Compressed Sensing

David Donoho (2006)
Dept. of Statistics
Stanford University

Presented by Alankrita Bhatt
CSE 254A
October 25, 2018
Problem Statement

- Interested in measuring a signal $x \in \mathbb{R}^m$
  - Could be a physical quantity of interest
  - Can we reconstruct with less than $m$ measurements?
Problem Statement

- Interested in measuring a signal $x \in \mathbb{R}^m$
  - Could be a physical quantity of interest
  - Can we reconstruct with less than $m$ measurements?
  - In general, obviously not.

$||x||_p \leq R$ for some $p \in (0, 2)$ and $R > 0$

Can we then do better than $m$ measurements?
Problem Statement

- Interested in measuring a signal $x \in \mathbb{R}^m$
  - Could be a physical quantity of interest
  - Can we reconstruct with less than $m$ measurements?
  - In general, obviously not.

- What if the signal is sparse?

$$\|x\|_p = \left( \sum_{i=1}^{m} |x_i|^p \right)^{\frac{1}{p}} \leq R$$

for some $p \in (0, 2)$ and $R > 0$
Problem Statement

- Interested in measuring a signal $x \in \mathbb{R}^m$
  - Could be a physical quantity of interest
  - Can we reconstruct with less than $m$ measurements?
  - In general, obviously not.

- What if the signal is sparse?
  \[ ||x||_p = \left( \sum_{i=1}^{m} |x_i|^p \right)^{\frac{1}{p}} \leq R \]
  for some $p \in (0, 2)$ and $R > 0$

- Can we then do better than $m$ measurements?
Recovering Sparse $x$

- $X_{p,m}(R)$ is the class of signals of interest

\[ X_{p,m}(R) = \{ x : ||x||_p \leq R \} \]
Recovering Sparse $x$

$X_{p,m}(R)$ is the class of signals of interest

$$X_{p,m}(R) = \{ x : \|x\|_p \leq R \}$$

**Punchline:** If $\Phi$ is a $n \times m$ random matrix (here $n << m$), we can recover $x$ from $\Phi x$
Recovering Sparse $x$

- $X_{p,m}(R)$ is the class of signals of interest

$$X_{p,m}(R) = \{x : \|x\|_p \leq R\}$$

- **Punchline**: If $\Phi$ is a $n \times m$ random matrix (here $n << m$), we can recover $x$ from $\Phi x$

**Theorem 1.7 of [Donoho2006]**

∃ a reconstruction operator $A_n$ running in polynomial time such that for matrices $\Phi$ satisfying certain properties and a constant $C_p$, we have the minmax error

$$E_n(X_{m,p}(R)) := \inf_{A_n} \sup_{x \in X_{m,p}(R)} \|x - A_n(\Phi x)\|_2 \leq C_p \cdot R \cdot (n/\log m)^{1/2 - 1/p}$$
Recovering Sparse $x$

- $X_{p,m}(R)$ is the class of signals of interest

$$X_{p,m}(R) = \{ x : ||x||_p \leq R \}$$

- **Punchline:** If $\Phi$ is a $n \times m$ random matrix (here $n \ll m$), we can recover $x$ from $\Phi x$

---

**Theorem 1.7 of [Donoho2006]**

There exists a reconstruction operator $A_n$ running in polynomial time such that for matrices $\Phi$ satisfying certain properties and a constant $C_p$, we have the minmax error

$$E_n(X_{m,p}(R)) := \inf_{A_n} \sup_{x \in X_{m,p}(R)} ||x - A_n(\Phi x)||_2 \leq C_p \cdot R \cdot (n/\log m)^{1/2-1/p}$$

- If $x_N$ denotes the vector with everything except $N$ largest components set to 0 then

$$||x - x_N||_2 \leq \zeta_{2,p} \cdot ||x||_p \cdot (N + 1)^{1/2-1/p}$$
Recovering Sparse $x$

- $X_{p,m}(R)$ is the **class of signals** of interest

$$X_{p,m}(R) = \{ x : ||x||_p \leq R \}$$

- If $\Phi$ is a $n \times m$ random matrix, we can still recover $x$ from $\Phi x$

**Theorem 1 of [Donoho2006]**

$\exists$ a (possibly nonlinear) reconstruction operator $A_n$ running in **polynomial time** such that for matrices $\Phi$ satisfying certain properties and a constant $C_p$, we have the minmax error

$$E_n(X_{m,p}(R)) := \inf_{A_n} \sup_{x \in X_{m,p}(R)} ||x - A_n(\Phi x)||_2 \leq C_p \cdot R \cdot \left( \frac{n}{\log m} \right)^{1/2 - 1/p}$$

- If $x_N$ denotes the vector with everything except $N$ largest components set to 0 then

$$||x - x_N||_2 \leq \zeta_{2,p} \cdot ||x||_p \cdot (N + 1)^{1/2 - 1/p}$$
Recovering Sparse $x$

- $X_{p,m}(R)$ is the class of signals of interest

$$X_{p,m}(R) = \{x : \|x\|_p \leq R\}$$

- If $\Phi$ is a $n \times m$ random matrix, we can still recover $x$ from $\Phi x$

**Theorem 1 of [Donoho2006]**

\[ E_n(X_{m,p}(R)) := \inf_{A_n} \sup_{x \in X_{m,p}(R)} \|x - A_n(\Phi x)\|_2 \leq C_p \cdot R \cdot (n/\log m)^{1/2-1/p} \]

- If $x_N$ denotes the vector with everything except $N$ largest components set to 0 then

$$\|x - x_N\|_2 \leq \zeta_{2,p} \cdot \|x\|_p \cdot (N+1)^{1/2-1/p}$$

- Taking $\approx N \log(m)$ pieces of nonadaptive information comparable to this

- Matching information-theoretic lower bound shows that the $(n/\log m)^{1/2-1/p}$ scaling is optimal
Recovering sparse $x$

A quick illustration

- Let $x$ be such that $\|x\|_0 \leq k$

- Then asking an oracle to give us the $k$ largest coefficients of $x$ recovers it exactly
Recovering sparse $x$

A quick illustration

- Let $x$ be such that $\|x\|_0 \leq k$

- Then asking an oracle to give us the $k$ largest coefficients of $x$ recovers it exactly

- As per the result above, $\approx k \log m$ measurements given by $\Phi x$ can also exactly recover $x$!
Recovering sparse $x$

A quick illustration

- Let $x$ be such that $||x||_0 \leq k$

- Then asking an oracle to give us the $k$ largest coefficients of $x$ recovers it exactly

- As per the result above, $\approx k \log m$ measurements given by $\Phi x$ can also exactly recover $x$!

Two main questions:

- What are the special properties of this matrix $\Phi$?

- What are the polynomial time methods to recover $x$ from $\Phi x$?
Part I: Sampling operators $\Phi$
Constructing optimal sampling operators

- Consider a matrix $\Phi$ the following three “good” properties

1. **CS–1**: Minimal singular value of submatrices (with $\rho n / \log(m)$ columns) is $\eta > 0$ — Quantifies linear independence

2. **CS–2**: For any $v \in$ the subspace spanned by each submatrix (with $\rho n / \log(m)$ columns), we have $\|v\|_2 \leq c \sqrt{n} \|v\|_1$ — Note that we always have $\sqrt{n} \|v\|_1 \leq \|v\|_2$.

3. **CS–3**: (A technical condition on the quotient norm)

Theorem 7 of [Donoho2006]

For a $\Phi$ satisfying CS–1-3 and a constant $C$,

$$\inf_{A_n} \sup_{x \in X^m, p \in \mathbb{R}} \left\| x - A_n(\Phi x) \right\|_2 \leq C \left( \frac{n}{\log m} \right)^{1/2 - 1/2}$$

$$\left( \frac{7}{15} \right)$$
Constructing optimal sampling operators

- Consider a matrix $\Phi$ the following three “good” properties
  - CS–1: Minimal singular value of submatrices (with $< \rho n / \log(m)$ columns) is $> \eta_1 > 0$
    — Quantifies linear independence

Theorem 7 of [Donoho2006]

For a $\Phi$ satisfying CS–1–3 and a constant $C$, $
\inf A \sup x \in X m, p (R) ||x - A_n(\Phi x)||_2 \leq C \cdot \left( \frac{n}{\log m} \right)^{1/2 - 1/p} \frac{7}{15}$
Consider a matrix $\Phi$ the following three “good” properties

- **CS–1:** Minimal singular value of submatrices (with $< \rho n / \log(m)$ columns) is $> \eta_1 > 0$
  
  —Quantifies linear independence

- **CS–2:** For any $v \in$ the subspace spanned by each submatrix (with $< \rho n / \log(m)$ columns), we have
  
  $||v||_2 \leq \frac{c}{\sqrt{n}} ||v||_1$

- **CS–3:** (A technical condition on the quotient norm)
  
  Theorem 7 of [Donoho2006] For a $\Phi$ satisfying CS–1–3 and a constant $C$

  \[
  \inf_A \sup_{x \in \mathbb{R}^m} \left| \left| x - A_n(\Phi x) \right| \right|_2 \leq C \left( \frac{n}{\log m} \right)^{1/2 - 1/p}
  \]
Consider a matrix $\Phi$ the following three “good” properties

- **CS–1:** Minimal singular value of submatrices (with $< \rho n / \log(m)$ columns) is $> \eta_1 > 0$ —Quantifies linear independence

- **CS–2:** For any $v \in$ the subspace spanned by each submatrix (with $< \rho n / \log(m)$ columns), we have

$$||v||_2 \leq \frac{c}{\sqrt{n}} ||v||_1$$

—Note that we always have $\frac{1}{\sqrt{n}} ||v||_1 \leq ||v||_2$.
Consider a matrix $\Phi$ the following three “good” properties

- **CS–1:** Minimal singular value of submatrices (with $< \rho n / \log(m)$ columns) is $> \eta_1 > 0$
  —Quantifies linear independence

- **CS–2:** For any $v \in$ the subspace spanned by each submatrix (with $< \rho n / \log(m)$ columns), we have
  $$||v||_2 \leq \frac{c}{\sqrt{n}} ||v||_1$$
  —Note that we always have $\frac{1}{\sqrt{n}} ||v||_1 \leq ||v||_2$.

- **CS–3:** (A technical condition on the quotient norm)
Consider a matrix Φ with the following three “good” properties:

- **CS–1**: Minimal singular value of submatrices (with $< \rho \frac{n}{\log(m)}$ columns) is $> \eta_1 > 0$—Quantifies linear independence
- **CS–2**: For any $v \in$ the subspace spanned by each submatrix (with $< \rho \frac{n}{\log(m)}$ columns), we have
  \[
  \|v\|_2 \leq \frac{c}{\sqrt{n}} \|v\|_1
  \]
  —Note that we always have $\frac{1}{\sqrt{n}} \|v\|_1 \leq \|v\|_2$.
- **CS–3**: (A technical condition on the quotient norm)

**Theorem 7 of [Donoho2006]**

For a Φ satisfying CS–1-3 and a constant $C$,

\[
\inf_{A_n} \sup_{x \in X_{m,p}(R)} \|x - A_n(\Phi x)\|_2 \leq C \cdot (n/ \log m)^{1/2 - 1/p}
\]
Do matrices satisfying the conditions CS–1-3 even exist?
  - Yes! And a random sampling method will almost surely yield a matrix satisfying CS–1-3
Finding a “good” $\Phi$

- Do matrices satisfying the conditions CS–1-3 even exist?
  - Yes! And a random sampling method will almost surely yield a matrix satisfying CS–1-3

- Algorithmically
  - Randomly generate each column of $\Phi$ from the uniform distribution over $S^{n-1}$ (n-dimensional unit sphere)
  - $P(\Phi$ doesn’t satisfy CS–1-3) decreases exponentially in $n$

- This is just one of many ways
Recap so far

- Want to reconstruct $x \in \mathbb{R}^m$ using fewer than $m$ measurements, given that $x$ is sparse

- Construct a matrix $\Phi$ by sampling each column randomly from $S^{n-1}$

- With high probability, we have
  \[
  \inf_{A_n} \sup_{\text{sparse } x} \|x - A_n(\Phi x)\|_2 \leq C \cdot (n/ \log m)^{1/2 - 1/p}
  \]

- Next up: Finding $A_n$ that are near-optimal and have low-complexity implementations
Part II: Reconstruction operators $A_n$
We have the measured $y_n$, and need to find $x \in X_{p,m}(R)$ such that $y_n = \Phi x$

Severely *undetermined* problem
We have the measured $y_n$, and need to find $x \in X_{p,m}(R)$ such that $y_n = \Phi x$

Severely **undetermined** problem

In practice, a good idea is min-$p$ norm solution

$$\text{minimize } ||x||_p \text{ subject to } \Phi x = y_n$$
Reconstruction Kernel

- We have the measured $y_n$, and need to find $x \in X_{p,m}(R)$ such that $y_n = \Phi x$

- Severely **undetermined** problem

- In practice, a good idea is min-$p$ norm solution

$$\text{minimize } ||x||_p \text{ subject to } \Phi x = y_n$$

- Don’t need to know $R$

- Need $p$

- For small $p$, highly **nonconvex** problem
What if we solve an easier problem?

P1: minimize $||x||_1$ subject to $\Phi x = y_n$

convex, well-studied, poly-time methods

Question: Is $\ell_1$ norm a good proxy for the $\ell_p$ norm?

Answer: Yes, if the matrix $\Phi$ satisfies the conditions CS-1–3!

Theorem 9 of [Donoho2006]

Let $y_n = \Phi x_0$. If $\Phi$ satisfies conditions CS-1–3, then a solution $\hat{x}_{1,n}$ of problem P1 satisfies

$||x_0 - \hat{x}_{1,n}||_2 \leq C_p ||x_0||_p \cdot \left(\frac{n}{\log m}\right)^{1/2 - 1/p}$. 

12/15
What if we solve an easier problem?

\[ \text{P1: minimize } ||x||_1 \text{ subject to } \Phi x = y_n \]

convex, well-studied, poly-time methods

Question: Is \( \ell_1 \) norm a good proxy for the \( \ell_p \) norm?

Answer: Yes, if the matrix \( \Phi \) satisfies the conditions CS-1–3!
What if we solve an easier problem?

\[ \text{P1: minimize } \|x\|_1 \text{ subject to } \Phi x = y_n \]

convex, well-studied, poly-time methods

Question: Is \( \ell_1 \) norm a good proxy for the \( \ell_p \) norm?

Answer: Yes, if the matrix \( \Phi \) satisfies the conditions CS-1–3!

**Theorem 9 of [Donoho2006]**

Let \( y_n = \Phi x_0 \). If \( \Phi \) satisfies conditions CS-1–3, then a solution \( \hat{x}_{1,n} \) of problem P1 satisfies

\[ \|x_0 - \hat{x}_{1,n}\|_2 \leq C_p \|x_0\|_p \cdot \left( \frac{n}{\log m} \right)^{1/2-1/p} . \]
A simple illustrative theorem

For \( \Phi \) satisfying CS1–3,

\[ ||v||_2 \leq c \sqrt{n} ||v||_1 \]

Fact 2: If \( v \in \ker(\Phi) \), then

\[ ||v^T||_1 \leq ||v||_1 \frac{1}{2} \]

for \( T \leq n \frac{1}{4} \).

Theorem

If \( \Phi \) satisfies CS1–3, \( \Phi x = y \) and

\[ ||x||_0 \leq n \frac{1}{16} c^2 \]

then \( x \) is the uniquely optimal solution to

\[ P_1 \]

Proof:

Let \( w = x + v \), for \( v \in \ker(\Phi) \). Set \( T = \text{supp}(x) \). We then have

\[ ||w||_1 \geq ||x||_1 - 2 ||v^T||_1 + ||v||_1 2 \text{ (invoking aforementioned fact)} \geq ||x||_1 \]
A simple illustrative theorem

For Φ satisfying CS1–3,

- **Fact 1:** If \( v \in \text{ker}(\Phi) \), then \( ||v||_2 \leq \frac{c}{\sqrt{n}} ||v||_1 \) for a constant \( c \).

- **Fact 2:** If \( v \in \text{ker}(\Phi) \), then \( ||v_T||_1 \leq \frac{||v||_1}{4} \) for \( T \leq \frac{n}{16c^2} \).
A simple illustrative theorem

For $\Phi$ satisfying CS1–3,

- **Fact 1:** If $v \in \ker(\Phi)$, then $||v||_2 \leq \frac{c}{\sqrt{n}} ||v||_1$ for a constant $c$

- **Fact 2:** If $v \in \ker(\Phi)$, then $||v_T||_1 \leq \frac{||v||_1}{4}$ for $T \leq \frac{n}{16c^2}$.

**Theorem**

If $\Phi$ satisfies CS1–3, $\Phi x = y_n$ and $||x||_0 \leq n/16c^2$, then $x$ is the uniquely optimal solution to $P_1$. 
A simple illustrative theorem

For Φ satisfying CS1–3,

- **Fact 1**: If \( v \in \ker(\Phi) \), then \( ||v||_2 \leq \frac{c}{\sqrt{n}} ||v||_1 \) for a constant \( c \)

- **Fact 2**: If \( v \in \ker(\Phi) \), then \( ||v T||_1 \leq \frac{||v||_1}{4} \) for \( T \leq \frac{n}{16c^2} \).

**Theorem**

If Φ satisfies CS1–3, \( \Phi x = y_n \) and \( ||x||_0 \leq n/16c^2 \), then \( x \) is the uniquely optimal solution to \( P_1 \).

**Proof**: Let \( w = x + v \), for \( v \in \ker(\Phi) \). Set \( T = \text{supp}(x) \). We then have
A simple illustrative theorem

For \( \Phi \) satisfying CS1–3,

- **Fact 1:** If \( v \in \ker(\Phi) \), then \( ||v||_2 \leq \frac{c}{\sqrt{n}}||v||_1 \) for a constant \( c \)

- **Fact 2:** If \( v \in \ker(\Phi) \), then \( ||v_T||_1 \leq \frac{||v||_1}{4} \) for \( T \leq \frac{n}{16c^2} \).

**Theorem**

If \( \Phi \) satisfies CS1–3, \( \Phi x = y_n \) and \( ||x||_0 \leq n/16c^2 \), then \( x \) is the uniquely optimal solution to \( P_1 \).

**Proof:** Let \( w = x + v \), for \( v \in \ker(\Phi) \). Set \( T = \text{supp}(x) \). We then have

\[
||w||_1 = ||w_T||_1 + ||w_{\overline{T}}||_1
\]
A simple illustrative theorem

For $\Phi$ satisfying CS1–3,

- **Fact 1:** If $v \in \ker(\Phi)$, then $\|v\|_2 \leq \frac{c}{\sqrt{n}} \|v\|_1$ for a constant $c$.

- **Fact 2:** If $v \in \ker(\Phi)$, then $\|v_T\|_1 \leq \frac{\|v\|_1}{4}$ for $T \leq \frac{n}{16c^2}$.

**Theorem**

If $\Phi$ satisfies CS1–3, $\Phi x = y_n$ and $\|x\|_0 \leq n/16c^2$, then $x$ is the uniquely optimal solution to $P_1$.

**Proof:** Let $w = x + v$, for $v \in \ker(\Phi)$. Set $T = \text{supp}(x)$. We then have

$$
\|w\|_1 = \|w_T\|_1 + \|w_{\overline{T}}\|_1 \\
\geq \|x\|_1 - \|v_T\|_1 + \|v_{\overline{T}}\|_1
$$
A simple illustrative theorem

For $\Phi$ satisfying CS1–3,

- **Fact 1:** If $v \in \ker(\Phi)$, then $||v||_2 \leq \frac{c}{\sqrt{n}}||v||_1$ for a constant $c$

- **Fact 2:** If $v \in \ker(\Phi)$, then $||v_T||_1 \leq \frac{||v||_1}{4}$ for $T \leq \frac{n}{16c^2}$.

**Theorem**

If $\Phi$ satisfies CS1–3, $\Phi x = y_n$ and $||x||_0 \leq n/16c^2$, then $x$ is the uniquely optimal solution to $P_1$.

**Proof:** Let $w = x + v$, for $v \in \ker(\Phi)$. Set $T = \supp(x)$. We then have

\[
||w||_1 = ||w_T||_1 + ||w_{\overline{T}}||_1 \\
\geq ||x||_1 - ||v_T||_1 + ||v_{\overline{T}}||_1 \\
= ||x||_1 - 2||v_T||_1 + ||v||_1 \\
\geq ||x||_1 + \frac{||v||_1}{2} \quad \text{(invoking aforementioned fact)}
\]
A simple illustrative theorem

For \( \Phi \) satisfying CS1–3,

- **Fact 1:** If \( v \in \ker(\Phi) \), then \( \|v\|_2 \leq \frac{c}{\sqrt{n}} \|v\|_1 \) for a constant \( c \)

- **Fact 2:** If \( v \in \ker(\Phi) \), then \( \|v_T\|_1 \leq \frac{\|v\|_1}{4} \) for \( T \leq \frac{n}{16c^2} \).

**Theorem**

If \( \Phi \) satisfies CS1–3, \( \Phi x = y_n \) and \( \|x\|_0 \leq n/16c^2 \), then \( x \) is the uniquely optimal solution to \( P_1 \).

**Proof:** Let \( w = x + v \), for \( v \in \ker(\Phi) \). Set \( T = \text{supp}(x) \). We then have

\[
\|w\|_1 = \|w_T\|_1 + \|w_{\overline{T}}\|_1 \\
\geq \|x\|_1 - \|v_T\|_1 + \|v_{\overline{T}}\|_1 \\
= \|x\|_1 - 2\|v_T\|_1 + \|v\|_1 \\
\geq \|x\|_1 + \frac{\|v\|_1}{2} \quad \text{(invoking aforementioned fact)} \\
\geq \|x\|_1.
\]
Some real-world applications

- MRI: Most significantly impacted—scan times are now much shorter

- Photography: Used in mobile phone camera sensors, can reduce image acquisition energy by as much as a factor of 15

- Electron microscopy, radio astronomy, ....
Set out to understand

\[ E_n(X_m, p(R)) := \inf_{A_n, \Phi} \sup_{x \in X_m, p(R)} ||x - A_n(\Phi x)||_2 \]
Set out to understand

\[ E_n(X_m, p(R)) := \inf_{A_n, \Phi} \sup_{x \in X_m, p(R)} \| x - A_n(\Phi x) \|_2 \]

Established the fundamental limit

\[ E_n(X_m, p(R)) = \Theta((n/ \log(m))^{1/2 - 1/p}) \]
Summary

- Set out to understand

\[ E_n(X_m,p(R)) := \inf_{A_n, \Phi} \sup_{x \in X_m,p(R)} \|x - A_n(\Phi x)\|_2 \]

- Established the fundamental limit

\[ E_n(X_m,p(R)) = \Theta((n/\log(m))^{1/2 - 1/p}) \]

- Listed some properties that characterize “good” \( \Phi \) achieving this limit

- Listed explicit constructions of good \( \Phi \)—one way is to randomly sample columns from \( S^{n-1} \)
Summary

- Set out to understand

\[ E_n(X_m, p(R)) := \inf_{A_n, \Phi} \sup_{x \in X_m, p(R)} ||x - A_n(\Phi x)||_2 \]

- Established the fundamental limit

\[ E_n(X_m, p(R)) = \Theta((n/\log(m))^{1/2 - 1/p}) \]

- Listed some properties that characterize “good” \( \Phi \) achieving this limit

- Listed explicit constructions of good \( \Phi \)—one way is to randomly sample columns from \( S^{n-1} \)

- Found a reconstruction operator \( A_n \) implementable in poly-time, \( A_n \) solves

\[ P1: \text{minimize } ||x||_1 \text{ subject to } \Phi x = y_n \]

- The above combination of \( \Phi \) and \( A_n \) achieves the fundamental limit