Compressive Sensing

I-Liang Chern

Department of Mathematics
National Taiwan University

Fall, 2016

Three parts of CS in this lecture

- Theory
- Optimization Algorithms
- Applications: students' presentations
    - CS + Image Science,
    - CS + Brain and Neuroscience,
    - CS + Data Science
    - CS + PDEs
    - CS + ...

Issues of CS

- Looking for sparse solution x from the measurement $y = Ax$.
- A is $m \times N$ matrix, called measurement matrix. Usually $m << N$.
- x is assumed to be sparse, namely

$$\|x\|_0 := \#\{x_i \neq 0\} = s << N.$$

$$(P0) \quad \boxed{\min \|x\|_0 \text{ subject to } Ax = y.}$$

- Issues:
  - For what's kind of A can one recover x exactly? Or how do we design measurement matrix?
  - Provide an algorithm to reconstruct the sparse vector x.

References: books

Theory

- Simon Foucart, Holger Rauhut, A Mathematical Introduction to Compressive Sensing

Optimation Algorithms

- Boyd and Vandenberghe, Convex Optimization
- Neal Parikh and Stephen Boyd, Proximal Algorithms

Applications

- Vishal M. Patel, Rama Chellappa, Sparse Representations and Compressive Sensing for Imaging and Vision
- H. Boche, R. Calderbank, G. Kutyniok and J. Vybiral, Compressive Sensing and Its Applications
- Y. Eldar and G. Kutyniok, Compressive Sensing: Theory

References: Webpages

- compressive sensing resources
- Tutorial: see Compressive Sensing Resources
- Codes: `http://web.stanford.edu/~boyd/papers/prox_algs.html`
- Candes lecture: Stats 330 (CME 362) An Introduction to Compressed Sensing `http://statweb.stanford.edu/~candes/stats330/index.shtml`

Motivations

An invitation to Compressive Sensing

- Sampling Theory
- Sparse Approximation
- Error Correction
- Statistics and Machine Learning
- Low-Rank Matrix Recovery and Matrix Completion
- ...
- See more from Compressive Sensing Resources

Theory: Outline

- Three Algorithms:
  - Basis Pursuit
  - Matching Pursuit (greedy algorithm)
  - Thresholding-based methods
- Conditions on A for possible recovery of sparse vector
  - Mutual incoherence
  - Restricted isometry property
- What kinds of A for possible recovery of sparse vector
  - Subgaussian Random matrices (Gaussian, Bernoulli, ...)
  - Random sampling BOS (Fourier, wavelets, etc.)

1

[1] Ref: Foucart and Rauhut's book (2013); Papers of Donoho; Candes, Tao; Cai.

Three kinds of Algorithms

Problem: Suppose x is a sparse vector and measured through A by $y = Ax$. The problem is to recovery x from y and A:

$$\boxed{\text{(P0)} \quad \min \|z\|_0 \text{ subject to } Az = y.}$$

This is an NP hard problem. It is not practical to solve it directly.

Instead, three algorithms (polynomial computational complexity) are proposed:

- Basis Pursuit
- Matching Pursuit (greedy algorithm)
- Thresholding-based methods

Basis Pursuit

Solve a convex relaxation problem

$$\boxed{\text{(P1)} \qquad \min \|z\|_1 \text{ subject to } Az = y}$$

Question: What kinds of A for possible recovery of sparse vector via basis pursuit.

Orthogonal Matching Pursuit

OMP algorithm: [2]

- $S^{n+1} = S^n \cup \{j_{n+1}\}$, $j_{n+1} = \text{argmax}_{j \in \overline{S^n}} |\langle a_j, (y - Ax^n) \rangle|$
- $x^{n+1} = \text{argmin}_z \{ \|(y - Az)\|^2 \ | \text{supp}\,(z) \subset S^{n+1} \}$

Question: What kinds of A for possible recovery of sparse vector via Orthogonal Matching Pursuit?

[2] $[N] = \{1, ..., N\}$, $S \subset [N]$, $\bar{S} = [N] \setminus S$.

Thresholding-based methods

Suppose the sparse s is known. Given s and the measured data y, [3]

- $S^{\#} := L_s(A^* y)$
- $x^{\#} = \text{argmin}_z \{ \|y - Az\| \mid \text{supp}\,(z) \subset S^{\#} \}$

Question: What kinds of A for possible recovery of sparse vector via Thresholding-based method?

[3] $L_s(x)$ is the index set of x whose absolute values are s-largest.

Conditions on A for possible recovery sparse vector

- Null space property: necessary & sufficient algebraic conditions, but difficult to verify
- Mutual Incoherence: simple sufficient condition, but not sharp
- Restrict Isometry Property (RIP): sharp sufficient condition, but may be hard to verify.

Algebraic Conditions on measurement matrix A

- null space property: for exact recovery of sparse vector;
- stable null space property: for stable recovery of compressible vector;
- robust null space property: for robust recovery (under small perturbation of measurement).

Exact recovery

- Null space property: A ($m \times N$ matrix) satisfies the null-space property of order s if for any index set S with $|S| \leq s$, it satisfies

$$\|v_S\|_1 < \|v_{\bar{S}}\|_1 \text{ for all } v \in \mathrm{Ker}A \setminus \{0\}$$

Theorem

Given $m \times N$ matrix A. Every s-sparse vector x can be recover by (P1) iff A satisfies the null space property of order s.

Stability

- Compressibility: $\sigma_s(x)_p := \min_z\{\|z - x\|_p \mid \|z\|_0 \le s\}$
- Stable null space property: There exists $\rho < 1$ s.t. for any S with $|S| \le s$,

$$\|v_S\|_1 \le \rho \|v_{\bar{S}}\|_1 \text{ for all } v \in \text{Ker}A \setminus \{0\}$$

Theorem

Let A satisfies the stable null space property. Then the solution $x^\#$ of (P1) satisfies

$$\|x^\# - x\|_1 \le \frac{2(1+\rho)}{1-\rho}\sigma_s(x)_1$$

Robustness

- A is satisfies robust null space property of order s if there exist constants $0 < \rho < 1$ and $\tau > 0$ such that for any index set S with $|S| \leq s$, we have

$$\|v_S\|_1 \leq \rho \|v_{\bar{S}}\|_1 + \tau \|Av\| \text{ for all } v \in \mathbb{C}^N$$

Theorem

Let A satisfies the robust null space property and $y = Ax + e$. Then the solution $x^\#$ of (P1) satisfies

$$\|x^\# - x\|_1 \leq \frac{2(1+\rho)}{1-\rho} \sigma_s(x)_1 + \frac{4\tau}{1-\rho} \|e\|$$

Condition on A: Mutual Incoherence

- Let $A = [a_1, \cdots, a_N]$, a_j normalized column m-vector. [4]
- Suppose $\text{supp}\,(x) = S$. Then
  solving $Az = Ax$ can recover x uniquely
  $\Leftrightarrow A_S : \mathbb{C}^s \to \mathbb{C}^m$ is 1-1
  $\Leftrightarrow A_S^* A_S : \mathbb{C}^s \to \mathbb{C}^s$ is invertible,
  where $A_S^* A_S = (\langle a_i, a_j \rangle)_{i,j \in S}$.

[4] $A_S = [a_{j_1}, \cdots, a_{j_s}]$, $S = \{j_1, ..., j_s\}$

Measure the coherence

Let $\mathbf{A} = [a_1, ..., a_N]$ be an $m \times N$ matrix with $\|a_j\|_2 = 1 \ \forall j$.

Definition

1. Coherence of $\mathbf{A}$ is defined to be

$$\mu(\mathbf{A}) = \max_{i \neq j} |\langle a_i, a_j \rangle|.$$

2. The ℓ_1-coherence function: for $1 \leq s \leq N - 1$

$$\mu_1(s) := \max_{i \in [N]} \max\{\sum_{j \in S} |\langle a_i, a_j \rangle|, S \subset [N], |S| = s, i \notin S\}$$

Question: How small of μ or $\mu_1(s)$ leads to (P1) $\Leftrightarrow$ (P0)?

Theorem

We have: for all s-sparse vector x

$$(1 - \mu_1(s-1)) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \mu_1(s-1)) \|x\|_2^2.$$

Equivalently, the spectrum

$$\sigma(A_S^* A_S) \subset [1 - \mu_1(s-1), 1 + \mu_1(s-1)]$$

for all S with $|S| \leq s$. In particular, $A_S^* A_S$ is invertible for all $|S| \leq s$ if

$$\mu_1(s-1) < 1.$$

Mutual Incoherence $\Rightarrow$ Exact Recovery

Theorem

If $\mu_1(s) + \mu_1(s-1) < 1$ *or* $\mu < 1/(2s-1)$, *then both basis pursuit and orthogonal matching pursuit are successful to recover* s*-sparse vector.*

Theorem

If $2\mu_1(s) + \mu_1(s-1) < 1$ *or* $\mu < 1/(3s-1)$, *then hard thresholding pursuit can recover* s*-sparse vector* x *after* s *step.*

Matrices with small coherence

Def. The normalized column vectors $(a_1, \cdots, a_N)$ are

- Equiangular: if there exists a c such that

$$|\langle a_i, a_j \rangle| = c \text{ for } i \neq j.$$

- Tight frame: if there exists a $\lambda > 0$ s.t.
  $\|x\|^2 = \lambda \sum_{j=1}^{N} |\langle x, a_j \rangle|^2$ for all x

Theorem

It holds $\mu \geq \sqrt{\frac{N-m}{m(N-1)}}$. *The equality holds iff* $(a_1, \cdots, a_N)$ *are equiangular tight frame.*

Small coherence

- $(a_1, \cdots, a_N)$ are equiangular implies $N \leq m^2$.
- The condition $\mu < 1/(2s-1)$ is too restrictive in applications. Because for the smallest conference,
  - for large N, smallest coherence $\mu \sim 1/\sqrt{m}$,
  - $\frac{1}{\sqrt{m}} \sim \mu < 1/(2s-1)$ leads to $m \geq s^2$;
- The optimal m is $m \sim s \ln(N/s)$ (from RIP). This means that those which satisfy incoherence condition is very limited.

Restricted Isometry Property

- Def. $\delta_s(A)$ is the smallest δ such that

$$(1-\delta)\|x\|^2 \le \|Ax\|^2 \le (1+\delta)\|x\|^2.$$

  for all s-sparse vector x.
- A satisfies RIP of order s if δ_s is small.
- Thms. Basis Pursuit, Orthogonal Matching Pursuit , Iterative Hard Pursuit and Hard Threasholding Pursuit are successful if

BP	IHP	HTP	OMP
$\delta_{2s} < 0.6248$	$\delta_{3s} < 0.5773$	$\delta_{3s} < 0.5773$	$\delta_{13s} < 0.1666$

What kind of A satisfying RIP

- Given an $m \times N$ matrix A with $N \leq m^2$. $\delta_s(A)$ has upper and lower estimates

$$\sqrt{cs}/\sqrt{m} \leq \delta_s \leq cs/\sqrt{m}$$

  There is a sufficient gap between the two bounds.
- In fact, certain random matrices satisfy $\delta_s \leq \delta$ with high probability provided

$$\boxed{m \geq \frac{C}{\delta^2} s \ln(eN/s)}$$

- Further, any matrix A with $\delta_s \leq \delta$ requires

$$m \geq C_\delta s \ln(eN/s).$$

What's kind of matrices satisfying RIP

- Random matrices with
  - iid Gaussian entries
  - iid Bernoulli entries $(+/-1)$
  - iid subgaussian entries
  - random Fourier ensemble
  - random ensemble in bounded orthogonal systems
- In each case, $m = O(s \ln N)$, they satisfy RIP with very high probability $(1 - e^{-Cm})$..

RPI for subgaussian matrices

Theorem

Let A *be a subgaussian matrix. Then there exists a constant* C *such that the RIP constant* δ_s *of the normalized matrix* $\frac{1}{\sqrt{m}}A$ *satisfies* $\delta_s \leq \delta$ *with probability at least* $1 - 2\exp(-\delta^2 m/(2C))$, *provided*

$$m \geq \frac{2C}{\delta} s \ln(eN/s).$$

- A random variable X is called subgaussian if $P(|X| \geq t) \leq \beta e^{-\kappa t^2}$.
- A random matrix A is called subgaussian if each entry is iid subgaussian (mean 0, variance 1).

Random sampling in bounded orthonormal system

- Bounded orthonormal system: $\{\phi_j : \mathcal{D} \mapsto \mathbb{C}\}$ be orthonormal system in $L^2(\mathcal{D}, \nu)$, and $\|\phi_j\|_\infty \leq K$, $\forall\ j$.
- $\{t_i,\ i = 1, ..., m\}$ are independent random variables with range in $\mathcal{D}$.
- $A = (\phi_j(t_i))_{m \times N}$ is a random matrix.

Theorem

Let x be s-sparse and A be random sampling from BOS with constant K. If

$$m \geq CK^2 s \ln^2(6N/\epsilon),$$

then with probability at least $1 - \epsilon$, we have exact recovery from basis pursuit.

Concentration Lemma

Lemma (Concentration Inequality)

Let A be iid subgaussian $m \times N$ matrix. Then for any $x \in \mathbb{R}^N$ and for any $\delta \in (0,1)$,

$$P\left(|m^{-1}\|Ax\|^2 - \|x\|^2| \geq \delta\|x\|^2\right) \leq 2\exp(-ct^2 m),$$

where c depends on the subgaussian parameter only.

Connection to Johnson-Lindenstrauss Lemma

Lemma (Johnson-Lindenstrauss)

Given $x_1, ..., x_M \in \mathbb{R}^N$ *arbitrary. Given* $\delta > 0$*. If* $m > C\delta^{-2} \ln M$*, then there exists a linear map* $A : \mathbb{R}^N \to \mathbb{R}^m$ *such that*

$$(1-\delta)\|x_j - x_\ell\|^2 \le \|A(x_j - x_\ell)\|^2 \le (1+\delta)\|x_j - x_\ell\|^2$$

for any $1 \le j, \ell \le M$*.*

Remarks

- It means we can project high dimension to low dimension with A being nearly relative isometry.
- The construction is probabilistic.

Optimization Algorithms

Problem to solve (Assume convexity)

- $\min f(x)$ subject to $y = Ax$

Main References:

- Boyd and Vandenberghe, Convex Optimization. This book can be downloaded. It provides a thorough material about optimization. Both of them have slides. They can also be downloaded from websites.
- Neal Parikh and Stephen Boyd, Proximal Algorithms
- Vandenberghe, Convex Optimization (slides)

Convex Optimization Algorithms

- Basic convex analysis
- Gradient methods and Newton's methods
- Proximal algorithms
- Augmented Lagrange Method (ALM) and Alternative Direction Method of Multipliers (ADMM)