Uniqueness of solution on sparse recovery problem

Condition of Kruskal rank and incoherent on the $\mathbf{A}$ matrix

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft: Janurary 8, 2020
Last update : February 5, 2020

Overview

Solving linear system : old and new

- Theme : given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, find $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$
- Traditional wisdom (before year 2000) : if $m < n$, there are ∞-ly many solution, no unique solution
  - i.e., we have a vector $\mathbf{y}$ such that $\mathbf{A}\mathbf{y} = \mathbf{b}$ with $\mathbf{y} = \mathbf{x} + \mathbf{z}$, where $\mathbf{z} \in$Null$(\mathbf{A})$. The size of the set $\{\mathbf{z}\}$ is ∞-ly big
  - In this case, no guarantee on the recovered solution $\mathbf{x}$ is exactly the $\mathbf{x}$ that generate the $\mathbf{b}$
  - To recover $\mathbf{x}$ uniquely, we need at least n equations for n unknowns (m has to be equal to n)
- New situation (around year 2000 to 2010) : we can recover the unique $\mathbf{x}$ if $\mathbf{x}$ has additional structure — $\mathbf{x}$ is *sparse*

Sparsity

- Measure of *sparsity* of a vector : the number of non-zero element in the vector
- **Support** : the *support* of a vector is a set that contains the indices of non-zero elements

$$\text{supp}(\mathbf{x}) := \Big\{ i \mid i \in \{1, 2, \ldots, n\}, x_i \neq 0 \Big\}$$

- L_0-**norm**, denoted as $\|\mathbf{x}\|_0$, is the size of the support of $\mathbf{x}$

$$\|\mathbf{x}\|_0 := |\text{supp}(\mathbf{x})| = \text{number of non-zero elements in } \mathbf{x}$$

  - L_0-norm is not a norm : it does not satisfy $\|\lambda\mathbf{x}\|_0 = |\lambda|\|\mathbf{x}\|_0$ for all $\lambda \neq 0$, as $\|\lambda\mathbf{x}\|_0 = \|-\lambda\mathbf{x}\|_0 = \|\mathbf{x}\|_0$ for $\lambda \neq 0$

Sparse recovery problem

- Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, find $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x}$ is *as sparse as possible*

- Mathematically,

$$(\mathcal{P}_0) \quad : \quad \min_{x \in \mathbb{R}^n} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{Ax} = \mathbf{b}.$$

- Suppose we have an algorithm that solves $(\mathcal{P}_0)$, the question is : suppose $\mathbf{x}_0$ is the ground truth factor that generate $\mathbf{b}$, what is the condition that the solution produced by the algorithm is exactly $\mathbf{x}_0$?

- In other words, what is the condition on $\mathbf{x}$ (and possibly on $\mathbf{A}$, $\mathbf{b}$, and also on the algorithm itself), such that the solution to $(\mathcal{P}_0)$ is unique ?

Structural conditions on $\mathbf{A}$

- Turns out that sparsity of $\mathbf{x}$ is not enough, we need extra condition on the matrix $\mathbf{A}$
- Based on different conditions on $\mathbf{A}$, we have different uniqueness guarantee
- This document focuses on two conditions on $\mathbf{A}$, namely the *Kruskal rank* and *incoherence*
- These two conditions describe the "collective behavior" on the columns of $\mathbf{A}$ on "similarity". That is, how "similar" are these columns to each other. Or, "how non-similar the columns can be"

Kruskal rank and incoherence

- **Kruskal rank** (k-rank) : k-rank$(\mathbf{A})$ = the largest number r such that any subset of columns of $\mathbf{A}$ has at most r-linear independent vectors
  - It is like "no matter how you select the columns, at most r of them are linear independent"
  - It is like "at most r of them are non-similar"

- **Incoherence** : for $\mu > 0$, $\mathbf{A}$ is *μ-incoherent* if the magnitude of the inner product of any two columns in $\mathbf{A}$ is upper bounded by μ, provided that columns of $\mathbf{A}$ have unit norm
  - Mathematically, it is $|\langle \mathbf{A}_i, \mathbf{A}_j \rangle| \leq \mu$, where $\mathbf{A}_j$ is the j-th column of $\mathbf{A}$
  - Inner product represents "similarity" of vectors, so incoherence is like "any two columns of $\mathbf{A}$ are at most μ-similar"
  - It is like an lower bound on "how non-similar these columns can be"
  - In general, columns of $\mathbf{A}$ do not have unit norm, in this case, the more general definition of μ-incoherent is

$$\frac{|\langle \mathbf{A}_i, \mathbf{A}_j \rangle|}{\|\mathbf{A}_i\| \|\mathbf{A}_j\|} \leq \mu$$

Uniqueness theorems on sparse recovery problem

$$(\mathcal{P}_0) \ : \ \min_{x \in \mathbb{R}^n} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}$$

- Uniqueness based on k-rank of $\mathbf{A}$
  If k-rank$(\mathbf{A}) \geq r$ and $\|\mathbf{x}_0\|_0 = \frac{r}{2}$ with $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$, then $\mathbf{x}_0$ is the unique solution to $(\mathcal{P}_0)$.

- Uniqueness based on incoherence of $\mathbf{A}$
  If $\mathbf{A}$ is μ-incoherent with unit column vectors, and $\|\mathbf{x}_0\|_0 < \frac{1}{2\mu}$ with $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$, then $\mathbf{x}_0$ is the unique solution to $(\mathcal{P}_0)$.

The remainder of this document gives the proofs of these two theorems. The proofs are contradiction-based : assumes there is another solution $\mathbf{y} \neq \mathbf{x}_0$ of $(\mathcal{P}_0)$, try to arrive at a contradiction to disprove the existence of $\mathbf{y}$.

Theorem Given $\mathbf{A} \in \mathbb{R}^{m\times n}$ has k-rank at least r, $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{x}_0 \in \mathbb{R}^n$ is $\frac{r}{2}$-sparse with $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$, then $\mathbf{x}_0$ is the unique solution to

$$(\mathcal{P}_0) \; : \; \min_{x\in\mathbb{R}^n} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}.$$

Proof. Let

$$\mathbf{y} \neq \mathbf{x}_0 \tag{1}$$

be a solution of $(\mathcal{P}_0)$.

Consider the difference between $\mathbf{x}_0$ and $\mathbf{y}$, let $\mathbf{z} = \mathbf{x}_0 - \mathbf{y} \overset{(1)}{\neq} 0$, then

$$\begin{aligned}
\mathbf{A}\mathbf{z} &= \mathbf{A}(\mathbf{x}_0 - \mathbf{y}) = \mathbf{A}\mathbf{x}_0 - \mathbf{A}\mathbf{y} \\
&= \mathbf{b} - \mathbf{A}\mathbf{y} && \text{by assumption} \\
&= \mathbf{b} - \mathbf{b} && \text{by assumption } \mathbf{y} \text{ solves } (\mathcal{P}_0) \\
&= 0
\end{aligned}$$

Hence, we have $\mathbf{z} \in \text{Null}(\mathbf{A})$.

Uniqueness based on k-rank of $\mathbf{A}$ ... (2/3)

$\mathbf{z} \in$Null$(\mathbf{A})$ means the linear combination of columns of $\mathbf{A}$ using $\mathbf{z}$ as coefficient is zero, i.e.,

$$z_1\mathbf{A}_1 + z_2\mathbf{A}_2 + \cdots + z_n\mathbf{A}_n = 0. \tag{2}$$

This is like "$\mathbf{z}$ is picking columns of $\mathbf{A}$ to give a zero sum".

By assumption, k-rank$(\mathbf{A}) \geq r$, meaning (at least) r columns in $\mathbf{A}$ are linearly independent.

Linear combination of linearly independent vectors can never be zero.

For (2) to hold true, $\mathbf{z}$ needs to have at least $r+1$ non-zero so that $\mathbf{z}$ selects more than those r independent columns in $\mathbf{A}$ to make the sum in (2) zero.

Mathematically,

$$\mathbf{z} \in \text{Null}(\mathbf{A}) \;\text{ and }\; \text{k-rank}(\mathbf{A}) \geq r \implies \|\mathbf{z}\|_0 \geq r+1.$$

Uniqueness based on k-rank of $\mathbf{A}$ ... (3/3)

We have $\|\mathbf{z}\|_0 \geq r+1$, $\mathbf{z} = \mathbf{x}_0 - \mathbf{y}$ and $\|\mathbf{x}_0\|_0 = \frac{r}{2}$ (by assumption), we can now consider $\|\mathbf{y}\|_0$

$$\begin{array}{rcll}
\|\mathbf{y}\|_0 & = & \|\mathbf{x}_0 - \mathbf{z}\|_0 & \\
& = & \|\mathbf{z} - \mathbf{x}_0\|_0 & L_0 \text{ norm invariant to sign} \\
& \geq & \|\mathbf{z}\|_0 - \|\mathbf{x}_0\|_0 & \text{standard inequality} \\
& = & r + 1 - \frac{r}{2} & \text{first line this page} \\
& = & \frac{r}{2} + 1 & \\
& > & \frac{r}{2} & \\
& = & \|\mathbf{x}_0\|_0 &
\end{array}$$

Hence we arrived at $\|\mathbf{y}\|_0 > \|\mathbf{x}_0\|_0$, the objective value at $\mathbf{y}$ is higher than that at $\mathbf{x}_0$, the objective value at $\mathbf{y}$ is not optimal, contradicting to the assumption that $\mathbf{y}$ solves $(\mathcal{P}_0)$.

So such $\mathbf{y}$ cannot exist, the solution of $(\mathcal{P}_0)$ is unique and it is $\mathbf{x}_0$. $\square$

Uniqueness based on incoherence of $\mathbf{A}$ ... (1/4)

Theorem Given $\mathbf{A} \in \mathbb{R}^{m\times n}$ is μ-incoherent with unit column vectors, $\mathbf{b} \in \mathbb{R}^m$ and $\|\mathbf{x}_0\|_0 < \frac{1}{2\mu}$ with $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$, then $\mathbf{x}_0$ is the unique solution to

$$(\mathcal{P}_0) \; : \; \min_{x\in\mathbb{R}^n} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}.$$

Proof. Let

$$\mathbf{y} \neq \mathbf{x}_0 \tag{3}$$

be a solution of $(\mathcal{P}_0)$ with $\|\mathbf{y}\|_0 < \frac{1}{2\mu}$.

Consider the difference between $\mathbf{x}_0$ and $\mathbf{y}$: let $\mathbf{z} = \mathbf{x}_0 - \mathbf{y} \overset{(3)}{\neq} 0$.

$$\begin{aligned}
\|\mathbf{z}\|_0 &= \|\mathbf{x}_0 - \mathbf{y}\|_0 \\
&= \|\mathbf{x}_0\|_0 + \|-\mathbf{y}\|_0 && \text{triangle inequality} \\
&= \|\mathbf{x}_0\|_0 + \|\mathbf{y}\|_0 && L_0 \text{ norm invariant to sign} \\
&< \frac{1}{2\mu} + \frac{1}{2\mu} && \text{by assumption} \\
&= \frac{1}{\mu}
\end{aligned}$$

So we have $\|\mathbf{z}\|_0 < \frac{1}{\mu}$.

Uniqueness based on incoherence of $\mathbf{A}$ ... (2/4)

Similar to the proof in the k-rank case, we have $\mathbf{A}\mathbf{z} = 0$.
So $\langle \mathbf{A}\mathbf{z}, \mathbf{A}\mathbf{z} \rangle = \langle \mathbf{A}^\top \mathbf{A}\mathbf{z}, \mathbf{z} \rangle = \mathbf{z}^\top \mathbf{A}^\top \mathbf{A}\mathbf{z} = 0$.
Let $\mathbf{M} = \mathbf{A}^\top \mathbf{A}$, we have $\mathbf{z}^\top \mathbf{M}\mathbf{z} = 0$.
$\|\mathbf{z}\|_0 < \frac{1}{\mu}$ means $\mathbf{z}$ is sparse and some elements of $\mathbf{z}$ are zero. Let $S = \text{supp}(\mathbf{z})$, we can restrict $\mathbf{z}^\top \mathbf{M}\mathbf{z} = 0$ to $\mathbf{z}_S^\top \mathbf{M}_S \mathbf{z}_S = 0$, where the subscript indicates the sub-vector of $\mathbf{z}$ restricted to the indices in S, and the sub-block matrix $\mathbf{M}_S$ restricted to the column and row with indices in S. An illustration example, suppose $n = 3$ and $z_3 = 0$, so $S = \{1, 2\}$ and

$$\begin{pmatrix} z_1 \\ z_2 \\ 0 \end{pmatrix}^\top \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ 0 \end{pmatrix} = 0 \iff \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}^\top \underbrace{\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}}_{\mathbf{M}_S} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 0$$

On $\mathbf{z}_S^\top \mathbf{M}_S \mathbf{z}_S = 0$, note that $\mathbf{z}_S$ is a full non-zero vector, all elements in $\mathbf{z}_S$ are non-zero. Hence for $\mathbf{z}_S^\top \mathbf{M}_S \mathbf{z}_S$ equal to zero, the columns or rows in $\mathbf{M}_S$ have to be linear dependent, and therefore

$$\text{Matrix } \mathbf{M}_S \text{ is singular.} \tag{4}$$

Uniqueness based on incoherence of $\mathbf{A}$ ... (3/4)

We haven't use the assumption $\mathbf{A}$ is μ-incoherent yet, now we use it.

Note that $\mathbf{M}_S$ is the sub-block matrix of $\mathbf{M} = \mathbf{A}^\top \mathbf{A}$, in which the elements are basically the inner product $\langle \mathbf{A}_i, \mathbf{A}_j \rangle$, so the diagonal of $\mathbf{M}$ will be $\langle \mathbf{A}_i, \mathbf{A}_i \rangle$ and the off-diagonal elements will be $\langle \mathbf{A}_i, \mathbf{A}_j \rangle, j \neq i$.

By assumption the columns of $\mathbf{A}$ have unit norm, we get $\langle \mathbf{A}_i, \mathbf{A}_i \rangle = 1$.

By assumption $\mathbf{A}$ is μ-incoherent, we get $\langle \mathbf{A}_i, \mathbf{A}_j \rangle \leq |\langle \mathbf{A}_i, \mathbf{A}_j \rangle| \leq \mu$.

Together we have

$$\mathbf{M}_S = \begin{pmatrix} 1 & * & * & \dots \\ * & 1 & * & \dots \\ * & * & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \; |*| \leq \mu$$

The next step is tricky : to link the elements of $\mathbf{M}_S$ with eigenvalues of $\mathbf{M}_S$, and to show $\mathbf{M}_S$ is non-singular.

$$\mathbf{M}_S = \begin{pmatrix} 1 & * & * & \dots \\ * & 1 & * & \dots \\ * & * & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \ |*| \le \mu$$

By **Gershgorin Circle Theorem** : the eigenvalues $\lambda_i(\mathbf{M}_S)$ will be located within a circle in the complex plane centered at 1 with radius $\mu|S| < 1$.

As the circle has radius < 1, so the circle never touch the origin $(0, 0)$,
$\implies$ none of the eigenvalue of $\mathbf{M}_S$ will be zero $\implies$ $\mathbf{M}_S$ is non-singular.

This contradict to (4), hence, such $\mathbf{y}$ does not exists, and thus the solution to $(\mathcal{P}_0)$ is unique and it is $\mathbf{x}_0$. □

What is Gershgorin Circle Theorem : All eigenvalue of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ lies within at least one of the Gershgorin discs $D(a_{ii}, \sum_{j \neq i} |a_{ij}|)$, for $i \in \{1, \dots, n\}$.

$$(\mathcal{P}_0) \ : \ \min_{x \in \mathbb{R}^n} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}$$

- Uniqueness based on k-rank of $\mathbf{A}$
  If k-rank$(\mathbf{A}) \geq r$ and $\|\mathbf{x}_0\|_0 = \frac{r}{2}$ with $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$, then $\mathbf{x}_0$ is the unique solution to $(\mathcal{P}_0)$.

- Uniqueness based on incoherence of $\mathbf{A}$
  If $\mathbf{A}$ is μ-incoherent with unit column vectors, and $\|\mathbf{x}_0\|_0 < \frac{1}{2\mu}$ with $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$, then $\mathbf{x}_0$ is the unique solution to $(\mathcal{P}_0)$.

- Remarks
  - Problem $(\mathcal{P}_0)$ is a non-convex problem, but solving it yields unique solution (the two theorems above)
  - However, problem $(\mathcal{P}_0)$ is NP-Hard to solve.

End of document