ENGG 5501: Foundations of Optimization **2024–25 First Term**

Handout B: Linear Algebra Cheat Sheet

Instructor: Anthony Man–Cho So Updated: August 24, 2024

The purpose of this handout is to give a brief review of some of the basic concepts and results in linear algebra. If you are not familiar with the material and/or would like to do some further reading, you may consult, e.g., the books [1, 2, 3].

1 Basic Notation, Definitions, and Results

1.1 Vectors and Matrices

We denote the set of real numbers (also referred to as **scalars**) by $\mathbb{R}$. For positive integers $m, n \geq 1$, we use $\mathbb{R}^{m\times n}$ to denote the set of $m \times n$ arrays whose components are from $\mathbb{R}$. In other words, $\mathbb{R}^{m\times n}$ is the set of n–dimensional real matrices, and an element $A \in \mathbb{R}^{m\times n}$ can be written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \tag{1}$$

where $a_{ij} \in \mathbb{R}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. A **row vector** is a matrix with $m = 1$, and a **column vector** is a matrix with $n = 1$. The word *vector* will always mean a column vector unless otherwise stated. The set of all n–dimensional real vectors is denoted by $\mathbb{R}^n$, and an element $x \in \mathbb{R}^n$ can be written as $x = (x_1, \ldots, x_n)$. Note that we still view $x = (x_1, \ldots, x_n)$ as a column vector, even though typographically it does not appear so. The reason for such a notation is simply to save space. Now, given an $m \times n$ matrix A of the form (1), its **transpose** A^T is defined as the following $n \times m$ matrix:

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

An $m \times m$ real matrix A is said to be **symmetric** if $A = A^T$. The set of $m \times m$ real symmetric matrices is denoted by $\mathcal{S}^m$.

We use $x \geq \mathbf{0}$ to indicate that all the components of x are non–negative, and $x \geq y$ to mean that $x - y \geq \mathbf{0}$. The notations $x > \mathbf{0}$, $x \leq \mathbf{0}$, $x < \mathbf{0}$, $x > y$, $x \leq y$, and $x < y$ are to be interpreted accordingly.

We say that a finite collection $\mathcal{C} = \{x^1, x^2, \ldots, x^m\}$ of vectors in $\mathbb{R}^n$ is

- **linearly dependent** if there exist scalars $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$, not all of them zero, such that $\sum_{i=1}^m \alpha_i x^i = \mathbf{0}$;
- **affinely dependent** if the collection $\mathcal{C}' = \{x^2 - x^1, x^3 - x^1, \ldots, x^m - x^1\}$ is linearly dependent.

The collection $\mathcal{C}$ (resp. $\mathcal{C}'$) is said to be **linearly independent** (resp. **affinely independent**) if it is not linearly dependent (resp. affinely dependent).

1.2 Inner Product and Vector Norms

Given two vectors $x, y \in \mathbb{R}^n$, their **inner product** is defined as

$$x^T y \equiv \sum_{i=1}^{n} x_i y_i.$$

We say that x and y are **orthogonal** if $x^T y = 0$. The **Euclidean norm** of $x \in \mathbb{R}^n$ is defined as

$$\|x\|_2 \equiv \sqrt{x^T x} = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}.$$

A fundamental inequality that relates the inner product of two vectors and their respective Euclidean norms is the **Cauchy–Schwarz inequality**:

$$\left| x^T y \right| \leq \|x\|_2 \cdot \|y\|_2.$$

Equality holds iff the vectors x and y are linearly dependent; i.e., $x = \alpha y$ for some $\alpha \in \mathbb{R}$.

Note that the Euclidean norm is not the only norm one can define on $\mathbb{R}^n$. In general, a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is called a **vector norm** on $\mathbb{R}^n$ if for all $x, y \in \mathbb{R}^n$, we have

(a) **(Non–Negativity)** $\|x\| \geq 0$;

(b) **(Positivity)** $\|x\| = 0$ iff $x = \mathbf{0}$;

(c) **(Homogeneity)** $\|\alpha x\| = |\alpha| \cdot \|x\|$ for all $\alpha \in \mathbb{R}$;

(d) **(Triangle Inequality)** $\|x + y\| \leq \|x\| + \|y\|$.

For instance, for $p \geq 1$, the ℓ_p**–norm** on $\mathbb{R}^n$, which is given by

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p},$$

is a vector norm on $\mathbb{R}^n$. It is well known that

$$\|x\|_\infty = \lim_{p \to \infty} \|x\|_p = \max_{1 \leq i \leq n} |x_i|.$$

1.3 Matrix Norms

We say that a function $\|\cdot\| : \mathbb{R}^{n \times n} \to \mathbb{R}$ is a **matrix norm** on the set of $n \times n$ matrices if for any $A, B \in \mathbb{R}^{n \times n}$, we have

(a) **(Non–Negativity)** $\|A\| \geq 0$;

(b) **(Positivity)** $\|A\| = 0$ iff $A = \mathbf{0}$;

(c) **(Homogeneity)** $\|\alpha A\| = |\alpha| \cdot \|A\|$ for all $\alpha \in \mathbb{R}$;

(d) **(Triangle Inequality)** $\|A + B\| \leq \|A\| + \|B\|$;

(e) **(Submultiplicativity)** $\|AB\| \leq \|A\| \cdot \|B\|$.

As an example, let $\|\cdot\|_v : \mathbb{R}^n \to \mathbb{R}$ be a vector norm on $\mathbb{R}^n$. Define the function $\|\cdot\| : \mathbb{R}^{n\times n} \to \mathbb{R}$ via

$$\|A\| = \max_{x\in\mathbb{R}^n : \|x\|_v = 1} \|Ax\|_v.$$

Then, it is straightforward to verify that $\|\cdot\|$ is a matrix norm on the set of $n \times n$ matrices.

1.4 Linear Subspaces and Bases

A non–empty subset S of $\mathbb{R}^n$ is called a **(linear) subspace** of $\mathbb{R}^n$ if $\alpha x + \beta y \in S$ whenever $x, y \in S$ and $\alpha, \beta \in \mathbb{R}$. Clearly, we have $\mathbf{0} \in S$ for any subspace S of $\mathbb{R}^n$.

The **span** (or **linear hull**) of a finite collection $\mathcal{C} = \{x^1, \ldots, x^m\}$ of vectors in $\mathbb{R}^n$ is defined as

$$\text{span}(\mathcal{C}) \equiv \left\{ \sum_{i=1}^{m} \alpha_i x^i : \alpha_1, \ldots, \alpha_m \in \mathbb{R} \right\}.$$

In particular, every vector $y \in \text{span}(\mathcal{C})$ is a **linear combination** of the vectors in $\mathcal{C}$. It is easy to verify that $\text{span}(\mathcal{C})$ is a subspace of $\mathbb{R}^n$.

We can extend the above definition to an arbitrary (i.e., not necessarily finite) collection $\mathcal{C}$ of vectors in $\mathbb{R}^n$. Specifically, we define $\text{span}(\mathcal{C})$ as the set of all *finite* linear combinations of the vectors in $\mathcal{C}$. Equivalently, we can define $\text{span}(\mathcal{C})$ as the intersection of all subspaces containing $\mathcal{C}$. Note that when $\mathcal{C}$ is finite, this definition coincides with the one given above.

Given a subspace S of $\mathbb{R}^n$ with $S \neq \{\mathbf{0}\}$, a **basis** $\mathcal{B}$ of S is a linearly independent collection of vectors whose span is equal to S. If the vectors in $\mathcal{B}$ are orthogonal to each other and have unit norm, then we call $\mathcal{B}$ an **orthonormal basis** of S. Recall that every basis of a given subspace S has the same number of vectors. This number is called the **dimension** of the subspace S and is denoted by $\dim(S)$. By definition, the dimension of the subspace $\{\mathbf{0}\}$ is zero. The **orthogonal complement** $S^\perp$ of S is defined as

$$S^\perp = \left\{ y \in \mathbb{R}^n : x^T y = 0 \text{ for all } x \in S \right\}.$$

It can be verified that $S^\perp$ is a subspace of $\mathbb{R}^n$, and that if $\dim(S) = k \in \{0, 1, \ldots, n\}$, then we have $\dim(S^\perp) = n - k$. Moreover, we have $S^{\perp\perp} = (S^\perp)^\perp = S$. Finally, every vector $x \in \mathbb{R}^n$ can be uniquely decomposed as $x = x^1 + x^2$, where $x^1 \in S$ and $x^2 \in S^\perp$.

Now, let A be an $m \times n$ real matrix. The **column space** of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A. It is also known as the **range** of A (viewed as a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$) and is denoted by

$$\text{range}(A) \equiv \{Ax : x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m.$$

Similarly, the **row space** of A is the subspace of $\mathbb{R}^n$ spanned by the rows of A. It is well known that the dimension of the column space is equal to the dimension of the row space, and this number is known as the **rank** of the matrix A (denoted by $\text{rank}(A)$). In particular, we have

$$\text{rank}(A) = \dim(\text{range}(A)) = \dim(\text{range}(A^T)).$$

Moreover, we have $\text{rank}(A) \leq \min\{m, n\}$, and if equality holds, then we say that A has **full rank**. The **nullspace** of A is the set $\text{null}(A) \equiv \{x \in \mathbb{R}^n : Ax = \mathbf{0}\}$. It is a subspace of $\mathbb{R}^n$ and has dimension $n - \text{rank}(A)$. The following summarizes the relationships among the subspaces $\text{range}(A)$, $\text{range}(A^T)$, $\text{null}(A)$, and $\text{null}(A^T)$:

$$\begin{aligned} (\text{range}(A))^{\perp} &= \text{null}(A^T), \\ \left(\text{range}(A^T)\right)^{\perp} &= \text{null}(A). \end{aligned}$$

The above implies that given an $m \times n$ real matrix A of rank $r \leq \min\{m, n\}$, we have $\text{rank}(AA^T) = \text{rank}(A^T A) = r$. This fact will be frequently used in the course.

1.5 Affine Subspaces

Let S_0 be a subspace of $\mathbb{R}^n$ and $x^0 \in \mathbb{R}^n$ be an arbitrary vector. Then, the set $S = \{x^0\} + S_0 = \{x + x^0 : x \in S_0\}$ is called an **affine subspace** of $\mathbb{R}^n$, and its dimension is equal to the dimension of the underlying subspace S_0.

Now, let $\mathcal{C} = \{x^1, \ldots, x^m\}$ be a finite collection of vectors in $\mathbb{R}^n$, and let $x^0 \in \mathbb{R}^n$ be arbitrary. By definition, the set $S = \{x^0\} + \text{span}(\mathcal{C})$ is an affine subspace of $\mathbb{R}^n$. Moreover, it is easy to verify that every vector $y \in S$ can be written in the form

$$y = \sum_{i=1}^{m} \left[\alpha_i(x^0 + x^i) + \beta_i(x^0 - x^i)\right]$$

for some $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m \in \mathbb{R}$ such that $\sum_{i=1}^{m} (\alpha_i + \beta_i) = 1$; i.e., the vector $y \in \mathbb{R}^n$ is an **affine combination** of the vectors $x^0 \pm x^1, \ldots, x^0 \pm x^m \in \mathbb{R}^n$. Conversely, let $\mathcal{C} = \{x^1, \ldots, x^m\}$ be a finite collection of vectors in $\mathbb{R}^n$, and define

$$S = \left\{ \sum_{i=1}^{m} \alpha_i x^i : \alpha_1, \ldots, \alpha_m \in \mathbb{R}, \sum_{i=1}^{m} \alpha_i = 1 \right\}$$

to be the set of affine combinations of the vectors in $\mathcal{C}$. We claim that S is an affine subspace of $\mathbb{R}^n$. Indeed, it can be readily verified that

$$S = \{x^1\} + \text{span}\left(\{x^2 - x^1, \ldots, x^m - x^1\}\right).$$

This establishes the claim.

Given an arbitrary (i.e., not necessarily finite) collection $\mathcal{C}$ of vectors in $\mathbb{R}^n$, the **affine hull** of $\mathcal{C}$, denoted by $\text{aff}(\mathcal{C})$, is the set of all *finite* affine combinations of the vectors in $\mathcal{C}$. Equivalently, we can define $\text{aff}(\mathcal{C})$ as the intersection of all affine subspaces containing $\mathcal{C}$.

1.6 Some Special Classes of Matrices

The following classes of matrices will be frequently encountered in this course.

- **Invertible Matrix.** An $n \times n$ real matrix A is said to be **invertible** if there exists an $n \times n$ real matrix A^{-1} (called the **inverse** of A) such that $A^{-1}A = I$, or equivalently, $AA^{-1} = I$. Note that the inverse of A is unique whenever it exists. Morever, recall that $A \in \mathbb{R}^{n \times n}$ is invertible iff $\text{rank}(A) = n$.

Now, let A be a non–singular $n \times n$ real matrix. Suppose that A is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where $A_{ii} \in \mathbb{R}^{n_i \times n_i}$ for $i = 1, 2$, with $n_1 + n_2 = n$. Then, *provided that the relevant inverses exist*, the inverse of A has the following form:

$$A^{-1} = \begin{bmatrix} \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} & A_{11}^{-1}A_{12}\left(A_{21}A_{11}^{-1}A_{12} - A_{22}\right)^{-1} \\ \left(A_{21}A_{11}^{-1}A_{12} - A_{22}\right)^{-1}A_{21}A_{11}^{-1} & \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} \end{bmatrix}.$$

- **Submatrix of a Matrix.** Let A be an $m \times n$ real matrix. For index sets $\alpha \subseteq \{1, 2, \ldots, m\}$ and $\beta \subseteq \{1, 2, \ldots, n\}$, we denote the submatrix that lies in the rows of A indexed by α and the columns indexed by β by $A(\alpha, \beta)$. If $m = n$ and $\alpha = \beta$, the matrix $A(\alpha, \alpha)$ is called a **principal submatrix** of A and is denoted by $A(\alpha)$. The determinant of $A(\alpha)$ is called a **principal minor** of A.

  Now, let A be an $n \times n$ matrix, and let $\alpha \subseteq \{1, 2, \ldots, n\}$ be an index set such that $A(\alpha)$ is non–singular. We set $\alpha' = \{1, 2, \ldots, n\} \backslash \alpha$. The following is known as the **Schur determinantal formula**:

  $$\det(A) = \det(A(\alpha))\det\left[A(\alpha') - A(\alpha', \alpha)A(\alpha)^{-1}A(\alpha, \alpha')\right].$$

- **Orthogonal Matrix.** An $n \times n$ real matrix A is called an **orthogonal matrix** if $AA^T = A^TA = I$. Note that if $A \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, then for any $u, v \in \mathbb{R}^n$, we have $u^Tv = (Au)^T(Av)$; i.e., orthogonal transformations preserve inner products.

- **Positive Semidefinite/Definite Matrix.** An $n \times n$ real matrix A is **positive semidefinite** (resp. **positive definite**) if A is symmetric and for any $x \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, we have $x^TAx \geq 0$ (resp. $x^TAx > 0$). We use $A \succeq \mathbf{0}$ (resp. $A \succ \mathbf{0}$) to denote the fact that A is positive semidefinite (resp. positive definite). We remark that although one can define a notion of positive semidefiniteness for real matrices that are not necessarily symmetric, we shall not pursue that option in this course.

- **Projection Matrix.** An $n \times n$ real matrix A is called a **projection matrix** if $A^2 = A$. Given a projection matrix $A \in \mathbb{R}^{n \times n}$ and a vector $x \in \mathbb{R}^n$, the vector $Ax \in \mathbb{R}^n$ is called the **projection of** $x \in \mathbb{R}^n$ **onto the subspace range**(A). Note that a projection matrix need not be symmetric. As an example, consider

  $$A = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

  We say that A defines an **orthogonal projection** onto the subspace $S \subseteq \mathbb{R}^n$ if for every $x = x^1 + x^2 \in \mathbb{R}^n$, where $x^1 \in S$ and $x^2 \in S^\perp$, we have $Ax = x^1$. Note that if A defines an orthogonal projection onto S, then $I - A$ defines an orthogonal projection onto $S^\perp$. Furthermore, it can be shown that A is an orthogonal projection onto S iff A is a symmetric projection matrix with range$(A) = S$.

  As an illustration, consider an $m \times n$ real matrix A, with $m \leq n$ and rank$(A) = m$. Then, the projection matrix corresponding to the orthogonal projection onto the nullspace of A is given by $P_{null(A)} = I - A^T(AA^T)^{-1}A$.

2 Eigenvalues and Eigenvectors

Let A be an $n \times n$ real matrix. We say that $\lambda \in \mathbb{C}$ is an **eigenvalue** of A with corresponding **eigenvector** $u \in \mathbb{C}^n \backslash \{\mathbf{0}\}$ if $Au = \lambda u$. Note that the zero vector $\mathbf{0} \in \mathbb{R}^n$ *cannot* be an eigenvector, although zero *can* be an eigenvalue. Also, recall that given an $n \times n$ real matrix A, there are exactly n eigenvalues (counting multiplicities).

The set of eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ of an $n \times n$ matrix A is closely related to the **trace** and **determinant** of A (denoted by $\mathrm{tr}(A)$ and $\det(A)$, respectively). Specifically, we have

$$\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i \quad \text{and} \quad \det(A) = \prod_{i=1}^{n} \lambda_i.$$

These formulae can be established by considering the **characteristic polynomial** $\lambda \mapsto p_A(\lambda) = \det(\lambda I - A)$ of A. Moreover, we have the following results:

(a) The eigenvalues of A^T are the same as those of A.

(b) For any $c \in \mathbb{R}$, the eigenvalues of $cI + A$ are $c + \lambda_1, \ldots, c + \lambda_n$.

(c) For any integer $k \geq 1$, the eigenvalues of A^k are $\lambda_1^k, \ldots, \lambda_n^k$.

(d) If A is invertible, then the eigenvalues of A^{-1} are $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$.

2.1 Spectral Properties of Real Symmetric Matrices

The **Spectral Theorem for Real Symmetric Matrices** states that an $n \times n$ real matrix A is symmetric iff there exists an orthogonal matrix $U \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ such that

$$A = U \Lambda U^T. \tag{2}$$

If the eigenvalues of A are $\lambda_1, \ldots, \lambda_n$, then we can take $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and u^i, the i–th column of U, to be the eigenvector associated with the eigenvalue λ_i for $i = 1, \ldots, n$. In particular, the eigenvalues of a real symmetric matrix are all real, and their associated eigenvectors are orthogonal to each other. Note that (2) can be equivalently written as

$$A = \sum_{i=1}^{n} \lambda_i u^i (u^i)^T,$$

and the rank of A is equal to the number of non–zero eigenvalues.

Note that the set of eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ of A is unique. Specifically, if $\{\gamma_1, \ldots, \gamma_n\}$ is another set of eigenvalues of A, then there exists a permutation $\pi = (\pi_1, \ldots, \pi_n)$ of $\{1, \ldots, n\}$ such that $\lambda_i = \gamma_{\pi_i}$ for $i = 1, \ldots, n$. This follows from the fact that the eigenvalues of A are the solutions to the characteristic polynomial

$$\det(\lambda I - A) = 0.$$

On the other hand, the set of unit–norm eigenvectors $\{u^1, \ldots, u^n\}$ of A is not unique. A simple reason is that if u is a unit–norm eigenvector, then $-u$ is also a unit–norm eigenvector. However, there is a deeper reason. Suppose that A has repeated eigenvalues, say, $\lambda_1 = \cdots = \lambda_k = \bar{\lambda}$ for some $k > 1$, with corresponding eigenvectors $u^1, \ldots, u^k$. Then, it can be verified that any vector in the k–dimensional subspace $\bar{\mathcal{L}} = \mathrm{span}\{u^1, \ldots, u^k\}$ is an eigenvector of A with eigenvalue $\bar{\lambda}$. Moreover,

all the remaining eigenvectors are orthogonal to $\bar{\mathcal{L}}$. Consequently, each orthonormal basis of $\bar{\mathcal{L}}$ gives rise to a set of k eigenvectors of A whose associated eigenvalue is $\bar{\lambda}$. It is worth noting that if $\{v^1, \ldots, v^k\}$ is an orthonormal basis of $\bar{\mathcal{L}}$, then we can find an orthogonal matrix $P_1^k \in \mathbb{R}^{k \times k}$ such that $V_1^k = U_1^k P_1^k$, where U_1^k (resp. V_1^k) is the $n \times k$ matrix whose i–th column is u^i (resp. v^i), for $i = 1, \ldots, k$. In particular, if $A = U\Lambda U^T = V\Lambda V^T$ are two spectral decompositions of A with

$$\Lambda = \begin{bmatrix} \lambda_{i_1} I_{n_1} & & & \\ & \lambda_{i_2} I_{n_2} & & \\ & & \ddots & \\ & & & \lambda_{i_l} I_{n_l} \end{bmatrix},$$

where $\lambda_{i_1}, \ldots, \lambda_{i_l}$ are the distinct eigenvalues of A, I_k denotes a $k \times k$ identity matrix, and $n_1 + n_2 + \cdots + n_l = n$, then there exists an orthogonal matrix P with the block diagonal structure

$$P = \begin{bmatrix} P_{n_1} & & & \\ & P_{n_2} & & \\ & & \ddots & \\ & & & P_{n_l} \end{bmatrix},$$

where P_{n_j} is an $n_j \times n_j$ orthogonal matrix for $j = 1, \ldots, l$, such that $V = UP$.

Now, suppose that we order the eigenvalues of A as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then, the **Courant–Fischer theorem** states that the k–th largest eigenvalue λ_k, where $k = 1, \ldots, n$, can be found by solving the following optimization problems:

$$\lambda_k = \min_{w^1, \ldots, w^{k-1} \in \mathbb{R}^n} \max_{\substack{x \neq \mathbf{0}, x \in \mathbb{R}^n \\ x \perp w^1, \ldots, w^{k-1}}} \frac{x^T A x}{x^T x} = \max_{w^1, \ldots, w^{n-k} \in \mathbb{R}^n} \min_{\substack{x \neq \mathbf{0}, x \in \mathbb{R}^n \\ x \perp w^1, \ldots, w^{n-k}}} \frac{x^T A x}{x^T x}. \tag{3}$$

2.2 Properties of Positive Semidefinite Matrices

By definition, a real positive semidefinite matrix is symmetric, and hence it has the properties listed above. However, much more can be said about such matrices. For instance, the following statements are equivalent for an $n \times n$ real symmetric matrix A:

(a) A is positive semidefinite.

(b) All the eigenvalues of A are non–negative.

(c) There exists a unique $n \times n$ positive semidefinite matrix $A^{1/2}$ such that $A = A^{1/2} A^{1/2}$.

(d) There exists an $k \times n$ matrix B, where $k = \text{rank}(A)$, such that $A = B^T B$.

Similarly, the following statements are equivalent for an $n \times n$ real symmetric matrix A:

(a) A is positive definite.

(b) A^{-1} exists and is positive definite.

(c) All the eigenvalues of A are positive.

(d) There exists a unique $n \times n$ positive definite matrix $A^{1/2}$ such that $A = A^{1/2} A^{1/2}$.

Sometimes it would be useful to have a criterion for determining the positive semidefiniteness of a matrix from a block partitioning of the matrix. Here is one such criterion. Let

$$A = \begin{bmatrix} X & Y \\ Y^T & Z \end{bmatrix}$$

be an $n \times n$ real symmetric matrix, where both X and Z are square. Suppose that Z is invertible. Then, the **Schur complement** of the matrix A is defined as the matrix $S_A = X - YZ^{-1}Y^T$. If $Z \succ \mathbf{0}$, then it can be shown that $A \succeq \mathbf{0}$ iff $X \succeq \mathbf{0}$ and $S_A \succeq \mathbf{0}$. There is of course nothing special about the block Z. If X is invertible, then we can similarly define the Schur complement of A as $S'_A = Z - Y^T X^{-1} Y$. If $X \succ \mathbf{0}$, then we have $A \succeq \mathbf{0}$ iff $Z \succeq \mathbf{0}$ and $S'_A \succeq \mathbf{0}$.

3 Singular Values and Singular Vectors

Let A be an $m \times n$ real matrix of rank $r \geq 1$. Then, there exist orthogonal matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ such that

$$A = U \Lambda V^T, \tag{4}$$

where $\Lambda \in \mathbb{R}^{m \times n}$ has $\Lambda_{ij} = 0$ for $i \neq j$ and $\Lambda_{11} \geq \Lambda_{22} \geq \cdots \geq \Lambda_{rr} > \Lambda_{r+1,r+1} = \cdots = \Lambda_{qq} = 0$ with $q = \min\{m, n\}$. The representation (4) is called the **Singular Value Decomposition (SVD)** of A; cf. (2). The entries $\Lambda_{11}, \ldots, \Lambda_{qq}$ are called the **singular values** of A, and the columns of U (resp. V) are called the **left** (resp. **right**) **singular vectors** of A. For notational convenience, we write $\sigma_i \equiv \Lambda_{ii}$ for $i = 1, \ldots, q$. Note that (4) can be equivalently written as

$$A = \sum_{i=1}^{r} \sigma_i u^i (v^i)^T,$$

where u^i (resp. v^i) is the i–th column of the matrix U (resp. V), for $i = 1, \ldots, r$. The rank of A is equal to the number of non–zero singular values.

Now, suppose that we order the singular values of A as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_q$, where $q = \min\{m, n\}$. Then, the **Courant–Fischer theorem** states that the k–th largest singular value σ_k, where $k = 1, \ldots, q$, can be found by solving the following optimization problems:

$$\sigma_k = \min_{w^1, \ldots, w^{k-1} \in \mathbb{R}^n} \max_{\substack{x \neq \mathbf{0}, x \in \mathbb{R}^n \\ x \perp w^1, \ldots, w^{k-1}}} \frac{\|Ax\|_2}{\|x\|_2} = \max_{w^1, \ldots, w^{n-k} \in \mathbb{R}^n} \min_{\substack{x \neq \mathbf{0}, x \in \mathbb{R}^n \\ x \perp w^1, \ldots, w^{n-k}}} \frac{\|Ax\|_2}{\|x\|_2}. \tag{5}$$

The optimization problems (3) and (5) suggest that singular value and eigenvalue are closely related notions. Indeed, if A is an $m \times n$ real matrix, then

$$\lambda_k(A^T A) = \lambda_k(A A^T) = \sigma_k^2(A) \qquad \text{for } k = 1, \ldots, q,$$

where $q = \min\{m, n\}$. Moreover, the columns of U and V are the eigenvectors of AA^T and $A^T A$, respectively. In particular, our discussion in Section 2.1 implies that the set of singular values of A is unique, but the sets of left and right singular vectors are not. Finally, we note that the largest singular value function induces a matrix norm, which is known as the **spectral norm** and is sometimes denoted by

$$\|A\|_2 = \sigma_1(A).$$

Given an SVD of an $m \times n$ matrix A as in (4), we can define another $n \times m$ matrix $A^\dagger$ by

$$A^\dagger = V \Lambda^\dagger U^T,$$

where $\Lambda^\dagger \in \mathbb{R}^{n \times m}$ has $\Lambda^\dagger_{ij} = 0$ for $i \neq j$ and

$$\Lambda^\dagger_{ii} = \begin{cases} 1/\Lambda_{ii} & \text{for } i = 1, \ldots, r, \\ 0 & \text{otherwise.} \end{cases}$$

The matrices A and $A^\dagger$ possess the following nice properties:

(a) $AA^\dagger$ and $A^\dagger A$ are symmetric.

(b) $AA^\dagger A = A$.

(c) $A^\dagger AA^\dagger = A^\dagger$.

(d) $A^\dagger = A^{-1}$ if A is square and non–singular.

The matrix $A^\dagger$ is known as the **Moore–Penrose generalized inverse** of A. It can be shown that $A^\dagger$ is uniquely determined by the conditions (a)–(c) above.

References

[1] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.

[2] G. Strang. *Introduction to Linear Algebra*. Wellesley–Cambridge Press, Wellesley, Massachusetts, third edition, 2003.

[3] G. Strang. *Linear Algebra and Its Applications*. Brooks/Cole, Boston, Massachusetts, fourth edition, 2006.