## Computational Principles for High-dim Data Analysis (Lecture Two)

#### Yi Ma

EECS Department, UC Berkeley

August 31, 2021



EECS208, Fall 2021

## **Sparse Signal Models**

#### Applications

Medical Imaging Image Processing Face Recognition

#### 2 Recovering a Sparse Signal

Norms for Sparsity The  $\ell^0$  Norm Minimizing the  $\ell^0$  for the Sparsest Solution Computational Complexity of  $\ell^0$  Minimization Recovering Signals from (Linear) Measurements



#### Different "structures" in x:



Ma (EECS Department, UC Berkeley)

EECS208, Fall 2021

August 31, 2021 3 / 27

イロト イポト イヨト イヨト

(1)

## Applications

3

<ロト <問ト < 目と < 目と

## Application I: Magnetic Resonance Imaging (Chap. 10)



Figure: Left: Key components. Right: The three-axis gradient coils.

X coils

Primary Electromagnetic Coi

Compute

(日) (四) (日) (日) (日)

## Mathematical Model of MRI

Simplified mathematical model for MRI:

$$y = \mathcal{F}[I](\boldsymbol{u}) = \int_{\boldsymbol{v}} I(\boldsymbol{v}) \exp(-\mathfrak{i} \, 2\pi \, \boldsymbol{u}^* \boldsymbol{v}) \, d\boldsymbol{v}, \quad \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$$
(2)  
$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \mathcal{F}[I](\boldsymbol{u}_1) \\ \vdots \\ \mathcal{F}[I](\boldsymbol{u}_m) \end{bmatrix} \doteq \mathcal{F}_{\mathsf{U}}[I], \quad \boldsymbol{m} \ll N^2.$$
(3)



Figure: Recovering MRI image from Fourier measurements.

EECS208, Fall 2023

< ロ > < 同 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

## Exploit Structures of MRI Images

Express I as a superposition of basis functions  $\Psi = \{\psi_1, \dots, \psi_{N^2}\}$ :

$$\displaystyle I_{ ext{image}} = \sum_{i=1}^{N^-} oldsymbol{\psi}_i imes oldsymbol{x}_i$$
 .  $extsf{transformula}_i$  the coefficient

Approximate I with the k largest coefficients  $J = \{i_1, \ldots, i_k\}$ :

$$I_{\text{target image}} \approx \tilde{I}_k = \sum_{i \in \mathsf{J}} \psi_i x_i. \tag{5}$$

superposition of k basis functions



Figure: MRI image and its wavelet coefficients

(4)

 $= \mathcal{F}_{U}[I],$ 

## Recovering Image from a Under-determined Linear System

$$oldsymbol{y}$$
 observed Fourier coefficients

$$= \mathcal{F}_{\mathsf{U}} \Big[ \psi_{1} x_{1} + \dots + \psi_{N^{2}} x_{N^{2}} \Big],$$

$$= \mathcal{F}_{\mathsf{U}} [\psi_{1}] x_{1} + \dots + \mathcal{F}_{\mathsf{U}} [\psi_{N^{2}}] x_{N^{2}},$$

$$= \Big[ \mathcal{F}_{\mathsf{U}} [\psi_{1}] | \dots | \mathcal{F}_{\mathsf{U}} [\psi_{N^{2}}] \Big] x,$$

$$\max \mathbf{x} \in \mathbb{R}^{m \times N^{2}}, \ m \ll N^{2}.$$

$$= \mathbf{A} x. \qquad (6)$$

#### x is sparse or approximately sparse!

$$I ~pprox \sum_{i\in \mathsf{J}} oldsymbol{\psi}_i x_i ~=~ oldsymbol{\Psi}_{N^2 imes N^2} oldsymbol{\mathsf{matrix}}_{\mathsf{sparse vector}} oldsymbol{x},$$

where  $x_i = 0$  for  $i \notin J$ , and  $k = |J| \ll N^2$ .

(7)

## Application II: Image Denoising

$$I_{\text{noisy}} = I_{\text{clean}} + z_{\text{noise}}$$
(8)  

$$I_{\text{clean}} \text{ into patches } y_{1\text{clean}}, \dots, y_{p_{\text{clean}}} :$$
  

$$y_{i} = y_{i\text{clean}} + z_{i} = A_{\text{patch dictionary}} \times x_{i} + z_{i}.$$
(9)  

$$I_{\text{patch dictionary}} = I_{\text{patch dictionary}} + z_{i} + z_{i}.$$
(9)

Figure: Left: input image; middle: denoised; right: dictionary for image patches. Sparse representation for color image restoration, Mairal, Elad, and Sapiro. TIP, 2008

Break

EECS208, Fall 2021

## Application II: Image Super-Resolution

Given a pair of corresponding low-resolution and high-resolution dictionaries  $({\bm A}_{\sf low}, {\bm A}_{\sf high})$ :

$$y_{i \mid \mathsf{low}} = A_{\mathsf{low}} x_i \quad \xrightarrow{\mathsf{Lifting}} y_{i \mid \mathsf{high}} = A_{\mathsf{high}} x_i.$$
 (10)



Figure: Super resolution for face images

Image Super-Resolution via Sparse Representation of Raw Image Patches, Yang, Wright, Huang, and Yi Ma, CVPR, 2008.

Ma (EECS Department, UC Berkeley)

EECS208, Fall 2021

August 31, 2021 10 / 27

## Application III: Robust Face Recognition (Chap. 13)



Concatenate gallery images of  $\boldsymbol{n}$  subjects into a large "dictionary":

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 \mid \boldsymbol{B}_2 \mid \dots \mid \boldsymbol{B}_n \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad n = \sum_i n_i. \tag{12}$$
  
all training images

< A >

# Application III: Robust Face Recognition (Chap. 13)

Find sparse solutions (x, e) to the linear system:

$$y = Bx + e = [B, I] \begin{bmatrix} x \\ e \end{bmatrix}.$$
(13)



Robust Face Recognition via Sparse Representation, Wright, Yang, Ganesh, Sastry, and Ma, TPAMI, 2009.

Ma (EECS Department, UC Berkeley)

EECS208, Fall 2021

## Robust Face Recognition: Rigorous Justification



**Theorem 1.** For any  $\delta > 0$ ,  $\exists \nu_0(\delta) > 0$  such that if  $\nu < \nu_0$  and  $\rho < 1$ , in weak proportional growth, with error support J and signs  $\sigma$  chosen uniformly at random,

$$\lim_{m \to \infty} P_{A,J,\sigma} \left[ \ell^1 \text{-recoverability at} (I,J,\sigma) \ \forall I \in \binom{[n]}{k_1} \right] = 1.$$

 ${}^{*}\ell^{1}$  recovers any sparse signal from almost any error with density less than 1  ${}^{"}$ 

Dense Error Correction via  $\ell^1$  Minimization, Wright and Ma, Trans. Info. Theory, 2010.

< □ > < □ > < □ > < □ > < □ > < □ >

## Many Many Applications...

#### Part III of the Textbook:

- Magnetic Resonnance Imaging (Chapter 10)
- Wideband Spectrum Sensing (Chapter 11)
- Scientific Imaging Problems (Chapter 12)
- Robust Face Recognition (Chapter 13)
- Robust Photometric Stereo (Chapter 14)
- Structured Texture Recovery (Chapter 15)
- Deep Networks for Classification (Chapter 16)

## Your final projects?

### How to Find a Sparse Solution x?



- B - - B

< 47 ▶

## Vector Spaces and Linear Algebra

Topics in Linear Algebra (Appendix A):

- Vector Space, Linear Independence, Basis
- Linear Mappings, Subspaces, Matrix Representation
- Linear Systems and Conjugate Gradient (over-determined, under-determined)
- Eigenvalue Decomposition and Singular Value Decomposition
- Norms and Matrix Norms



## Norms

A norm on a vector space  $\mathbb V$  over  $\mathbb R$  is a function  $\|\cdot\|:\mathbb V\to\mathbb R$  that is

**1** positive definite:  $||x|| \ge 0$ , and ||x|| = 0 if and only if x = 0; **2** nonnegatively homogeneous:

$$\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|, \quad \forall \boldsymbol{x} \in \mathbb{V}, \alpha \in \mathbb{R};$$

**3** subadditive:  $\|\cdot\|$  satisfies the triangle inequality

$$\|\boldsymbol{x} + \boldsymbol{y}\| \le \|\boldsymbol{x}\| + \|\boldsymbol{y}\|, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}.$$

**Example:** for  $\mathbb{V} = (\mathbb{R}^n, \mathbb{R})$ , a *p*-norm:

$$\|\boldsymbol{x}\|_{p} \doteq \left(\sum_{i} |x_{i}|^{p}\right)^{1/p}, \quad p \in [1, \infty].$$
(15)

イロト イポト イヨト イヨト 二日

## Sparse Vectors and Norms

 $\ell^2$  norm or "Euclidean norm":

$$\|x\|_2 = \sqrt{\sum_i |x_i|^2} = \sqrt{x^*x}.$$
 (16)

 $\ell^1$  norm:

$$\|\boldsymbol{x}\|_1 = \sum_i |x_i|,$$



 $\ell^{\infty}$  norm:

$$\|\boldsymbol{x}\|_{\infty} = \max_{i} |x_{i}|.$$
 (18) Figure: The  $\ell^{p}$  balls.

(17)

 $\ell^p$  unit ball:

$$\mathsf{B}_p \doteq \{ \boldsymbol{x} \mid \|\boldsymbol{x}\|_p \le 1 \} \quad 0 
(19)$$

< A I

**Example:** relative volumes of the  $\ell^p$  balls when n is large.

## Measure of Sparsity: the $\ell^0$ Norm

 $\ell^0$  "norm":

$$\|\boldsymbol{x}\|_{0} = \#\{i \mid \boldsymbol{x}(i) \neq 0\}.$$
 (20)

Not a norm, and only subadditive:

$$\forall x, x', \qquad ||x+x'||_0 \leq ||x||_0 + ||x'||_0.$$
 (21)

"Limit" of  $\ell^p$ -norm as  $p \searrow 0$ :

$$\lim_{p \searrow 0} \|\boldsymbol{x}\|_{p}^{p} = \sum_{i=1}^{n} \lim_{p \searrow 0} |\boldsymbol{x}(i)|^{p} = \sum_{i=1}^{n} \mathbb{1}_{\boldsymbol{x}(i) \neq 0} = \|\boldsymbol{x}\|_{0}.$$
 (22)

< 1 k

## Minimizing the $\ell^0$ Norm

Given  $y = Ax_o$ , to find  $x_o$  as the sparsest solution to y = Ax, we can:

$$\hat{\boldsymbol{x}} = \arg\min \|\boldsymbol{x}\|_0$$
 s.t.  $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ . (23)

The "support" of the nonzero entries:

$$\operatorname{supp}(\boldsymbol{x}) = \{i \mid \boldsymbol{x}(i) \neq 0\} \subset \{1, \dots, n\}, \quad k = |\operatorname{supp}(\boldsymbol{x}_o)|.$$
 (24)

Brute force exhaustive search:

$$\boldsymbol{A}_{\mathsf{I}}\boldsymbol{x}_{\mathsf{I}} = \boldsymbol{y}? \quad \forall \mathsf{I} \subseteq \{1, \dots, n\}, \ |\mathsf{I}| \le k.$$
(25)

## Uniqueness of the Sparsest Solution

#### Definition (Kruskal Rank)

The *Kruskal rank* of a matrix A, written as krank(A), is the largest number r such that every subset of r columns of A is linearly independent.

Theorem ( $\ell^0$  Recovery)

Suppose that  $oldsymbol{y} = oldsymbol{A} oldsymbol{x}_o$ , with

$$\|\boldsymbol{x}_o\|_0 \leq \frac{1}{2} \operatorname{krank}(\boldsymbol{A}).$$
(26)

Then  $x_o$  is the unique optimal solution to the  $\ell^0$  minimization problem

$$\min \|\boldsymbol{x}\|_0 \quad s.t. \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}. \tag{27}$$

Proof:  $A\hat{x} = y \Rightarrow A(\hat{x} - x_o) = A\hat{x} - Ax_o = y - y = 0.$ 

イロト 不得下 イヨト イヨト 二日

## Minimizing the $\ell^0$ Norm: Exhaustive Search

#### Algorithm ( $\ell^0$ Minimization via Exhaustive Search):

1: Input: a matrix 
$$A \in \mathbb{R}^{m \times n}$$
 and a vector  $y \in \mathbb{R}^{m}$ .  
2: for  $k = 0, 1, 2, ..., n$ ,  
3: for each  $I \subseteq \{1, ..., n\}$  of size  $k$ ,  
4: if the system of equations  $A_{I}z = y$  has a solution  $z$ ,  
5: set  $x_{I} = z$ ,  $x_{I^{c}} = 0$ .  
6: return  $x$ .  
7: end if  
8: end for

9: end for

**Example:** computational complexity in finding the sparsest solution via the exhaustive search algorithm.

## Minimizing the $\ell^0$ Norm: Simulations

Solve: 
$$\min \|\boldsymbol{x}\|_0$$
 s.t.  $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ . (28)

A is of size  $5 \times 12$ . Fraction of success across 100 trials.



## Computational Complexity of $\ell^0$ Minimization

Theorem (Hardness of  $\ell^0$  Minimization)

The  $\ell^0$ -minimization problem  $\min \|x\|_0$  s.t. Ax = y is (strongly) NP-hard.

Proof: Reducible to an Exact 3-Set Cover (E3C) problem.



Figure: Exact 3-Set Cover as a Sparse Representation Problem. Left: a universe  $S = \{1, \ldots, 6\}$  and four subsets  $U_1, \ldots, U_4 \subseteq S$ .  $\{U_1, U_3\}$  is an exact 3-set cover. **Right:** the same problem as a linear system of equations. The columns of A are the incidence vectors for the sets  $U_1, U_2, U_3, U_4$ . The Exact 3-Cover  $\{U_1, U_3\}$  corresponds to a solution x to the system Ax = y with only m/3 = 2 nonzero entries.

## Computational Complexity of $\ell^p$ Minimization (0 )

Theorem (Hardness of  $\ell^p$  Minimization)

For any fixed  $0 , the <math>\ell^p$ -minimization problem

 $\min \|m{x}\|_p^p$  subject to  $m{A}m{x} = m{y}$ 

is (strongly) NP-hard.

#### Corollary (Hardness of Smoothed $\ell^p$ Minimization)

For any fixed  $0 and <math>\varepsilon > 0$ , the smoothed  $\ell^p$ -minimization problem  $\min \|x + \varepsilon\|_p^p$  s.t. Ax = y is (strongly) NP-hard.

Nevertheless, computing a local minimizer of the problem can be done in polynomial time. (... what about  $\ell^1$ ?)

A Note on the Complexity of  $L_p$  Minimization, D. Ge, X. Jiang, and Y. Ye, Mathematical Programming, vol. 129, 2011.

Ma (EECS Department, UC Berkeley)

EECS208, Fall 2021

August 31, 2021 25 / 27

## Two Fundamental Questions

To find the correct k-sparse solution to the under-determined linear system:

$$y = \frac{A}{m \times n, m \ll n} \times \frac{x}{k \text{ sparse}}$$
(29)

we want to know:

- sample complexity: how many measurements are needed for the problem to become computationally tractable? (Part I)
- 2 computational complexity: once tractable, what is the precise computational complexity in finding the correct solution? (Part II)

"It is quite probable that our mathematical insights and understandings are often used to achieve things that could in principle also be achieved computationally – but where blind computation without much insight may turn out to be so inefficient that it is unworkable."

- Roger Penrose, Shadows of the Mind

< ロ > < 同 > < 回 > < 回 > < 回 > <

## Assignments

- Reading: Section 2.1 and 2.2 of Chapter 2.
- Reading: Appendix B.
- Written Homework # 1.

э

4 E

< ∃ ►

< 1 k