Introduction to Optimization

K. R. Sahasranand

Data Science

sahasranand@iitpkd.ac.in

Full-rank factorization

a matrix of rank r = a matrix of full column rank r $\times$ a matrix of full row rank r.

Full-rank factorization

a matrix of rank r = a matrix of full column rank r $\times$ a matrix of full row rank r.

Helper lemma – Let $A \in \mathbb{R}^{m \times n}$, $\text{rank } A = r \leq \min\{m, n\}$. Then, there exist matrices $B \in \mathbb{R}^{m \times r}$ and $C \in \mathbb{R}^{r \times n}$ such that

$$A = BC$$

with

$$\text{rank } A = \text{rank } B = \text{rank } C = r.$$

Full-rank factorization

a matrix of rank r = a matrix of full column rank r $\times$ a matrix of full row rank r.

Helper lemma – Let $A \in \mathbb{R}^{m \times n}$, $\text{rank } A = r \leq \min\{m, n\}$. Then, there exist matrices $B \in \mathbb{R}^{m \times r}$ and $C \in \mathbb{R}^{r \times n}$ such that

$$A = BC$$

with

$$\text{rank } A = \text{rank } B = \text{rank } C = r.$$

Proof – WLOG, $a_1, \ldots, a_r$ (columns of A) be linearly independent

Full-rank factorization

a matrix of rank r = a matrix of full column rank r $\times$ a matrix of full row rank r.

Helper lemma – Let $A \in \mathbb{R}^{m \times n}$, $\mathrm{rank}\ A = r \leq \min\{m, n\}$. Then, there exist matrices $B \in \mathbb{R}^{m \times r}$ and $C \in \mathbb{R}^{r \times n}$ such that

$$A = BC$$

with

$$\mathrm{rank}\ A = \mathrm{rank}\ B = \mathrm{rank}\ C = r.$$

Proof – WLOG, $a_1, \ldots, a_r$ (columns of A) be linearly independent

$$B = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_r \\ | & & | \end{bmatrix}$$

Full-rank factorization

a matrix of rank r = a matrix of full column rank r $\times$ a matrix of full row rank r.

Helper lemma – Let $A \in \mathbb{R}^{m \times n}$, $\text{rank } A = r \leq \min\{m, n\}$. Then, there exist matrices $B \in \mathbb{R}^{m \times r}$ and $C \in \mathbb{R}^{r \times n}$ such that

$$A = BC$$

with

$$\text{rank } A = \text{rank } B = \text{rank } C = r.$$

Proof – WLOG, $a_1, \ldots, a_r$ (columns of A) be linearly independent

$$B = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_r \\ | & & | \end{bmatrix} \quad ; \quad C = \begin{bmatrix} 1 & \cdots & 0 & c_{1,r+1} & \cdots & c_{1,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & c_{m,r+1} & \cdots & c_{m,n} \end{bmatrix}$$

An example

$$A = \begin{bmatrix} 2 & 1 & -2 & 5 \\ 1 & 0 & -3 & 2 \\ 3 & -1 & -13 & 5 \end{bmatrix}$$

rank $A = ?$

An example

$$A = \begin{bmatrix} 2 & 1 & -2 & 5 \\ 1 & 0 & -3 & 2 \\ 3 & -1 & -13 & 5 \end{bmatrix}$$

rank $A = \ 2$

An example

$$A = \begin{bmatrix} 2 & 1 & -2 & 5 \\ 1 & 0 & -3 & 2 \\ 3 & -1 & -13 & 5 \end{bmatrix}$$

rank $A = \ 2$

$$B = \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ 3 & -1 \end{bmatrix}$$

An example

$$A = \begin{bmatrix} 2 & 1 & -2 & 5 \\ 1 & 0 & -3 & 2 \\ 3 & -1 & -13 & 5 \end{bmatrix}$$

rank $A = \ 2$

$$B = \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ 3 & -1 \end{bmatrix} \qquad ; \qquad C = \begin{bmatrix} 1 & 0 & -3 & 2 \\ 0 & 1 & 4 & 1 \end{bmatrix}$$

Generalized inverse

Definition – Given a matrix $A \in \mathbb{R}^{m\times n}$, a matrix $A^\dagger \in \mathbb{R}^{n\times n}$ is called a generalized inverse of A if

$$AA^\dagger A = A$$

and there exist matrices $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ such that

$$A^\dagger = UA^\top \qquad \text{and} \qquad A^\dagger = A^\top V.$$

Generalized inverse

Definition – Given a matrix $A \in \mathbb{R}^{m\times n}$, a matrix $A^\dagger \in \mathbb{R}^{n\times n}$ is called a generalized inverse of A if

$$AA^\dagger A = A$$

and there exist matrices $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ such that

$$A^\dagger = UA^\top \qquad \text{and} \qquad A^\dagger = A^\top V.$$

Tall – $A \in \mathbb{R}^{m\times n}, m \geq n, \mathrm{rank}\ A = n$: $\quad A^\dagger = (A^\top A)^{-1}A^\top$

Generalized inverse

Definition – Given a matrix $A \in \mathbb{R}^{m \times n}$, a matrix $A^\dagger \in \mathbb{R}^{n \times n}$ is called a generalized inverse of A if

$$AA^\dagger A = A$$

and there exist matrices $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ such that

$$A^\dagger = UA^\top \qquad \text{and} \qquad A^\dagger = A^\top V.$$

Tall – $A \in \mathbb{R}^{m \times n}, m \geq n, \text{rank } A = n$: $\quad A^\dagger = (A^\top A)^{-1} A^\top$

Fat – $A \in \mathbb{R}^{m \times n}, m \leq n, \text{rank } A = m$: $\quad A^\dagger = A^\top (AA^\top)^{-1}$

Generalized inverse

Definition – Given a matrix $A \in \mathbb{R}^{m\times n}$, a matrix $A^\dagger \in \mathbb{R}^{n\times n}$ is called a generalized inverse of A if

$$AA^\dagger A = A$$

and there exist matrices $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ such that

$$A^\dagger = UA^\top \qquad \text{and} \qquad A^\dagger = A^\top V.$$

Tall – $A \in \mathbb{R}^{m\times n}, m \geq n, \text{rank } A = n$: $\quad A^\dagger = (A^\top A)^{-1}A^\top$

Fat – $A \in \mathbb{R}^{m\times n}, m \leq n, \text{rank } A = m$: $\quad A^\dagger = A^\top (AA^\top)^{-1}$

What are U and V here?

Generalized inverse

Definition – Given a matrix $A \in \mathbb{R}^{m\times n}$, a matrix $A^\dagger \in \mathbb{R}^{n\times n}$ is called a generalized inverse of A if

$$AA^\dagger A = A$$

and there exist matrices $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ such that

$$A^\dagger = UA^\top \qquad \text{and} \qquad A^\dagger = A^\top V.$$

Tall – $A \in \mathbb{R}^{m\times n}, m \geq n, \operatorname{rank} A = n$: $\quad A^\dagger = (A^\top A)^{-1}A^\top$

Fat – $A \in \mathbb{R}^{m\times n}, m \leq n, \operatorname{rank} A = m$: $\quad A^\dagger = A^\top(AA^\top)^{-1}$

What are U and V here?

If $A^\dagger$ exists, it is *unique*. Proof?

Generalized inverse

General form

Theorem – Let $A \in \mathbb{R}^{m\times n}$, $\mathrm{rank}\ A = r \leq \min\{m, n\}$ have a full rank factorization

$$A = BC$$

with $B \in \mathbb{R}^{m\times r}, C \in \mathbb{R}^{r\times n}$, and

$$\mathrm{rank}\ A = \mathrm{rank}\ B = \mathrm{rank}\ C = r.$$

Then,

$$A^\dagger = C^\dagger B^\dagger.$$

Generalized inverse

General form

Theorem – Let $A \in \mathbb{R}^{m \times n}$, $\mathrm{rank}\ A = r \leq \min\{m, n\}$ have a full rank factorization

$$A = BC$$

with $B \in \mathbb{R}^{m \times r}, C \in \mathbb{R}^{r \times n}$, and

$$\mathrm{rank}\ A = \mathrm{rank}\ B = \mathrm{rank}\ C = r.$$

Then,

$$A^{\dagger} = C^{\dagger} B^{\dagger}.$$

Proof –

1. Verify that $AC^{\dagger}B^{\dagger}A = A$

Generalized inverse

General form

Theorem – Let $A \in \mathbb{R}^{m\times n}$, $\mathrm{rank}\ A = r \leq \min\{m,n\}$ have a full rank factorization

$$A = BC$$

with $B \in \mathbb{R}^{m\times r}, C \in \mathbb{R}^{r\times n}$, and

$$\mathrm{rank}\ A = \mathrm{rank}\ B = \mathrm{rank}\ C = r.$$

Then,

$$A^\dagger = C^\dagger B^\dagger.$$

Proof –

1. Verify that $AC^\dagger B^\dagger A = A$
2. Let

$$U = C^\top (CC^\top)^{-1}(B^\top B)^{-1}(CC^\top)^{-1}C$$
$$V = B(B^\top B)^{-1}(CC^\top)^{-1}(B^\top B)^{-1}B^\top$$

Generalized inverse

General form

Theorem – Let $A \in \mathbb{R}^{m\times n}$, $\operatorname{rank} A = r \leq \min\{m, n\}$ have a full rank factorization

$$A = BC$$

with $B \in \mathbb{R}^{m\times r}, C \in \mathbb{R}^{r\times n}$, and

$$\operatorname{rank} A = \operatorname{rank} B = \operatorname{rank} C = r.$$

Then,

$$A^\dagger = C^\dagger B^\dagger.$$

Proof –

1. Verify that $AC^\dagger B^\dagger A = A$
2. Let

$$U = C^\top (CC^\top)^{-1}(B^\top B)^{-1}(CC^\top)^{-1}C$$
$$V = B(B^\top B)^{-1}(CC^\top)^{-1}(B^\top B)^{-1}B^\top$$

3. Verify that $C^\dagger B^\dagger = UA^\top$ and $A^\dagger = A^\top V$

Generalized inverse

$A^\dagger = C^\dagger B^\dagger \quad ; \quad B^\dagger = (B^\top B)^{-1} B^\top \quad ; \quad C^\dagger = C^\top (C C^\top)^{-1}$

$$A^\dagger = C^\top (C C^\top)^{-1} (B^\top B)^{-1} B^\top$$

Generalized inverse

$A^\dagger = C^\dagger B^\dagger \quad ; \quad B^\dagger = (B^\top B)^{-1} B^\top \quad ; \quad C^\dagger = C^\top (CC^\top)^{-1}$

$$\begin{aligned} A^\dagger &= C^\top (CC^\top)^{-1} (B^\top B)^{-1} B^\top \\ &= C^\top (B^\top A C^\top)^{-1} B^\top \end{aligned}$$

Generalized inverse

$A^\dagger = C^\dagger B^\dagger \quad ; \quad B^\dagger = (B^\top B)^{-1} B^\top \quad ; \quad C^\dagger = C^\top (CC^\top)^{-1}$

$$\begin{aligned} A^\dagger &= C^\top (CC^\top)^{-1} (B^\top B)^{-1} B^\top \\ &= C^\top (B^\top A C^\top)^{-1} B^\top \end{aligned}$$

Verify using $A = BC$

Generalized inverse

$A^\dagger = C^\dagger B^\dagger \quad ; \quad B^\dagger = (B^\top B)^{-1} B^\top \quad ; \quad C^\dagger = C^\top (CC^\top)^{-1}$

$$\begin{aligned} A^\dagger &= C^\top (CC^\top)^{-1} (B^\top B)^{-1} B^\top \\ &= C^\top (B^\top A C^\top)^{-1} B^\top \end{aligned}$$

Verify using $A = BC$

Caution – $A^\dagger = C^\dagger B^\dagger$ *not valid* if $A = BC$ is *not* a full rank factorization.

See Example 12.11 in textbook.

Equivalent definition

Definition – Given a matrix $A \in \mathbb{R}^{m \times n}$, a matrix $A^\dagger \in \mathbb{R}^{n \times n}$ is called a generalized inverse of A if

$$AA^\dagger A = A$$

and there exist matrices $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ such that

$$A^\dagger = UA^\top \qquad \text{and} \qquad A^\dagger = A^\top V.$$

Equivalent definition

Definition – Given a matrix $A \in \mathbb{R}^{m\times n}$, a matrix $A^\dagger \in \mathbb{R}^{n\times n}$ is called a generalized inverse of A if

$$AA^\dagger A = A$$

and there exist matrices $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ such that

$$A^\dagger = UA^\top \qquad \text{and} \qquad A^\dagger = A^\top V.$$

equivalently, the Penrose definition:

Equivalent definition

Definition – Given a matrix $A \in \mathbb{R}^{m \times n}$, a matrix $A^\dagger \in \mathbb{R}^{n \times n}$ is called a generalized inverse of A if

$$AA^\dagger A = A$$

and there exist matrices $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ such that

$$A^\dagger = UA^\top \qquad \text{and} \qquad A^\dagger = A^\top V.$$

equivalently, the Penrose definition:

Definition – Given a matrix $A \in \mathbb{R}^{m \times n}$, a matrix $A^\dagger \in \mathbb{R}^{n \times n}$ is called a generalized inverse of A if

1. $AA^\dagger A = A$
2. $A^\dagger AA^\dagger = A^\dagger$
3. $(AA^\dagger)^\top = AA^\dagger$
4. $(A^\dagger A)^\top = A^\dagger A.$

Equivalent definition

Definition – Given a matrix $A \in \mathbb{R}^{m \times n}$, a matrix $A^\dagger \in \mathbb{R}^{n \times n}$ is called a generalized inverse of A if

$$AA^\dagger A = A$$

and there exist matrices $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ such that

$$A^\dagger = UA^\top \qquad \text{and} \qquad A^\dagger = A^\top V.$$

equivalently, the Penrose definition:

Definition – Given a matrix $A \in \mathbb{R}^{m \times n}$, a matrix $A^\dagger \in \mathbb{R}^{n \times n}$ is called a generalized inverse of A if

1. $AA^\dagger A = A$
2. $A^\dagger A A^\dagger = A^\dagger$
3. $(AA^\dagger)^\top = AA^\dagger$
4. $(A^\dagger A)^\top = A^\dagger A.$

(Verify the equivalence)

Properties of $A^\dagger$

Verify:

$$(A^\top)^\dagger = (A^\dagger)^\top$$

$$(A^\dagger)^\dagger = A$$

Properties of $A^\dagger$

Verify:

$$(A^\top)^\dagger = (A^\dagger)^\top$$
$$(A^\dagger)^\dagger = A$$

However,

$$(BC)^\dagger \neq C^\dagger B^\dagger \quad \text{in general}^*$$

(construct an example)

*Note that if it were true that

$$(BC)^\dagger = C^\dagger B^\dagger \qquad \text{for all } B, C,$$

then we need not have proved that $C^\dagger B^\dagger$ is the generalized inverse of A since it would have followed from

$$A = BC.$$

Least squares, minimum norm

Theorem – Consider a system of linear equations

$$Ax = b,\ A \in \mathbb{R}^{m \times n},\ \text{rank } A = r.$$

The vector

$$x^* = A^\dagger b$$

minimizes $\|Ax - b\|^2$ on $\mathbb{R}^n$.

Furthermore, among all vectors in $\mathbb{R}^n$ that minimize $\|Ax - b\|^2$, the vector

$$x^* = A^\dagger b$$

is the unique vector with minimal norm.

(proof on board)