L_0 - Minimization

The problem:

$$\begin{aligned} &\min \|x\|_0 \\ &\text{s.t. } Ax = b \end{aligned} \qquad \text{———————} \quad (P_0)$$

Remark: if we relax the constant to $\|Ax-b\|^2 \leq \varepsilon$, then under appropriate choices of ε and λ,

$$\begin{aligned} &\min \|x\|_0 \\ &\text{s.t. } \|Ax-b\| \leq \varepsilon \end{aligned} \quad = \quad \min_x \|Ax-b\|^2 + \lambda\|x\|_0 .$$

Meaning of (P_0): We know that there are infinitely many solutions for (P_0). But there is one with the fewest non-zero. Our goal is to find out this solution.

Compare (P_0) with LASSO: Typically in LASSO, we assume $y \cong X\beta + \varepsilon$. There is nothing called "infinitely many solution" because the goal is to find approximation (or fitting) to y.

What is $\|x\|_0$? Note that $\|x\|_0$ is only a quasi-norm, and is not a norm.

Triangle inequality: (Yes)

$$\|x+y\|_0 \leq \|x\|_0 + \|y\|_0$$

Scalar multiplication: (No)

$$\|tx\|_0 \neq t\|x\|_0 .$$

Properties of L_0-Minimization

Mutual Coherence

Two-ortho System

$A = [\Psi, \Phi]$,
where Ψ and Φ orthonormal

$$\boxed{\mu(A) \overset{\text{def}}{=} \max_{i,j} |\psi_i^T \phi_j|}$$

General System

Any A

$$\boxed{\mu(A) \overset{\text{def}}{=} \max_{i,j} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2}}$$

Why care about mutual coherence?
- We need a way to quantify the solution of (P_0).
- $\mu(A)$ provides some measurable method for this purpose.

Important Theorem (Elad 2.1)
Let α and β be such that

$$y = \Psi\alpha \quad \text{and} \quad y = \Phi\beta.$$

then

$$\|\alpha\|_0 + \|\beta\|_0 \geq \frac{2}{\mu(A)}, \quad \text{where } A = [\Psi, \Phi].$$

Why is this result important?
- α and β are the "sub-solution" to (P_0).
- Theorem says the sparsity of the solution cannot be arbitrarily low.

So do we want $\mu(A)$ large or small?

Small $\mu(A) \Rightarrow \|\alpha\|_0 + \|\beta\|_0 \gtrsim$ large #.

So # non-zero will be large. i.e. NOT sparse.

large $\mu(A) \Rightarrow$ fewer # non-zero.

i.e. sparse.

$$\boxed{\begin{array}{l}\underline{\text{Property of } \mu(A)} \\ \text{if } A = [\Psi, \Phi], \text{ then} \\ \quad \frac{1}{\sqrt{n}} \leq \mu(A) \leq 1\end{array}}$$

proof: "$\mu(A) \leq 1$":

(optional) $|\psi_i^T \phi_j|^2 \leq \|\psi_i\|^2 \|\phi_j\|^2$, by Cauchy

$= 1, \quad \forall i, j.$

So $\max_{ij} |\psi_i^T \phi_j| \leq 1.$

"$\mu(A) \geq \frac{1}{\sqrt{n}}$"

Suppose $\mu(A) < \frac{1}{\sqrt{n}}$, then $\max_{ij} |\psi_i^T \phi_j| < \frac{1}{\sqrt{n}} \ \forall i,j$

$\Rightarrow |\psi_i^T \phi_j| < \frac{1}{\sqrt{n}} \quad \forall i,j$

choose $\phi_j = [0 \ldots \underset{\substack{\uparrow \\ j^{th}}}{1} \ldots 0]$, then $|\psi_i^T \phi_j| = |\psi_{ij}|$

Since $|\psi_i^T \phi_j| < \frac{1}{\sqrt{n}} \quad \forall i,j$, we have $|\psi_{ij}| < \frac{1}{\sqrt{n}}$.

choose Ψ = Fourier. Then

$\|\psi_i\|^2 = \sum_{j=1}^{n} \psi_{ij}^2 < \sum_{j=1}^{n} \left(\frac{1}{\sqrt{n}}\right)^2 = 1.$ Contradiction. So must have $\mu(A) \geq \frac{1}{\sqrt{n}}$.

The result of Theorem Elad 2.1 is not enough.

(i) it is about "sub-solutions" of a system

(ii) it does not address the problem who is the "most sparse solution".

Theorem Elad 2.3 $\quad A = [\Psi, \Phi]$

If a solution $Ax = y$ has # less than $\frac{1}{\mu(A)}$ non-zeros, then it is necessarily the sparest solution

---

__proof__: (outline) First assume that α and β are both

$\cancel{\Psi\alpha = y \text{ and } \Phi\beta = y}$

solutions. Then

$$A\alpha = y \quad \text{and} \quad A\beta = y$$

$\Rightarrow \; A(\alpha - \beta) = 0$, then we let $e = \alpha - \beta$, and $e \in Null(A)$.

$$\Rightarrow \; \begin{bmatrix} \Psi & \Phi \end{bmatrix} \begin{bmatrix} e_\psi \\ e_\phi \end{bmatrix} = 0$$

$\Rightarrow \; \Psi e_\psi = -\Phi e_\phi$ So by Elad 2.1 we have

something

$$\underbrace{\|e_\psi\|_0 + \|e_\phi\|_0}_{= \|e\|_0 \text{ because } e = \begin{bmatrix} e_\psi \\ e_\phi \end{bmatrix}} \geq \frac{2}{\mu(A)}.$$

and then

$$\|e\|_0 \leq \|\alpha\|_0 + \|\beta\|_0 \text{ by triangle inequality.}$$

So we have

$$\|\alpha\|_0 + \|\beta\|_0 \geq \frac{2}{\mu(A)}.$$

Now, if $\|\alpha\|_0 \leq \frac{1}{\mu(A)}$, then $\|\beta\|_0 \geq \frac{1}{\mu(A)}$.

So ~~that~~ α is the most sparse solution.

Is mutual coherence good?

- Not quite. Limited to two-ortho so far.

Spark

The spark of a matrix A is the smallest number of columns that are linearly dependent

spark VS rank:

rank = largest number of columns that are linearly independent.

Computing spark is NP-hard!!

Theorem Elad 2.4

If a system $Ax = y$ has a solution such that $\|x\|_0 < \text{spark}(A)/2$, then x is necessarily the most sparse solution.

Good thing about Elad 2.4:

- works for any matrix A
- global optimal certificate

Bad thing:

- NP hard to verify

not necessarily $[\Phi, \Phi]$.

For any matrix A,

$$\text{spark}(A) \geq 1 + \frac{1}{\mu(A)}. \qquad (*)$$

$\boxed{\begin{array}{l}\text{Theorem 2.5 Elad} \\ \text{If a system has a solution } x \text{ s.t.} \\ \qquad \|x\|_0 < \frac{1}{2}\left(1+\frac{1}{\mu(A)}\right), \text{ then} \\ x \text{ is necessarily the most sparse solution}\end{array}}$

proof: x is most sparse if

$$\|x\|_0 < \text{spark}(A)/_2 .$$

~~So if $\|x\|$~~ Now, since

$$\text{spark}(A) \geq 1 + \frac{1}{\mu(A)} .$$

So if $\|x\|_0 < \frac{1}{2}\left(1+\frac{1}{\mu(A)}\right)$ then $\|x\|_0 < \frac{\text{spark}(A)}{2}$. .

Summary

① (P_0) is NP hard problem. Solution exists, but very difficult to find.

② $\mu(A)$ and $\text{Spark}(A)$ are quantities to verify whether a solution is the most sparse or not.

③ $\mu(A)$ is easier to compute.
$\text{Spark}(A)$ is NP hard to compute.