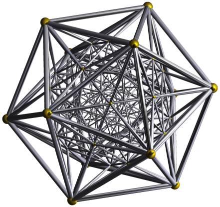
Computational Principles for High-dim Data Analysis

(Lecture Two)

Yi Ma

EECS Department, UC Berkeley

August 31, 2021



Sparse Signal Models

1. Applications
   - Medical Imaging
   - Image Processing
   - Face Recognition

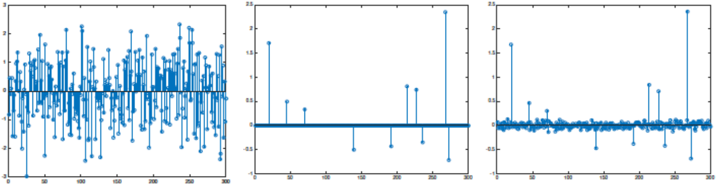
2. Recovering a Sparse Signal
   - Norms for Sparsity
   - The ℓ^0 Norm
   - Minimizing the ℓ^0 for the Sparsest Solution
   - Computational Complexity of ℓ^0 Minimization

Recovering Signals from (Linear) Measurements

$$\underset{\text{observation}}{\boldsymbol{y}} = \underset{\text{measurement matrix}}{\boldsymbol{A}} \; \underset{\text{unknown}}{\boldsymbol{x}} . \tag{1}$$

$$\boldsymbol{y} \in \mathbb{R}^m, \quad \boldsymbol{x} \in \mathbb{R}^n, \quad \boldsymbol{A} \in \mathbb{R}^{m \times n}, m \ll n.$$

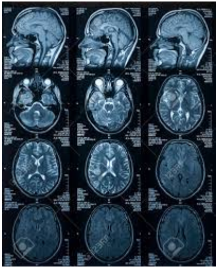
Different “structures” in $\boldsymbol{x}$:

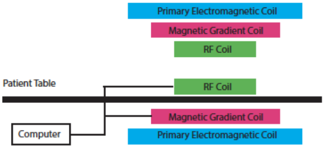


dense vector sparse vector compressible vector

Applications

Application I: Magnetic Resonance Imaging (Chap. 10)





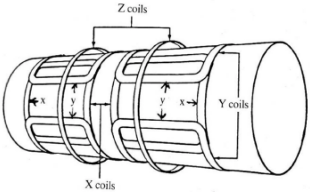


Figure: Left: Key components. Right: The three-axis gradient coils.

Mathematical Model of MRI

Simplified mathematical model for MRI:

$$y = \mathcal{F}[I](\boldsymbol{u}) = \int_{\boldsymbol{v}} I(\boldsymbol{v}) \exp(-\mathrm{i}\, 2\pi\, \boldsymbol{u}^* \boldsymbol{v})\, d\boldsymbol{v}, \quad \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2 \tag{2}$$

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \mathcal{F}[I](\boldsymbol{u}_1) \\ \vdots \\ \mathcal{F}[I](\boldsymbol{u}_m) \end{bmatrix} \doteq \mathcal{F}_{\mathsf{U}}[I], \quad m \ll N^2. \tag{3}$$

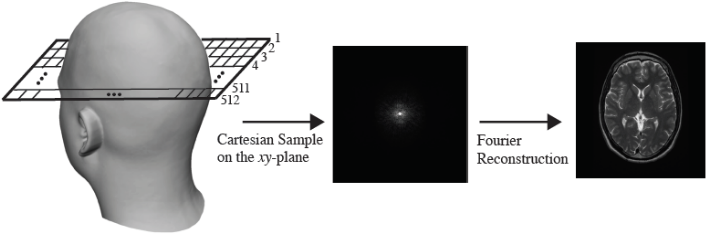


Figure: Recovering MRI image from Fourier measurements.

Exploit Structures of MRI Images

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

$$\underbrace{I}_{\text{image}} = \sum_{i=1}^{N^2} \underbrace{\psi_i}_{i\text{-th basis signal}} \times \underbrace{x_i}_{i\text{-th coefficient}}. \tag{4}$$

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

$$\underbrace{I}_{\text{target image}} \approx \underbrace{\tilde{I}_k = \sum_{i \in \mathsf{J}} \psi_i x_i}_{\text{superposition of } k \text{ basis functions}}. \tag{5}$$

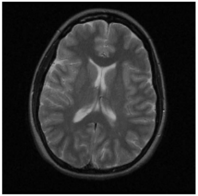
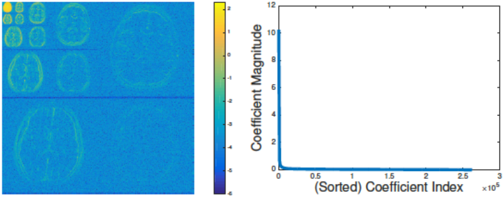


image $I(v)$



wavelet coefficients x: $I = \Psi[x]$.

Figure: MRI image and its wavelet coefficients

Recovering Image from a Under-determined Linear System

$$\begin{aligned} \underbrace{\boldsymbol{y}}_{\text{observed Fourier coefficients}} &= \mathcal{F}_{\mathsf{U}}[I], \\ &= \mathcal{F}_{\mathsf{U}}\Big[\, \boldsymbol{\psi}_1 x_1 + \cdots + \boldsymbol{\psi}_{N^2} x_{N^2} \,\Big], \\ &= \mathcal{F}_{\mathsf{U}}[\boldsymbol{\psi}_1]\, x_1 + \cdots + \mathcal{F}_{\mathsf{U}}[\boldsymbol{\psi}_{N^2}]\, x_{N^2}, \\ &= \underbrace{\Big[\mathcal{F}_{\mathsf{U}}[\boldsymbol{\psi}_1] \mid \cdots \mid \mathcal{F}_{\mathsf{U}}[\boldsymbol{\psi}_{N^2}]\Big]}_{\text{matrix } \boldsymbol{A} \in \mathbb{R}^{m \times N^2},\ m \ll N^2.} \boldsymbol{x}, \\ &= \boldsymbol{A}\boldsymbol{x}. \end{aligned} \tag{6}$$

x is sparse or approximately sparse!

$$I \approx \sum_{i \in \mathsf{J}} \boldsymbol{\psi}_i x_i = \underbrace{\boldsymbol{\Psi}}_{N^2 \times N^2 \text{ matrix}} \underbrace{\boldsymbol{x},}_{\text{sparse vector}} \tag{7}$$

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

Application II: Image Denoising

$$I_{\text{noisy}} = \underbrace{I_{\text{clean}}}_{\text{target image}} + \underbrace{z}_{\text{noise}}. \tag{8}$$

Break I_{clean} into patches $y_{1\text{clean}}, \ldots, y_{p\text{clean}}$:

$$y_i = y_{i\text{clean}} + z_i = \underbrace{A}_{\text{patch dictionary}} \times \underbrace{x_i}_{\text{sparse coefficient vector}} + z_i. \tag{9}$$

Figure: Left: input image; middle: denoised; right: dictionary for image patches.

Sparse representation for color image restoration, Mairal, Elad, and Sapiro. TIP, 2008

Application II: Image Super-Resolution

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

$$\boldsymbol{y}_{i\,\text{low}} = \boldsymbol{A}_{\text{low}}\boldsymbol{x}_i \quad \xrightarrow{\textbf{Lifting}} \boldsymbol{y}_{i\,\text{high}} = \boldsymbol{A}_{\text{high}}\boldsymbol{x}_i. \tag{10}$$

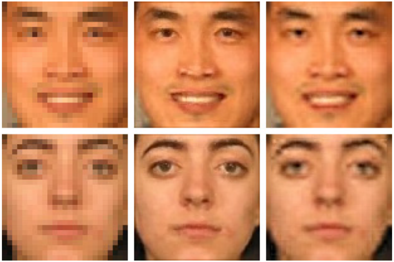
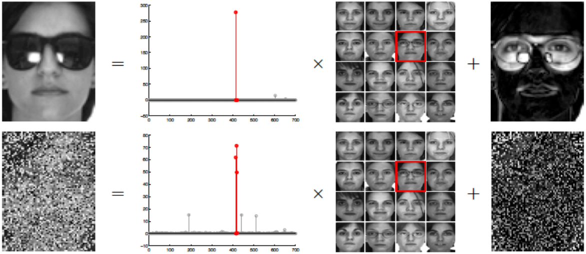


Figure: Super resolution for face images

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

Application III: Robust Face Recognition (Chap. 13)



$$\underbrace{\boldsymbol{y}}_{\text{observation}} = \underbrace{\boldsymbol{y}_o}_{\text{clean data}} + \underbrace{\boldsymbol{e}}_{\text{sparse error}} \in \mathbb{R}^m. \tag{11}$$

Concatenate gallery images of n subjects into a large "dictionary":

$$\boldsymbol{B} = \underbrace{[\boldsymbol{B}_1 \mid \boldsymbol{B}_2 \mid \cdots \mid \boldsymbol{B}_n]}_{\text{all training images}} \in \mathbb{R}^{m \times n}, \quad n = \sum_i n_i. \tag{12}$$

Application III: Robust Face Recognition (Chap. 13)

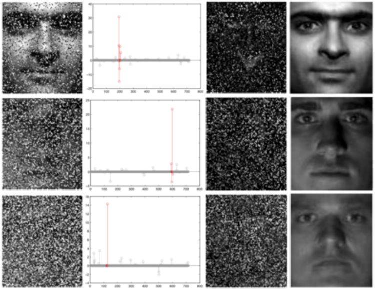
Find sparse solutions $(\boldsymbol{x}, \boldsymbol{e})$ to the linear system:

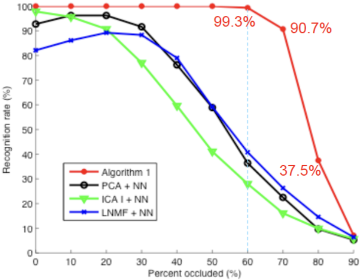
$$\boldsymbol{y} \;=\; \boldsymbol{B}\boldsymbol{x} + \boldsymbol{e} \;=\; [\boldsymbol{B}, \boldsymbol{I}]\left[\begin{smallmatrix}\boldsymbol{x}\\ \boldsymbol{e}\end{smallmatrix}\right]. \tag{13}$$

Extended Yale B Database
(38 subjects)

Training: subsets 1 and 2 (717 images)
Testing: subset 3 (453 images)

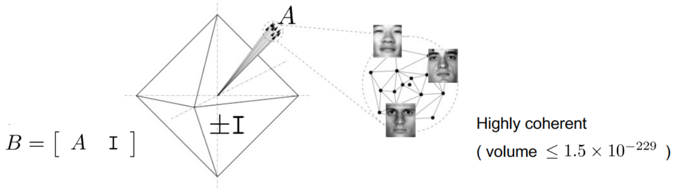
$\boldsymbol{y}$ $\quad \hat{\boldsymbol{x}}_1 \quad \hat{\boldsymbol{e}}_1 \quad \hat{\boldsymbol{y}}_0 = A\hat{\boldsymbol{x}}_1$





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

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 $\boldsymbol{\sigma}$ chosen uniformly at random,*

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

> *"ℓ^1 recovers any sparse signal from almost any error with density less than 1"*

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

Many Many Applications...

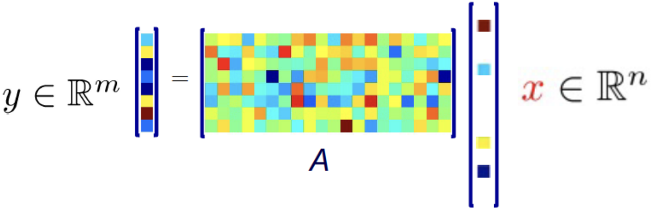
Part III of the Textbook:

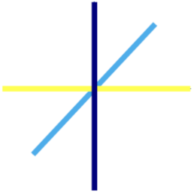
- Magnetic Resonnance Imaging (Chapter 10)
- Wideband Spectrum Sensing (Chapter 11)
- Scientific Imaging Problems (Chapter 12)
- Robust Face Recognition (Chapter 13)
- Robust Photometric Stereo (Chapter 14)
- Structured Texture Recovery (Chapter 15)
- Deep Networks for Classification (Chapter 16)

Your final projects?

How to Find a Sparse Solution x?

$$\boldsymbol{y} = \underset{m\times n,\; m\ll n}{\boldsymbol{A}} \times \underset{k \text{ sparse}}{\boldsymbol{x}}. \tag{14}$$

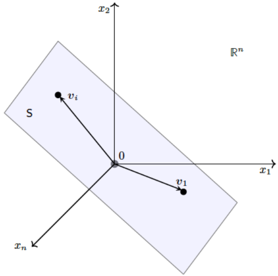




Vector Spaces and Linear Algebra

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**: $\|\boldsymbol{x}\| \geq 0$, and $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$;
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
$$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}.$$

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

$$\|\boldsymbol{x}\|_p \doteq \left(\sum_i |x_i|^p \right)^{1/p}, \quad p \in [1, \infty]. \tag{15}$$

Sparse Vectors and Norms

ℓ^∞ norm or the "cube" norm:

$$\|\boldsymbol{x}\|_\infty = \max_i |x_i|. \tag{16}$$

ℓ^2 norm or the "Euclidean norm":

$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_i |x_i|^2} = \sqrt{\boldsymbol{x}^*\boldsymbol{x}}. \tag{17}$$

ℓ^1 norm:

$$\|\boldsymbol{x}\|_1 = \sum_i |x_i|, \tag{18}$$

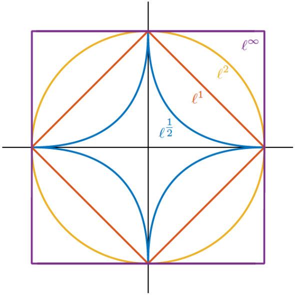


Figure: The ℓ^p balls.

ℓ^p unit ball:

$$\mathsf{B}_p \doteq \{\boldsymbol{x} \mid \|\boldsymbol{x}\|_p \leq 1\} \quad 0 < p \leq \infty. \tag{19}$$

Example: relative volumes of the ℓ^p balls when n is large.

Measure of Sparsity: the ℓ^0 Norm

ℓ^0 "norm":

$$\|\boldsymbol{x}\|_0 = \#\{i \mid \boldsymbol{x}(i) \neq 0\}. \tag{20}$$

Not a norm, and only subadditive:

$$\forall\, \boldsymbol{x},\ \boldsymbol{x}', \qquad \left\|\boldsymbol{x} + \boldsymbol{x}'\right\|_0 \ \leq\ \|\boldsymbol{x}\|_0 + \left\|\boldsymbol{x}'\right\|_0. \tag{21}$$

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

$$\lim_{p \searrow 0} \|\boldsymbol{x}\|_p^p = \sum_{i=1}^{n} \lim_{p \searrow 0} |\boldsymbol{x}(i)|^p = \sum_{i=1}^{n} \mathbb{1}_{\boldsymbol{x}(i) \neq 0} = \|\boldsymbol{x}\|_0. \tag{22}$$

Minimizing the ℓ^0 Norm

Given $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_o$, to find $\boldsymbol{x}_o$ as the sparsest solution to $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, we can:

$$\hat{\boldsymbol{x}} = \arg\min \|\boldsymbol{x}\|_0 \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}. \tag{23}$$

The "support" of the nonzero entries:

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

Brute force exhaustive search:

$$\boldsymbol{A}_{\mathsf{I}}\boldsymbol{x}_{\mathsf{I}} = \boldsymbol{y}? \quad \forall \mathsf{I} \subseteq \{1, \ldots, n\}, \ |\mathsf{I}| \leq k. \tag{25}$$

Uniqueness of the Sparsest Solution

Definition (Kruskal Rank)

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

Theorem (ℓ^0 Recovery)

Suppose that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_o$, *with*

$$\|\boldsymbol{x}_o\|_0 \;\le\; \tfrac{1}{2}\,\mathrm{krank}(\boldsymbol{A}). \tag{26}$$

Then $\boldsymbol{x}_o$ *is the unique optimal solution to the* ℓ^0 *minimization problem*

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

Proof: $\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{y} \Rightarrow \boldsymbol{A}\left(\hat{\boldsymbol{x}} - \boldsymbol{x}_o\right) = \boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{A}\boldsymbol{x}_o = \boldsymbol{y} - \boldsymbol{y} \;=\; \boldsymbol{0}.$

Minimizing the ℓ^0 Norm: Exhaustive Search

Algorithm (ℓ^0 Minimization via Exhaustive Search):

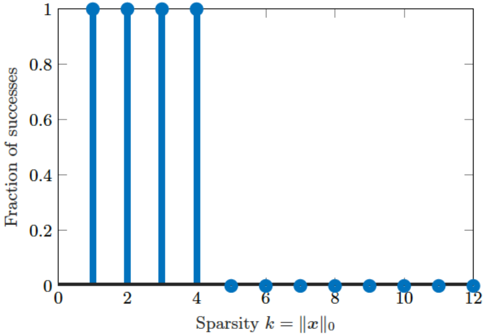
Input: a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and a vector $\boldsymbol{y} \in \mathbb{R}^{m}$.
for $k = 0, 1, 2, \ldots, n$,
    for each $\mathsf{I} \subseteq \{1, \ldots, n\}$ of size k,
        if the system of equations $\boldsymbol{A}_{\mathsf{I}} \boldsymbol{z} = \boldsymbol{y}$ has a solution $\boldsymbol{z}$,
            set $\boldsymbol{x}_{\mathsf{I}} = \boldsymbol{z}$, $\boldsymbol{x}_{\mathsf{I}^c} = \boldsymbol{0}$.
            return $\boldsymbol{x}$.
        end if
    end for
end for

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

Minimizing the ℓ^0 Norm: Simulations

$$\textbf{Solve:} \quad \min \|\boldsymbol{x}\|_0 \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}. \tag{28}$$

$\boldsymbol{A}$ is of size 5×12. Fraction of success across 100 trials.

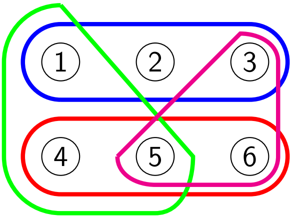


Computational Complexity of ℓ^0 Minimization

Theorem (Hardness of ℓ^0 Minimization)

The ℓ^0-minimization problem $\min \|\boldsymbol{x}\|_0$ s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ is (strongly) **NP***-hard.*

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



$$\underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{x}}$$

Figure: **Exact 3-Set Cover as a Sparse Representation Problem. Left:** a universe $\mathsf{S} = \{1, \ldots, 6\}$ and four subsets $\mathsf{U}_1, \ldots, \mathsf{U}_4 \subseteq \mathsf{S}$. $\{\mathsf{U}_1, \mathsf{U}_3\}$ is an exact 3-set cover. **Right:** the same problem as a linear system of equations. The columns of $\boldsymbol{A}$ are the incidence vectors for the sets $\mathsf{U}_1, \mathsf{U}_2, \mathsf{U}_3, \mathsf{U}_4$. The Exact 3-Cover $\{\mathsf{U}_1, \mathsf{U}_3\}$ corresponds to a solution $\boldsymbol{x}$ to the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ with only $m/3 = 2$ nonzero entries.

Computational Complexity of ℓ^p Minimization $(0 < p < 1)$

Theorem (Hardness of ℓ^p Minimization)

For any fixed $0 < p < 1$, the ℓ^p-minimization problem

$$\min \|\boldsymbol{x}\|_p^p \quad \textit{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

is (strongly) **NP***-hard.*

Corollary (Hardness of Smoothed ℓ^p Minimization)

For any fixed $0 < p < 1$ and $\boldsymbol{\varepsilon} > \mathbf{0}$, the smoothed ℓ^p-minimization problem $\min \|\boldsymbol{x} + \boldsymbol{\varepsilon}\|_p^p$ s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ is (strongly) **NP***-hard.*

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

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

Two Fundamental Questions

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

$$\boldsymbol{y} = \underset{m\times n,\ m\ll n}{\boldsymbol{A}} \times \underset{k\ \text{sparse}}{\boldsymbol{x}}, \tag{29}$$

we want to know:

1. **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

Assignments

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