Compressed sensing: Variations on a theme

Dustin G. Mixon

AFIT
Air Force Institute of Technology

Mathematical Physics meets Sparse Recovery
April 13 – 19, 2014
WARNING:
Prelude

Image compression and denoising

---

Scriabin, Op. 11, No. 22
image from en.wikipedia.org
Big idea: Natural images have low entropy

Lena

DWT of Lena

Model: Wavelet decomposition has few significant components
Model (more or less):

\[
\left\{ \text{natural images} \right\} \subseteq \left\{ W\alpha : \|\alpha\|_0 \leq K \right\}
\]

The set of $K$-sparse vectors is a union of $K$-dimensional subspaces.

How else can we leverage this low dimensionality?
Denoising: Given $y = x + e$, estimate $x \in S$

Recipe: Pick the closest point in $S$ to $y$ (think least squares)

---

Barbara + noise
SNR = 10dB

Barbara denoised
Shrinkage in DWT domain

Donoho, Johnstone, Biometrika, 1994
images from webee.technion.ac.il/people/orih/
A conventional MRI scan can take up to 2 hours

Donoho’s Question (paraphrased)
Why go to so much effort to acquire all the data when most of it will be thrown away in compression? Can’t we just directly measure the important part?

images from newscenter.philips.com and brandonremler.blogspot.com
1st movement

Compressed sensing

\[ \text{\( \downarrow = \text{ca. 90} \) } \]

\[ \text{\( \text{\( p \)} \) } \]
An experiment

An experiment

Intuition

You want to solve an NP-hard problem:

$$\minimize \|\alpha\|_0 \quad \text{s.t.} \quad AW\alpha = y$$

Instead, relax to a convex problem:

$$\minimize \|\alpha\|_1 \quad \text{s.t.} \quad AW\alpha = y$$

The pointy $\ell_1$ ball ensures a sparse minimizer
A toy example

\[ \Phi \alpha = \begin{bmatrix}
13 & -4 & 6 & 4 & 11 \\
-1 & -11 & -3 & -15 & 5 \\
4 & -31 & -2 & -8 & -13 \\
\end{bmatrix}
\begin{bmatrix}
0 \\
1 \\
0 \\
0 \\
2 \\
\end{bmatrix}
= \begin{bmatrix}
18 \\
-1 \\
-57 \\
\end{bmatrix}
= y
\]

\[ \| \alpha \|_0 \text{ s.t. } \Phi \alpha = y \]

\[ \| \alpha \|_1 \text{ s.t. } \Phi \alpha = y \]
Déjà vu?

Minimizing $\|\alpha\|_1$ s.t. $\Phi \alpha = y$ is an old trick for

- deconvolution
- regression
- various sparse approximation problems

But these applications are invariably stuck with a particular $\Phi$, and the performance guarantees suffer accordingly

Taylor, Bank, McCoy, Geophys., 1979
Levy, Fullagar, Geophys., 1981
The freedom of choice

The Philosophy of Compressed Sensing

Choose the matrix $\Phi$ so as to

- innovate the sensing process
- obtain optimal guarantees for convex recovery

MRI: Nyquist sampling vs. 10$\times$ undersampling w/CS

Single-pixel camera for photon-limited imaging

---

The freedom of choice

The Philosophy of Compressed Sensing

Choose the matrix $\Phi$ so as to

- innovate the sensing process
- obtain optimal guarantees for convex recovery

\[ \ell_1 \text{ recovery } \iff \text{null space of } \Phi \text{ is slanted properly} \]

(convex geometry, random matrix theory, probabilistic method)
Different types of guarantees

Non-uniform guarantee
Fix $\alpha$, draw $\Phi$ at random — when is the null space good?

Number of rows in $\Phi$ vs. sparsity level $K$

Phase transition derived using conic geometry

Different types of guarantees

**Uniform guarantee**
When is the null space of $\Phi$ good for every $K$-sparse vector simultaneously? (This is possible!)

**Definition**
An $M \times N$ matrix $\Phi$ has the $K$-null space property if every nonzero $h$ in the null space of $\Phi$ satisfies $\|h_S\|_1 < \|h_{S^c}\|_1$ for every $S \subseteq \{1, \ldots, N\}$ with $|S| \leq K$.

**Theorem**
*Every $K$-sparse vector $x$ uniquely minimizes $\|\tilde{x}\|_1$ s.t. $\Phi\tilde{x} = \Phi x$ if and only if $\Phi$ satisfies the $K$-null space property.*

Different types of guarantees

Uniform guarantee with stability

Definition
A matrix $\Phi$ has the $(K, \delta)$-restricted isometry property if

$$(1 - \delta)\|x\|^2 \leq \|\Phi x\|^2 \leq (1 + \delta)\|x\|^2$$

whenever $x$ has at most $K$ nonzero entries.

Theorem
Let $\Phi$ be $(2K, \delta)$-RIP with $\delta \leq \sqrt{2} - 1$. Then given $y = \Phi x + z$ with $\|z\|_2 \leq \epsilon$,

$$\hat{x} = \arg\min \|\tilde{x}\|_1 \text{ s.t. } \|y - \Phi \tilde{x}\|_2 \leq \epsilon$$

obeys

$$\|\hat{x} - x\|_2 \leq \frac{C_0}{\sqrt{K}}\|x - x_K\|_1 + C_1 \epsilon.$$
Intuition: A $(2K, \delta)$-RIP matrix simultaneously preserves all pairwise distances between $K$-sparse vectors.
**Definition**
A random $M \times N$ matrix $P$ is an $(\varepsilon, \eta)$-Johnson–Lindenstrauss projection if for every $x \in \mathbb{R}^N$,

$$(1 - \varepsilon)\|x\|^2 \leq \|Px\|^2 \leq (1 + \varepsilon)\|x\|^2$$

with probability $\geq 1 - \eta$.

**Theorem**

(a) A JL projection is RIP with high probability.

(b) Randomly negating columns of an RIP matrix yields a JL projection.

Removing some choice

Real-world sensors must satisfy certain design constraints

Many random matrix constructions satisfy RIP:

- JL projections (e.g., iid subgaussian entries)
- Random rows of DFT
- Random samples of a bounded orthogonal system
- Rows of a circulant matrix with Bernoulli seed
- Gabor synthesis matrix with Bernoulli seed
- Consecutive entries of the Legendre symbol

References:

Rauhut, Theoretical Foundations and Numerical Methods for Sparse Recovery, 2010
Bandeira, Fickus, M., Moreira, in preparation
Open problems

Removing more choice, getting closer to the real world

- Deterministic RIP matrices
  - finding hay in a haystack

- Coherence in the sensing matrix
  - super-resolution, regression

- Other signal classes
  - sparsity in redundant dictionaries
And now for something completely different

In 2006, Netflix offered a US$1M prize to improve its movie rating prediction algorithm

Put movie ratings in a matrix, #movies by #users

Most entries are blank ... Can you guess what they would be?
An experiment

<table>
<thead>
<tr>
<th>⭐⭐⭐⭐⭐</th>
<th>⭐⭐⭐⭐⭐</th>
<th>⭐⭐</th>
<th>⭐⭐⭐⭐⭐</th>
<th>⭐</th>
<th>⭐⭐⭐⭐⭐</th>
<th>⭐⭐⭐⭐⭐</th>
</tr>
</thead>
<tbody>
<tr>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
</tbody>
</table>
An experiment

Overview:

- $5 \times 7$ matrix of ratings
- Singular values: 19.52, 8.04, 2.83, 1.53, 0.00
- Keep the top 3 singular values $\Rightarrow$ Correct rankings $\pm 1$ star

Conclusion: Movie rankings have (approximately) low rank
2nd movement

Low-rank matrix completion and recovery

Mozart, Piano Sonata No. 16 in C major
A familiar solution

Consider the vector space of $m \times n$ matrices

Reported entries $\mathcal{A}(X)$ are linear in the underlying matrix $X$

You want to solve an NP-hard problem:

$$\text{minimize } \text{rank } X \quad \text{s.t. } \mathcal{A}(X) = y$$

Instead, relax to a convex problem:

$$\text{minimize } \|X\|_* \quad \text{s.t. } \mathcal{A}(X) = y$$
Non-uniform guarantees

Random entries fail to distinguish $e_1 e_1^*$ from the zero matrix

Incoherence between $X$ and the sampling basis is necessary

**Theorem (informally)**

Take a basis $\{W_i\}_{i=1}^{n^2}$ of the space of $n \times n$ matrices, and randomly select a subcollection $\Omega$ of this basis to sample with:

$$(A(X))[i] = \text{Tr}[W_i^* X] \quad \forall i \in \Omega.$$  

Then nuclear-norm minimization recovers $X$ whp provided

- the sampling basis is sufficiently incoherent with $X$
- $|\Omega| = O(nr \log^2 n)$

Also: Phase transition if we measure with iid Gaussian matrices

---

Another application: Phase retrieval

What does the diffraction pattern say about the object?

---

image from optics.rochester.edu/workgroups/fienup/
Another application: Phase retrieval

Diffraction pattern can shed light on nanoscale structures:

- 1962 Nobel Prize (Watson, Crick, Wilkins) Deduced DNA’s double helix structure
- 1985 Nobel Prize (Hauptman, Karle) Ad hoc “shake-and-bake” algorithm determined structures of small proteins and antibiotics
Another application: Phase retrieval

Modern goal: Find a way to systematically win Nobel Prizes to recover the object $u$ from its diffraction pattern $|Fu|^2$.

The phase retrieval step is severely underdetermined, so more information is necessary:

- A priori knowledge about object
- Additional measurements
Another application: Phase retrieval

Modulate the X-rays to change the object’s appearance

Claim: If chosen properly, masks $\{\mu_i\}_{i=1}^P$ give complete information

To solve: $|\Phi^* x|^2 \mapsto x$, where $\Phi^* = [\mathcal{F}\mu_1; \mathcal{F}\mu_2; \ldots; \mathcal{F}\mu_P]$
Another application: Phase retrieval

Let $\Phi = \{\varphi_n\}_{n=1}^{N}$ be arbitrary (Philosophy of CS)

Define $A(x) := |\Phi^* x|^2 = \{|\langle x, \varphi_n \rangle|^2\}_{n=1}^{N}$

Goal: Recover any $x$ up to global phase from $A(x)$
Another application: Phase retrieval

Let $\Phi = \{\varphi_n\}_{n=1}^{N}$ be arbitrary (Philosophy of CS)

Define $A(x) := |\Phi^* x|^2 = \{|\langle x, \varphi_n \rangle|^2\}_{n=1}^{N}$

Goal: Recover any $x$ up to global phase from $A(x)$

How to lift:

$$|\langle x, \varphi_n \rangle|^2 = x^* \varphi_n \varphi_n^* x = \text{Tr}[x^* \varphi_n \varphi_n^* x] = \text{Tr}[xx^* \varphi_n \varphi_n^*]$$

Overall, each $|\langle x, \varphi_n \rangle|^2$ is a linear measurement of $X := xx^*$, and we seek to recover the rank-1 matrix $X$ from these measurements

---

Another application: Phase retrieval

Solution: **PhaseLift** (i.e., nuclear-norm minimization)

Guarantees for PhaseLift (iid Gaussian measurement vectors):
- Non-uniform with stability, $O(N \log N)$ measurements
- Uniform, $O(N)$ measurements (stability is open)

Guarantees with randomly masked Fourier transforms:
- Non-uniform, $O(\log^2 N)$ masks (stability is open)
- Uniform, $O(\log N)$ masks, but not PhaseLift (stability is open)

---

Bandeira, Chen, M., Inform. Inference, to appear
Open problems

Removing randomness from sensing process
  ▶ Uniform guarantee with stability for masked Fourier transforms
  ▶ Phase retrieval from short-time Fourier transform magnitudes

Lifted space is huge, SDPs are slow
  ▶ Simplex method for SDPs?
  ▶ Speedups for special cases?
3rd movement

Blind deconvolution and calibration

Haydn, Symphony No. 65
Convolution of a rapidly decaying function with a sparse function:

Do you think you can separate them?

This is called **blind deconvolution**
Blind deconvolution

Problem: Given $x \ast y$, determine $x \in S$ and $y \in T$

Note that each entry of $x \ast y$ is a bilinear form $B_n(x, y)$

$$(x \ast y)[n] = y^\top B_n x = \text{Tr}[y^\top B_n x] = \text{Tr}[xy^\top B_n]$$

Thus, entries of $x \ast y$ are linear measurements of $xy^\top$

Theorem (Informally)

Let $S$ and $T$ be random subspaces of sufficiently small dimension. Then for any $x \in S$ and $y \in T$, nuclear-norm minimization recovers $x$ and $y$ up to a scalar factor with high probability.

Now suppose you are trying to find a sparse $x$ such that $y = Ax$

Sometimes, you only know $A$ up to a parametrized family $\{A(\theta)\}$

This is called **calibration**
Now suppose you are trying to find a sparse $x$ such that $y = Ax$

Sometimes, you only know $A$ up to a parametrized family $\{A(\theta)\}$

This is called **calibration**

If $A(\theta)$ is a linear, then each entry of $A(\theta)x$ is a bilinear form
Another application of “bilinear compressed sensing”

Shake your camera while taking a picture

\[ \text{State-of-the-art blind deconvolution:} \]

Ito, Sankaranarayanan, Veeraraghavan, Baraniuk, ACM Trans. Graph., submitted
Another application of “bilinear compressed sensing”

New idea: Take multiple exposures in rapid succession

Each exposure is blurred by a different sparse kernel $k_i$

**BlurBurst:**

$$\text{minimize} \sum_i \left\| y_i - x \ast k_i \right\|_2^2 + \lambda_1 \left\| x \right\|_{TV} + \lambda_2 \sum_i \left\| k_i \right\|_1$$

Better performance?

Ito, Sankaranarayanan, Veeraraghavan, Baraniuk, ACM Trans. Graph., submitted
Another application of “bilinear compressed sensing”

Ito, Sankaranarayanan, Veeraraghavan, Baraniuk, ACM Trans. Graph., submitted
Open problems

Performance guarantees for bilinear compressed sensing

Can BlurBurst be improved using nuclear-norm minimization?

What are the right questions to ask?
Questions?

Shameless plug

google short fat matrices for my research blog