# Compressive Sensing, Lecture 2

## Yesterday

- Undetermined systems of equations and $\ell_1$ minimization
- Compressive sensing
  - Probabilistic approach: isotropy & incoherence
  - Deterministic approach: restricted isometry property

## Today

- Ideas of the main proofs
- Main computational tool: convex optimization

# Recap

## Compressive sampling approach

- measure $b = A\bar{x}$
- obtain $\hat{x}$ via $\ell_1$ minimization: $\hat{x} := \text{argmin}_x\{\|x\|_1 : Ax = b\}$.

## Probabilistic approach

- fix $\bar{x} \in \mathbb{R}^n$ arbitrary
- randomize $A$
- with high probability $\hat{x}$ recovers $\bar{x}$ or $\bar{x}_s$

## Deterministic approach: RIP

- find $m \times n$ matrix $A$ satisfying RIP
- $\hat{x}$ recovers $\bar{x}$ or $\bar{x}_s$ for all $\bar{x} \in \mathbb{R}^n$

# RIP and exact recovery

## Recall

Given a sensing matrix $A$, $\delta_k$ is smallest $\delta$ such that

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2$$

for all $k$-sparse vector $x$.

## Observe

$\delta_{2s}$ is the smallest $\delta$ such that

$$(1-\delta)\|x_1 - x_2\|_2^2 \leq \|A(x_1 - x_2)\|_2^2 \leq (1+\delta)\|x_1 - x_2\|_2^2$$

for all $s$-sparse vectors $x_1, x_2$.

Therefore if $\delta_{2s} < 1$ in principle we can recover $\bar{x}$ from $b = A\bar{x}$, e.g., via $\ell_0$ minimization.

# RIP and signal recovery (special case)

### Theorem

*Assume* $\bar{x} \in \mathbb{R}^n$ *and* $A$ *satisfies RIP with* $\delta_{2s} \leq \sqrt{2} - 1$*. Then the* $\ell_1$ *solution* $\hat{x}$ *satisfies*

$$\|\hat{x} - \bar{x}\|_2 \leq C \cdot \frac{\|\bar{x} - \bar{x}_s\|_1}{\sqrt{s}}$$

*for some constant* $C$*.*

### Proof

Let $h := \hat{x} - \bar{x}$. Put $T_0 :=$ indexes of $s$ largest entries of $|h|$, $T_1 :=$ indexes of $s$ largest entries of $|h_{T_0^c}|$, etc.

Let $\Delta := \frac{\|\bar{x} - \bar{x}_s\|_1}{\sqrt{s}}$.

By construction of the $T_j$s:

$$\sum_{j\geq 2} \|h_{T_j}\|_2 \leq \frac{\|h_{T_0^c}\|_1}{\sqrt{s}}.$$

By optimality of $\hat{x}$:

$$\|h_{T_0^c}\|_1 \leq \|h_{T_0}\|_1 + 2 \cdot \sqrt{s} \cdot \Delta.$$

By RIP:

$$(1-\delta_{2s})\|h_{T_0\cup T_1}\|_2 \leq \sqrt{2} \cdot \delta_{2s} \cdot \sum_{j\geq 2} \|h_{T_j}\|_2.$$

Hence $\|h_{T_0\cup T_1}\|_2 \leq \frac{2\rho\cdot\Delta}{1-\rho}$ for $\rho := \frac{\sqrt{2}\cdot\delta_{2s}}{1-\delta_{2s}}$. Therefore

$$\|h\|_2 \leq \frac{2(1+\rho)}{1-\rho} \cdot \Delta$$

□

# Gaussian matrices and RIP

### Theorem

*Let $M$ be an $m \times n$ Gaussian matrix and $A := \frac{1}{\sqrt{m}} M$. If $m \geq \frac{k \log(en/k)}{\delta^2}$ for $\delta \in (0, 1/3)$ and $1 \leq k \leq n$, then with probability at least $1 - 2e^{-\delta^2 m}$*

$$1 - 3\delta \leq \sigma_{\min}(A_T) \leq \sigma_{\max}(A_T) \leq 1 + 3\delta \text{ for all } |T| = k.$$

*In particular, $A$ satisfies RIP with high probability.*

# Related property of random projections

**Theorem (Johnson-Lindenstrauss Lemma)**

*Assume* $x_1, \ldots, x_n \in \mathbb{R}^d$. *If* $k \geq \frac{8\delta \log n}{\epsilon^2(1-2\epsilon/3)}$ *for some* $\epsilon \in (0,1)$ *and* $\delta \geq 1$, *then a random projection* $\Pi : \mathbb{R}^d \to \mathbb{R}^k$ *satisfies*

$$(1-\epsilon)\frac{k}{d}\|x_i - x_j\|_2^2 \leq \|\Pi x_i - \Pi x_j\|_2^2 \leq (1+\epsilon)\frac{k}{d}\|x_i - x_j\|_2^2, \ \forall i \neq j$$

*with probability at least* $1 - \frac{n(n-1)}{n^{2\delta}}$.

Key lemmas (for both Johnson-Lindenstrauss and Gaussian RIP):

### Lemma (Borell, Tsirelson-Ibragimov-Sudakov)

*Let* $X \sim N(0, I_d)$ *and* $f : \mathbb{R}^d \to \mathbb{R}$ *be* $L$*-Lipschitz. Then for* $t \geq 0$

$$\mathbb{P}\left(f(X) - \mathbb{E}[f(X)] > t\right) \leq e^{-t^2/2L^2}.$$

### Lemma (Sudakov-Fernique)

*Let* $(X_t)_{t \in I}$ *and* $(Y_t)_{t \in I}$ *be Gaussian processes. If* $\mathbb{E}X_t = \mathbb{E}Y_t$ *and* $\mathbb{E}(X_s - X_t)^2 \leq \mathbb{E}(Y_s - Y_t)^2$ *for all* $s, t \in I$ *then*

$$\mathbb{E} \sup_{t \in I} X_t \leq \mathbb{E} \sup_{t \in I} Y_t.$$

# Proof of Gaussian RIP Theorem

Assume $|T| = k$ and $t \geq 0$.

By Sudakov-Fernique:

$$\mathbb{E}(\sigma_{\max}(M_T)) \leq \sqrt{m} + \sqrt{k}$$

$$\mathbb{E}(\sigma_{\min}(M_T)) \geq \sqrt{m} - \sqrt{k}.$$

By Borell, Tsirelson-Ibragimov-Sudakov:

$$\mathbb{P}(\sigma_{\max}(M_T) \geq \sqrt{m} + \sqrt{k} + t) \leq e^{-t^2/2}$$

$$\mathbb{P}(\sigma_{\min}(M_T) \leq \sqrt{m} - \sqrt{k} - t) \leq e^{-t^2/2}.$$

Hence

$$\mathbb{P}\left(\max_{|T|=k} \sigma_{\max}(A_T) \geq 1 + \frac{\sqrt{k} + t}{\sqrt{m}}\right) \leq \binom{n}{k} e^{-t^2/2} \leq (en/k)^k e^{-t^2/2}$$
$$= \exp\left(k \log(en/k) - t^2/2\right).$$

To finish, take $t = 2\sqrt{m} \cdot \delta$. (Similar for $\sigma_{\min}$.) □

# Signal recovery for probabilistic ("RIPless") approach

## Probabilistic approach

- Suppose $\bar{x} \in \mathbb{R}^n$ is $s$-sparse.
- Pick $A \in \mathbb{R}^{m \times n}$ and measure $b = A\bar{x}$.

## Question

How likely it is that the solution $\hat{x}$ to the $\ell_1$ minimization problem

$$\begin{array}{ll} \min & \|x\|_1 \\ & Ax = b \end{array}$$

recovers $\bar{x}$?

# Optimality conditions for $\ell_1$ minimization

$$\begin{aligned} \min \quad & \|x\|_1 \\ & Ax = b \end{aligned}$$

## Optimality conditions

A feasible $x \in \mathbb{R}^n$ is optimal iff there exists $v = A^*\lambda$ such that

- $v_i = \operatorname{sgn}(x_i)$ for $i \in T := \{i : x_i \neq 0\}$.
- $|v_i| \leq 1$ for $i \in T^c := \{i : x_i = 0\}$.

## Sufficient condition for uniqueness

If, in addition, $|v_i| < 1$ for all $i \in T^c$ and $A_T$ is full column rank then $x$ is the unique solution.

# Strategy to prove exact recovery via $\ell_1$ minimization

Suppose $\bar{x}$ has support $T$, i.e., $T := \{i : x_i \neq 0\}$.

Take

$$v := A^* A_T (A_T^* A_T)^{-1} \mathrm{sgn}(x_T).$$

By construction $v = A^* \lambda$ and $v_i = \mathrm{sgn}(x_i)$ for $i \in T$.

We would be done if we can show that $|v_i| < 1$ for $i \in T^c$.

An easy probabilistic result:

**Theorem**

*For $A$ Gaussian, achieve exact recovery with probability at least $1 - 3/\sqrt{n}$ provided $m \geq 4s \log n$.*

# Proof of Theorem

Put $w := A_T(A_T^* A_T)^{-1}\mathrm{sgn}(x_T)$.

Observe: $w, A_i$ are independent for $i \in T^c$.

Thus $v_i | w \sim \mathcal{N}(0, \|w\|_2^2)$. Hence

$$\mathbb{P}(|v_i| \geq 1 | w) \leq 2e^{-1/2\|w\|_2^2}.$$

On the other hand, as in the RIP Theorem,

$$\mathbb{P}(\sigma_{\min}(A_T) \leq \sqrt{m} - \sqrt{s} - t) \leq e^{-t^2/2}.$$

Therefore with probability at least $1 - e^{-t^2/2}$

$$\|w\| \leq \frac{\sqrt{s}}{\sqrt{m} - \sqrt{s} - t} := B.$$

Consequently

$$\mathbb{P}\left(\max_{i \in T^c} |v_i| \geq 1\right) \leq 2ne^{-1/2B^2} + e^{-mt^2/2}.$$

To finish, take $t := \sqrt{\log n}$.

# Optimality of Compressive Sensing

## Back to Gaussian sensing

$m$ Gaussian measurements and $\ell_1$ decoding:

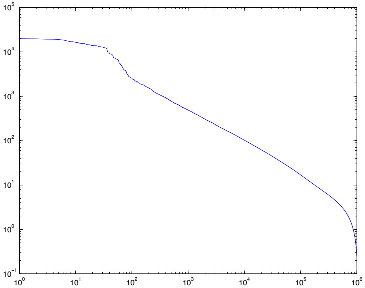
$$\|\hat{x} - x\|_2 \lesssim \frac{\|x - x_s\|_1}{\sqrt{s}}, \ s \approx \frac{m}{\log(n/m)}.$$

## Question

Can we do better with other measurements or other algorithms?

# Signals with power law

image




sorted wavelet coefficients

Power law decay: $|x|_{(1)} \geq |x|_{(2)} \geq \cdots \geq |x|_{(n)}$

$$|x|_{(k)} \leq \frac{C}{k^p}$$

## Model

$\ell_p$ ball $\mathcal{B}_p := \{x : \|x\|_p \leq 1\}$.
Discuss case $p = 1$ but same discussion applies to $0 \leq p \leq 1$.

# Recovery of $\ell_1$ ball $\mathcal{B}$

## Gaussian sensing

- Suppose unknown vector is in $\mathcal{B}$
- Take $m$ Gaussian measurements

$$\|\hat{x} - x\|_2 \lesssim \sqrt{\frac{\log(n/m) + 1}{m}}.$$

## Ideal sensing

Best we can hope from $m$ linear measurements:

$$E_m(\mathcal{B}) = \inf_{D,F} \sup_{x \in \mathcal{B}} \|x - D(F(x))\|.$$

# Gelfand widths

## Theorem (Donoho)

$$d_m(\mathcal{B}) \leq E_m(\mathcal{B}) \leq C \cdot d_m(\mathcal{B}),$$

*where* $d_m(\mathcal{B})$ *is the* $m$*-width of* $\mathcal{B}$*:*

$$d_m(\mathcal{B}) := \inf_V \left\{ \sup_{x \in \mathcal{B}} \|P_V x\|_2 : \mathit{codim}(V) < m \right\}$$

## Theorem (Kashin, Garnaev-Gluskin)

*For* $\ell_1$ *ball*

$$C_1 \cdot \sqrt{\frac{\log(n/m)+1}{m}} \leq d_m(\mathcal{B}) \leq C_2 \cdot \sqrt{\frac{\log(n/m)+1}{m}}.$$

Compressive sensing achieves the limits of performance.

# Convex functions and sets

### Subdifferential

Assume $f : \mathbb{R}^n \to \mathbb{R}$ convex and $x \in \mathbb{R}^n$. A vector $g \in \mathbb{R}^n$ is a subgradient of $f$ at $x$ if

$$f(y) \geq f(x) + \langle g, y - x \rangle, \quad \text{for all } y \in \mathbb{R}^n.$$

Subdifferential $\partial f(x) := \{g : g \text{ subgradient of } f \text{ at } x\}$.

### Normal cone

Assume $S \subseteq \mathbb{R}^n$ is convex and $x \in S$.
Normal cone to $S$ at $x$:

$$N_S(x) := \{d : \langle d, y - x \rangle \leq 0 \;\text{ for all }\; y \in S\}.$$

# Convex optimization

Problem of the form

$$\begin{array}{ll} \min & f(x) \\ & x \in S. \end{array}$$

where $f$ and $S$ are convex.

## Sufficient optimality conditions

A point $\bar{x} \in S$ is a solution to the above problem if

$$-\partial f(\bar{x}) \cap N_S(\bar{x}) \neq \emptyset.$$

# Special convex optimization problems

## Linear programming

Objective function is linear: $f(x) = \langle c, x \rangle$
Constraint set is polyhedral: $S = \{x : Ax = b,\ Bx \geq d\}$.

## Optimality conditions for linear programming

$$c = A^* y + B^* z,\ z \geq 0,\ \langle z, Bx - d \rangle = 0.$$

# Special convex optimization problems

Consider the vector space $\mathbb{S}^n$ : $n \times n$ symmetric matrices with inner product

$$\langle X, Z \rangle = \text{trace}(XZ).$$

Cone of positive semidefinite matrices

$$\mathbb{S}^n_+ := \{X \in \mathbb{S}^n : \lambda(X) \geq 0\} = \{X : u^\mathsf{T} X u \geq 0 \; \forall u \in \mathbb{R}^n\}.$$

Write $X \succeq Z$ for $X - Z \in \mathbb{S}^n_+$.

## Semidefinite programming

Objective function: $f(X) = \langle C, X \rangle$
Constraint set: $S = \{X \in \mathbb{S}^n : \mathcal{A}(X) = b, \; \mathcal{B}(X) \succeq D\}$ for some linear maps $\mathcal{A}, \mathcal{B}$.

## Sufficient optimality conditions for semidefinite programming

$$C = \mathcal{A}^*(y) + \mathcal{B}^*(Z), \; Z \succeq 0, \; \langle Z, \mathcal{B}(X) - D \rangle = 0.$$

## What is so special about linear and semidefinite programming?

- They have powerful duality properties
- They can be solved efficiently (via interior-point methods)
- Popular matlab-based solvers: `SeDuMi`, `SDPT3`
- Matlab toolbox CVX serves as a wrapper for these solvers

# CVX examples

To solve

$$\min_x \quad \|x\|_1 \qquad \rightsquigarrow \qquad \min_{x,t} \quad \sum_{i=1}^{n} t_i$$

$$Ax = b \qquad\qquad x \le t$$

$$-x \le t$$

$$Ax = b$$

Use CVX code

```
cvx_begin
  variable x(n);
  variable t(n);
  minimize(sum(t));
  subject to
    x <= t;
    -x <= t;
    A*x == b;
cvx_end
```

CVX does some standard transformations.

To solve

$$\begin{aligned} \min \quad & \|x\|_1 \\ & Ax = b \end{aligned}$$

Use CVX code

```
cvx_begin
  variable x(n);
  minimize(norm(x,1));
  subject to
    A*x == b;
cvx_end
```

## More CVX examples

To solve

$$\begin{aligned} \min \quad & \langle I, X \rangle \\ & \langle A, X \rangle = b \\ & X \succeq 0 \end{aligned}$$

use CVX code

```
cvx_begin
  variable X(n,n) symmetric;
  minimize( trace( I * X ) );
  subject to
    trace( A * X ) == b;
    X == semidefinite(n);
cvx_end
```

# Main references for today's material

- Slides for this minicourse:
`http://andrew.cmu.edu/user/jfp/UNencuentro`


- E. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris,* Ser. I 346, pp. 589-592, 2008.

- R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, "A Simple Proof of the Restricted Isometry Property for Random Matrices," *Constructive Approximation,* 2008.

- D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory,* vol 52, no. 4, pp. 1289–1306, April 2006.

- M. Grant and S. Boyd, `CVX`*: Matlab Software for Disciplined Convex Programming,* `http://cvxr.com/cvx/`