Sparse Recovery

Dhanus M Lal

April 29, 2022

Introduction

- The number of linear measurements required to completely understand a signal in $\mathbb{R}^n$ is n.

Introduction

- The number of linear measurements required to completely understand a signal in $\mathbb{R}^n$ is n.
- If it is known that all but the first coordinate of it is zero, only one measurement, $y = \langle e_1, x \rangle$ is required.

Introduction

- The number of linear measurements required to completely understand a signal in $\mathbb{R}^n$ is n.
- If it is known that all but the first coordinate of it is zero, only one measurement, $y = \langle e_1, x \rangle$ is required.
- If we know that at most one coordinate is non-zero, can we learn the signal with fewer than n measurements?

Introduction

- The number of linear measurements required to completely understand a signal in $\mathbb{R}^n$ is n.
- If it is known that all but the first coordinate of it is zero, only one measurement, $y = \langle e_1, x \rangle$ is required.
- If we know that at most one coordinate is non-zero, can we learn the signal with fewer than n measurements?
- Interesting result is that such a signal can be recovered in $\mathcal{O}(\log n)$ measurements.

Introduction

- The number of linear measurements required to completely understand a signal in $\mathbb{R}^n$ is n.
- If it is known that all but the first coordinate of it is zero, only one measurement, $y = \langle e_1, x \rangle$ is required.
- If we know that at most one coordinate is non-zero, can we learn the signal with fewer than n measurements?
- Interesting result is that such a signal can be recovered in $\mathcal{O}(\log n)$ measurements.

Definition (Sparsity)

The number of non-zero coordinates of a signal in $\mathbb{R}^n$. Denoted by $\|x\|_0$.

Sparse recovery problem

- Recover s–sparse signal x from linear measurement $y = Ax$, where $A \in \mathbb{R}^{m \times n}$ known.

Sparse recovery problem

- Recover s–sparse signal x from linear measurement $y = Ax$, where $A \in \mathbb{R}^{m \times n}$ known.
- Are there measurement matrices $A \in \mathbb{R}^{m \times n}$ that make recovery possible?

Sparse recovery problem

- Recover s–sparse signal x from linear measurement $y = Ax$, where $A \in \mathbb{R}^{m \times n}$ known.
- Are there measurement matrices $A \in \mathbb{R}^{m \times n}$ that make recovery possible?
- Answer is yes!

Sparse recovery problem

- Recover s–sparse signal x from linear measurement $y = Ax$, where $A \in \mathbb{R}^{m \times n}$ known.
- Are there measurement matrices $A \in \mathbb{R}^{m \times n}$ that make recovery possible?
- Answer is yes!

Proposition

If $A \in \mathbb{R}^{m \times n}$ has $m > 2s$ and every subset of $2s$ columns of A is linearly independent, then sparse recovery problem has a unique solution.

Sparse recovery problem

- Recover s–sparse signal x from linear measurement $y = Ax$, where $A \in \mathbb{R}^{m \times n}$ known.
- Are there measurement matrices $A \in \mathbb{R}^{m \times n}$ that make recovery possible?
- Answer is yes!

Proposition

If $A \in \mathbb{R}^{m \times n}$ has $m > 2s$ and every subset of $2s$ columns of A is linearly independent, then sparse recovery problem has a unique solution.

Proof.

If x_1 and x_2 are two s–sparse vectors with $Ax_1 = Ax_2$, then $A(x_1 - x_2) = 0$. Now, $x_1 \neq x_2$ would mean that a subset of $2s$ columns is linearly dependent. $\square$

Basis pursuit

- If solution is unique, the following optimization problem gives the actual signal.

$$x = \arg\min \|x'\|_0 \quad \text{s.t } Ax' = y$$

Basis pursuit

- If solution is unique, the following optimization problem gives the actual signal.

$$x = \arg\min \|x'\|_0 \quad \text{s.t } Ax' = y$$

- Solving this is computationally hard: requires solving $\binom{n}{s}$ linear systems.

Basis pursuit

- If solution is unique, the following optimization problem gives the actual signal.

$$x = \arg\min \|x'\|_0 \quad \text{s.t } Ax' = y$$

- Solving this is computationally hard: requires solving $\binom{n}{s}$ linear systems.
- Solve the convex relaxation instead

$$\hat{x} = \arg\min \|x'\|_1 \quad \text{s.t } Ax' = y$$

Basis pursuit

- If solution is unique, the following optimization problem gives the actual signal.

$$x = \arg\min \|x'\|_0 \quad \text{s.t } Ax' = y$$

- Solving this is computationally hard: requires solving $\binom{n}{s}$ linear systems.
- Solve the convex relaxation instead

$$\hat{x} = \arg\min \|x'\|_1 \quad \text{s.t } Ax' = y$$

- This is a convex optimization problem called basis pursuit, it is computationally efficient.

Basis pursuit

- If solution is unique, the following optimization problem gives the actual signal.

$$x = \arg\min \|x'\|_0 \quad \text{s.t } Ax' = y$$

- Solving this is computationally hard: requires solving $\binom{n}{s}$ linear systems.
- Solve the convex relaxation instead

$$\hat{x} = \arg\min \|x'\|_1 \quad \text{s.t } Ax' = y$$

- This is a convex optimization problem called basis pursuit, it is computationally efficient.
- Questions:
  - How close is $\hat{x}$ to x?
  - Exact recovery when?

We will now sample measurement matrices from the class $\mathcal{A}(m, n)$

Definition

$\mathcal{A}(m, n)$ is the class of random matrices in $\mathbb{R}^{m \times n}$ with isotropic, sub-gaussian, independent rows.

We will now sample measurement matrices from the class $\mathcal{A}(m, n)$

Definition

$\mathcal{A}(m, n)$ is the class of random matrices in $\mathbb{R}^{m \times n}$ with isotropic, sub-gaussian, independent rows.

A random vector $X \in \mathbb{R}^n$ is:

Isotropic if $\mathbb{E}XX^T = I_n$

We will now sample measurement matrices from the class $\mathcal{A}(m, n)$

Definition

$\mathcal{A}(m, n)$ is the class of random matrices in $\mathbb{R}^{m \times n}$ with isotropic, sub-gaussian, independent rows.

A random vector $X \in \mathbb{R}^n$ is:

Isotropic if $\mathbb{E}XX^T = I_n$

Sub-gaussian if the tail of $\langle X, x \rangle$ decays faster than some gaussian for any $x \in S^{n-1}$

Recovery using random matrices

expected ℓ_2 error

Several useful properties of $\mathcal{A}(m, n)$ are known.

Several useful properties of $\mathcal{A}(m,n)$ are known.

Theorem

Let $A \in \mathcal{A}(m,n)$, *and let* $E = \ker(A)$, *then*

$$\mathbb{E} diam(E \cap B_1^n) \leq C\sqrt{\frac{\log n}{m}}$$

Several useful properties of $\mathcal{A}(m, n)$ are known.

Theorem

Let $A \in \mathcal{A}(m, n)$, and let $E = \ker(A)$, then

$$\mathbb{E} diam(E \cap B_1^n) \leq C\sqrt{\frac{\log n}{m}}$$

This can be used to get

$$\mathbb{E}\|x - \hat{x}\|_2 \leq C\sqrt{\frac{s \log n}{m}}\|x\|_2$$

Exact recovery for $A \in \mathcal{A}$

Another property of matrices in $\mathcal{A}$.

Theorem

If $A \in \mathcal{A}(m, n)$ with $m \geq cs \log n$, then $\ker(A) \cap S^{n-1} \cap 2\sqrt{s}B_1^n$ is empty with probability at least $1 - 2\exp(-cm)$.

Exact recovery for $A \in \mathcal{A}$

Another property of matrices in $\mathcal{A}$.

Theorem

If $A \in \mathcal{A}(m,n)$ with $m \geq cs \log n$, then $\ker(A) \cap S^{n-1} \cap 2\sqrt{s}B_1^n$ is empty with probability at least $1 - 2\exp(-cm)$.

- From theorem, $\hat{x} = x$ with high probability provided $m \geq cs \log n$.
- Because, if $h = \hat{x} - x \neq 0$, then $h/\|h\|_2 \in \ker(A) \cap S^{n-1} \cap 2\sqrt{s}B_1^n$, which is empty with high probability.

Exact recovery for $A \in \mathcal{A}$

Another property of matrices in $\mathcal{A}$.

Theorem

If $A \in \mathcal{A}(m, n)$ with $m \geq cs \log n$, then $\ker(A) \cap S^{n-1} \cap 2\sqrt{s}B_1^n$ is empty with probability at least $1 - 2\exp(-cm)$.

- From theorem, $\hat{x} = x$ with high probability provided $m \geq cs \log n$.
- Because, if $h = \hat{x} - x \neq 0$, then $h/\|h\|_2 \in \ker(A) \cap S^{n-1} \cap 2\sqrt{s}B_1^n$, which is empty with high probability.
- Exact recovery using basis pursuit is possible!

RIP implies exact recovery

Definition (RIP)

A matrix A satisfies Restricted Isometry Property with parameter s if the following holds for every $3s$–sparse vectors:

$$0.9\|x\|_2 \leq \|Ax\|_2 \leq 1.1\|x\|_2$$

RIP implies exact recovery

Definition (RIP)

A matrix A satisfies Restricted Isometry Property with parameter s if the following holds for every $3s$–sparse vectors:

$$0.9\|x\|_2 \leq \|Ax\|_2 \leq 1.1\|x\|_2$$

- RIP implies exact recovery for s–sparse vectors. Proof of this only involves triangle inequality and Cauchy-Schwarz inequality.

RIP implies exact recovery

Definition (RIP)

A matrix A satisfies Restricted Isometry Property with parameter s if the following holds for every $3s$–sparse vectors:

$$0.9\|x\|_2 \leq \|Ax\|_2 \leq 1.1\|x\|_2$$

- RIP implies exact recovery for s–sparse vectors. Proof of this only involves triangle inequality and Cauchy-Schwarz inequality.
- All $m \times 3s$ sub-matrices must be almost isometry: difficult condition.

RIP implies exact recovery

Definition (RIP)

A matrix A satisfies Restricted Isometry Property with parameter s if the following holds for every $3s$–sparse vectors:

$$0.9\|x\|_2 \leq \|Ax\|_2 \leq 1.1\|x\|_2$$

- RIP implies exact recovery for s–sparse vectors. Proof of this only involves triangle inequality and Cauchy-Schwarz inequality.
- All $m \times 3s$ sub-matrices must be almost isometry: difficult condition.
- Surprisingly, $A \in \mathcal{A}(m, n)$ satisfies RIP if $m \geq cs \log n$ with high probability ($\geq 1 - 2\exp(-cm)$).

Summary

- Goal: to recover sparse signal with small number of measurements.

Summary

- Goal: to recover sparse signal with small number of measurements.

- Basis pursuit can recover sparse signal under some conditions.

Summary

- Goal: to recover sparse signal with small number of measurements.
- Basis pursuit can recover sparse signal under some conditions.
- Number of measurements depend linearly on sparsity and logarithmically on the dimension.

References

R. Vershynin
High Dimensional Probability

T. Tao
Compressed Sensing: robust recovery of sparse signals from limited measurements