

EE6617: Rudiments of Linear Algebra

- Review some basic concepts from linear algebra.
- These concepts will heavily be used throughout the course, so they are very important.
- All the matrix and vectors are assumed to be real.

1 Vector and Matrix Spaces

- **Linear vector space:** Denote $\mathbb{R}^{m \times 1}$ as the set of all real m -dimensional column vectors $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T \in \mathbb{R}^{m \times 1}$. The transpose of $\mathbf{x}$ is $\mathbf{x}^T = [x_1 \ x_2 \ \dots \ x_m] \in \mathbb{R}^{1 \times m}$.
 - $\mathbb{R}^{m \times 1}$ is a **linear vector space** because, for any real scalar $a \in \mathbb{R}$ and any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m \times 1}$, it is **close** under scalar multiplication and vector addition

$$a\mathbf{x} \in \mathbb{R}^{m \times 1} \text{ and } \mathbf{x} + \mathbf{y} \in \mathbb{R}^{m \times 1}.$$

- **Inner product between two vectors:**

$$\mathbf{x}^T \mathbf{y} = [x_1 \ x_2 \ \dots \ x_m] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \sum_{i=1}^m x_i y_i.$$

- **Euclidean space:** The linear space $\mathbb{R}^{m \times 1}$ in which the Euclidean norm (vector 2-norm) of $\mathbf{x}$

$$\|\mathbf{x}\| = (\mathbf{x}^T \mathbf{x})^{1/2} = \left(\sum_{i=1}^m x_i^2 \right)^{1/2}$$

is defined, is called m -dimensional Euclidean space.

- **Euclidean distance:** Once with the norm, one can define a distance

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \left(\sum_{i=1}^m (x_i - y_i)^2 \right)^{1/2}.$$

- **Orthogonal and orthonormal:** If $\mathbf{x}^T \mathbf{y} = 0$, we say that $\mathbf{x}$ and $\mathbf{y}$ are **orthogonal**. If $\mathbf{x}^T \mathbf{y} = 0$ and $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$, we say that $\mathbf{x}$ and $\mathbf{y}$ are **orthonormal**.

- **Linear matrix space:** Denote $\mathbb{R}^{m \times n}$ as the set of all real $m \times n$ -dimensional matrices

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n] \in \mathbb{R}^{m \times n}. \text{ The transpose of } \mathbf{X} \text{ is } \mathbf{X}^T = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} \in \mathbb{R}^{n \times m}. \text{ It is also easy}$$

to see that $\mathbb{R}^{m \times n}$ is a **linear matrix space** (why?).

2 Linear Subspace

- **Linear combination** of vectors $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ using a vector $\mathbf{a} = [a_1 \ \dots \ a_n]^T \in \mathbb{R}^n$ is

$$\mathbf{x} = \sum_{i=1}^n a_i \mathbf{x}_i = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + \dots + a_n \mathbf{x}_n = \mathbf{X} \mathbf{a}.$$

We can interpret the linear combination as an inner product between $\mathbf{X}$ and $\mathbf{a}$.

- **Linearly dependent:** The vectors $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ are **linearly dependent** if

$$\text{there exists } \mathbf{a} \neq \mathbf{0} \text{ such that } \mathbf{X} \mathbf{a} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + \dots + a_n \mathbf{x}_n = \mathbf{0}.$$

In this case, one vector can be expressed as a linear combination of the other vectors, e.g.,

$$\mathbf{x}_1 = - \sum_{i=2}^n \frac{a_i}{a_1} \mathbf{x}_i, \quad (1)$$

implying the linear dependency.

- **Linearly independent:** The vectors $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ are **linearly independent** if

$$\mathbf{X} \mathbf{a} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + \dots + a_n \mathbf{x}_n = \mathbf{0}, \text{ only if } \mathbf{a} = \mathbf{0}.$$

- **Span:** The **span** of a set of vectors $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$, $\mathbf{x}_i \in \mathbb{R}^m$ (these vectors are not necessarily independent) is a set

$$\mathcal{S}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) = \left\{ \mathbf{x} \mid \mathbf{x} = \sum_{i=1}^p a_i \mathbf{x}_i, \ a_i \in \mathbb{R} \right\}.$$

The span $\mathcal{S}$ is a **subspace** of $\mathbb{R}^m$.

- **Basis:** If $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ of $\mathcal{S}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ are all linearly independent, then the set $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ is called a **basis** of the subspace $\mathcal{S}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$.
 - How about if the set $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ includes some linearly dependent vectors?
 - A **basis** of $\mathcal{S}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ is then formed by removing those dependent vectors and re-labeling the linearly independent vectors such that $(\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(r)})$, $r < p$.
 - When $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ are **orthogonal**, $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ is called an **orthogonal basis** of $\mathcal{S}$.
- **Rank:** The **rank** of a matrix $\mathbf{X}$ is the number of independent columns of $\mathbf{X}$ and we write $\text{rank}(\mathbf{X}) \geq 0$.
- **Dimension:** The dimension of the subspace $\mathcal{S}(\mathbf{x}_1, \dots, \mathbf{x}_p)$, $\dim(\mathcal{S}(\mathbf{x}_1, \dots, \mathbf{x}_p))$, is the number of linearly independent vectors in the set $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$, i.e., the number of vectors in the basis of $\mathcal{S}(\mathbf{x}_1, \dots, \mathbf{x}_p)$. Note that with the construction $\mathbf{X} = [\mathbf{x}_1 \ \dots \ \mathbf{x}_p]$, we have the identity

$$\text{rank}(\mathbf{X}) = \dim(\mathcal{S}(\mathbf{x}_1, \dots, \mathbf{x}_p)). \quad (2)$$

3 Orthogonal Subspace Decomposition

- **Range space:** The **range** of $\mathbf{X} \in \mathbb{R}^{m \times n}$, denoted as $\mathcal{R}(\mathbf{X})$, is the set of all vectors in $\mathbb{R}^m$ that can be written as linear combinations of the columns of $\mathbf{X}$, i.e.,

$$\mathcal{R}(\mathbf{X}) = \{\mathbf{X}\mathbf{a} | \mathbf{a} \in \mathbb{R}^n\}$$

- The range $\mathcal{R}(\mathbf{X})$ is a subspace of $\mathbb{R}^m$ (why?).
- The dimension of $\mathcal{R}(\mathbf{X})$ is $\text{rank}(\mathbf{X}) \leq \min(m, n)$. Formally speaking, $\dim(\mathcal{R}(\mathbf{X})) = \text{rank}(\mathbf{X})$.

- **Null space:** The **nullspace** of $\mathbf{X} \in \mathbb{R}^{m \times n}$, denoted as $\mathcal{N}(\mathbf{X})$, is the set of all vectors in $\mathbb{R}^n$ mapped into zero by $\mathbf{X}$, i.e.,

$$\mathcal{N}(\mathbf{X}) = \{\mathbf{a} | \mathbf{X}\mathbf{a} = \mathbf{0}\}.$$

The nullspace is a subspace of $\mathbb{R}^n$.

- **Orthogonal complement:** If $\mathcal{V}$ is a subspace of $\mathbb{R}^n$, its orthogonal complement, denoted $\mathcal{V}^\perp$ is defined as

$$\mathcal{V}^\perp = \{\mathbf{x} | \mathbf{z}^T \mathbf{x} = 0 \text{ for all } \mathbf{z} \in \mathcal{V}\}$$

- **Orthogonal subspace decomposition:** For any $\mathbf{X} \in \mathbb{R}^{m \times n}$, we have

$$\mathcal{N}(\mathbf{X}) = \mathcal{R}(\mathbf{X}^T)^\perp \text{ or } \mathcal{N}(\mathbf{X}) \perp \mathcal{R}(\mathbf{X}^T).$$

Any vectors from $\mathbf{a} \in \mathcal{N}(\mathbf{X})$ and $\mathbf{b} \in \mathcal{R}(\mathbf{X}^T)$ are orthogonal, i.e., $\mathbf{a}^T \mathbf{b} = 0$.

Similarly, we have

$$\mathcal{R}(\mathbf{X}) = \mathcal{N}(\mathbf{X}^T)^\perp \text{ or } \mathcal{R}(\mathbf{X}) \perp \mathcal{N}(\mathbf{X}^T).$$

This result is often stated as $\mathcal{N}(\mathbf{X}) \cup \mathcal{R}(\mathbf{X}^T) = \mathbb{R}^n$ and $\mathcal{R}(\mathbf{X}) \cup \mathcal{N}(\mathbf{X}^T) = \mathbb{R}^m$.

- Suppose $\dim(\mathcal{R}(\mathbf{X})) = r$. Then, it follows $\dim(\mathcal{N}(\mathbf{X}^T)) = m - r$. Hence,

$$\dim(\mathcal{R}(\mathbf{X})) + \dim(\mathcal{N}(\mathbf{X}^T)) = m.$$

- Similarly, $\dim(\mathcal{R}(\mathbf{X}^T)) = r$ and then $\dim(\mathcal{N}(\mathbf{X})) = n - r$. Hence,

$$\dim(\mathcal{R}(\mathbf{X}^T)) + \dim(\mathcal{N}(\mathbf{X})) = n.$$

4 Square Matrix

A **square** matrix is $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m = n$.

- The **trace** of a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the sum of its diagonal elements,

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii} \quad (3)$$

where $a_{ij} = [\mathbf{A}]_{ij}$ denotes the i th row and j th column element of $\mathbf{A}$.

- The **determinant** of a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is denoted as $\det(\mathbf{A})$ and is defined by

$$\det(\mathbf{A}) = \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) \prod_{i=1}^n a_{i, \sigma_i}. \quad (4)$$

Here the sum is computed over all permutations σ of the set $\{1, 2, \dots, n\}$. For instance, if $n = 3$, all permutations are $\mathcal{S}_3 = \{\{1, 2, 3\}, \{2, 1, 3\}, \{3, 2, 1\}, \{1, 3, 2\}, \{2, 3, 1\}, \{3, 1, 2\}\}$. For each $\sigma \in \mathcal{S}_n$, σ_i denotes the i th element of σ . For instance, if $\sigma = \{2, 1, 3\}$, then $\sigma_2 = 1$. The $\text{sgn}(\sigma)$ is computed by $(-1)^\tau$ where τ denote the number of permutations applied to form σ from $\{1, 2, \dots, n\}$. For instance, $\text{sgn}(\{2, 1, 3\}) = (-1)^1 = -1$ and $\text{sgn}(\{2, 3, 1\}) = (-1)^2 = 1$.

For $n = 1$, we have $\mathbf{A} = [a_{11}]$, $\mathcal{S}_1 = \{\{1\}\}$, and

$$\det(\mathbf{A}) = \sum_{\sigma \in \mathcal{S}_1} \text{sgn}(\sigma) \prod_{i=1}^1 a_{i, \sigma_i} = a_{11}$$

For $n = 2$, we have $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $\mathcal{S}_2 = \{\{1, 2\}, \{2, 1\}\}$, and

$$\begin{aligned} \det(\mathbf{A}) &= \sum_{\sigma \in \mathcal{S}_2} \text{sgn}(\sigma) \prod_{i=1}^2 a_{i, \sigma_i} \\ &= \text{sgn}(\{1, 2\})a_{11}a_{22} + \text{sgn}(\{2, 1\})a_{12}a_{21} \\ &= a_{11}a_{22} - a_{12}a_{21} \end{aligned}$$

For $n = 3$, we have $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$,

$\mathcal{S}_3 = \{\{1, 2, 3\}, \{2, 1, 3\}, \{3, 2, 1\}, \{1, 3, 2\}, \{2, 3, 1\}, \{3, 1, 2\}\}$, and

$$\begin{aligned} \det(\mathbf{A}) &= \sum_{\sigma \in \mathcal{S}_3} \text{sgn}(\sigma) \prod_{i=1}^3 a_{i, \sigma_i} \\ &= \text{sgn}(\{1, 2, 3\})a_{11}a_{22}a_{33} + \text{sgn}(\{2, 1, 3\})a_{12}a_{21}a_{33} + \text{sgn}(\{3, 2, 1\})a_{13}a_{22}a_{31} \\ &\quad \text{sgn}(\{1, 3, 2\})a_{11}a_{23}a_{32} + \text{sgn}(\{2, 3, 1\})a_{12}a_{23}a_{31} + \text{sgn}(\{3, 1, 2\})a_{13}a_{21}a_{32} \\ &= a_{11}a_{22}a_{33} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \end{aligned}$$

The computation for $n > 3$ is possible, but we will not treat this case during the course. For $n > 3$, just try Matlab.

Note: The determinant is only defined for a square matrix (why?).

- The **inverse** of a square $\mathbf{A} \in \mathbb{R}^{n \times n}$ is $\mathbf{A}^{-1} \in \mathbb{R}^{n \times n}$ such that

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n \quad (5)$$

where $\mathbf{I}_n = \text{diag}(1, 1, \dots, 1) \in \mathbb{R}^{n \times n}$, identity matrix. For instance, $\mathbf{I}_3 = \text{diag}(1, 1, 1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Notice that the definition of the inverse matrix in (5) is only valid for a square matrix.

The inverse of $\mathbf{A} \in \mathbb{R}^{n \times n}$ can be computed via

$$\mathbf{A}^{-1} = \frac{\mathbf{C}^T}{\det(\mathbf{A})} \quad (6)$$

where $\mathbf{C} \in \mathbb{R}^{n \times n}$ is the cofactor of $\mathbf{A}$, which is defined by

$$[\mathbf{C}]_{ij} = (-1)^{i+j} M_{ij} \quad (7)$$

where M_{ij} is the determinant of the submatrix of $\mathbf{A}$ formed by deleting the i th row and j th column of $\mathbf{A}$. M_{ij} is termed as the minor of a_{ij} .

For instance of $n = 2$, $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $\mathbf{A}^{-1}$ is given by

$$\begin{aligned} \mathbf{A}^{-1} &= \frac{1}{\det(\mathbf{A})} \begin{bmatrix} M_{11} & -M_{21} \\ -M_{12} & M_{22} \end{bmatrix} \\ &= \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \end{aligned}$$

Computation of $\mathbf{A}^{-1}$ for $n > 2$ is possible, but we will not treat this case in the course.

- From (6), the square matrix $\mathbf{A}$ is **invertible** if and only if $\det(\mathbf{A}) \neq 0$.

- Some useful formulae: For $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$

$$\begin{aligned}
(\mathbf{AB})^T &= \mathbf{B}^T \mathbf{A}^T, \mathbf{A} \in \mathbb{R}^{m \times n} \text{ and } \mathbf{B} \in \mathbb{R}^{n \times p} \\
\left((\mathbf{AB})^T\right)^T &= \mathbf{AB}, \mathbf{A} \in \mathbb{R}^{m \times n} \text{ and } \mathbf{B} \in \mathbb{R}^{n \times p} \\
(\mathbf{A}^T)^{-1} &= (\mathbf{A}^{-1})^T, \text{ square matrix } \mathbf{A} \in \mathbb{R}^{n \times n} \\
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1} \mathbf{A}^{-1}, \text{ square matrix } \mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n} \\
\det(\mathbf{A}^T) &= \det(\mathbf{A}), \text{ square matrix } \mathbf{A} \in \mathbb{R}^{n \times n} \\
\det(c\mathbf{A}) &= c^n \det(\mathbf{A}), \text{ square matrix } \mathbf{A} \in \mathbb{R}^{n \times n} \\
\det(\mathbf{AB}) &= \det(\mathbf{A}) \det(\mathbf{B}), \text{ square matrix } \mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n} \\
\det(\mathbf{I}) &= 1 \\
\det(\mathbf{A}^{-1}) &= \frac{1}{\det(\mathbf{A})}, \text{ square matrix } \mathbf{A} \in \mathbb{R}^{n \times n} \\
\text{tr}(\mathbf{AB}) &= \text{tr}(\mathbf{BA}), \text{ square matrix } \mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n} \\
\text{tr}(\mathbf{A}^T \mathbf{B}) &= \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ij}, \mathbf{A} \in \mathbb{R}^{m \times n} \text{ and } \mathbf{B} \in \mathbb{R}^{m \times n} \\
\text{tr}(\mathbf{A}^T \mathbf{A}) &= \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 = \|\mathbf{A}\|_F^2, \mathbf{A} \in \mathbb{R}^{m \times n}
\end{aligned}$$

where $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$ is called the Frobenius norm.

- An **eigenvector** of a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a vector $\mathbf{v} \in \mathbb{R}^{n \times 1}$ satisfying an equality

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \quad (8)$$

where a scalar λ is the **eigenvalue** of $\mathbf{A}$. The $\mathbf{v}$ is referred to as the eigenvector associated with the eigenvalue of λ . It is assumed that the eigenvector is normalized to have unit norm $\|\mathbf{v}\|^2 = \mathbf{v}^T \mathbf{v} = 1$. The relation (8) can be extended to

$$\mathbf{A}[\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n] = [\lambda_1 \mathbf{v}_1 \ \lambda_2 \mathbf{v}_2 \ \dots \ \lambda_n \mathbf{v}_n]$$

or this can be rewritten by

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}$$

where $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n] \in \mathbb{R}^{n \times n}$ and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^{n \times n}$. Unless otherwise stated, we assume eigenvalues are ordered such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. We often write $\lambda_{\max}(\mathbf{A}) = \lambda_1(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A}) = \lambda_n(\mathbf{A})$. This also can be rewritten as

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} \quad (9)$$

and we call the **eigenvalue decomposition (EVD)** of $\mathbf{A}$.

- Numerous useful interpretations can be made by the EVD.

$$\begin{aligned}
\det(\mathbf{A}) &= \det(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}) \\
&= \det(\mathbf{V}) \det(\mathbf{\Lambda}) \det(\mathbf{V}^{-1}) \\
&= \det(\mathbf{\Lambda}) = \prod_{i=1}^n \lambda_i \\
\mathbf{A}^{-1} &= \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^{-1} \\
\det(\mathbf{A}^{-1}) &= \prod_{i=1}^n \lambda_i^{-1} \\
\text{tr}(\mathbf{A}) &= \text{tr}(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}) \\
&= \text{tr}(\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{V}) \\
&= \text{tr}(\mathbf{\Lambda}) = \sum_{i=1}^n \lambda_i \\
\text{tr}(\mathbf{A}^{-1}) &= \sum_{i=1}^n \lambda_i^{-1}
\end{aligned} \tag{10}$$

From (6), the square matrix $\mathbf{A}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$. This means, from (10), the square matrix $\mathbf{A}$ must have non-zero eigenvalues and $\det(\mathbf{V}) \neq 0$.

- **Having non-zero eigenvalues** has very special meaning in the perspective of **linear combination**. To see this suppose the definition of the eigenvalue and eigenvector in (8), i.e., $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Having a zero eigenvalue means that there exists a vector $\mathbf{v} \neq \mathbf{0}$ such that $\mathbf{A}\mathbf{v} = \mathbf{0}$, meaning the columns of $\mathbf{A}$ are linearly dependent and $\text{rank}(\mathbf{A}) < n$. Thus, we can state the following:

A square matrix $\mathbf{A}$ is **invertible** if and only if

- (a) $\det(\mathbf{A}) \neq 0$
- (b) $\lambda_i \neq 0, i = 1, \dots, n$
- (b) columns of $\mathbf{A}$ are linearly independent
- (c) $\text{rank}(\mathbf{A}) = n$.

- When $\text{rank}(\mathbf{A}) = n$ or $\det(\mathbf{A}) \neq 0$, we say $\mathbf{A}$ has a **full rank**. Otherwise, $\mathbf{A}$ is said to be **singular**.
- A square matrix $\mathbf{A}$ is **positive definite** if $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$. If $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$ for $\mathbf{x} \neq \mathbf{0}$, $\mathbf{A}$ is said to be **positive semidefinite**.

5 Symmetric Matrix

- A square matrix is said to be **symmetric** if $\mathbf{A} = \mathbf{A}^T$. Denote $\mathbb{S}^n$ the set of symmetric $n \times n$ matrices. Then, $\mathbf{A} \in \mathbb{S}^n$.
- Suppose $\mathbf{A} \in \mathbb{S}^n$. Then, EVD of $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \quad (11)$$

where $\mathbf{V} \in \mathbb{R}^{n \times n}$ is orthonormal, i.e., satisfies $\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_n$. From (9) and (11), $\mathbf{V}^{-1} = \mathbf{V}^T$ for the symmetric $\mathbf{A}$ and

$$\begin{aligned} \mathbf{A} &= \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^T \\ \mathbf{A}^{-1} &= \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^T = \sum_{i=1}^n \lambda_i^{-1} \mathbf{v}_i \mathbf{v}_i^T. \end{aligned}$$

- **Rayleigh quotient theorem:** Suppose $\mathbf{x} \in \mathbb{R}^{n \times 1}$ and $\mathbf{A} \in \mathbb{S}^n$. The ratio between the norm of $\mathbf{x}$ and the quadratic form $\mathbf{x}^T \mathbf{A} \mathbf{x}$ is called **Rayleigh quotient**,

$$\frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|_2} = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}.$$

The maximum of Rayleigh quotient is the dominant eigenvalue λ_1 of $\mathbf{A}$, i.e.,

$$\lambda_1 = \max_{\mathbf{x}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \quad (12)$$

and this is achieved by the dominant eigenvector $\mathbf{x} = \mathbf{v}_1$ associated with λ_1 .

6 Discussions

- **Invertibility of a rectangular matrix:** Is a rectangular matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\text{rank}(\mathbf{A}) = \min(m, n)$ invertible?
 - Answer:
 - However, we can still define non-genuine (or pseudo) inverse of $\mathbf{A}$.
- **Pseudo inverse**
 - If $m \geq n$ of $\mathbf{A} \in \mathbb{R}^{m \times n}$, one can design the left inverse (or pseudo inverse) matrix $\mathbf{A}^\dagger \in \mathbb{R}^{n \times m}$ of $\mathbf{A} \in \mathbb{R}^{m \times n}$ such that $\mathbf{A}^\dagger \mathbf{A} = \mathbf{I}_n$. This can be done by looking at the equality $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A} = \mathbf{A}^\dagger \mathbf{A} = \mathbf{I}_n$. Thus, we choose $\mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ so that $\mathbf{A}^\dagger \mathbf{A} = \mathbf{I}_n$.
 - Similarly, if $m \leq n$, there exist a right pseudo inverse matrix $\mathbf{A}^\dagger \in \mathbb{R}^{n \times m} = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1}$ such that $\mathbf{A} \mathbf{A}^\dagger = \mathbf{I}_m$.

We shall call $\mathbf{A}^\dagger$ the **pseudo inverse of $\mathbf{A}$** . The pseudo inverse exists if and only if $\mathbf{A} \mathbf{A}^T$ or $\mathbf{A}^T \mathbf{A}$ are full rank.

- **EVD of a singular square matrix:** Can we still define an EVD of a singular square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$?
 - Answer:
 - Suppose that $\mathbf{A}$ has $\text{rank}(\mathbf{A}) = r \leq n$. Then, we can always find r eigenvalues $\lambda_1, \dots, \lambda_r$ and the corresponding eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_r$. Let $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_r) \in \mathbb{R}^{r \times r}$ and $\mathbf{V} = [\mathbf{v}_1 \ \dots \ \mathbf{v}_r] \in \mathbb{R}^{n \times r}$. Then, one can always write

$$\mathbf{A} = [\mathbf{V} \ \tilde{\mathbf{V}}] \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} [\mathbf{V} \ \tilde{\mathbf{V}}]^{-1}$$

where $\tilde{\mathbf{V}} \in \mathbb{R}^{n \times (n-r)}$ is any matrix such that $\det([\mathbf{V} \ \tilde{\mathbf{V}}]) \neq 0$.

- If $\mathbf{A} \in \mathbb{S}^n$ (symmetric matrix), we have the EVD of the singular matrix $\mathbf{A}$

$$\mathbf{A} = [\mathbf{V} \ \tilde{\mathbf{V}}] \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} [\mathbf{V} \ \tilde{\mathbf{V}}]^T$$

where in this case, $[\mathbf{V} \ \tilde{\mathbf{V}}]$ is orthonormal matrix, i.e., $[\mathbf{V} \ \tilde{\mathbf{V}}]^T [\mathbf{V} \ \tilde{\mathbf{V}}] = [\mathbf{V} \ \tilde{\mathbf{V}}] [\mathbf{V} \ \tilde{\mathbf{V}}]^T = \mathbf{I}_n$.

- **Existence of EVD of a rectangular matrix:** Can we still define an eigenvalue decomposition of a rectangular matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\text{rank}(\mathbf{A}) = r \leq \min(m, n)$?
 - Answer:
 - There are many other useful decompositions which can substitute EVD. One of the most useful and that we will most frequently use is the **singular value decomposition (SVD)**.

7 Singular Value Decomposition (SVD)

- A matrix $\mathbf{A}$ with $\text{rank}(\mathbf{A}) = r \leq \min(m, n)$ can be factorized as

$$\mathbf{A} = \mathbf{V}\mathbf{\Sigma}\mathbf{W}^T, \quad (13)$$

where $\mathbf{V} \in \mathbb{R}^{m \times r}$ satisfies $\mathbf{V}^T \mathbf{V} = \mathbf{I}_r$, $\mathbf{W} \in \mathbb{R}^{n \times r}$ satisfies $\mathbf{W}^T \mathbf{W} = \mathbf{I}_r$, and $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_r)$, with

$$\sigma_1 \geq \sigma_2 \geq \dots \sigma_r > 0.$$

The factorization in (13) is called the **singular value decomposition (SVD)** of $\mathbf{A}$. The $\mathbf{V}$ is called left singular matrix of $\mathbf{A}$, the $\mathbf{W}$ is right singular matrix of $\mathbf{A}$, and the numbers σ_i are the singular values of $\mathbf{A}$.

The singular value decomposition can be rewritten as

$$\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{v}_i \mathbf{w}_i^T,$$

where $\mathbf{v}_i$ is the i th column of $\mathbf{V}$ and $\mathbf{w}_i$ is the i th column of $\mathbf{W}$.

- The number of non-zero singular values $\sigma_1 \geq \sigma_2 \geq \dots \sigma_r > 0$ is the rank of $\mathbf{A}$.
- The $\mathbf{V}$ is the orthonormal basis of the left (or column) subspace of $\mathbf{A}$.
- The $\mathbf{W}$ is the orthonormal basis of the right (or row) subspace of $\mathbf{A}$.
- *Proof of the existence of SVD:* Suppose the eigenvalue decomposition of the symmetric matrix $\mathbf{A}\mathbf{A}^T = \begin{bmatrix} \mathbf{V} & \tilde{\mathbf{V}} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V} & \tilde{\mathbf{V}} \end{bmatrix}^T$ with $\mathbf{V} \in \mathbb{R}^{m \times r}$, $\tilde{\mathbf{V}} \in \mathbb{R}^{m \times (m-r)}$, and $\mathbf{\Lambda} \in \mathbb{R}^{r \times r}$. Then, it is clear that

$$\begin{aligned} \mathbf{V}^T \mathbf{A} \mathbf{A}^T \mathbf{V} &= \mathbf{\Lambda} \\ \tilde{\mathbf{V}}^T \mathbf{A} \mathbf{A}^T \tilde{\mathbf{V}} &= \tilde{\mathbf{V}}^T \mathbf{A} \left(\tilde{\mathbf{V}}^T \mathbf{A} \right)^T = \mathbf{0} \Leftrightarrow \tilde{\mathbf{V}}^T \mathbf{A} = \mathbf{0}. \end{aligned} \quad (14)$$

Now, suppose an unitary matrix $\begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix} \in \mathbb{R}^{n \times n}$ with $\mathbf{W} \in \mathbb{R}^{n \times r}$ and $\tilde{\mathbf{W}} \in \mathbb{R}^{n \times (n-r)}$ such that $\begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix} \begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix}^T = \begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix}^T \begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix} = \mathbf{I}_n$. We can write

$$\mathbf{V}^T \mathbf{A} \mathbf{A}^T \mathbf{V} = \mathbf{V}^T \mathbf{A} \begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix} \begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix}^T \mathbf{A}^T \mathbf{V} = \mathbf{\Lambda}$$

and, without any loss of any generality, we can set

$$\underbrace{\mathbf{V}^T \mathbf{A} \begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix}}_{= [\mathbf{\Lambda}^{1/2} \ \mathbf{0}]} \underbrace{\begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix}^T \mathbf{A}^T \mathbf{V}}_{= [\mathbf{\Lambda}^{1/2} \ \mathbf{0}]^T} = \mathbf{\Lambda}.$$

This leads to

$$\mathbf{V}^T \mathbf{A} \begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix} = \begin{bmatrix} \Lambda^{1/2} & \mathbf{0} \end{bmatrix} \Rightarrow \begin{bmatrix} \mathbf{V}^T \\ \tilde{\mathbf{V}}^T \end{bmatrix} \mathbf{A} \begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix} = \begin{bmatrix} \Lambda^{1/2} & \mathbf{0} \\ \tilde{\mathbf{V}}^T \mathbf{A} \begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \Lambda^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (15)$$

where we use the fact that $\tilde{\mathbf{V}}^T \mathbf{A} = \mathbf{0}$ from (14). Since $\begin{bmatrix} \mathbf{V} & \tilde{\mathbf{V}} \end{bmatrix} \begin{bmatrix} \mathbf{V} & \tilde{\mathbf{V}} \end{bmatrix}^T = \mathbf{I}_m$ and $\begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix} \begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix}^T = \mathbf{I}_n$, the (15) is equivalently rewritten as

$$\mathbf{A} = \begin{bmatrix} \mathbf{V} & \tilde{\mathbf{V}} \end{bmatrix} \begin{bmatrix} \Lambda^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{W} & \tilde{\mathbf{W}} \end{bmatrix}^T = \mathbf{V} \Lambda^{1/2} \mathbf{W}^T. \quad (16)$$

Taking $\Lambda^{1/2} = \Sigma$ in (13) concludes the proof.

- SVD exists for all kinds of matrices (square matrix, symmetric matrix, rectangular matrix, etc.), so it is universal.
- For a square matrix $\mathbf{A}$, singular values of $\mathbf{A}$ are same as the eigenvalues of $\mathbf{A}$, i.e., $\sigma_1 = \lambda_1, \dots, \sigma_r = \lambda_r$?
- For a square matrix $\mathbf{A}$, singular vectors and eigenvectors are equivalent?
- What are the differences between SVD and EVD?

8 Gram Matrix and Gram Determinant

Are there any convenient ways to check whether the vectors $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ are linearly independent or not?

Answering to this question was the major theme of the fields in Numerical Analysis and Linear Algebra. We will shortly overview some of them.

- **Gram matrix and Gram determinant:** Denote $\mathbf{X} = [\mathbf{x}_1 \ \dots \ \mathbf{x}_p] \in \mathbb{R}^{n \times p}$. The Gram matrix for $\mathbf{X}$ is

$$\mathbf{G} = \mathbf{X}^T \mathbf{X}, \quad g_{ij} = \mathbf{x}_i^T \mathbf{x}_j.$$

The determinant of the Gram matrix is the Gram determinant, $\det(\mathbf{G})$.

- **Theorem:** The columns of $\mathbf{X}$ are linearly independent if and only if the Gram determinant is nonzero.
- *Proof of the theorem:* We show it by contrast, i.e., the Gram determinant is zero if and only if the columns of $\mathbf{X}$ are linearly dependent.

if: Suppose the columns of $\mathbf{X}$ are linearly dependent and denote the k th column of $\mathbf{G}$ as $\mathbf{g}_k = \mathbf{X}^T \mathbf{x}_k$. Then, there exists $\mathbf{a} \neq \mathbf{0}$ such that

$$\sum_{k=1}^p a_k \mathbf{g}_k = \mathbf{X}^T \sum_{k=1}^p a_k \mathbf{x}_k = \mathbf{0},$$

because $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ are linearly dependent. This means $\sum_{k=1}^p a_k \mathbf{g}_k = \mathbf{0} \Rightarrow (\mathbf{g}_1, \dots, \mathbf{g}_p)$ are linearly dependent $\Rightarrow \det(\mathbf{G}) = 0$.

only if: Suppose $\det(\mathbf{G}) = 0 \Rightarrow (\mathbf{g}_1, \dots, \mathbf{g}_p)$ are linearly dependent,

$$\begin{aligned} \sum_{k=1}^p a_k \mathbf{g}_k &= \mathbf{X}^T \sum_{k=1}^p a_k \mathbf{x}_k = \mathbf{0} \Rightarrow \mathbf{x}_i^T \sum_{k=1}^p a_k \mathbf{x}_k = 0, i = 1, \dots, p, \\ &\Rightarrow \sum_{i=1}^p a_i \mathbf{x}_i^T \sum_{k=1}^p a_k \mathbf{x}_k = \left\| \sum_{k=1}^p a_k \mathbf{x}_k \right\|^2 = 0 \end{aligned}$$

$\Rightarrow \sum_{k=1}^p a_k \mathbf{x}_k = \mathbf{0} \Rightarrow (\mathbf{x}_1, \dots, \mathbf{x}_p)$ are linearly dependent.

- Hence, the above theorem tells us to evaluate $\det(\mathbf{G})$ in order to check linear independency of $(\mathbf{x}_1, \dots, \mathbf{x}_p)$.

- Next question: How to efficiently calculate $\det(\mathbf{G})$?
- Directly computing $\det(\mathbf{G})$ using (4) is prohibitively complex.

- **Computation of Gram determinant (using Schur complement):** Consider a procedure that sequentially testing each vector in $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ to see if $\mathbf{x}_k$ is linearly independent of the preceding vectors $(\mathbf{x}_1, \dots, \mathbf{x}_{k-1})$.

Suppose that the first $k-1$ vectors have passed the test for independence. If $\mathbf{x}_k$ is linearly independent, we have

$$\mathbf{c}_k = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + \dots + a_{k-1} \mathbf{x}_{k-1} + \mathbf{x}_k \neq \mathbf{0}, \quad (17)$$

where, without loss of generality, $a_k = 1$. Let us rewrite it as

$$\mathbf{c}_k = [\mathbf{X}_{k-1} \ \mathbf{x}_k] \mathbf{a}_k, \text{ with } \mathbf{X}_{k-1} = [\mathbf{x}_1 \ \dots \ \mathbf{x}_{k-1}], \mathbf{a}_k = [a_1 \ \dots \ a_{k-1} \ 1]^T.$$

The norm square of $\mathbf{c}_k$ is

$$\begin{aligned} \sigma_k^2 = \mathbf{c}_k^T \mathbf{c}_k &= \mathbf{a}_k^T \begin{bmatrix} \mathbf{X}_{k-1}^T \\ \mathbf{x}_k^T \end{bmatrix} [\mathbf{X}_{k-1} \ \mathbf{x}_k] \mathbf{a}_k = \mathbf{a}_k^T \begin{bmatrix} \mathbf{X}_{k-1}^T \mathbf{X}_{k-1} & \mathbf{X}_{k-1}^T \mathbf{x}_k \\ \mathbf{x}_k^T \mathbf{X}_{k-1} & \mathbf{x}_k^T \mathbf{x}_k \end{bmatrix} \mathbf{a}_k \\ &= \mathbf{a}_k^T \begin{bmatrix} \mathbf{G}_{k-1} & \mathbf{h}_k \\ \mathbf{x}_k^T & g_{kk} \end{bmatrix} \mathbf{a}_k, \end{aligned}$$

where $\mathbf{G}_{k-1} = \mathbf{X}_{k-1}^T \mathbf{X}_{k-1}$ is the Gram matrix of $(\mathbf{x}_1, \dots, \mathbf{x}_{k-1})$. Now, by (17), the vectors $(\mathbf{x}_1, \dots, \mathbf{x}_k)$ are linearly independent iff $\sigma_k^2 > 0$. If $\sigma_k^2 = 0$, $\mathbf{x}_k$ is linearly dependent.

To see whether or not $\sigma_k^2 > 0$, it is of interest what is the minimum possible value of σ_k^2 . To this end, we solve

$$\hat{\sigma}_k^2 = \min_{\mathbf{a}_k} \mathbf{a}_k^T \begin{bmatrix} \mathbf{G}_{k-1} & \mathbf{h}_k \\ \mathbf{x}_k^T & g_{kk} \end{bmatrix} \mathbf{a}_k \text{ subject to } a_k = 1.$$

Since the problem is convex, we from the Lagrangian function, take the gradient with respect to $\mathbf{a}_k$, and equate to zero, i.e.,

$$\begin{aligned} L(\mathbf{a}_k, \lambda) &= \mathbf{a}_k^T \begin{bmatrix} \mathbf{G}_{k-1} & \mathbf{h}_k \\ \mathbf{x}_h^T & g_{kk} \end{bmatrix} \mathbf{a}_k + 2\lambda(\mathbf{a}_k^T \boldsymbol{\delta} - 1), \quad \boldsymbol{\delta} = [0, \dots, 0, 1]^T \\ \frac{\partial L(\mathbf{a}_k, \lambda)}{\partial \mathbf{a}_k} &= \begin{bmatrix} \mathbf{G}_{k-1} & \mathbf{h}_k \\ \mathbf{x}_h^T & g_{kk} \end{bmatrix} \mathbf{a}_k + \lambda \boldsymbol{\delta} = 0 \\ \Rightarrow &\begin{bmatrix} \mathbf{G}_{k-1} & \mathbf{h}_k \\ \mathbf{x}_h^T & g_{kk} \end{bmatrix} \mathbf{a}_k = -\lambda \boldsymbol{\delta} = \begin{bmatrix} \mathbf{0} \\ -\lambda \end{bmatrix}. \end{aligned} \quad (18)$$

Hence, $\hat{\sigma}_k^2 = \mathbf{a}_k^T \begin{bmatrix} \mathbf{G}_{k-1} & \mathbf{h}_k \\ \mathbf{x}_h^T & g_{kk} \end{bmatrix} \mathbf{a}_k = -\lambda a_k = -\lambda$ and (18) is

$$\begin{bmatrix} \mathbf{G}_{k-1} & \mathbf{h}_k \\ \mathbf{h}^T & g_{kk} \end{bmatrix} \mathbf{a}_k = \begin{bmatrix} \mathbf{0} \\ \hat{\sigma}_k^2 \end{bmatrix}. \quad (19)$$

We have

$$\begin{aligned} \mathbf{G}_{k-1}[a_1 \ \cdots \ a_{k-1}]^T + \mathbf{h}_k &= \mathbf{0} \Rightarrow [a_1 \ \cdots \ a_{k-1}]^T = -\mathbf{G}_{k-1}^{-1} \mathbf{h}_k \\ g_{kk} + \mathbf{h}_k^T [a_1 \ \cdots \ a_{k-1}]^T &= \hat{\sigma}_k^2 \\ \Rightarrow \hat{\sigma}_k^2 &= g_{kk} - \mathbf{h}_k^T \mathbf{G}_{k-1}^{-1} \mathbf{h}_k. \end{aligned} \quad (20)$$

The quantity $\sigma_k^2 = g_{kk} - \mathbf{h}_k^T \mathbf{G}_{k-1}^{-1} \mathbf{h}_k$ is the famous **Schur complement** of the Gram matrix $\mathbf{G}_k$. If the Schur complement $\sigma_k^2 = 0$, then $\mathbf{x}_k$ is a linearly dependent vector. Otherwise, the columns of $\mathbf{X}_k = [\mathbf{X}_{k-1} \ \mathbf{x}_k]$ are linearly independent.

Once with the linear independence, one can efficiently calculate $\det(\mathbf{G})$. From (19), we have

$$\mathbf{G}_k \mathbf{a}_k = \sigma_k^2 \boldsymbol{\delta} \Rightarrow \mathbf{a}_k = \sigma_k^2 \mathbf{G}_k^{-1} \boldsymbol{\delta} \quad (21)$$

and premultiplying $\boldsymbol{\delta}^T$ at the both sides of (21) yields

$$\boldsymbol{\delta}^T \mathbf{a}_k = 1 = \sigma_k^2 \boldsymbol{\delta}^T \mathbf{G}_k^{-1} \boldsymbol{\delta}.$$

Since $\mathbf{G}_k^{-1} = \frac{\mathbf{C}^T}{\det(\mathbf{G}_k)}$ by (6),

$$\boldsymbol{\delta}^T \mathbf{G}_k^{-1} \boldsymbol{\delta} = \frac{\boldsymbol{\delta}^T \mathbf{C}^T \boldsymbol{\delta}}{\det(\mathbf{G}_k)} = \frac{M_{nn}}{\det(\mathbf{G}_k)} = \frac{\det(\mathbf{G}_{k-1})}{\det(\mathbf{G}_k)},$$

meaning

$$1 = \sigma_k^2 \frac{\det(\mathbf{G}_{k-1})}{\det(\mathbf{G}_k)} \Rightarrow \det(\mathbf{G}_k) = \sigma_k^2 \det(\mathbf{G}_{k-1}).$$

Then, iterating this formula results in

$$\det(\mathbf{G}_k) = \sigma_k^2 \det(\mathbf{G}_{k-1}) = \sigma_k^2 \sigma_{k-1}^2 \det(\mathbf{G}_{k-2}) = \prod_{i=1}^k \sigma_i^2.$$

$$\Rightarrow \det(\mathbf{G}_k) = \prod_{i=1}^k \sigma_i^2. \quad (22)$$

Hence, the determinant of Gram matrix $\mathbf{G}_k$ can be efficiently computed by (22).

If $\det(\mathbf{G}_k) = 0$, $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ are linearly dependent and if $\det(\mathbf{G}_k) \neq 0$, $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ are linearly independent.

9 Gram-Schmidt Orthogonalization

There is another way to check for the linear independence of vectors $(\mathbf{x}_1, \dots, \mathbf{x}_p)$:

- (a) transform a set of vectors $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ into orthogonal vectors
- (b) check to see how many orthogonal vectors can be generated.

The number of orthogonal vectors is the number of bases of $\mathcal{S}(\mathbf{x}_1, \dots, \mathbf{x}_p)$. The transform of $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ into orthogonal vectors can be done by Gram-Schmidt orthogonalization.

• Gram-Schmidt procedure:

- Given a set of vectors $(\mathbf{x}_1, \dots, \mathbf{x}_p)$, start with the initialization

$$\mathbf{y}_1 = \mathbf{x}_1, \quad \mathbf{u}_1 = \frac{\mathbf{y}_1}{\|\mathbf{y}_1\|}, \quad r_{11} = \|\mathbf{y}_1\|$$

- Construct $\mathbf{u}_2$ as follows:

$$\mathbf{y}_2 = \mathbf{x}_2 - (\mathbf{u}_1^T \mathbf{x}_2) \mathbf{u}_1, \quad \mathbf{u}_2 = \frac{\mathbf{y}_2}{\|\mathbf{y}_2\|}, \quad r_{12} = (\mathbf{u}_1^T \mathbf{x}_2), \quad r_{11} = \|\mathbf{y}_2\|$$

The $\mathbf{u}_2 = \mathbf{0}$ iff $\mathbf{x}_2 = a\mathbf{x}_1$, meaning that $\mathbf{x}_2$ is linearly dependent on $\mathbf{x}_1$. If $\mathbf{u}_2 \neq \mathbf{0}$, $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal (i.e., linearly independent). This can be easily checked by computing $\mathbf{y}_1^T \mathbf{y}_2$,

$$\begin{aligned} \|\mathbf{y}_1\| \|\mathbf{y}_2\| \mathbf{u}_1^T \mathbf{u}_2 = \mathbf{y}_1^T \mathbf{y}_2 &= \mathbf{x}_1^T \mathbf{x}_2 - (\mathbf{x}_2^T \mathbf{u}_1) \mathbf{x}_1^T \mathbf{u}_1 \\ &= \mathbf{x}_1^T \mathbf{x}_2 - \left(\mathbf{x}_2^T \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|} \right) \mathbf{x}_1^T \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|} \\ &= \mathbf{x}_1^T \mathbf{x}_2 - (\mathbf{x}_2^T \mathbf{x}_1) \frac{\mathbf{x}_1^T \mathbf{x}_1}{\|\mathbf{x}_1\|^2} \\ &= 0 \Rightarrow \mathbf{u}_1^T \mathbf{u}_2 = 0. \end{aligned}$$

- Continue this procedure, generating each new vector $\mathbf{u}_k$:

$$\mathbf{y}_k = \mathbf{x}_k - \sum_{i=1}^{k-1} r_{ik} \mathbf{u}_i, \quad \mathbf{u}_k = \frac{\mathbf{y}_k}{\|\mathbf{y}_k\|}, \quad r_{ik} = \mathbf{u}_i^T \mathbf{x}_k, \quad r_{kk} = \|\mathbf{y}_k\|.$$

The vector $\mathbf{u}_k \neq \mathbf{0}$ (i.e., $\mathbf{y}_k \neq \mathbf{0}$) if the vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ are linearly independent and, otherwise, they are dependent.

After p -times iteration, the algorithm will yield the orthogonal basis $(\mathbf{u}_1, \dots, \mathbf{u}_p)$.

- **QR decomposition:**

One important byproduct of the Gram-Schmidt orthogonalization is the QR decomposition. After successful p -times Gram-Schmidt procedure, we have the following decomposition of $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_p]$:

$$\mathbf{X} = \mathbf{Q}\mathbf{R},$$

where $\mathbf{Q} = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_p]$ and

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ 0 & r_{22} & & r_{2p} \\ & & \ddots & \vdots \\ 0 & 0 & \cdots & r_{pp} \end{bmatrix}.$$

Note that $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_p$ and $\mathbf{R}$ is the upper triangular matrix.

In the representation of what Gram-Schmidt procedure delivers, each r_{ii} is nonzero indicates that $\mathbf{x}_i$ is linearly independent of $(\mathbf{x}_1, \dots, \mathbf{x}_{i-1})$ and $\mathbf{Q}$ represents the left (or column) subspace of $\mathbf{X}$.