Chapter 1

Mathematical Preliminaries

In this chapter, we review some of the most important topics from Calculus, Linear Algebra, and Matrix Analysis that will be required in the subsequent chapters throught this course. Throught this book, $\mathbb{R}$ will denote the set of all real numbers.

1.1 Review of Calculus

We start with the review of some definitions and theorems in Calculus including limit, continuity, derivative, and Riemann integral. We shall emphasize the various forms of Taylor's theorem.

1.1.1 Limit, Continuity, Derivative, and Integral

Unless otherwise stated, f will denote a real-valued function defined in a domain $\Omega \subseteq \mathbb{R}$, i.e., $f : \Omega \to \mathbb{R}$.

Definition 1.1 (Limit) *We say f has the limit L at c, written as $\lim_{x\to c} f(x) = L$, if*

$$\forall\, \varepsilon > 0,\ \exists\, \delta > 0,\ \ \textit{such that}\ \ |x - c| < \delta \Rightarrow |f(x) - L| < \varepsilon.$$

If there is no number L with this property, then we say the limit of f at c does not exist.

Definition 1.2 (Continuous) *The function f is said to be continuous at c if*

$$\lim_{x\to c} f(x) = f(c).$$

Definition 1.3 *Let $\{x_i\}_{i=1}^{\infty}$, where $x_i \in \mathbb{R}$ or $\mathbb{C}$, be an infinite sequence. The sequence converges to $x^\star$, write $\lim_{i\to\infty} x_i = x^\star$ or $x_i \to x^\star$ as $i \to \infty$, if for any $\varepsilon > 0$, there exists a positive integer $N(\varepsilon)$ such that $i > N$ implies $|x_i - x^\star| < \varepsilon$.*

Theorem 1.1 *Suppose $f : \Omega \to \mathbb{R}$, where $\Omega \subseteq \mathbb{R}$, is a function and $x^\star \in X$. Then the following are equivalent.*

1. *f is continuous at $x^\star$.*

2. *If $\{x_i\}_{i=1}^{\infty}$ is any sequence in X and $\lim_{i\to\infty} x_i = x^\star$, then $\lim_{i\to\infty} f(x_i) = f(x^\star)$.*

Definition 1.4 (Derivative) *The derivative of the function f at c is the limit, if it exists, defined by*

$$f'(c) = \lim_{x\to c} \frac{f(x)-f(c)}{x-c}.$$

If $f'(c)$ exists, then we say f is defferentiable at c.

Remarks 1.1 *Another form of derivative:*

$$f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}.$$

Theorem 1.2 *If f is differentiable at c, then f must be continuous at c. But the converse is not true.*

Definition 1.5 *By convention, we use the following definitions and notations.*

- *$C(\Omega)$: the set of all functions that are continuous everywhere in Ω.*
- *$C^1(\Omega)$: the set of all functions f for which f' exists and is continuous everywhere in Ω.*
- *$C^n(\Omega)$: the set of all functions f for which $f^{(n)}$ exists and is continuous everywhere in Ω.*
- *$C^\infty(\Omega)$: the set of all functions each of whose derivatives is continuous.*

Property 1.1 $C^\infty(\Omega) \subset \ldots \subset C^2(\Omega) \subset C^1(\Omega) \subset C(\Omega)$

Remarks 1.2 *Similarly, we define $C^n[a,b]$ to be the set of all functions f defined on $[a,b]$ such that $f^{(n)}$ exists and is continuous everywhere in the interval (a,b).*

Theorem 1.3 (Rolle's Theorem) *If $f \in C[a,b]$, f' exists on (a,b), and $f(a) = f(b) = 0$, then there exists at least one $\xi \in (a,b)$ such that $f'(\xi) = 0$.*

Theorem 1.4 (Generalized Roll's Theorem) *Suppose $f \in C[a,b]$ is n times differentiable on (a,b). If $f(x) = 0$ for $x = x_0, x_1, \ldots, x_n$, where $x_i \in [a,b]$, are distinct, then there is a $\xi \in (a,b)$ such that $f^{(n)}(\xi) = 0$.*

Theorem 1.5 (Mean Value Theorem) *If $f \in C[a,b]$ and f' exists on (a,b), then for any $x, c \in [a,b]$, there exist a number ξ between x and c, such that*

$$f(x) = f(c) + f'(\xi)(x-c),$$

or, equivalently,

$$f'(\xi) = \frac{f(x)-f(c)}{x-c}.$$

Remarks 1.3 *Rolle's theorem is a special case of the Mean Value theorem.*

Theorem 1.6 (Intermediate Value Theorem) *If $f \in C[a,b]$ and η is some number between $f(a)$ and $f(b)$, then there exists $\xi \in (a,b)$ such that $f(\xi) = \eta$.*

Theorem 1.7 (Extreme Value Theorem) *If $f \in C[a,b]$, then there exist ξ and η in $[a,b]$ such that $f(\xi) \leq f(x) \leq f(\eta)$ for all $x \in [a,b]$. If, in addition, f is differentiable on (a,b), then ξ and η occur either at the end-points of $[a,b]$ or where f' is zero.*

Definition 1.6 (Riemann Integral) *Suppose f is a function defined on $[a,b]$. The Riemann integral of f on $[a,b]$ is the limit, if it exists,*

$$\int_a^b f(x)dx = \lim_{\max \triangle x_i \to 0} \sum_{i=1}^n f(z_i) \triangle x_i,$$

where $\triangle x_i = x_i - x_{i-1}$, $z_i \in [x_{i-1}, x_i]$ is arbitrary, and $a = x_0 \leq x_1 \leq \cdots \leq x_n = b$ is a partition of $[a,b]$.

Remarks 1.4 *In the case of equally spaced partition and $z_i = x_i$, one has*

$$\int_a^b f(x)dx = \lim_{n \to \infty} \frac{b-a}{n} \sum_{i=1}^n f(x_i),$$

where $x_i = a + i(b-1)/n$.

Theorem 1.8 (Mean Value Theorem for Integral) *If f is continuous on $[a,b]$ and g is integrable and does not change sign on $[a,b]$, then there exists $\xi \in (a,b)$, such that*

$$\int_a^b f(x)g(x)dx = f(\xi) \int_a^b g(x)dx.$$

In the special case when $g(x) \equiv 1$ on $[a,b]$, then one has

$$\int_a^b f(x)dx = f(\xi)(b-a).$$

1.1.2 Taylor's Theorem

Theorem 1.9 (Taylor's Theorem with Lagrange Formula) *Suppose $f \in C^n[a,b]$ and $f^{(n+1)}$ exists on (a,b). Then for any x and c in $[a,b]$, there exists a number ξ, depending on x, between x and c such that*

$$f(x) = P_n(x) + E_n(x), \tag{1.1}$$

where

$$P_n(x) = \sum_{k=0}^n \frac{1}{k!} f^{(k)}(c)(x-c)^k \tag{1.2}$$

is called the n-th Taylor polynomial for f centered at c, and

$$E_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-c)^{n+1} \tag{1.3}$$

is called the remainder term, error term, or truncation error associated with $P_n(x)$*. The infinite series by taking the limit of* $P_n(x)$ *as* $n \to \infty$ *is called the Taylor series for* $f(x)$*.*

Remarks 1.5 *The mean-value theorem is a special case of Tayler's theorem by taking* $n = 0$*.*

Theorem 1.10 (Taylor's Theorem with Integral Form) *If* $f \in C^{n+1}[a,b]$*, then for any* x *and* c *in* $[a,b]$*,*

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(c)(x-c)^k + R_n(x), \tag{1.4}$$

where

$$R_n(x) = \frac{1}{n!} \int_c^x f^{(n+1)}(t)(x-t)^n dt. \tag{1.5}$$

Theorem 1.11 (Other forms of Taylor's Theorem) *If* $f \in C^{n+1}[a,b]$*, then for any* $x \in (a,b)$ *and* $h \geq 0$ *in a neighborhood of* x*, there exists* ξ *such that*

$$f(x+h) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(x)h^k + \frac{1}{(n+1)!} f^{(n+1)}(\xi)h^{n+1}, \tag{1.6}$$

if $x < \xi < x+h$*, and*

$$f(x-h) = \sum_{k=0}^{n} (-1)^k \frac{1}{k!} f^{(k)}(x)h^k + \frac{(-1)^{n+1}}{(n+1)!} f^{(n+1)}(\xi)h^{n+1}, \tag{1.7}$$

if $x - h < \xi < x$*.*

Theorem 1.12 (Taylor's Theorem in Two Variables) *Suppose that* $f(x,y)$ *and all its partial derivatives of order less than or equal to* $n+1$ *are continuous on* $D = \{(x,y) | a \leq x \leq b, c \leq y \leq d\}$*, and let* $(x_0, y_0) \in D$*. For every* $(x,y) \in D$*, there exists* ξ *between* x *and* x_0 *and* μ *between* y *and* y_0 *with*

$$f(x,y) = P_n(x,y) + R_n(x,y), \tag{1.8}$$

where

$$\begin{aligned} P_n(x,y) &= f(x_0,y_0) + \left[(x-x_0)\frac{\partial f}{\partial x}(x_0,y_0) + (y-y_0)\frac{\partial f}{\partial y}(x_0,y_0)\right] \\ &+ \left[\frac{(x-x_0)^2}{2}\frac{\partial^2 f}{\partial x^2}(x_0,y_0) + (x-x_0)(y-y_0)\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0) \right. \\ &\left. + \frac{(y-y_0)^2}{2}\frac{\partial^2 f}{\partial y^2}(x_0,y_0)\right] + \cdots\cdots \\ &+ \left[\frac{1}{n!}\sum_{j=0}^{n}\binom{n}{j}(x-x_0)^{n-j}(y-y_0)^j \frac{\partial^n f}{\partial x^{n-j}\partial y^j}(x_0,y_0)\right] \end{aligned} \tag{1.9}$$

and

$$R_n(x,y) = \frac{1}{(n+1)!}\sum_{j=0}^{n+1}\binom{n+1}{j}(x-x_0)^{n+1-j}(y-y_0)^j\frac{\partial^{n+1} f}{\partial x^{n+1-j}\partial y^j}(\xi,\mu). \tag{1.10}$$

The function $P_n(x,y)$ is called the n-th Taylor polynomial in two variables for thefunction $f(x,y)$, and $R_n(x,y)$ is the remainder term associated with $P_n(x,y)$.

Example 1.1 *Consider the function* $f(x) = \ln x$ *in the interval* $[1,2]$. *Since* $f'(x) = x^{-1}$, $f''(x) = -x^{-2}$, $f'''(x) = 2x^{-3}, \cdots$ *and so on. In general,* $f^{(k)}(x) = (-1)^{k-1}(k-1)!x^{-k}$. *Hence, with* $c = 1$, *Taylor's theorem 1.9 gives,*

$$\ln x = \sum_{k=1}^{n}\frac{(-1)^{k-1}}{k}(x-1)^k + \frac{(-1)^n}{n+1}\xi^{-(n+1)}(x-1)^{n+1},$$

where $1 \le x \le 2$ *and* $1 < \xi < x$. *The error term is*

$$|E_n| = \frac{1}{n+1}\xi^{-(n+1)}(x-1)^{n+1} \le \frac{1}{n+1}(x-1)^{n+1},$$

since $\xi > 1$ *and* $\xi^{-(n+1)} < 1$. *To estimate* $\ln 2$ *with an error less than* 10^{-8}, *it would require*

$$|E_n| \le \frac{1}{n+1} \le 10^{-8}.$$

This means $n \ge 10^8 - 1$. *Similarly, to estimate* $\ln 1.5$ *with the same precision, it requires*

$$|E_n| \le \frac{1}{n+1}\left(\frac{1}{2}\right)^{n+1} \le 10^{-8}.$$

It is sufficient to attain the accuracy if $n \ge 26$. ■

Example 1.2 *Consider the function* $f(x) = \cos x$ *defined on* $\mathbb{R}$. *Since* $f \in C^\infty(\mathbb{R})$, *Taylor's theorem can be applied for any* $n > 0$. *With* $n = 2$ *and* $c = 0$, *we have*

$$f(x) = \cos x = 1 - \frac{1}{2}x^2 + \frac{1}{6}x^3 \sin\xi,$$

where $0 < \xi < x$. *To estimate* $\cos(0.01)$ *using the Taylor's theorem, one has*

$$\cos(0.01) = 1 - \frac{1}{2}(0.01)^2 + \frac{1}{6}(0.01)^3\sin\xi = 0.99995 + \frac{1}{6}\times 10^{-6}\sin\xi,$$

where $0 < \xi < 0.01$. *Since* $|\sin\xi| < 1$, *the error is*

$$|\cos(0.01) - 0.99995| \le \frac{1}{6}\times 10^{-6} \approx 0.167\times 10^{-6}.$$

This means that the approximation 0.99995 *matches at least the first five digits of* $\cos(0.01)$. *On the other hand, with* $n = 3$, *we also have*

$$f(x) = \cos x = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 \cos \xi.$$

The apporximation of $\cos(0.01)$ *with this formula remains the same, still* 0.99995, *but we now have much better accuracy assurance since the error is now*

$$|\frac{1}{24}x^4 \cos \xi| \leq \frac{1}{24}(0.01)^4 \approx 0.42 \times 10^{-9}.$$

This means that the approximation 0.99995 *matches at least the first eight digits when compared to the 11-digit accuracy* $\cos(0.01) = 0.99995000042$. ■

These examples illustrate the two objectives of numerical analysis. The first is to find approximation, which both Taylor polynomials provide. The second objective is to determine the accuracy of the approximation. In the previous example, the later analysis is much more informative, even though both gave the same approximation.

1.2 Review of Linear Algebra and Matrix Analysis

1.2.1 Matrices and Vectors

We let $\mathbb{R}^n$ denote the set of all real n-vectors (column vectors):

$$x \in \mathbb{R}^n \quad \Longleftrightarrow \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad x_i \in \mathbb{R}, \quad \forall\, i = 1, 2, \ldots, n,$$

in which x_i is referred as the i-th component of x. By convection, $\mathbb{R}^{n\times 1} = \mathbb{R}^n$ denotes the set of column vectors and $\mathbb{R}^{1\times n}$ the row vectors, $x \in \mathbb{R}^{1\times n}$, $x = [x_1, x_2, \ldots, x_n]$, $x_i \in \mathbb{R}$.

$\mathbb{R}^{m\times n}$ will denote the set of all $m \times n$ matrices:

$$A \in \mathbb{R}^{m\times n} \quad \Longleftrightarrow \quad A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

in which $a_{ij} \in \mathbb{R}$ is called the (i, j) component of A. Frequently, we use the notation

$$A = [a_1, a_2, \ldots, a_n],$$

in which

$$a_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \in \mathbb{R}^m$$

denotes the j-th column of A. Analogously, we use $\mathbb{C}^n$ and $\mathbb{C}^{m\times n}$ to denote the set of complex n-vectors and complex $m \times n$ matrices, respectively.

The transpose of matrix A is $A^T \in \mathbb{R}^{n\times m}$ and

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

For complex matrix $A \in \mathbb{C}^{m\times n}$, the Hermitian of A is

$$A^H = \begin{bmatrix} \bar{a}_{11} & \bar{a}_{21} & \cdots & \bar{a}_{m1} \\ \bar{a}_{12} & \bar{a}_{22} & \cdots & \bar{a}_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{a}_{1n} & \bar{a}_{2n} & \cdots & \bar{a}_{mn} \end{bmatrix},$$

where $a_{ij} \in \mathbb{C}$ and$\bar{a}_{ij}$ denotes the complex conjugate of a_{ij}.

Some basic matrix-vector operations include

- **dot-product:** for x, $y \in \mathbb{R}^n$,

$$x^T y = x_1 y_1 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i \ \in \mathbb{R}.$$

- **outer product:** for $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$,

$$xy^T = [x_i y_j] = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix} \in \mathbb{R}^{m\times n}.$$

- **matrix-vector multiplication:** for $A = [a_1, a_2, \ldots, a_n] \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$,

$$y = Ax = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n, \quad \text{or} \quad y_i = \sum_{k=1}^{n} a_{ik} x_k, \quad i = 1, 2, \ldots, n.$$

- **matrix-matrix multiplication:** for $A \in \mathbb{R}^{m\times p}$ and $B \in \mathbb{R}^{p\times n}$,

$$C = AB \in \mathbb{R}^{m\times n} \quad \text{with} \quad c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}.$$

- **saxpy:**

$$z = \alpha x + y, \qquad \text{or} \qquad z_i = \alpha x_i + y_i$$

with x, y, $z \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$.

- **outer-product update:**

$$A \leftarrow A + xy^T \qquad \text{or} \qquad a_{ij} \leftarrow a_{ij} + x_i y_j$$

with $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$.

- **Gaxpy:**

$$z = Ax + y \qquad \text{or} \qquad z_i = \sum_{k=1}^{n} a_{ik}x_k + y_i$$

with $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $z \in \mathbb{R}^m$.

The identity matrix of order n, $I_n = [\delta_{ij}]$, is a special matrix with entries

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The columns of I_n,

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \cdots, \qquad e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

are called the standard unit vectors. When the size of I_n is clear from the content, this matrix is generally written simply as I.

A square matrix $A \in \mathbb{R}^{n\times n}$ is said to be nonsingular (or invertible) if there exists $B \in \mathbb{R}^{n\times n}$ such that $BA = AB = I_n$, the $n \times n$ identity matrix. The matrix B is called the inverse of A and is denoted as A^{-1}.

When A and B are both n-by-n non-singular matrices, we have

$$(AB)^{-1} = B^{-1}A^{-1} \quad \text{and} \quad (A^{-1})^T = (A^T)^{-1} = A^{-T}.$$

Suppose $A \in \mathbb{R}^{n\times n}$ is nonsingular and A^{-1} is known. The the inverse $(A + xy^T)^{-1}$, where $x, y \in \mathbb{R}^n$, can be computed by the so-called Sherman-Morrison-Woodbury formula

$$(A + xy^T)^{-1} = A^{-1} - \frac{A^{-1}xy^TA^{-1}}{1 + y^TA^{-1}x}$$

if $1 + y^TA^{-1}x \neq 0$. The extension to rank-k formula for $U, V \in \mathbb{R}^{n\times k}$ is

$$(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + V^TA^{-1}U)^{-1}V^TA^{-1}$$

if $I + V^TA^{-1}U$ is nonsingular.

We list in the following some matrices which has either special structure or property.

Definition 1.7 *Let $A = [a_{ij}]$ be a square or rectangular matrix, A is called*

- *diagonal if* $a_{ij} = 0, \quad \forall\, i \neq j$;
- *tridiagonal if* $a_{ij} = 0$ *if* $|i - j| > 1$;
- *upper bidiagonal if* $a_{ij} = 0, \quad \forall\, i > j$ *or* $j > i + 1$;
- *upper triangular if* $a_{ij} = 0, \quad \forall\, i > j$;
- *strictly upper triangular if* $a_{ij} = 0, \quad \forall\, i \geq j$;
- *upper Hessenberg matrix if* $a_{ij} = 0, \quad \forall\, i > j + 1$;
- *symmetric if* $A^T = A$;
- *skew symmetric if* $A^T = -A$;
- *if positive if* $a_{ij} > 0, \quad \forall\, i,\ j$;
- *positive definite if* $x^T A x > 0, \quad \forall\, x \neq 0, \ \ x \in \mathbb{R}^n$;
- *non-negative definite or positive semi-definite if* $x^T A x \geq 0, \quad \forall\, x \in \mathbb{R}^n$;
- *indefinite if there exist* $x, y \in \mathbb{R}^n$ *such that* $(x^T A x)(y^T A y) < 0$;
- *idempotent if* $A^2 = A$.
- *nilpotent if* $A^k = 0$ *for some positive integer* k;
- *diagonal dominant if* $|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \ \ \forall\, i$;
- *orthogonal if* $A^T A = I$ *(that is,* $A^T = A^{-1}$*) when* $A \in \mathbb{R}^{n \times n}$ *and unitary if* $A^H A = I$ *when* $A \in \mathbb{C}^{n \times n}$;
- *normal if* $A^T A = A A^T$ *when* $A \in \mathbb{R}^{n \times n}$ *and* $A^H A = A A^H$ *when* $A \in \mathbb{C}^{n \times n}$.

1.2.2 Vector Space, Range Space, and Null Space

Definition 1.8 (Linearly Independent) *We say vectors $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$ are linearly independent if $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$*

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m = 0 \quad \Longrightarrow \quad \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0.$$

If there exists any $\alpha_k \neq 0$ but $\sum_{i=1}^{m} \alpha_i x_i = 0$, then $x_1, x_2, \ldots, x_m$ are said to be linearly dependent.

A subspace spanned by $x_1, x_2, \ldots, x_m$ is denoted and defined as

$$\operatorname{span}\{x_1, x_2, \ldots, x_m\} = \left\{ y \in \mathbb{R}^n \mid y = \sum_{i=1}^{m} \alpha_i x_i \text{for some} \alpha_i \in \mathbb{R} \right\}.$$

Suppose $S \subseteq \mathbb{R}^n$ is a subspace and $B = \{b_1, \ldots, b_m\} \subseteq S$ is a subset. If $b_1, \ldots, b_m$ are linearly independent and $\operatorname{span}(B) = S$, then B is called a basis for S and $\dim(S) = m$.

Definition 1.9 (Range and Null Space) *Let $A = [a_1, \ldots, a_n] \in \mathbb{R}^{m \times n}$, where $a_i \in \mathbb{R}^m$ are columns of A. The range of A is denoted and defined as*

$$\mathcal{R}(A) = \{y \in \mathbb{R}^m |\; y = Ax, \textit{ for some } x \in \mathbb{R}^n\} = \operatorname{span}\{a_1, a_2, \ldots, a_n\}, \tag{1.11}$$

and the null space of A is denoted and defined as

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n | Ax = 0\}. \tag{1.12}$$

Definition 1.10 (Rank and Nullity) *The rank of a matrix A, $\operatorname{rank}(A)$, is the dimension of $\mathcal{R}(A)$ and the nullity of A, $\operatorname{nullity}(A)$, is the dimension of $\mathcal{N}(A)$.*

Note 1.1 $\operatorname{rank}(A) = \operatorname{rank}(A^T)$ *is the maximal number of linearly independent columns (or rows) in A.*

Note 1.2 *Suppose $A \in \mathbb{R}^{m \times n}$ and $m \geq n$, then*

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n.$$

Note 1.3 *Suppose $A \in \mathbb{R}^{n \times n}$. Then*

$$\begin{aligned}
& A \text{ is non-singular (invertible, } A^{-1} \text{ exists)} \\
\iff & Ax = 0 \Rightarrow x = 0 \\
\iff & Ax = b \text{ has a unique solution} \\
\iff & \operatorname{rank}(A) = n \\
\iff & \det(A) \neq 0
\end{aligned}$$

1.2.3 Orthogonality

Definition 1.11 (Orthogonal and Orthonormal) *A set of vectors $x_1, \ldots, x_m \in \mathbb{R}^n$ are said to be orthogonal if*

$$x_i^T x_j = 0, \; \forall\, i \neq j,$$

and orthonormal if

$$x_i^T x_j = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Definition 1.12 *Two subspaces S_1 and S_2 of $\mathbb{R}^n$ are said to be orthogonal if $x^T y = 0$ for all $x \in S_1$ and $y \in S_2$.*

Definition 1.13 (Orthogonal Complement) *Suppose $S \subseteq \mathbb{R}^n$ is a subspace. The orthogonal complement of S is defined as*

$$S^\perp = \left\{ y \in \mathbb{R}^n \mid y^T x = 0,\ \forall\, x \in S \right\}$$

Definition 1.14 (Direct Sum) *Suppose U and V are subspaces of a vector space S. If each $s \in S$ can be written uniquely as a sum $u + v$, where $u \in U$ and $v \in V$, then we say that S is a direct sum of U and V and we write $S = U \oplus V$.*

Theorem 1.13 *Suppose $A \in \mathbb{R}^{m \times n}$. Then*

1. $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$;
2. $\mathcal{N}(A) = \mathcal{R}(A^T)^\perp$;
3. $\mathbb{R}^n = \mathcal{R}(A^T) \oplus \mathcal{N}(A)$.

Proof: For any $y \in \mathcal{R}(A)^\perp$, $y^T x = 0$, $\forall\, x \in \mathcal{R}(A)$. Since x is in $\mathcal{R}(A)$, there exists $z \in \mathbb{R}^n$ such that $x = Az$. Then

$$0 = y^T x = y^T (Az) = (A^T y)^T z.$$

Since $z = A^T y$ is arbitrary, it must be $A^T y = 0$. That is, $y \in \mathcal{N}(A^T)$. Hence

$$\mathcal{R}(A)^\perp \subseteq \mathcal{N}(A^T).$$

Conversely, suppose $y \in \mathcal{N}(A^T)$. Then $A^T y = 0$ and hence $(A^T y)^T x = y^T (Ax) = 0$ for any $x \in \mathbb{R}^n$. This means $y \in \mathcal{R}(A)^\perp$. Therefore

$$\mathcal{N}(A^T) \subseteq \mathcal{R}(A)^\perp.$$

■

1.3 Norms

1.3.1 Vector Norm Definition and Properties

Definition 1.15 *A vector norm is a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ satisfying the following conditions for all $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.*

1. $\|x\| \geq 0 \quad (\|x\| = 0 \Leftrightarrow x = 0)$;
2. $\|x + y\| \leq \|x\| + \|y\|$;
3. $\|\alpha x\| = |\alpha| \|x\|$.

Definition 1.16 *For $x \in \mathbb{R}^n$, some of the most frequently used vector norms are*

- **1-norm:**

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| = |x_1| + |x_2| + \ldots + |x_n|. \tag{1.13}$$

- **2-norm:**

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} = \left(|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2\right)^{1/2} = \sqrt{x^T x}. \tag{1.14}$$

- **∞-norm:**

$$\|x\|_\infty = \max_{1 \le i \le n} |x_i|. \tag{1.15}$$

- **p-norm:**

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} = \left(|x_1|^p + |x_2|^p + \ldots + |x_n|^p\right)^{1/p}. \tag{1.16}$$

Definition 1.17 (unit vector) *$x \in \mathbb{R}^n$ is called a unit vector if $\|x\| = 1$ with respect to some vector norm.*

Property 1.2 *For any $x, y \in \mathbb{R}^n$, the following two inequalities hold.*

- *Hölder inequality:*

$$|x^T y| \le \|x\|_p \|y\|_q, \quad \text{where} \quad \frac{1}{p} + \frac{1}{q} = 1.$$

- *Cauchy-Schwartz inequality:*

$$|x^T y| \le \|x\|_2 \|y\|_2.$$

Definition 1.18 *Two vector norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are equivalent if there exist constants $c_1, c_2 \in \mathbb{R}$ such that*

$$c_1 \|x\|_\alpha \le \|x\|_\beta \le c_2 \|x\|_\alpha$$

for any $x \in \mathbb{R}^n$.

In fact, all vector norms on $\mathbb{R}^n$ are equivalent, and it is easy to show the following equivalence properties for vector norms.

Property 1.3 *For all* $x \in \mathbb{R}^n$,

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2. \tag{1.17}$$
$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty. \tag{1.18}$$
$$\|x\|_\infty \leq \|x\|_1 \leq n\|x\|_\infty. \tag{1.19}$$

Definition 1.19 (absolute error and relative error) *Suppose* $x \in \mathbb{R}^n$ *is the exact solution of some problem and* $\tilde{x}$ *is an approximation to* x. *We define*

$$\textit{absolute error} = \|x - \tilde{x}\| \tag{1.20}$$

and

$$\textit{relative error} = \frac{\|x - \tilde{x}\|}{\|x\|}, \quad \textit{if } x \neq 0. \tag{1.21}$$

Definition 1.20 (convergence) *Suppose* $\{x_k\}$ *is a sequence of vectors in* $\mathbb{R}^n$. *We say* x_k *converges to* $x^\star$

$$x_k \longrightarrow x^\star \quad \textit{iff} \quad \lim_{k \to \infty} \|x_k - x^\star\| = 0,$$

for some vector norm.

1.3.2 Matrix Norm Definition and Properties

Definition 1.21 *A matrix norm is a function* $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ *satisfying the following conditions for all* $A, B \in \mathbb{R}^{m\times n}$ *and* $\alpha \in \mathbb{R}$.

1. $\|A\| \geq 0 \quad (\|A\| = 0 \Leftrightarrow A = 0)$;
2. $\|A + B\| \leq \|A\| + \|B\|$;
3. $\|\alpha A\| = |\alpha|\|A\|$.

Definition 1.22 *Some of the most frequently used matrix norms are*

- **Frobenius norm:**

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}. \tag{1.22}$$

- **2-norm:**

$$\|A\|_2 = \max_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2 = 1} \|Ax\|_2. \tag{1.23}$$

- **1-norm:**

$$\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|. \tag{1.24}$$

- **∞-norm:**

$$\|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|. \tag{1.25}$$

- ***p*-norm:**

$$\|A\|_p = \max_{\substack{x\in\mathbb{R}^n\\ x\neq 0}} \frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|_p=1} \|Ax\|_p. \tag{1.26}$$

Remarks 1.6 *Frobenius norm, 1-norm, and ∞-norm are easy to compute, but not the 2-norm or general p-norm.*

Theorem 1.14 *Suppose $A \in \mathbb{R}^{m\times n}$. Then there exists $z \in \mathbb{R}^n$, $\|z\|_2 = 1$, such that $A^TAz = \mu^2 z$, where $\mu = \|A\|_2$.*

Proof: Let $z \in \mathbb{R}^n$, $\|z\|_2 = 1$, be a unit vector that satisfies $\|A\|_2 = \|Az\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$. Define

$$g(x) = \frac{1}{2}\frac{x^TA^TAx}{x^Tx} = \frac{1}{2}\frac{\|Ax\|_2^2}{\|x\|_2^2} = \frac{1}{2}\left(\frac{\|Ax\|_2}{\|x\|_2}\right)^2,$$

for $x \in \mathbb{R}^n$. Then z is a maximizer of $g(x)$ which impliese $\bigtriangledown g(z) = 0$. Since

$$\bigtriangledown g(x) = \frac{(x^Tx)(A^TAx) - (x^TA^TAx)(x)}{(x^Tx)^2}.$$

Hence

$$\begin{aligned}
\bigtriangledown g(z) = 0 \quad &\Rightarrow \quad (z^Tz)(A^TAz) - (z^TA^TAz)z = 0 \\
&\Rightarrow \quad \|z\|_2^2(A^TAz) - \|Az\|_2^2 z = 0 \\
&\Rightarrow \quad A^TAz = \|A\|_2^2 z = \mu^2 z.
\end{aligned}$$

■

Remarks 1.7 *$\|A\|_2$ is the square root of the largest eigenvalue of A^TA. When A is symmetric, $\|A\|_2$ is the absolute value of the largest eigenvalue in magnitude.*

Definition 1.23 (submultiplicative property) *A matrix norm $\|\cdot\|$ is said to have the submultiplicative property if for any matrices $A \in \mathbb{R}^{m\times p}$ and $B \in \mathbb{R}^{p\times n}$ such that*

$$\|AB\| \le \|A\|\|B\|.$$

Not all matrix norms satisfy the submultiplicative property, e.g., define $\|A\|_\triangle = \max_{i,j} |a_{ij}|$, and let $A = B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. But we conly consider matrix norms that satisfy the submultiplicative property.

Property 1.4 *p-norm satisfies the submultiplicative property.*

$$\begin{aligned} \|Ax\|_p &\leq \|A\|_p\|x\|_p && (1.27) \\ \|AB\|_p &\leq \|A\|_p\|B\|_p && (1.28) \end{aligned}$$

Property 1.5 *All matrix norms satisfying the submultiplicative property are equivalent.*

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|A\|_2. \tag{1.29}$$

$$\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\|A\|_\infty. \tag{1.30}$$

$$\frac{1}{\sqrt{m}}\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\|A\|_1. \tag{1.31}$$

$$\max_{i,j} |a_{ij}| \leq \|A\|_2 \leq \sqrt{mn} \max_{i,j} |a_{ij}|. \tag{1.32}$$

Property 1.6 *Vector 2-norm is invariant under orthogonal transformation, that is, if Q is an n-by-n orthogonal matrix, then*

$$\|Qx\|_2 = \|x\|_2 \qquad \forall\ x \in \mathbb{R}^n. \tag{1.33}$$

Proof: Since

$$\|Qx\|_2^2 = (Qx)^T(Qx) = x^TQ^TQx = x^Tx = \|x\|_2^2.$$

∎

Property 1.7 *Matrix 2-norm and Frobenius norm are invariant under orthogonal transformation, that is, if Q is an n-by-n orthogonal matrix, then*

$$\begin{aligned} \|QA\|_2 &= \|A\|_2 && (1.34) \\ \|QA\|_F &= \|A\|_F && (1.35) \end{aligned}$$

for all $A \in \mathbb{R}^{n\times m}$.

Proof: Since

$$\|QA\|_2 = \max_{\|x\|_2=1} \|QAx\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \|A\|_2.$$

∎

Property 1.8 *If $A \in \mathbb{R}^{m\times n}$, then $\|A\|_2 \leq \sqrt{\|A\|_1\|A\|_\infty}$.*

Proof: From Theorem 1.14, there exists a vector $z \in \mathbb{R}^n$, $z \neq 0$, such that $\|A\|_2^2 z = A^T A z$. This implies

$$\|A\|_2^2 \|z\|_1 = \|A^T A z\|_1 \leq \|A^T\|_1 \|A\|_1 \|z\|_1 = \|A\|_\infty \|A\|_1 \|z\|_1.$$

Hence

$$\|A\|_2^2 \leq \|A\|_\infty \|A\|_1.$$

■

Theorem 1.15 *Suppose that* $A \in \mathbb{R}^{n \times n}$ *and* $\|\cdot\|$ *is a submultiplicative matrix norm. If* $\|A\| < 1$*, then* $I - A$ *is nonsingular and*

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k \tag{1.36}$$

with

$$\|(I - A)^{-1}\| \leq \frac{1}{1 - \|A\|}. \tag{1.37}$$

Proof: Suppose $I - A$ is singular. Then there exists an $x \in \mathbb{R}^n$, $x \neq 0$ such that $(I - A)x = 0$. This implies

$$Ax = x \quad \Longrightarrow \quad \|Ax\| = \|x\| \quad \Longrightarrow \quad \frac{\|Ax\|}{\|x\|} = 1 \quad \Longrightarrow \quad \|A\| = \max_{y \neq 0} \frac{\|Ay\|}{\|y\|} \geq 1.$$

This contradicts to the assumption $\|A\| < 1$. Hence $I - A$ must be nonsingular. Moreover

$$(I - A)\left(\sum_{k=0}^{N} A^k\right) = I - A^{N+1}.$$

Since $\|A\| < 1$ and $\|\cdot\|$ is a submultiplicative matrix norm, $\|A^N\| \leq \|A\|^N$ and $\lim_{N \to \infty} A^N = 0$. Therefore

$$(I - A)\left(\lim_{N \to \infty} \sum_{k=0}^{N} A^k\right) = I.$$

It follows that

$$(I - A)^{-1} = \lim_{N \to \infty} \sum_{k=0}^{N} A^k = \sum_{k=0}^{\infty} A^k$$

and

$$\|(I - A)^{-1}\| = \left\| \sum_{k=0}^{\infty} A^k \right\| \leq \sum_{k=0}^{\infty} \|A^k\| \leq \sum_{k=0}^{\infty} \|A\|^k = \frac{1}{1 - \|A\|}.$$

■

1.4 SVD: The Singular Value Decomposition

The singular value decomposition (SVD) is a matrix factorization whose computation is a step in many algorithms. Many problems of linear algebra can be better understood by considering the SVD.

Theorem 1.16 (Existence of SVD) *If $A \in \mathbb{R}^{m\times n}$, then there exists orthogonal matrices*

$$U = [u_1, u_2, \dots, u_m] \in \mathbb{R}^{m\times m} \quad \textit{and} \quad V = [v_1.v_2, \dots, v_n] \in \mathbb{R}^{n\times n}$$

such that

$$A = U\Sigma V^T, \tag{1.38}$$

where

$$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \dots, \sigma_p), \qquad p = \min(m, n),$$

with

$$\sigma_1 \ge \sigma_2 \ge \dots \ge \sigma_p \ge 0.$$

Proof: Let $\sigma_1 = \|A\|_2 \neq 0$. By the compactness arguement there exists $v_1 \in \mathbb{R}^n$, $\|v_1\|_2 = 1$ such that $\|Av_1\|_2 = \sigma_1$. Let $u_1 = \frac{1}{\sigma_1}Av_1$. Then $Av_1 = \sigma_1 u_1$, $\|u_1\|_2 = 1$, and $u_1^T Av_1 = \sigma_1$.

Let $V_1 = [v_1, \widehat{V}_1] \in \mathbb{R}^{n\times n}$ and $U_1 = [u_1, \widehat{U}_1] \in \mathbb{R}^{m\times m}$ be the orthogonal matrices using the extension of v_1 and u_1 to form orthonormal bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Note that this can always be done by Gram-Schmidt process. Then

$$\widehat{U}_1^T Av_1 = \widehat{U}_1^T(\sigma_1 u_1) = \sigma_1(\widehat{U}_1^T u_1) = 0$$

and

$$U_1^T AV_1 = \begin{bmatrix} u_1^T \\ \widehat{U}_1^T \end{bmatrix} A \left[v_1, \widehat{V}_1\right] = \begin{bmatrix} u_1^T Av_1 & u_1^T A\widehat{V}_1 \\ \widehat{U}_1^T Av_1 & \widehat{U}_1^T A\widehat{V}_1 \end{bmatrix} \equiv \begin{bmatrix} \sigma_1 & w^T \\ 0 & B \end{bmatrix} \equiv A_1.$$

Next we show that $w^T \equiv u_1^T A\widehat{V}_1 = 0$. Since

$$A_1 \begin{bmatrix} \sigma \\ w \end{bmatrix} = \begin{bmatrix} \sigma & w^T \\ 0 & B \end{bmatrix} \begin{bmatrix} \sigma \\ w \end{bmatrix} = \begin{bmatrix} \sigma^2 + w^T w \\ Bw \end{bmatrix},$$

hence

$$\begin{aligned} \|A_1\|_2 = \max \frac{\|A_1 x\|_2}{\|x\|_2} \quad &\ge \quad \frac{\left\| A_1 \begin{bmatrix} \sigma \\ w \end{bmatrix} \right\|_2}{\left\| \begin{bmatrix} \sigma \\ w \end{bmatrix} \right\|_2} = \frac{\left\| \begin{bmatrix} \sigma^2 + w^T w \\ Bw \end{bmatrix} \right\|_2}{\left\| \begin{bmatrix} \sigma \\ w \end{bmatrix} \right\|_2} \\ &= \quad \frac{((\sigma_1^2 + w^T w)^2 + w^T B^T Bw)^{1/2}}{(\sigma_1^2 + w^T w)^{1/2}} \\ &\ge \quad \frac{\left(\sigma_1^2 + w^T w\right)}{\left(\sigma_1^2 + w^T w\right)^{1/2}} = \left(\sigma_1^2 + w^T w\right)^{1/2} \ge \sigma_1. \end{aligned}$$

But

$$\|A_1\|_2 = \|U_1^T A V_1\|_2 = \|A\|_2 = \sigma_1.$$

Hence

$$w^T w = 0 \quad \Rightarrow \quad w = 0.$$

Therefore

$$U_1^T A V_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix} \quad \Rightarrow \quad A = U_1 \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix} V_1^T.$$

Using induction argument, there exist orthogonal matrices $U_2 \in \mathbb{R}^{(m-1)\times(m-1)}$ and $V_2 \in \mathbb{R}^{(n-1)\times(n-1)}$ such that $B = U_2 \Sigma_2 V_2^T$, where $\Sigma_2 \in \mathbb{R}^{(m-1)\times(n-1)}$ is diagonal. Then

$$\begin{aligned} A &= U_1 \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix} V_1^T \\ &= U_1 \begin{bmatrix} \sigma_1 & 0 \\ 0 & U_2 \Sigma_2 V_2^T \end{bmatrix} V_1^T \\ &= U_1 \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix}^T V_1^T \\ &\equiv U \Sigma V^T, \end{aligned}$$

where

$$U = U_1 \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix}, \quad V = V_1 \begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}.$$

This proves the theorem. ∎

The σ_i are called the singular values of A and the vectors u_i and v_i the i-th left singular vector and the i-th right singular vector, respectively. We usually use $\sigma_{max}(A)$ to denote the largest singular value of A and $\sigma_{min}(A)$ the smallest singular value of A.

Suppose $A \in \mathbb{R}^{m\times n}$ and $\text{rank}(A) = r$. Then

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \sigma_p = 0,$$

and A can be expressed as the summation of r rank-one matrices

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T. \tag{1.39}$$

The SVD gives complete information about the four fundamental subspaces associated with A:

$$\begin{aligned} \mathcal{R}(A) &= \text{span}\{u_1, u_2, \ldots, u_r\}, && (1.40) \\ \mathcal{N}(A) &= \text{span}\{v_{r+1}, v_{r+2}, \ldots, v_n\}, && (1.41) \\ \mathcal{R}(A^T) &= \text{span}\{v_1, v_2, \ldots, v_r\}, && (1.42) \\ \mathcal{N}(A^T) &= \text{span}\{u_{r+1}, u_{r+2}, \ldots, u_m\}. && (1.43) \end{aligned}$$

The matrix 2-norm and Frobenius norm can be characterized in terms of singular values.

$$\begin{aligned} \|A\|_2 &= \sigma_1, && (1.44) \\ \|A\|_F &= \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}. && (1.45) \end{aligned}$$

If we write

$$A = U\Sigma V^T = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^T,$$

where

$$\Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \in \mathbb{R}^{r\times r},$$

then

$$A^TA = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T\Sigma V^T = V \begin{bmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{bmatrix} V^T$$

and

$$AA^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T = U \begin{bmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{bmatrix} U^T.$$

Thus $\sigma_1^2, \dots, \sigma_r^2$ are the nonzero eigenvalues of A^TA and AA^T, and v_j and u_j are the corresponding eigenvectors. When A is symmetric, $A^TA = AA^T = A^2$, and has real eigenvalues λ_i, then the singular values are given by $\sigma_i = |\lambda_i|$ for $i = 1, \dots, r$. In principal, the SVD can be found from the eigenvalue decomposition of A^TA and AA^T. However, this does not lead to a stable algorithm for computing the SVD. Finally, when A is $n \times n$ and square, then $|\det(A)| = \sigma_1 \cdots \sigma_n$.