to a NP-complete problem
- \( \ell_p \), with \( p < 1 \) is not a norm (triangular inequality). Also not practical.
- \( \ell_2 \) computationally easy but does not lead to sparse solutions.
  The unique solution of minimum \( \ell_2 \) norm is (pseudo-inverse)

\[
x = A'(AA')^{-1}b
\]
Sparsity and the $\ell_1$-norm (2D case)

Example

$$a_1 x_1 + a_2 x_2 = b_1$$
**Example — $l_2$**

\[
\min_{x_1, x_2} \sqrt{x_1^2 + x_2^2} \quad \text{subject to} \quad a_1 x_1 + a_2 x_2 = b_1
\]
**Sparsity and the $\ell_1$-norm (2D case)**

**Example – $\ell_1$**

$$\min_{x_1, x_2} |x_1| + |x_2| \quad \text{subject to} \quad a_1 x_1 + a_2 x_2 = b_1$$
Recall Parseval’s Formula:
\[ f(t) = \sum_{k=0}^{N} x_k \phi_k(t), \] with \( \phi_k \) orthonormal in \( L_2 \).

\[ \|f\|_2^2 = \sum_{k=0}^{N} |x_k|^2. \]

Also, \( \ell_2 \) penalizes heavily large values, while small values don’t affect the norm significantly. In general will not give a sparse representation!

See matlab experiment! (Test-l1-l2.m)
MINIMIZING $\|x\|_{\ell_1}$

- Matlab experiment! (Test-l1-l2.m)
- Note: solution may not be unique!
- Solve an optimization problem (in practice $O(N^3)$ operations).
- Several codes are available for CS see:
  http://www.dsp.ece.rice.edu/cs/
\[ f(t) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} x_k \sin(\pi kt) \]

\( N = 1024 \), number of samples: \( m = 50 \)
A SIMPLE EXAMPLE

System of equations:

\[ f(t_j) = \frac{1}{\sqrt{1024}} \sum_{k=1}^{1024} x_k \sin(\pi kt_j), \quad j = 1 \ldots 50 \]

SOLVE:

\[ \min \| x \|_{\ell_1} \quad \text{subject to } Ax = b, \]

where \( A \) has 50 rows and 1024 columns.

\[ A_{j,k} = \frac{1}{\sqrt{1024}} \sin(\pi kt_j), \quad b_j = f(t_j). \]

Matlab code on Blackboard: "SineExample.m" (uses CVX)
Recovery of coefficients is accurate to almost machine precision!

$$\frac{\|x - x_0\|_2}{\|x_0\|_2} = 7.9611 \ldots \times 10^{-11}$$
WHY SPARSITY?

Sparsity is often a good regularization criteria because most signals have structure.

Take a picture! (this one has \(512 \times 512\) pixels)
WHY SPARSITY?

Sparsity is often a good regularization criteria because most signals have structure.

Gray scale please!
Why sparsity?

Sparsity is often a good regularization criteria because most signals have structure.

Find wavelet coefficients. Daubechies(6,2), 3 vanish. moments
Sparsity is often a good regularization criteria because most signals have structure.

Make 75% of the coefficients zero.
<table>
<thead>
<tr>
<th>Introduction</th>
<th>Incoherence</th>
<th>RIP</th>
<th>Polynomial Matrices</th>
<th>Dynamical Systems</th>
</tr>
</thead>
</table>

**Why Sparsity?**

Sparsity is often a good regularization criteria because most signals have structure.

![Image](image_url)

Restored image from 25% of the coefficients.
Sparsity is often a good regularization criteria because most signals have structure.

Relative error $\approx 3\%$. 
Why sparsity?

Sparsity is often a good regularization criteria because most signals have structure.

Keep only 2% of the coefficients, set 98% to zero.
**WHY SPARSITY?**

Sparsity is often a good regularization criteria because most signals have structure.

Reconstructed image from **2%** of the coefficients.
Example: $A$ is a finite difference matrix $A$ maps a sparse vector $x$ into another sparse vector $y$.

\[
\begin{bmatrix}
0 \\
0 \\
\vdots \\
1 \\
-1 \\
0 \\
\vdots
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & 0 & \ldots & 0 \\
-1 & 1 & 0 & \ldots & 0 \\
0 & -1 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & -1 & 1 \\
0 & 0 & \ldots & \ldots & \ldots \\
\end{bmatrix}
\begin{bmatrix}
0 \\
0 \\
\vdots \\
1 \\
0 \\
\vdots
\end{bmatrix}
\]
Example: $A$ is a finite difference matrix. $A$ maps a sparse vector $x$ into another sparse vector $y$.

$$
\begin{bmatrix}
0 \\
0 \\
\vdots \\
1 \\
-1 \\
\vdots \\
0
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
-1 & 1 & 0 & \cdots & 0 \\
0 & -1 & 1 & \cdots & 0 \\
\vdots \\
0 & 0 & \cdots & -1 & 1 \\
\vdots 
\end{bmatrix}
\begin{bmatrix}
0 \\
0 \\
\vdots \\
1 \\
0 \\
\vdots 
\end{bmatrix}
$$

A few samples of $y$ are likely to be all zeros!
Sparsity is not sufficient for CS to work!

The image below is sparse in physical domain and Haar wavelet coefficients.
A general approach

Sample coefficients in a representation by random vectors.

\[ y = \sum_{k=1}^{N} < y, \psi_k > \psi_k, \]

\( \psi_k \) are obtained from orthogonalized Gaussian matrices.

\[ Ax = y \implies \psi' Ax = \psi' y \implies \Theta x = z \]
Incoherence + Sparsity is needed

Incoherence

Sparse representation

Sample here
Incoherence + Sparsity is needed

Incoherence

Assume that $x$ is $S$-sparse and that we are given $K$ Fourier coefficients with frequencies selected uniformly at random. Suppose that the number of observations obeys

$$K \geq C \cdot S \cdot \log N.$$ 

Then minimizing $\ell_1$ reconstructs $x$ exactly with overwhelming probability. In details, if the constant $C$ is of the form $22(\delta + 1)$, then the probability of success exceeds $1 - O(N^{-\delta})$. 

Theorem (Candès, Romberg, Tao)
INCOHERENCE + SPARSITY IS NEEDED

NUMERICAL EXPERIMENT

Signal recovered from Fourier coefficients:

Code "FourierSampling.m".
Let \((\Phi, \Psi)\) be orthonormal bases of \(\mathbb{R}^n\).

\[
f(t) = \sum_{i=1}^{n} x_i \psi_i(t) \quad \text{and} \quad y_k = \langle f, \varphi_k \rangle, \quad k = 1, \ldots, m.
\]

Representation matrix: \(\Psi = [\psi_1 \, \psi_2 \, \cdots \, \psi_n]\)

Sensing matrix: \(\Phi = [\varphi_1 \, \varphi_2 \, \cdots \, \varphi_n]\)

**COHERENCE BETWEEN \(\Phi\) AND \(\Psi\)**

\[
\mu(\Phi, \Psi) = \sqrt{n} \max_{1 \leq j, k \leq n} |\langle \varphi_k, \psi_j \rangle|.
\]

Remark: \(\mu(\Phi, \Psi) \in [1, \sqrt{n}]\)

Upper bound: Cauchy-Schwarz

Lower bound: \(\Psi^T \Phi\) is also orthonormal, hence

\[
\sum |\langle \varphi_k, \psi_j \rangle|^2 = 1 \Rightarrow \max_j |\langle \varphi_k, \psi_j \rangle| \geq 1/\sqrt{n}
\]
A general result for sparse recovery

\[ f(t) = \sum_{i=1}^{n} x_i \psi_i(t) \quad \text{and} \quad y_k = \langle f, \varphi_k \rangle, \quad k = 1, \ldots, m. \]

Consider the optimization problem:

\[ \min_{x \in \mathbb{R}^n} \| x \|_{\ell_1} \quad \text{subject to} \quad y_k = \langle \Psi x, \varphi_k \rangle, \quad k = 1, \ldots, m. \]

**Theorem (Candès and Romberg, 2007)**

Fix \( f \in \mathbb{R}^n \) and suppose that the coefficient sequence \( x \) of \( f \) in the basis \( \psi \) is \( s \)-sparse. Select \( m \) measurements in the \( \Phi \) domain uniformly at random. Then if

\[ m \geq C \mu^2(\Phi, \Psi) S \log(n/\delta) \]

for some positive constant \( C \), the solution of the problem above is exact with probability exceeding \( 1 - \delta \).
Examples of incoherent bases

- $\Phi$ is the identity ($\varphi_k(t) = \delta(t - k)$) and $\Psi$ is the Fourier basis. The time-frequency pair obeys $\mu(\Phi, \Psi) = 1$.
- Noiselets and Haar wavelets have coherence $\sqrt{2}$.
- Random matrices are largely incoherent with any fixed basis $\Psi$ (about $\sqrt{2 \log n}$).

Matlab example: 'measurementsl1.m'
Multiple solutions of min $\ell_1$-norm

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{N} a_k \cos(\pi kt) + \sum_{k=1}^{N} b_k \sin(\pi kt), \quad t \in [-1, 1]$$

Data: $f(-1) = 1$, $f(0) = 1$, $f(1) = 1$
**Multiple solutions of min \(\ell_1\)-norm**

\[ f(t) = \frac{a_0}{2} + \sum_{k=1}^{N} a_k \cos(\pi kt) + \sum_{k=1}^{N} b_k \sin(\pi kt), \quad t \in [-1, 1] \]

Data: \(f(-1) = 1, \ f(0) = 1, \ f(1) = 1\)

even function: \(b_k = 0\)

Solutions of min \(\ell_1\): \(\{a_2 = 1, \ a_k = 0(k \neq 2)\}, \ \{a_4 = 1, \ a_k = 0(k \neq 4)\}\),

---

**Introduction to Compressive Sensing**

**R. Platte**

**Mathematics and Statistics**
How about signals that are not exactly sparse?

**Isometry Constants**

For each $s = 1, 2, \ldots$, define $\delta_s$ of a matrix $A$ as the smallest number such that

$$(1 - \delta_s)\|x\|_{\ell_2}^2 \leq \|Ax\|_{\ell_2}^2 \leq (1 + \delta_s)\|x\|_{\ell_2}^2$$

holds for all $s$-sparse vectors $x$. 

**The Restricted Isometry Property (RIP)**
Theorem (Candès, 2007?)

Assume $\delta_{2s} < \sqrt{2} - 1$. Then

$$x^* := \arg\min_{x \in \mathbb{R}^n} \|x\|_{\ell_1} \quad \text{subject to} \quad y = Ax$$

$$\downarrow$$

$$\|x^* - x\|_{\ell_2} \leq C \|x - x_s\|_{\ell_1} \sqrt{s}.$$  

where $x_s$ is the vector $x$ with all but the largest $s$ components set to 0. If $x$ is $s$-sparse (exactly), then the recovery is exact.
**RIP - basic idea**

- want $A$ to preserve norm of $s$-sparse vectors.
- $\|Ax_1 - Ax_2\|_2^2$ should not be small for $s$-sparse vectors $x$.
- want $0 < c\|x_1 - x_2\|_2^2 \leq \|A(x_1 - x_2)\|_2^2$ for all $s$-sparse $x$.
- If $\delta_2 s = 1$, then $\|Az\|^2 = 0$ for a $2s$-sparse $z$.
  $z = x_1 - x_2$ with $x_1$ and $x_2$ both $s$-sparse.
the theorem above is deterministic
how does one show that column vectors taken from arbitrarily subsets are nearly orthogonal?

isometry constants are shown for random matrices (randomness is back)

for Fourier basis $m \geq C s \log^4 n$

RIP is too conservative (Donoho, Tanner 2010)
Back to Dr. Lai’s dynamical system problem:

\[
\frac{dx}{dt} = F(x(t)),
\]

with

\[
[F(x(t))]_j = \sum_{k_1} \sum_{k_2} \cdots \sum_{k_m} (a_j)^{k_1 k_2 \cdots k_m} x_1^{k_1}(t) \cdots x_m^{k_m}(t)
\]

- This does not fit in classical CS-results.
- monomial basis becomes ill-conditioned even for small powers
- we know condition numbers of Vandermonde depend on where $x$ is evaluated.
- Some CS results are available for orthogonal polynomials.
Orthogonal Polynomials

For Chebyshev polynomials expansions we have that

\[ f(x) \approx \sum_{k=0}^{N} \lambda_k \cos(k \arccos(x)) \]

If we let \( y = \arccos(x) \) or \( x = \cos(y) \),

\[ f(\cos(y)) \approx \sum_{k=0}^{N} \lambda_k \cos(ky) \]

A Chebyshev expansion is equivalent to a cosine expansion on the variable \( y \).

Results carry over from Fourier expansions but with samples chosen independently according to the Chebyshev measure

\[ d_\nu(x) = \pi^{-1}(1 - x^2)^{-1/2} \, dx \]
Rauhut and Ward (2010) proved that the same type sampling applies for Legendre expansions.

How about polynomial expansions as power series?
How about if we choose just a few function values?

- $\Phi_m$: $m$ randomly chosen rows of identity matrix.
- And assume that $x$ is $K$-sparse.
- $t_d$ according to some distribution in $(-1, 1)$.

$$y_m = \begin{bmatrix} f(t_{d_1}) \\ f(t_{d_2}) \\ \vdots \\ f(t_{d_m}) \end{bmatrix}_m = \Phi_m \begin{bmatrix} t_0^0 & t_0^1 & \ldots & t_0^N \\ t_1^0 & t_1^1 & \ldots & t_1^N \\ \vdots & \vdots & \ddots & \vdots \\ t_N^0 & t_N^1 & \ldots & t_N^N \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$
Robert Thompson’s experiments

Each pixel, 50 experiments: choose random polynomial with \( k \) non-zero Gaussian i.i.d coefficients, measure \( m \) samples, attempt to recover polynomial coefficients.

Sampling at Chebyshev points give (very) slightly better results than uniform points.
Consider linear combinations of Chebyshev polynomials:
\[ y = \sum_{i=1}^{N} T_i(t), \quad T_i(t) = \cos(i \arccos(t)) \]

- \( \Phi_m \): \( m \) randomly chosen rows of identity matrix.
- And assume that \( x \) is \( K \)-sparse.
- \( t_d \) according to some distribution in \((-1, 1)\).

\[
\begin{bmatrix}
  f(t_{d1}) \\
  f(t_{d2}) \\
    \vdots \\
  f(t_{dm}) \\
\end{bmatrix}_m = \Phi_m
\begin{bmatrix}
  T_0(t_0) & T_1(t_0) & \ldots & T_N(t_0) \\
  T_0(t_1) & T_1(t_1) & \ldots & T_N(t_1) \\
    \vdots & \vdots & \ddots & \vdots \\
  T_0(t_N) & T_1(t_N) & \ldots & T_N(t_N) \\
\end{bmatrix}_m
\begin{bmatrix}
  x_1 \\
  x_2 \\
    \vdots \\
  x_N \\
\end{bmatrix}
\]
Using Chebyshev basis functions, we realize improvement as $m$ increases.
Robert Thompson’s experiments

- Columns of $C$ are orthogonal.
- All vectors will be distinguishable if we use full $C$.
- If we use less than full $C$, orthogonality is lost, some vectors start to become indistinguishable.
Robert Thompson’s experiments

- What about 2-D polynomials?
- In natural basis: \( f(t, u) = \sum_{i+j=0..Q} x_{ij} t^i u^j \)
- \((t_d, u_d)\) according to some distribution in \((-1, 1) \times (-1, 1)\).

\[
y_m = \begin{bmatrix}
  f(t_{d_1}, u_{d_1}) \\
  f(t_{d_2}, u_{d_2}) \\
  \vdots \\
  f(t_{d_m}, u_{d_m})
\end{bmatrix}_{m} = \Phi_m 
\begin{bmatrix}
  1 & t_0 & u_0 & t_0u_0 & t_0^2 & u_0^2 & \ldots \\
  1 & t_1 & u_1 & t_1u_1 & t_1^2 & u_1^2 & \ldots \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
  1 & t_N & u_N & t_Nu_N & t_N^2 & u_N^2 & \ldots \\
\end{bmatrix} 
\begin{bmatrix}
  x_{00} \\
  x_{10} \\
  x_{01} \\
  x_{11} \\
  x_{20} \\
  \vdots
\end{bmatrix}
\]
Robert Thompson’s experiments

- Similar to 1-d results.
- Again increasing $m$ doesn’t change much.
ROBERT THOMPSON’S experiments (BACK TO DYNAMICAL SYSTEMS)

- $x_{n+1} = f(x_n) = rx_n(1 - x_n)$
- Coefficient vector: $(0, r, -r, 0, \ldots)$
- We can recover the system equation in chaotic regime taking about 10 sample pairs or more.

![Sampling the logistic map, $m = 10$](image1)

![Recovery error for logistic map, $r = 3.7$](image2)
Robert Thompson’s experiments (back to dynamical systems)

- Sensitive to the dynamics determined by $r$.
- (Bifurcation diagram: Wikipedia).
Final Remarks

- As previously pointed by Dr. Lai – recovery seems impractical with monomial basis of large degree. Change of basis to orthogonal polynomials result in full coefficients.
- Considering small degree expansions in high dimensions – what is the optimal sampling strategy?
- How about a system of PDEs? For example,

\[
\begin{align*}
    u_t &= u(1 - u) - uv + \triangle u \\
    v_t &= v(1 - v) + uv + \triangle v
\end{align*}
\]

- Thanks! In particular to Robert Thompson and Wen Xu.