| | |
|---|---|
| **Matrix Computations** | **Fall 2020** |

Lecture 8: Positive Semidefinite Matrices and Pesudo-inverse

*Lecturer: Prof. Yue Qiu & Prof. Ziping Zhao* — *Scribe: Zhihang Xu*

**Disclaimer**: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

# 1 Positive Semidefinite Matrices

## 1.1 Definitions

**Definition 1.** *A matrix* $\mathbf{A} \in \mathbb{S}^n$[1] *is said to be* ***positive semidefinite (PSD)*** *if* $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A}$ *is said to be* ***positive definite (PD)*** *if* $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$ *with* $\mathbf{x} \neq 0$, *and* $\mathbf{A}$ *is said to be* ***indefinite*** *if both* $\mathbf{A}$ *and* $-\mathbf{A}$ *are not PSD.*

We use $\mathbf{A} \succeq \mathbf{0}$ to denote $\mathbf{A}$ is PSD, $\mathbf{A} \succ \mathbf{0}$ to denote $\mathbf{A}$ is PD, and $\mathbf{A} \not\succeq \mathbf{0}$ to denote that $\mathbf{A}$ is indefinite. In this way, we can define **Matrix inequalities**: $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is PSD, $\mathbf{A} \succ \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is PD, and $\mathbf{A} \not\succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is indefinite. If $\mathbf{A}$ is PD, then it is also PSD by definition.

**Definition 2.** *For a vector* $\mathbf{x} \in \mathbb{R}^n$ *and a matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$, *the scalar function defined by*

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j A_{ij}$$

*is called a* ***quadratic form****. A quadratic form is said to be positive definite whenever* $\mathbf{A}$ *is a positive definite matrix, in that case,* $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$ *with* $\mathbf{x} \neq 0$.

**Exercise:**

1. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be PD. Show that $\mathbf{A}$ is nonsingular (i.e., invertible).
2. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be PSD (resp. PD). Show that its diagonal elements are non-negative (resp. positive).
3. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be PSD. Show that $\mathbf{B}^T\mathbf{A}\mathbf{B}$ is PSD for any $\mathbf{B} \in \mathbb{R}^{n\times m}$.
4. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be PD. Show that $\mathbf{B}^T\mathbf{A}\mathbf{B}$ is PD for any $\mathbf{B} \in \mathbb{R}^{n\times m}$ with linearly independent columns.
5. If $\mathbf{A} \succeq \mathbf{0}$, $\alpha \geq 0$, then $\alpha\mathbf{A} \succeq \mathbf{0}$.
6. If $\mathbf{A}, \mathbf{B} \succeq \mathbf{0}$, and $\alpha, \beta \geq 0 \in \mathbb{R}$, then $\alpha\mathbf{A} + \beta\mathbf{B} \succeq \mathbf{0}$.
7. If $\mathbf{A} \succeq \mathbf{B}$, $\mathbf{B} \succeq \mathbf{C}$, then $\mathbf{A} \succeq \mathbf{C}$.
8. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be PSD. Show that the eigenvalues of $\mathbf{A}$ are all non-negative.

[1] $\mathbb{S}^n$ denotes the set of $n \times n$ real symmetric matrices.

9. Orthogonal projection matrices are PSD.

**Solutions.**

1. Let $\mathbf{x} \in \mathcal{N}(\mathbf{A})$, then $\mathbf{A}\mathbf{x} = \mathbf{0}$, and therefore $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$. Hence $\mathbf{x} = \mathbf{0}$ since $\mathbf{A}$ is PD, which implies $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$. Therefore $\mathbf{A}$ is nonsingular.
2. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be PSD and let $\mathbf{x} = \mathbf{e}_i$ (the $i$-th unit vector), then
$$\mathbf{e}_i^T\mathbf{A}\mathbf{e}_i = \mathbf{A}_{ii} \geq 0\,, \quad i = 1, \ldots, n\,.$$
3. For any $\mathbf{y} \in \mathbb{R}^m$, let $\mathbf{x} = \mathbf{B}\mathbf{y} \in \mathbb{R}^n$, then
$$\mathbf{y}^T(\mathbf{B}^T\mathbf{A}\mathbf{B})\mathbf{y} = (\mathbf{B}\mathbf{y})^T\mathbf{A}(\mathbf{B}\mathbf{y}) = \mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0\,,$$
therefore $\mathbf{B}^T\mathbf{A}\mathbf{B}$ is PSD.
4. For any $\mathbf{y} \in \mathbb{R}^m$ with $\mathbf{y} \neq 0$, let $\mathbf{x} = \mathbf{B}\mathbf{y} \in \mathbb{R}^n$, since $\mathbf{B}$ has linearly independent columns, $\mathbf{y} \neq 0$ indicates that $\mathbf{x} \neq 0$, then
$$\mathbf{y}^T(\mathbf{B}^T\mathbf{A}\mathbf{B})\mathbf{y} = (\mathbf{B}\mathbf{y})^T\mathbf{A}(\mathbf{B}\mathbf{y}) = \mathbf{x}^T\mathbf{A}\mathbf{x} > 0\,,$$
therefore $\mathbf{B}^T\mathbf{A}\mathbf{B}$ is PD.
5. For every $\mathbf{x} \in \mathbb{R}^n$, we have $\mathbf{x}^T(\alpha\mathbf{A})\mathbf{x} = \alpha\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0\,$, therefore $\alpha\mathbf{A} \succeq \mathbf{0}$.
6. For every $\mathbf{x} \in \mathbb{R}^n$, we have $\mathbf{x}^T(\alpha\mathbf{A}+\beta\mathbf{B})\mathbf{x} = \alpha\mathbf{x}^T\mathbf{A}\mathbf{x}+\beta\mathbf{x}^T\mathbf{B}\mathbf{x} \geq 0\,$, therefore $\alpha\mathbf{A}+\beta\mathbf{B} \succeq \mathbf{0}$.
7. For every $\mathbf{x} \in \mathbb{R}^n$, we have $\mathbf{x}^T(\mathbf{A}-\mathbf{C})\mathbf{x} = \mathbf{x}^T(\mathbf{A}-\mathbf{B})\mathbf{x} + \mathbf{x}^T(\mathbf{B}-\mathbf{C})\mathbf{x} \geq 0\,$, therefore $\mathbf{A} \succeq \mathbf{C}$.
8. Let $(\lambda, \mathbf{x})$ be an eigenpair of $\mathbf{A}$, i.e., $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, therefore
$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\lambda\mathbf{x} = \lambda\mathbf{x}^T\mathbf{x} = \lambda\|\mathbf{x}\|_2^2 \geq 0\,,$$
since $\|\mathbf{x}\|_2^2$ are non-zero (eigenvectors are non-zero), we can get
$$\lambda = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \geq 0\,.$$
(The relation between eigenvalues and PSD matrices will be further introduced in the subsequent.)
9. For any vector $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^T\mathbf{P}\mathbf{x} = \mathbf{x}^T\mathbf{P}^2\mathbf{x} = \|\mathbf{P}\mathbf{x}\|_2^2 \geq 0\,$.

## 1.2 PSD matrices and eigenvalues

**Theorem 1.** *For matrix $\mathbf{A} \in \mathbb{S}^n$, let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{A}$, we have*

1. $\mathbf{A} \succeq \mathbf{0} \Leftrightarrow \lambda_i \geq 0$ *for* $i = 1, \ldots, n$.
2. $\mathbf{A} \succ \mathbf{0} \Leftrightarrow \lambda_i > 0$ *for* $i = 1, \ldots, n$.

*Proof.* Let $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ be the eigendecomposition of $\mathbf{A}$, then
$$\begin{aligned}\mathbf{A} \succeq \mathbf{0} &\Leftrightarrow \mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{x}\,, \quad \forall\mathbf{x} \in \mathbb{R}^n\\ &\Leftrightarrow \mathbf{z}^T\mathbf{\Lambda}\mathbf{z} \geq 0\,, \quad \forall\mathbf{z} \in \mathcal{R}(\mathbf{V}^T) = \mathbb{R}^n\\ &\Leftrightarrow \sum_{i=1}^n \lambda_i|z_i|^2 \geq 0\,, \quad \forall\mathbf{z} \in \mathbb{R}^n\\ &\lambda_i \geq 0\,, \quad i = 1, \ldots, n\end{aligned}$$
And the PD case can be proven by the same manner. $\square$

### 1.3 Symmetric Factorization

**Theorem 2** (Symmetric Factorization)**.** *A matrix* $\mathbf{A} \in \mathbb{S}^n$ *is PSD if and only if it can be factored as*

$$\mathbf{A} = \mathbf{B}^T\mathbf{B},$$

*for some* $\mathbf{B} \in \mathbb{R}^{m\times n}$ *with some integer* $m$*.*

*Proof.* 1. If $\mathbf{A}$ can be factored as $\mathbf{A} = \mathbf{B}^T\mathbf{B}$, then for any $x \in \mathbb{R}^n$,

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{B}^T\mathbf{B}\mathbf{x} = \|\mathbf{B}\mathbf{x}\|_2^2 \geq 0\,,$$

then $\mathbf{A}$ is PSD.

2. For PSD $\mathbf{A}$, $\mathbf{A}$ has eigendecomposition $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ with non-negative eigenvalues, then let $\mathbf{\Lambda}^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$,

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{V}^T = (\mathbf{V}\mathbf{\Lambda}^{1/2})\,,$$

therefore PSD $\mathbf{A}$ can be factored as $\mathbf{A} = \mathbf{B}^T\mathbf{B}$. □

Such factorization is *not* unique, since for $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ and any orthogonal $\mathbf{U} \in \mathbb{R}^{n\times n}$, $\mathbf{B} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^T$ is a factor fir $\mathbf{A} = \mathbf{B}^T\mathbf{B}$,

$$\mathbf{B}^T\mathbf{B} = (\mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{U}^T)(\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^T) = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{A}\,.$$

However, there exists one and only one lower-triangular ~~upper-triangular~~ matrix $\mathbf{G}$ with $\mathbf{G}_{ii} > 0$ such that $\mathbf{A} = \mathbf{G}\mathbf{G}^T$, which is the *Cholesky factorization* of $\mathbf{A}$ (details can be found in 1.3.1). $\mathbf{B} = \mathbf{A}^{1/2} = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{V}^T$ is the *unique PSD* factor for $\mathbf{A} = \mathbf{B}^T\mathbf{B}$.

#### 1.3.1 Cholesky decomposition revisited

In this subsection, we will focus on Cholesky decomposition (Recall Problem 4 in Homework 2).

**Lemma 3.** *If a nonsingular matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *has an LDU decomposition, then the LDU factors are uniquely determined.*

*Proof.* To prove the uniqueness, first we assume that $\mathbf{A}$ has two LDU decompositions as

$$\mathbf{A} = \mathbf{L}_1\mathbf{D}_1\mathbf{U}_1\,, \quad \mathbf{A} = \mathbf{L}_2\mathbf{D}_2\mathbf{U}_2\,,$$

and we try to prove that $\mathbf{L}_1 = \mathbf{L}_2$, $\mathbf{D}_1 = \mathbf{D}_2$, and $\mathbf{U}_1 = \mathbf{U}_2$.

*First*, the determinant of $\mathbf{A}$ satisfies $|\mathbf{A}| \neq 0$ since $\mathbf{A}$ is nonsingular. Besides, since $|\mathbf{A}| = |\mathbf{L}_1| \times |\mathbf{D}_1| \times |\mathbf{U}_1| = |\mathbf{L}_2| \times |\mathbf{D}_2| \times |\mathbf{U}_2|$, we have that $\mathbf{L}_1$, $\mathbf{D}_1$, $\mathbf{U}_1$, $\mathbf{L}_2$, $\mathbf{D}_2$, and $\mathbf{U}_2$ are all nonsingular (*i.e.*, invertible).

*Second*, multiplying $\mathbf{L}_2^{-1}$ to $\mathbf{L}_1$ gives

$$\mathbf{L}_2^{-1}\mathbf{L}_1 = \mathbf{D}_2\mathbf{U}_2(\mathbf{D}_1\mathbf{U}_1)^{-1}\,,$$

note that the left hand side $\mathbf{L}_2^{-1}\mathbf{L}_1$ is lower triangular, while the right hand side $\mathbf{D}_2\mathbf{U}_2(\mathbf{D}_1\mathbf{U}_1)^{-1}$ must be upper triangular. Hence, the left hand side and the right hand side are both diagonal matrices.

*Third*, the diagonal entries of the left hand side $\mathbf{L}_2^{-1}\mathbf{L}_1$ must be one, which implies $\mathbf{L}_2^{-1}\mathbf{L}_1 = \mathbf{I}$ and accordingly $\mathbf{L}_1 = \mathbf{L}_2$. Similarly, we can derive that $\mathbf{U}_1 = \mathbf{U}_2$ and $\mathbf{D}_1 = \mathbf{D}_2$, which concludes this proof. □

**Lemma 4.** *If* $\mathbf{A}$ *is a nonsingular symmetric matrix, then its LDU decomposition must be* $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^T$*, which is called LDL decomposition in this case.*

*Proof.* For symmetric and nonsingular $\mathbf{A}$, $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U}$, we have

$$\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U} = (\mathbf{L}\mathbf{D}\mathbf{U})^T = \mathbf{U}^T\mathbf{D}^T\mathbf{L}^T = \mathbf{U}^T\mathbf{D}\mathbf{L}^T ,$$

these are the LDU decomposition for $\mathbf{A}$ and since it is unique, we must have $\mathbf{U} = \mathbf{L}^T$. □

**Definition 3** (Cholesky Factorization)**.** *Given a PSD matrix* $\mathbf{A}$ *there exists a lower triangular matrix* $\mathbf{G}$ *such that* $\mathbf{A} = \mathbf{G}\mathbf{G}^T$*. The lower triangular matrix* $\mathbf{G}$ *is known as the Cholesky factor and* $\mathbf{A} = \mathbf{G}\mathbf{G}^T$ *is known as the Cholesky factorization of* $\mathbf{A}$*.*

**Theorem 5.** $\mathbf{A}$ *is positive definite matrix if and only if its Cholesky decomposition exists.*

*Proof.* We will prove the theorem from two directions.

1. First we try to prove that if $\mathbf{A}$ is a positive definite matrix, then its Cholesky decomposition exists. According to Exercise 1 in subsection 1.1, PD matrix $\mathbf{A}$ must be nonsingular. Then according to Lemma 4, $\mathbf{A}$ has LDL decomposition $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^T$. For any vector $\mathbf{x} \in \mathbb{R}^n$, there exists a vector $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{y} = \mathbf{L}^T\mathbf{x}$. Since $A$ is a positive definite matrix, we can derive that
$$\mathbf{y}^T\mathbf{D}\mathbf{y} = \mathbf{x}^T\mathbf{L}\mathbf{D}\mathbf{L}^T\mathbf{x} = \mathbf{x}^T\mathbf{A}\mathbf{x} > 0.$$
Hence, the diagonal entries of $\mathbf{D}$ are all positive. Let $\mathbf{G} = \mathbf{L}\mathbf{D}^{1/2}$ yields the Cholesky decomposition.

2. Second, we try to prove that if the Cholesky decomposition for $\mathbf{A}$ exists, i.e., there exists $\mathbf{G}$ such that $\mathbf{A} = \mathbf{G}\mathbf{G}^T$, then $\mathbf{A}$ is PD.

   (a) First we have $\mathbf{A}^T = (\mathbf{G}\mathbf{G}^T)^T = \mathbf{G}\mathbf{G}^T = \mathbf{A}$, therefore $\mathbf{A}$ is symmetric.

   (b) For any $\mathbf{x} \in \mathbb{R}^n$ with $\mathbf{x} \neq 0$, $\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{G}\mathbf{G}^T\mathbf{x} = \|\mathbf{G}^T\mathbf{x}\|_2 > 0$.

   Hence $\mathbf{A}$ is PD. □

**Theorem 6.** *If* $\mathbf{A}$ *is a PD matrix, then its Cholesky decomposition is uniquely determined.*

*Proof.* From Exercise 1 in subsection 1.1, first we have that PD matrix $\mathbf{A}$ must be nonsingular. To prove the uniqueness, suppose $\mathbf{A}$ has two Cholesky decompositions

$$\mathbf{A} = \mathbf{G}_1\mathbf{G}_1^T = \mathbf{G}_2\mathbf{G}_2^T .$$

where $\mathbf{G}_1$ and $\mathbf{G}_2$ are lower triangular matrices with positive diagonal entries. Then we have

$$\mathbf{G}_1\mathbf{G}_1^T = \mathbf{G}_2\mathbf{G}_2^T \Rightarrow \mathbf{G}_1 = \mathbf{G}_2\mathbf{G}_2^T(\mathbf{G}_1^T)^{-1} \Rightarrow \mathbf{G}_2^{-1}\mathbf{G}_1 = \mathbf{G}_2^T(\mathbf{G}_1^T)^{-1} ,$$

the left side is lower triangular and the right side is upper triangular, therefore both sides are diagonal matrices which we denote as $\mathbf{D}$,

$$\begin{aligned}\mathbf{G}_1\mathbf{G}_1^T = \mathbf{G}_2\mathbf{G}_2^T &\Rightarrow \mathbf{G}_1^T = \mathbf{G}_2\mathbf{G}_2^T(\mathbf{G}_1^T)^{-1} \\ &\Rightarrow \mathbf{I} = \mathbf{G}_1^{-1}\mathbf{G}_2\mathbf{G}_2^T(\mathbf{G}_1^T)^{-1} = (\mathbf{G}_1^{-1}\mathbf{G}_2)(\mathbf{G}_1^{-1}\mathbf{G}_2)^T \\ &\Rightarrow \mathbf{I} = \mathbf{D}\mathbf{D}^T ,\end{aligned}$$

also notice that both $\mathbf{G}_1$ and $\mathbf{G}_2$ have positive diagonal entries therefore $\mathbf{D}$ has positive diagonal entries and hence $\mathbf{D} = \mathbf{I}$. To sum up, $\mathbf{G}_1 = \mathbf{G}_2$ and the Cholesky decomposition of $\mathbf{A}$ is uniquely determined. □

**Exercise:**

1. Let $\mathbf{A}, \mathbf{B}$ be PSD matrices. Show that all eigenvalues of $\mathbf{AB}$ are non-negative.
2. Show that $\mathsf{rank}(\mathbf{A}) = \mathsf{rank}(\mathbf{A}^T) = \mathsf{rank}(\mathbf{A}^T\mathbf{A}) = \mathsf{rank}(\mathbf{A}\mathbf{A}^T)$ for any $\mathbf{A}$.
3. $\mathbf{A} \in \mathbb{R}^{n\times n}$ is PSD with $\mathsf{rank}(\mathbf{A}) = r$ if and only if there exists a $\mathbf{B}$ with $\mathsf{rank}(\mathbf{B}) = r$ such that $\mathbf{A} = \mathbf{B}^T\mathbf{B}$.
4. $\mathbf{A}$ is PD if and only if there exists a nonsingular $\mathbf{R}$ such that $\mathbf{A} = \mathbf{B}^T\mathbf{B}$.
5. If $\mathbf{A}$ is PD, then $\mathbf{A}^{-1}$ is PD.
6. Let $\mathbf{A} \in \mathbb{R}^{m\times k}$ and $\mathbf{B} \in \mathbb{R}^{k\times n}$, and suppose that $\mathbf{B}$ has full row rank. Then
$$\mathcal{R}(\mathbf{AB}) = \mathcal{R}(\mathbf{A})\,.$$
7. If $\mathbf{R}$ is a PSD matrix with factorization $\mathbf{R} = \mathbf{B}\mathbf{B}^T$ for some full-column rank $\mathbf{B}$, then $\mathcal{R}(\mathbf{R}) = \mathcal{R}(\mathbf{B})$.

**Solution:**

1. Let $(\lambda, \mathbf{x})$ be an eigenpair for $\mathbf{AB}$, and let $\mathbf{B} = \mathbf{C}^T\mathbf{C}$ for some $\mathbf{C}$ since $\mathbf{B}$ is PSD. Then we have
$$\begin{aligned}\lambda\mathbf{x} &= \mathbf{ABx} = \mathbf{AC}^T\mathbf{Cx} \\ \Rightarrow \lambda\mathbf{Cx} &= \mathbf{CAC}^T\mathbf{Cx}\end{aligned}$$
which means $(\lambda, \mathbf{Cx})$ is an eigenpair for $\mathbf{CAC}^T$, we have proved that $\mathbf{CAC}^T$ is PSD (See Exercise 3 in Subsection 1.1), hence the eigenvalues $\lambda \geq 0$.
2. First we have
$$\mathbf{x} \in \mathcal{N}(\mathbf{A}) \Rightarrow \mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{A}^T\mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{x} \in \mathcal{N}(\mathbf{A}^T\mathbf{A}) \Rightarrow \mathcal{N}(\mathbf{A}) \subseteq \mathcal{N}(\mathbf{A}^T\mathbf{A})\,,$$
conversely,
$$\begin{aligned}\mathbf{x} \in \mathcal{N}(\mathbf{A}^T\mathbf{A}) \Rightarrow \mathbf{A}^T\mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} = 0 \Rightarrow \mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{x} \in \mathcal{N}(\mathbf{A}) \\ \Rightarrow \mathcal{N}(\mathbf{A}^T\mathbf{A}) \subseteq \mathcal{N}(\mathbf{A})\,,\end{aligned}$$
therefore $\mathcal{N}(\mathbf{A}^T\mathbf{A}) = \mathcal{N}(\mathbf{A})$ and $\mathsf{rank}(\mathbf{A}) = \mathsf{rank}(\mathbf{A}^T\mathbf{A})$. Similarly, $\mathsf{rank}(\mathbf{A}^T) = \mathsf{rank}(\mathbf{A}\mathbf{A}^T)$ and the result follows.
3. Given $\mathbf{A}$ is PSD with $\mathsf{rank}(\mathbf{A}) = r$, the eigendecomposition for $\mathbf{A}$ is given by
$$\mathbf{A} = \mathbf{V\Lambda V}^T\,, \quad \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)\,, \lambda_i > 0\,, i = 1, \ldots, r\,,$$
then let $\mathbf{\Lambda}^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_r^{1/2}, 0, \ldots, 0)$,
$$\mathbf{A} = \mathbf{V\Lambda V}^T = \mathbf{V\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{V}^T = \mathbf{B}^T\mathbf{B}\,,$$
where $\mathsf{rank}(\mathbf{B}) = \mathsf{rank}(\mathbf{\Lambda}^{1/2}) = r$. Conversely, if $\mathbf{A} = \mathbf{B}^T\mathbf{B}$, then $\mathsf{rank}(\mathbf{A}) = \mathsf{rank}(\mathbf{B}) = r$ and
$$\mathbf{x}^T\mathbf{Ax} = \mathbf{x}^T\mathbf{B}^T\mathbf{Bx} = \|\mathbf{Bx}\|_2^2 \geq 0\,.$$

4. From Exercise 1 in Subsection 1.1, we know that PD $\mathbf{A}$ indicates that $\mathbf{A}$ is nonsingular, therefore let $\mathsf{rank}(\mathbf{A}) = r = n$ in Exercise 3, we have $\mathsf{rank}(\mathbf{B}) = r$.

5. Write $\mathbf{A} = \mathbf{B}^T\mathbf{B}$, $\mathbf{B}$ is nonsingular for PD matrix $\mathbf{A}$, then $\mathbf{A}^{-1} = (\mathbf{B}^T\mathbf{B})^{-1} = \mathbf{B}^{-1}(\mathbf{B}^T)^{-1} = \mathbf{C}^T\mathbf{C}$ for some nonsingular $\mathbf{C}$. Therefore $\mathbf{A}^{-1}$ is nonsingular.

6. We have

$$\mathcal{R}(\mathbf{AB}) = \{\mathbf{y} = \mathbf{ABx} | \mathbf{x} \in \mathbb{R}^n\} \xlongequal{\text{full row rank } \mathbf{B}} \{\mathbf{y} = \mathbf{Az} | \mathbf{z} \in \mathcal{R}(\mathbf{B}) = \mathbb{R}^k\} = \mathcal{R}(\mathbf{A}).$$

7. Full-column rank $\mathbf{B}$ means full-row rank $\mathbf{B}^T$, then $\mathcal{R}(\mathbf{R}) = \mathcal{R}(\mathbf{BB}^T) = \mathcal{R}(\mathbf{B})$.

### 1.4 Summary

**Positive Definite Matrices**

For real-symmetric matrices $\mathbf{A}$, the following statements are equivalent, and any one can serve as the definition of a positive definite matrix.

1. $\mathbf{x}^T\mathbf{Ax} > 0$ for every nonzero $\mathbf{x} \in \mathbb{R}^n$ (most commonly used as the definition).
2. All eigenvalues of $\mathbf{A}$ are positive.
3. $\mathbf{A} = \mathbf{B}^T\mathbf{B}$ for some nonsingular $\mathbf{B}$.
   (a) While $\mathbf{B}$ is not unique, there is one and only one upper-triangular matrix $\mathbf{R}$ with positive diagonals such that $\mathbf{A} = \mathbf{R}^T\mathbf{R}$. This is the Cholesky factorization of $\mathbf{A}$.

**Positive Semi-Definite Matrices**

For real-symmetric matrices such that $\mathsf{rank}(\mathbf{A}_{n\times n}) = r$, the following statements are equivalent, so any one of them can serve as the definition of a positive semidefinite matrix.

1. $\mathbf{x}^T\mathbf{Ax} \geq 0$ for every nonzero $\mathbf{x} \in \mathbb{R}^n$ (most commonly used as the definition).
2. All eigenvalues of $\mathbf{A}$ are nonneagtive.
3. $\mathbf{A} = \mathbf{B}^T\mathbf{B}$ for some $\mathbf{B}$ with $\mathsf{rank}(\mathbf{B}) = r$.

## 2 Pesudo-inverse

**Definition 4.** *For any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{A}^\dagger$ is called the Moore-Penrose inverse, or pseudo-inverse of $\mathbf{A}$ if*

*1.* $\mathbf{AA}^\dagger\mathbf{A} = \mathbf{A}$.

*2.* $\mathbf{A}^\dagger\mathbf{AA}^\dagger = \mathbf{A}^\dagger$.

*3.* $(\mathbf{A}\mathbf{A}^\dagger)^T = \mathbf{A}\mathbf{A}^\dagger$.

*4.* $(\mathbf{A}^\dagger\mathbf{A})^T = \mathbf{A}^\dagger\mathbf{A}$.

The pseudo inverse is a generalization of the matrix inverse when the matrix may not be invertible. If $\mathbf{A}$ is invertible, then the pseudo inverse is equal to the matrix inverse.

**Theorem 7.** *For any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{A}^\dagger \in \mathbb{R}^{n\times m}$ exists and is unique.*

*Proof.* First we show the existence. Suppose the *full SVD* for $\mathbf{A}$ is given by,

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = [\mathbf{U}_1, \mathbf{U}_2]\begin{bmatrix}\mathbf{\Sigma}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^T\\ \mathbf{V}_2^T\end{bmatrix} = \underbrace{\mathbf{U}_1}_{\text{orthonormal basis for } \mathcal{R}(\mathbf{A})} \quad \mathbf{\Sigma}_1 \quad \underbrace{\mathbf{V}_1^T}_{\text{orthonormal basis for } \mathcal{N}(\mathbf{A})},$$

usually $\mathbf{A} = \mathbf{U}_1\mathbf{\Sigma}\mathbf{V}_1^T$ is called *thin SVD* for $\mathbf{A}$. Let $\mathsf{rank}(\mathbf{A}) = r$, then we have $\mathbf{U}_1 \in \mathbb{R}^{m\times r}, \mathbf{V}_1 \in \mathbb{R}^{n\times r}$ and $\mathbf{\Sigma}_1 \in \mathbb{R}^{r\times r}$ with $\mathbf{U}_1^T\mathbf{U}_1 = \mathbf{I}$ and $\mathbf{V}_1^T\mathbf{V}_1 = \mathbf{I}$. Then the pseudo-inverse for $\mathbf{A}$ is given by

$$\mathbf{A}^\dagger = \mathbf{V}_1\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^T \in \mathbb{R}^{n\times m}.$$

Check if $\mathbf{A}^\dagger$ satisfies the definition by

1. $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}(\mathbf{V}_1\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^T)\mathbf{A} = (\mathbf{U}_1\mathbf{\Sigma}\mathbf{V}_1^T)(\mathbf{V}_1\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^T)(\mathbf{U}_1\mathbf{\Sigma}\mathbf{V}_1^T) = \mathbf{U}_1\mathbf{\Sigma}\mathbf{V}_1^T = \mathbf{A}$.
2. $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = (\mathbf{V}_1\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^T)(\mathbf{U}_1\mathbf{\Sigma}\mathbf{V}_1^T)(\mathbf{V}_1\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^T) = \mathbf{V}_1\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^T = \mathbf{A}^\dagger$.
3. $(\mathbf{A}\mathbf{A}^\dagger)^T = (\mathbf{U}_1\mathbf{\Sigma}\mathbf{V}_1^T\mathbf{V}_1\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^T)^T = (\mathbf{U}_1\mathbf{U}_1^T)^T = \mathbf{U}_1\mathbf{U}_1^T = \mathbf{A}\mathbf{A}^\dagger$.
4. $(\mathbf{A}^\dagger\mathbf{A})^T = (\mathbf{V}_1\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^T\mathbf{U}_1\mathbf{\Sigma}\mathbf{V}_1^T)^T = (\mathbf{V}_1\mathbf{V}_1^T)^T = \mathbf{V}_1\mathbf{V}_1^T = \mathbf{A}^\dagger\mathbf{A}$.

Next we show the uniqueness. Suppose $\mathbf{B}, \mathbf{G} \in \mathbb{R}^{n\times m}$ are pseudo-inverses of $\mathbf{A}$, i.e., they both satisfy the 4 properties in definition of pseudo-inverse. Then

$$\begin{aligned}\mathbf{AB} &\xlongequal{\text{by property 1}} \underbrace{(\mathbf{AGA})}_{\mathbf{A}}\mathbf{B} = (\mathbf{AG})(\mathbf{AB}) \xlongequal{\text{by property 3}} (\mathbf{AG})^T(\mathbf{AB})^T\\ &= \mathbf{G}^T\mathbf{A}^T\mathbf{B}^T\mathbf{A}^T = \mathbf{G}^T(\mathbf{A}^T\mathbf{B}^T\mathbf{A}^T) = \mathbf{G}^T(\mathbf{ABA})^T \xlongequal{\text{by property 1}} \mathbf{G}^T\mathbf{A}^T\\ &= (\mathbf{AG})^T \xlongequal{\text{by property 3}} \mathbf{AG},\end{aligned}$$

and similarly, we can obtain

$$\begin{aligned}\mathbf{BA} &\xlongequal{\text{by property 1}} \mathbf{B}(\mathbf{AGA}) = (\mathbf{BA})(\mathbf{GA}) \xlongequal{\text{by property 4}} (\mathbf{BA})^T(\mathbf{GA})^T\\ &= (\mathbf{A}^T\mathbf{B}^T)(\mathbf{A}^T\mathbf{G}^T) = (\mathbf{A}^T\mathbf{B}^T\mathbf{A}^T)\mathbf{G}^T = (\mathbf{ABA})^T\mathbf{G}^T \xlongequal{\text{by property 1}} \mathbf{A}^T\mathbf{G}^T\\ &= (\mathbf{GA})^T \xlongequal{\text{by property 4}} \mathbf{GA}.\end{aligned}$$

Consequently, we have

$$\mathbf{G} \xlongequal{\text{by property 2}} \mathbf{GAG} = (\mathbf{GA})\mathbf{G} = (\mathbf{BA})\mathbf{G} = \mathbf{B}(\mathbf{AG}) = \mathbf{B}(\mathbf{AB}) \xlongequal{\text{by property 2}} \mathbf{B},$$

which proves the uniqueness of the pseudo-inverse. To sum up, the pseudo-inverse of any arbitrary matrix $\mathbf{A}$ exists and is unique. □

**Exercise.**

1. Show that $\mathbf{A}\mathbf{A}^\dagger$ and $\mathbf{A}^\dagger\mathbf{A}$ are orthogonal projections.
2. Show that $\mathbf{A}\mathbf{A}^\dagger$ is the orthogonal projection onto the range space of $\mathbf{A}$, and $\mathbf{A}^\dagger\mathbf{A}$ is the orthogonal projection on the orthogonal complement of $\mathcal{N}(\mathbf{A})$.

**Solution.**

1. Since $(\mathbf{A}\mathbf{A}^\dagger)(\mathbf{A}\mathbf{A}^\dagger) = (\mathbf{A}\mathbf{A}^\dagger\mathbf{A})(\mathbf{A}^\dagger) = \mathbf{A}\mathbf{A}^\dagger$, and $(\mathbf{A}\mathbf{A}^\dagger)^T = \mathbf{A}\mathbf{A}^\dagger$, therefore $\mathbf{A}\mathbf{A}^\dagger$ are orthogonal projections. $\mathbf{A}^\dagger\mathbf{A}$ can be proved in the same manner. To be more specific, $\mathbf{A}\mathbf{A}^\dagger = \mathbf{U}_1\mathbf{U}_1^T$ and $\mathbf{A}^\dagger\mathbf{A} = \mathbf{V}_1\mathbf{V}_1^T$.
2. First we have,
$$\mathbf{y} \in \mathcal{R}(\mathbf{A}\mathbf{A}^\dagger) \Rightarrow \mathbf{y} = \mathbf{A}\mathbf{A}^\dagger\mathbf{x} \text{ for some } \mathbf{x} \Rightarrow \mathbf{y} \in \mathcal{R}(\mathbf{A}) \Rightarrow \mathcal{R}(\mathbf{A}\mathbf{A}^\dagger) \subseteq \mathcal{R}(\mathbf{A}),$$
next we have
$$\mathbf{y} \in \mathcal{R}(\mathbf{A}) \Rightarrow \mathbf{y} = \mathbf{A}\mathbf{x} \text{ for some } \mathbf{x} \Rightarrow \mathbf{A}\mathbf{A}^\dagger\mathbf{y} = \mathbf{A}\mathbf{A}^\dagger\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x} = \mathbf{y} \Rightarrow \mathbf{y} \in \mathcal{R}(\mathbf{A}\mathbf{A}^\dagger),$$
therefore the image of $\mathbf{A}\mathbf{A}^\dagger$, i.e., $\mathcal{R}(\mathbf{A}\mathbf{A}^\dagger)$ is indeed $\mathcal{R}(\mathbf{A})$. Similarly,
$$\mathbf{x} \in \mathcal{N}(\mathbf{A}) \Rightarrow \mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{A}^\dagger\mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x} \in \mathcal{N}(\mathbf{A}^\dagger\mathbf{A}) \Rightarrow \mathcal{N}(\mathbf{A}) \subseteq \mathcal{N}(\mathbf{A}^\dagger\mathbf{A}),$$
next we have
$$\mathbf{x} \in \mathcal{N}(\mathbf{A}^\dagger\mathbf{A}) \Rightarrow \mathbf{A}^\dagger\mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{A}\mathbf{A}^\dagger\mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x} \in \mathcal{N}(\mathbf{A}) \Rightarrow \mathcal{N}(\mathbf{A}^\dagger\mathbf{A}) \subseteq \mathcal{N}(\mathbf{A}).$$
Therefore
$$\mathcal{N}((\mathbf{A}^\dagger\mathbf{A})^T) = \mathcal{N}(\mathbf{A}^\dagger\mathbf{A}) = \mathcal{N}(\mathbf{A}) \Rightarrow \mathcal{R}(\mathbf{A}^\dagger\mathbf{A}) = \mathcal{N}(\mathbf{A})^\perp$$
To sum up,
$$\mathcal{R}(\mathbf{A}\mathbf{A}^\dagger) = \mathcal{R}(\mathbf{A}), \quad \mathcal{R}(\mathbf{A}^\dagger\mathbf{A}) = \mathcal{N}(\mathbf{A})^\perp.$$
Combining the conclusion of Exercise 1 completes the proof.

## 2.1 Least square revisited

For $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, let $\boldsymbol{\epsilon} = \mathbf{A}\mathbf{x} - \mathbf{b}$, the least squares problem is to find a vector $\mathbf{x}$ that minimizes
$$\sum_{i=1}^m \boldsymbol{\epsilon}_i^2 = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2.$$
Any vector that provides a minimum value for this expression is called a ***least square solution***. (Recall Problem 6 of Homework 3).

**Theorem 8.** *A vector $\mathbf{x}^\star$ is an optimal solution to the LS problem if and only if it satisfies*
$$\mathbf{A}^T\mathbf{A}\mathbf{x}^\star = \mathbf{A}^T\mathbf{b}. \tag{1}$$

Equation (1) is called the normal equation.

*Proof.* **Proof by Convex Optimization** The LS objective function is given by
$$f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{b}^T\mathbf{A}^T\mathbf{x} + \mathbf{b}^T\mathbf{b},$$
and $f(\mathbf{x})$ is convex since the Hessian matrix of $f$ is PSD (you can verify it by yourself). Taking derivative of $f(\mathbf{x})$ with respect to $\mathbf{x}$ equals to zero gives
$$\nabla f(\mathbf{x}) = 2\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{A}^T\mathbf{b} = \mathbf{0} \Rightarrow \mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}.$$
Therefore $\mathbf{x}^\star$ is an optimal LS solution if and only if it satisfies $\mathbf{A}^T\mathbf{A}\mathbf{x}^\star = \mathbf{A}^T\mathbf{b}$. $\square$

**Theorem 9.** *Least Squares Solutions* $\mathbf{x}^\star$ *has the form of* $\mathbf{x}^\star \in \mathbf{A}^\dagger\mathbf{b} + \mathcal{N}(\mathbf{A})$.

*Proof.* When $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consitent (has at least one solution), suppose $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$ for some $\mathbf{x}_0$, then we have

$$\mathbf{A}\mathbf{x}_0 = (\mathbf{A}\mathbf{A}^\dagger\mathbf{A})\mathbf{x}_0 = (\mathbf{A}\mathbf{A}^\dagger)(\mathbf{A}\mathbf{x}_0) = (\mathbf{A}\mathbf{A}^\dagger)\mathbf{b} = \mathbf{A}(\mathbf{A}^\dagger\mathbf{b}) = \mathbf{b}\,,$$

therefore, $\mathbf{A}^\dagger\mathbf{b}$ solves $\mathbf{A}\mathbf{x} = \mathbf{b}$ and the general solution of the system is $\mathbf{A}^\dagger\mathbf{b} + \mathbf{z}$, $\mathbf{z} \in \mathcal{N}(\mathbf{A})$.

When $\mathbf{A}\mathbf{x} = \mathbf{b}$ is inconsitent, the least squares solutions are the soluution of the normal equation $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$, we can verify that

$$\begin{aligned}\mathbf{A}^T\mathbf{A}(\mathbf{A}^\dagger\mathbf{b}) = \mathbf{A}^T(\mathbf{A}\mathbf{A}^\dagger)\mathbf{b} &\overset{\text{by property 3}}{=\!=\!=\!=} \mathbf{A}^T(\mathbf{A}\mathbf{A}^\dagger)^T\mathbf{b} = \mathbf{A}^T(\mathbf{A}^\dagger)^T\mathbf{A}^T\mathbf{b} \\ &= (\mathbf{A}\mathbf{A}^\dagger\mathbf{A})^T\mathbf{b} = \mathbf{A}^T\mathbf{b}\,,\end{aligned}$$

therefore $\mathbf{A}^\dagger\mathbf{b}$ satisfies the normal equation. To sum up, the least squares solutions can be given by $\mathbf{x}^\star \in \mathbf{A}^\dagger\mathbf{b} + \mathcal{N}(\mathbf{A})$. □

**Theorem 10.** *For* $\mathbf{A} \in \mathbb{R}^{m\times n}(m > n)$, $\mathbf{A}^\dagger\mathbf{b}$ *is the solution to* $\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$ *of minimum 2-norm.*

*Proof.* According to Theorem 9, any solution to the LS problem can be written as

$$\mathbf{x} = \mathbf{A}^\dagger\mathbf{b} + \tilde{\mathbf{z}}\,, \quad \tilde{\mathbf{z}} \in \mathcal{N}(\mathbf{A})\,.$$

For any vector $\mathbf{z} \in \mathbb{R}^n$, the orthogonal projection onto $\mathcal{R}(\mathbf{A})$ is given by

$$\Pi_{\mathcal{N}(\mathbf{A})} = \mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\,,$$

and therefore we can rewrite $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{A}^\dagger\mathbf{b} + (\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\mathbf{z}\,, \quad \mathbf{z} \in \mathbb{R}^n\,.$$

Note that

$$[(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\mathbf{z}]^T(\mathbf{A}^\dagger\mathbf{b}) = \mathbf{z}^T[\mathbf{I} - (\mathbf{A}^\dagger\mathbf{A})^T](\mathbf{A}^\dagger\mathbf{b}) = \mathbf{z}^T\mathbf{A}^\dagger\mathbf{b} - \mathbf{z}^T\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger\mathbf{b} = 0\,,$$

which means $\mathbf{A}^\dagger\mathbf{b} \perp \tilde{\mathbf{z}}$, therefore we have

$$\|\mathbf{x}\|_2^2 = \|\mathbf{A}^\dagger\mathbf{b} + (\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\mathbf{z}\|_2^2 = \|\mathbf{A}^\dagger\mathbf{b}\|_2^2 + \|(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\mathbf{z}\|_2^2 \ge \|\mathbf{A}^\dagger\mathbf{b}\|_2^2\,.$$

Equality holds if and only if $\mathbf{z} = \mathbf{0}$. So $\mathbf{A}^\dagger\mathbf{b}$ is the unique minimum norm solution. □

### 2.2 Summary

**Pseudo-inverse**

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ admits a SVD $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, if $\mathsf{rank}(\mathbf{A}) = r$, then the thin SVD for $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = [\mathbf{U}_1, \mathbf{U}_2] \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} = \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{V}_1^T$$

1. Pseudo-inverse: $\mathbf{A}^\dagger = \mathbf{V}_1\boldsymbol{\Sigma}_1^{-1}\mathbf{U}_1^T$
2. LS solution: $\mathbf{x}_{\text{LS}} = \mathbf{A}^\dagger\mathbf{y} + \boldsymbol{\eta}$ for any $\boldsymbol{\eta} \in \mathcal{N}(\mathbf{A})\cancel{\mathcal{R}(\mathbf{A})}$.
3. Orthogonal projection:
   (a) $\mathbf{A}\mathbf{A}^\dagger = \mathbf{U}_1\mathbf{U}_1^T$ is the ortgonal projection onto $\mathcal{R}(\mathbf{A})$.
   (b) $\mathbf{A}^\dagger\mathbf{A} = \mathbf{V}_1\mathbf{V}_1^T$ is the ortgonal projection onto $\mathcal{N}(\mathbf{A})^\perp$.

**Least Square Solutions**

Each of the following four statements is equivalent to saying that $\mathbf{x}^\star$ is a least square solution for a possibly inconsistent linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$.

- $\|\mathbf{A}\mathbf{x}^\star + \mathbf{b}\|_2 = \min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$.
- $\mathbf{A}\mathbf{x}^\star = \Pi_{\mathcal{R}(\mathbf{A})}\mathbf{b}$ .
- $\mathbf{A}^T\mathbf{A}\mathbf{x}^\star = \mathbf{A}^T\mathbf{b}$.
- $\mathbf{x}^\star = \mathbf{A}^\dagger\mathbf{b} + \mathcal{N}(\mathbf{A})$. ($\mathbf{A}^\dagger\mathbf{b}$ is the minimal 2-norm LSS).

# Notes and Reference

1. Gene H G, Charles F. *Matrix computations* [J]. Johns Hopkins Universtiy Press, 3rd edtion, 1996.
2. Meyer C D. *Matrix analysis and applied linear algebra* [M]. Siam, 2000.