1

Mathematical Foundations

The basic requirements of this book are the fundamental knowledge of functions, basic calculus, and vector algebra. However, we will review here the most relevant fundamentals of functions, vectors, differentiation, and integration. Then, we will introduce some useful concepts such as eigenvalues, complexity, convexity, probability distributions, and optimality conditions.

1.1 Functions and Continuity

1.1.1 Functions

Loosely speaking, a function is a quantity (say y) which varies with another independent quantity or variable x in a deterministic way. For example, a simple quadratic function is

$$y = x^2. \tag{1.1}$$

For any given value of x, there is a unique corresponding value of y. By varying x smoothly, we can vary y in such a manner that the point (x, y) will trace out a curve on the x–y plane (see Figure 1.1). Thus, x is called the independent variable, and y is called the dependent variable or function. Sometimes, in order to emphasize the relationship as a function, we use $f(x)$ to express a generic function, showing that it is a function of x. This can also be written as $y = f(x)$.

The domain of a function is the set of numbers x for which the function $f(x)$ is valid (that is, $f(x)$ gives a valid value for a corresponding value of x). If a function is defined over a range $a \le x \le b$, we say its domain is $[a, b]$ that is called a closed interval. If both a and b are not included, we have $a < x < b$, which is denoted by (a, b), and we call this interval an open interval. If b is included, while a is not, we have $a < x \le b$, and we often write this half-open and half-closed interval as $(a, b]$. Thus, the domain of function $f(x) = x^2$ is the whole set of all real numbers $\mathbb{R}$, so we have

$$f(x) = x^2 \qquad (-\infty < x < +\infty). \tag{1.2}$$

Optimization Techniques and Applications with Examples, First Edition. Xin-She Yang.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

COPYRIGHTED MATERIAL

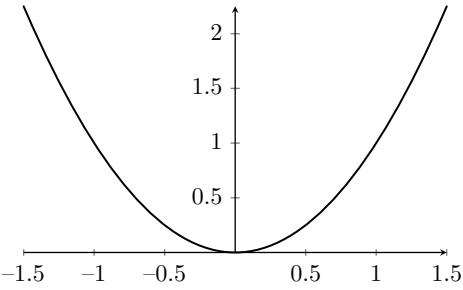

Figure 1.1 A simple quadratic function $y = x^2$.

Here, the notation ∞ means infinity. In this case, the domain of $f = x^2$ is all the real numbers, which can be simply written as $\mathbb{R}$, that is, $x \in \mathbb{R}$.

All the values that a function can take for a given domain form the range of the function. Thus, the range of $y = f(x) = x^2$ is $0 \leq y < +\infty$ or $[0, +\infty)$. Here, the closed bracket "[" means that the value (here 0) is included as part of the range, while the open round bracket ")" means that the value (here $+\infty$) is not part of the range.

1.1.2 Continuity

A function is called continuous if an infinitely small change δ of the independent variable x always lead to an infinitely small change in $f(x + \delta) - f(x)$. Alternatively, we can loosely view that the graph representing the function forms a single piece, unbroken curve. More formally, we can say that for any small change $\delta > 0$ of the independent variable in the domain, there is an $\epsilon > 0$ such that

$$|f(x + \delta) - f(x)| < \epsilon. \tag{1.3}$$

This is the continuity condition. Obviously, functions such as x, $|x|$, and x^2 are all continuous.

If a function does not satisfy the continuity condition, then the function is called discontinuous. For example, the Heaviside step function

$$H(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } 0 < 0, \end{cases} \tag{1.4}$$

is discontinuous at $x = 0$ as shown in Figure 1.2 where the solid dot means that $x = 0$ is included in the right branch $x \geq 0$, and the hollow dot means that it is not included. In this case, this function is called a right-continuous function.

1.1.3 Upper and Lower Bounds

For a given non-empty set $S \in \mathbb{R}$ of real numbers, we now introduce some important concepts such as the supremum and infimum. A number U is called

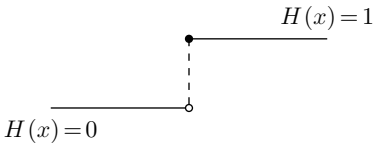

Figure 1.2 Discontinuity of the Heaviside step function at $x = 0$.

an upper bound for $\mathcal{S}$ if $x \leq U$ for all $x \in \mathcal{S}$. An upper bound β is said to be the least (or smallest) upper bound for $\mathcal{S}$, or the supremum, if $\beta \leq U$ for any upper bound U. This is often written as

$$\beta \equiv \sup_{x \in \mathcal{S}} x \equiv \sup \mathcal{S} \equiv \sup(\mathcal{S}). \tag{1.5}$$

All such notations are widely used in the literature of mathematical analysis. Here, "≡" denotes both equality and definition, which means "is identical to."

On the other hand, a number L is called a lower bound for $\mathcal{S}$ if $x \geq L$ for all x in $\mathcal{S}$ (that is, for all x, denoted by $\forall x \in \mathcal{S}$). A lower bound α is referred to as the greatest (or largest) lower bound if $\alpha \geq L$ for any lower bound L, which is written as

$$\alpha \equiv \sup_{x \in \mathcal{S}} x \equiv \inf \mathcal{S} \equiv \inf(\mathcal{S}). \tag{1.6}$$

In general, both the supremum β and the infimum α, if they exist, may or may not belong to $\mathcal{S}$.

Example 1.1 For example, any numbers greater than 5, say, 7.2 and 500 are an upper bound for the interval $-2 \leq x \leq 5$ or $[-2, 5]$. However, its smallest upper bound (or sup) is 5. Similarly, numbers such as -10 and -10^5 are lower bound of the interval, but -2 is the greatest lower bound (or inf). In addition, the interval $\mathcal{S} = [15, \infty)$ has an infimum of 15 but it has no upper bound. That is to say, its supremum does not exist, or $\sup \mathcal{S} \to \infty$.

There is an important completeness axiom which says that if a non-empty set $\mathcal{S} \in \mathbb{R}$ of real numbers is bounded above, then it has a supremum. Similarly, if a non-empty set of real numbers is bounded below, then it has an infimum.

Furthermore, the maximum for $\mathcal{S}$ is the largest value of all elements $s \in \mathcal{S}$, and often written as $\max(\mathcal{S})$ or $\max \mathcal{S}$, while the minimum, $\min(\mathcal{S})$ or $\min \mathcal{S}$, is the smallest value among all $s \in \mathcal{S}$. For the same interval $[-2, 5]$, the maximum of this interval is 5 which is equal to its supremum, while its minimum 5 is also equal to its infimum. Though the supremum and infimum are not necessarily part of the set $\mathcal{S}$, however, the maximum and minimum (if they exist) always belong to the set.

However, the concepts of supremum (or infimum) and maximum (or minimum) are not the same, and maximum/minimum may not always exist.

Example 1.2 For example, the interval $\mathcal{S} = [-2, 7)$ or $-2 \leq x < 7$ has the supremum of $\sup \mathcal{S} = 7$, but $\mathcal{S}$ has no maximum.

Similarly, the interval $(-10, 15]$ does not have a minimum, though its infimum is -10. Furthermore, the open interval $(-2, 7)$ has no maximum or minimum; however, its supremum is 7, and infimum is -2.

It is worth pointing out that the problems we will discuss in this book will always have at least a maximum or a minimum.

1.2 Review of Calculus

1.2.1 Differentiation

The gradient or first derivative of a function $f(x)$ at point x is defined by

$$f'(x) \equiv \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \lim_{\delta \to 0} \frac{f(x+\delta) - f(x)}{\delta}. \tag{1.7}$$

From the definition and basic function operations, it is straightforward to show that

$$(x^n)' = nx^{n-1} \quad (n = 1, 2, 3, \ldots). \tag{1.8}$$

In addition, for any two functions $f(x)$ and $g(x)$, we have

$$[af(x) + bg(x)]' = af'(x) + bg'(x), \tag{1.9}$$

where a, b are two real constants. Therefore, it is easy to show that

$$(x^3 - x + k)' = (x^3)' - x' + k' = 3x^2 - 1 + 0 = 3x^2 - 1, \tag{1.10}$$

where k is a constant. This means that a family of functions shifted by a different constant k will have the same gradient at the same point x, as shown in Figure 1.3.

Some useful differentiation rules are the product rule

$$[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x) \tag{1.11}$$

and the chair rule

$$\frac{\mathrm{d}f[g(x)]}{dx} = \frac{\mathrm{d}f(g)}{\mathrm{d}g} \cdot \frac{\mathrm{d}g(x)}{\mathrm{d}x}, \tag{1.12}$$

where $f(g(x))$ is a composite function, which means that f is a function of g, and g is a function of x.

Example 1.3 For example, from $[\sin(x)]' = \cos(x)$ and $(x^3)' = 3x^2$, we have

$$\frac{\mathrm{d}\sin(x^3)}{\mathrm{d}x} = \cos(x^3) \cdot (3x^2) = 3x^2 \sin(x^3).$$

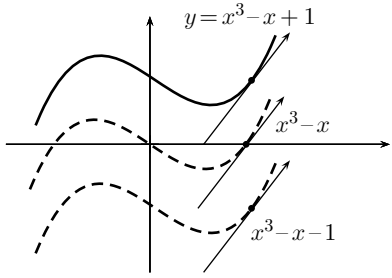

Figure 1.3 The gradients of a family of curves $y = x^3 - x + k$ (where $k = 0, \pm 1$) at any point x are the same $3x^2 - 1$.

Similarly, we have

$$\frac{\mathrm{d}\sin^n(x)}{\mathrm{d}x} = n\sin^{n-1}(x)\cdot\cos(x).$$

From the product rule (1.11), if we replace $g(x)$ by $1/g(x)$, we have

$$\frac{\mathrm{d}[g(x)^{-1}]}{\mathrm{d}x} = -1g^{-1-1}\cdot\frac{\mathrm{d}g(x)}{\mathrm{d}x} = -\frac{1}{[g(x)]^2}\frac{\mathrm{d}g(x)}{\mathrm{d}x} \tag{1.13}$$

and

$$\begin{aligned}\frac{\mathrm{d}[f(x)/g(x)]}{\mathrm{d}x} &= \frac{\mathrm{d}[f(x)g(x)^{-1}]}{\mathrm{d}x}\\ &= f'(x)g(x)^{-1} + f(x)\left\{\frac{-1}{[g(x)]^2}\frac{\mathrm{d}g(x)}{\mathrm{d}x}\right\} = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2},\end{aligned} \tag{1.14}$$

which is the well-known quotient rule. For example, we have

$$\begin{aligned}\frac{\mathrm{d}\tan(x)}{\mathrm{d}x} &= \frac{\mathrm{d}[\sin(x)/\cos(x)]}{\mathrm{d}x} = \frac{\cos(x)\sin'(x) - \sin(x)\cos'(x)}{\cos^2(x)}\\ &= \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)}.\end{aligned} \tag{1.15}$$

It is worth pointing out that a continuous function may not have well-defined derivatives. For example, the absolute or modulus function

$$f(x) = |x| = \begin{cases} x & \text{if } x \geq 0,\\ -x & \text{if } x < 0,\end{cases} \tag{1.16}$$

does not have a well-defined gradient at x because the gradient of $|x|$ is $+1$ if x approaches 0 from $x > 0$ (using notation 0^+). However, if x approaches 0 from $x < 0$ (using notation 0^-), the gradient of $|x|$ is -1. That is

$$\frac{\mathrm{d}|x|}{\mathrm{d}x}\Big|_{x\to 0^+} = +1, \quad \frac{\mathrm{d}|x|}{\mathrm{d}x}\Big|_{x\to 0^-} = -1. \tag{1.17}$$

In this case, we say that $|x|$ is not differentiable at $x = 0$. However, for $x \neq 0$, we can write

$$\frac{\mathrm{d}|x|}{\mathrm{d}x} = \frac{x}{|x|} \quad (x \neq 0). \tag{1.18}$$

Example 1.4 A nature extension of this is that for a function $f(x)$, we have

$$\frac{\mathrm{d}|f(x)|}{\mathrm{d}x} = \frac{f(x)}{|f(x)|}\frac{\mathrm{d}f(x)}{\mathrm{d}x}, \quad \text{if } f(x) \neq 0. \tag{1.19}$$

As an example, for $f(x) = |x^3|$, we have

$$\begin{aligned}\frac{\mathrm{d}|x^3|}{\mathrm{d}x} &= \frac{x^3}{|x^3|}\frac{\mathrm{d}x^3}{\mathrm{d}x} = \frac{x^3}{|x^3|}(3x^2) = \frac{3x^5}{|x^3|} \\ &= \frac{3x^5}{|x^2| \cdot |x|} = \frac{3x^5}{x^2|x|} = \frac{3x^3}{|x|}, \quad \text{if} \quad x \neq 0.\end{aligned}$$

Higher derivatives of a univariate real function can be defined as

$$f''(x) \equiv \frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} \equiv \frac{\mathrm{d}f'(x)}{\mathrm{d}x}, \; f'''(x) = [f''(x)]', \; \ldots, \; f^{(n)}(x) = \frac{\mathrm{d}^n f(x)}{\mathrm{d}x^n}, \tag{1.20}$$

for all positive integers ($n = 1, 2, \ldots$).

Example 1.5 The first, second, and third derivatives of $f(x) = x\mathrm{e}^{-x}$ are

$$f'(x) = x'\mathrm{e}^{-x} + x(\mathrm{e}^{-x})' = \mathrm{e}^{-x} + x(-1\mathrm{e}^{-x}) = \mathrm{e}^{-x} - x\mathrm{e}^{-x},$$

$$f''(x) = [\mathrm{e}^{-x} - x\mathrm{e}^{-x}]' = x\mathrm{e}^{-x} - 2\mathrm{e}^{-x},$$

and

$$f'''(x) = [x\mathrm{e}^{-x} - 2\mathrm{e}^{-x}]' = -x\mathrm{e}^{-x} + 3\mathrm{e}^{-x}.$$

For a continuous function $f(x)$, if its first derivatives are well defined at every point in the whole domain, the function is called differentiable or a continuously differential function. A continuously differentiable function is said to be class C^1 if its first derivative exists and is continuous. Similarly, a function is said to be class C^2 if its both first and second derivatives exist, and are continuous. This can be extended to class C^k in a similar manner. If the derivatives of all orders (all positive integers) everywhere in its domain, the function is called smooth.

It is straightforward to check that $f(x) = x$, $\sin(x)$, $\exp(x)$, and $x\mathrm{e}^{-x}$ are all smooth functions, but $|x|$ and $|x|\mathrm{e}^{-|x|}$ are not smooth. Some of these functions are shown in Figure 1.4.

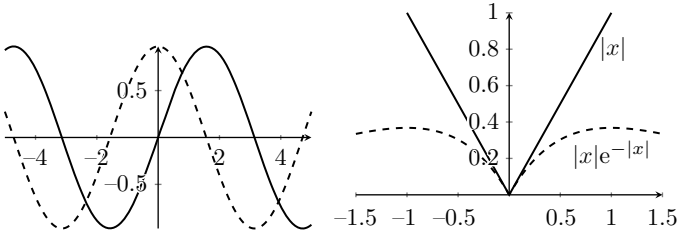

Figure 1.4 Smooth functions (left) and non-smooth (but continuous) functions (right).

1.2.2 Taylor Expansions

In numerical methods and some mathematical analysis, series expansions make some calculations easier. For example, we can write the exponential function e^x as a series about $x_0 = 0$ as

$$e^x = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \cdots + \alpha_n x^n. \tag{1.21}$$

Now let us try to determine these coefficients. At $x = 0$, we have

$$e^0 = 1 = \alpha_0 + \alpha_1 \times 0 + \alpha_2 \times 0^2 + \cdots + \alpha_n \times 0^n = \alpha_0, \tag{1.22}$$

which gives $\alpha_0 = 1$. In order to reduce the power or order of the expansion so that we can determine the next coefficient, we first differentiate both sides of Eq. (1.21) once; we have

$$e^x = \alpha_1 + 2\alpha_2 x + 3\alpha_3 x^2 + \cdots + n\alpha_n x^{n-1}. \tag{1.23}$$

By setting again $x = 0$, we have

$$e^0 = 1 = \alpha_1 + 2\alpha_2 \times 0 + \cdots + n\alpha_n \times 0^{n-1} = \alpha_1, \tag{1.24}$$

which gives $\alpha_1 = 1$. Similarly, differentiating it again, we have

$$e^x = (2 \times 1) \times \alpha_2 + 3 \times 2\alpha_3 x + \cdots + n(n-1)\alpha_n x^{n-2}. \tag{1.25}$$

At $x = 0$, we get

$$e^0 = (2 \times 1) \times \alpha_2 + 3 \times 2\alpha_3 \times 0 + \cdots + n(n-1)\alpha_n \times 0^{n-2} = 2\alpha_2, \tag{1.26}$$

or $\alpha_2 = 1/(2 \times 1) = 1/2!$. Here, $2! = 2 \times 1$ is the factorial of 2. In general, the factorial $n!$ is defined as $n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$.

Following the same procedure and differentiating it n times, we have

$$e^x = n!\alpha_n, \tag{1.27}$$

and $x = 0$ leads to $\alpha_n = 1/n!$. Therefore, the final series expansion can be written as

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n + \cdots, \tag{1.28}$$

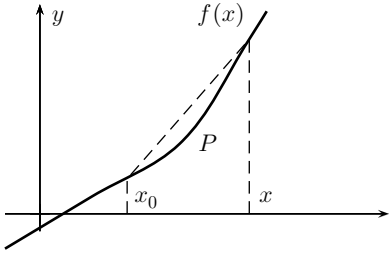

Figure 1.5 Expansion and approximations for $f(x) = f(x_0 + h)$, where $h = x - x_0$.

which is an infinite series. Obviously, we can follow a similar process to expand other functions. We have seen here the importance of differentiation and derivatives.

If we know the value of $f(x)$ at x_0, we can use some approximations in a small interval $h = x - x_0$ (see Figure 1.5). Following the same idea as Eq. (1.21), we can first write the approximation in the following general form:

$$f(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots + a_n(x - x_0)^n, \tag{1.29}$$

and then try to figure out the unknown coefficients $a_i (i = 0, 1, 2, \ldots)$. For the above approximation to be valid at $x = x_0$, we have

$$f(x_0) = a_0 + 0 \quad \text{(all the other terms are zeros)}, \tag{1.30}$$

so that $a_0 = f(x_0)$.

Now let us first take the first derivative of Eq. (1.29),

$$f'(x) = 0 + a_1 + 2a_2(x - x_0) + \cdots + na_n(x - x_0)^{n-1}. \tag{1.31}$$

By setting $x = x_0$, we have

$$f'(x_0) = 0 + a_1 + 0 + \cdots + na_n \times 0, \tag{1.32}$$

which gives

$$a_1 = f'(x_0). \tag{1.33}$$

Similarly, we differentiate Eq. (1.29) twice with respect to x and we have

$$f''(x) = 0 + 0 + a_2 \times 2 \times 1 + \cdots + n(n-1)a_n(x - x_0)^2. \tag{1.34}$$

Setting $x = x_0$, we have

$$f''(x_0) = 2!a_2, \quad \text{or} \quad a_2 = \frac{f''(x_0)}{2!}. \tag{1.35}$$

Following the same procedure, we have

$$a_3 = \frac{f'''(x_0)}{3!}, \quad a_4 = \frac{f''''(x_0)}{4!}, \quad \ldots, \quad a_n = \frac{f^{(n)}(x_0)}{n!}. \tag{1.36}$$

Thus, we finally obtain

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \frac{f'''(a)}{3!}(x - x_0)^3 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n, \tag{1.37}$$

which is the well-known Taylor series.

In a special case when $x_0 = 0$ and $h = x - x_0 = x$, the above Taylor series becomes zero centered, and such expansions are traditionally called Maclaurin series

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(n)}}{n!}x^n + \cdots, \tag{1.38}$$

named after mathematician Colin Maclaurin.

In theory, we can use as many terms as possible, but in practice, the series converges very quickly and only a few terms are sufficient. It is straightforward to verify that the exponential series for e^x is identical to the results given earlier. Now let us look at other examples.

Example 1.6 Let us expand $f(x) = \sin x$ about $x_0 = 0$. We know that

$$f'(x) = \cos x, \quad f''(x) = -\sin x, \quad f'''(x) = -\cos x, \quad \ldots,$$

or $f'(0) = 1$, $f''(0) = 0$, $f'''(0) = -1$, $f''''(0) = 0, \ldots$, which means that

$$\sin x = \sin 0 + xf'(0) + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots,$$

where the angle x is in radians.

For example, we know that $\sin 30° = \sin(\pi/6) = 1/2$. We now use the expansion to estimate it for $x = \pi/3 = 0.523\,598$,

$$\sin\frac{\pi}{6} \approx \frac{\pi}{6} - \frac{(\pi/6)^3}{3!} + \frac{(\pi/6)^5}{5!} \approx 0.523\,599 - 0.023\,92 + 0.000\,032\,8 \approx 0.500\,002\,132\,6,$$

which is very close to the true value $1/2$.

If we continue the process to infinity, we then reach the infinite power series and the error $f^{(n)}(0)x^n/n!$ becomes negligibly small if the series converges. For example, some common series are

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots + x^n + \cdots, \quad x \in (-1, 1), \tag{1.39}$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots, \quad \cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots, \quad x \in \mathbb{R}, \tag{1.40}$$

$$\tan(x) = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \cdots, \quad x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \tag{1.41}$$

and

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \cdots, \quad x \in (-1, 1]. \tag{1.42}$$

As an exercise, we leave the reader to prove the above series.

1.2.3 Partial Derivatives

For multivariate functions, we can define the partial derivatives with respect to an independent variable by assuming that other independent variables are constants. For example, for a function $f(x, y)$ with two independent variables, we can define

$$\frac{\partial f(x,y)}{\partial x} \equiv \frac{\partial f}{\partial x} \equiv \frac{\partial f}{\partial x}\Big|_y = \lim_{\delta \to 0, y=\text{constant}} \frac{f(x+\delta, y) - f(x,y)}{\delta}, \tag{1.43}$$

and

$$\frac{\partial f(x,y)}{\partial y} \equiv \frac{\partial f}{\partial y} \equiv \frac{\partial f}{\partial y}\Big|_x = \lim_{\delta \to 0, x=\text{constant}} \frac{f(x, y+\delta) - f(x,y)}{\delta}. \tag{1.44}$$

Similar to the ordinary derivatives, partial derivative operations are also linear.

Example 1.7 For $f(x, y) = x^2 + y^3 + 3xy^2$, its partial derivatives are

$$\frac{\partial f}{\partial x} = \frac{\partial x^2}{\partial x} + \frac{\partial y^3}{\partial x} + \frac{\partial (3xy^2)}{\partial x} = 2x + 0 + 3y^2 = 2x + 3y^2,$$

where we have treated y as a constant and also used the fact that $\mathrm{d}y/\mathrm{d}x = 0$ because x and y are both independent variables. Similarly, we have

$$\frac{\partial f}{\partial y} = \frac{\partial x^2}{\partial y} + \frac{\partial y^3}{\partial y} + \frac{\partial (3xy^2)}{\partial y} = 0 + 3y^2 + 3x(2y) = 3y^2 + 6xy.$$

We can define higher-order derivatives as

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \quad \frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right), \tag{1.45}$$

and

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right). \tag{1.46}$$

Let us revisit the previous example.

Example 1.8 For $f(x, y) = x^2 + y^3 + 3xy^2$, we have

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial (2x + 3y^2)}{\partial x} = 2 + 0 = 2, \quad \frac{\partial^2 f}{\partial y^2} = \frac{\partial (3y^2 + 6xy)}{\partial y} = 6y + 6x.$$

In addition, we have

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial(3y^2 + 6xy)}{\partial x} = 0 + 6y = 6y,$$

or

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial(2x + 3y^2)}{\partial y} = 0 + 3(2y) = 6y,$$

which shows that

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

Other higher-order partial derivatives can be defined in a similar manner.

1.2.4 Lipschitz Continuity

A very important concept related to optimization and convergence analysis is the Lipschitz continuity of a function $f(x)$,

$$|f(x_1) - f(x_2)| \leq L|x_1 - x_2|, \tag{1.47}$$

for any x_1 and x_2 in the domain of $f(x)$. Here, $L \geq 0$ is called the Lipschitz constant or the modulus of uniform continuity, which is independent of x_1 and x_2. This is equivalent to that case that the absolute derivative is finite, that is

$$\frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} \leq L < \infty, \tag{1.48}$$

which limits the rate of change of function. This means the small change in the independent variable (input) can lead to the arbitrarily small change in the function (output). However, when the Lipschitz constant is sufficiently large, a small change in x could lead to a much larger change in $f(x)$, but this Lipschitz estimate is an upper bound and the actual change can be much smaller. Therefore, any function with a finite or bounded first derivative is Lipschitz continuous. For example, $\sin(x)$, $\cos(x)$, x and x^2 are all Lipschitz, while the binary step function $H(x) = 1$ if $x \geq 0$ (otherwise $H(x) = 0$ if $x < 0$) is not Lipschitz at $x = 0$.

For example, function $f(x) = 3x^2$ is Lipschitz continuous in the domain $Q = [-5, 5]$. For any $x_1, x_2 \in Q$, we have

$$|f(x_1) - f(x_1)| = |3x_1^2 - 3x_2^2| = 3|(x_1^2 - x_2^2)| = 3|x_1 + x_2| \cdot |x_1 - x_2|. \tag{1.49}$$

Since $|x_1 + x_2| \leq |x_1| + |x_2| \leq 5 + 5 = 10$, we have

$$|f(x_1) - f(x_2)| \leq 30|x_1 - x_2|, \tag{1.50}$$

which means that the Lipschitz constant is 30. It is worth pointing out that $L = 30$ is the just one lower value. We can also use $L = 50$ or $L = 100$ such as

$$|f(x_1) - f(x_2) \leq 100|x_1 - x_2|. \tag{1.51}$$

Though Lipschitz constant should not depend on x_1 and x_2, it may depend on the size of the domain when we try to derive this constant.

1.2.5 Integration

Integration is a reverse operation of differentiation. For a univariate function $f(x)$, the meaning of its integral over a finite interval $[a, b]$ is the area enclosed by the curve $f(x)$ and the x-axis

$$I = \int_a^b f(x)\mathrm{d}x, \tag{1.52}$$

where a and b are the integration limits, and $f(x)$ is called the integrand. It is worth pointing out that the area above the x-axis is considered as positive, while the area under the x-axis is negative. If there is a function $F(x)$ whose first derivative is $f(x)$ (i.e. $F'(x) = f(x)$) in the same interval $[a, b]$, we can write

$$\int_a^b f(x)\mathrm{d}x = F(x)\Big|_a^b = F(b) - F(a). \tag{1.53}$$

The above integral is a definite integral with fixed integral limits. If there are no specific integration limits involved in the integral, we can, in general, write it as

$$\int f(x)\mathrm{d}x = F(x) + C, \tag{1.54}$$

which is an indefinite integral and C is the unknown integration constant. This integral constant comes from the fact that a function or curve shifted by a constant will have the same gradient.

One of the integration limits (or both) can be infinite.

Example 1.9 For example, we have

$$\int_0^\infty \mathrm{e}^{-x}\mathrm{d}x. \tag{1.55}$$

Since $[-\mathrm{e}^{-x}]' = \mathrm{e}^{-x}$, we have $F(x) = -\mathrm{e}^{-x}, f(x) = \mathrm{e}^{-x}$, and $F'(x) = f(x)$. Thus, we get

$$\int_0^\infty \mathrm{e}^{-x}\mathrm{d}x = [-\mathrm{e}^{-x}]\Big|_0^\infty = (-\mathrm{e}^{-\infty}) - (-\mathrm{e}^{-0}) = -0 - (-1) = 1. \tag{1.56}$$

A very useful formula that can be derived from the product rule of differentiation $[u(x)v(x)]' = u'(x)v(x) + u(x)v(x)$ is the integration by parts

$$\int \frac{\mathrm{d}[u(x)v(x)]}{\mathrm{d}x}\mathrm{d}x = \int [u'(x)v(x) + u(x)v'(x)]\mathrm{d}x$$
$$= \int v(x)u'(x)\mathrm{d}x + \int u(x)v'(x)\mathrm{d}x, \tag{1.57}$$

which leads to

$$\int u'(x)v(x)\mathrm{d}x = u(x)v(x) - \int u(x)v'(x)\mathrm{d}x. \tag{1.58}$$

Ignoring the integration constants, the above formula can be written in a more compact form as

$$\int v\mathrm{d}u = uv - \int u\mathrm{d}v. \tag{1.59}$$

Similarly, its corresponding definite integral becomes

$$\int_a^b v\mathrm{d}u = (uv)\Big|_a^b - \int_a^b u\mathrm{d}v. \tag{1.60}$$

Let us look at an example.

Example 1.10 To evaluate the integral

$$I = \int_0^\infty x\mathrm{e}^{-x}\mathrm{d}x,$$

we have $v(x) = x$ and $u'(x) = \mathrm{e}^{-x}$, which leads to

$$\mathrm{d}v = \mathrm{d}x, \quad v'(x) = 1, \quad u(x) = -\mathrm{e}^{-x}.$$

From the formula of integration by parts, we have

$$I = \int_0^\infty x\mathrm{e}^{-x}\mathrm{d}x = \left[x(-\mathrm{e}^{-x})\right]_0^\infty - \int_0^\infty 1 \cdot (-\mathrm{e}^{-x})\mathrm{d}x$$
$$= -(\infty)\mathrm{e}^{-\infty} - [-0\mathrm{e}^{-0}] + \int_0^\infty \mathrm{e}^{-x} = 0 + 0 + \int_0^\infty \mathrm{e}^{-x}\mathrm{d}x = 1,$$

where we have used the earlier result in Eq. (1.56) and the fact that

$$\lim_{K\to\infty} K\mathrm{e}^{-K} \to 0.$$

The multiple integral of a multivariate function can be defined in a similar manner. For example, for a function of $f(x, y)$ of two independent variables x and y, its double integral can be defined as

$$I = \iint f(x, y)\mathrm{d}x\mathrm{d}y. \tag{1.61}$$

In a rectangular domain $D = [a, b] \times [c, d]$ (that is $a \le x \le b$ and $c \le y \le d$), the following integral means the volume enclosed by the surface $f(x, y)$ over the rectangular domain. We have

$$I = \int\int_D f(x, y)\mathrm{d}x\mathrm{d}y = \int_c^d \left[\int_a^b f(x, y)\mathrm{d}x\right] \mathrm{d}y, \tag{1.62}$$

which is the same as

$$I = \int_a^b \left[\int_c^d f(x, y)\mathrm{d}y\right] \mathrm{d}x, \tag{1.63}$$

due to Fubini's theorem that is valid when

$$\int\int_D |f(x, y)|\mathrm{d}x\mathrm{d}y < \infty. \tag{1.64}$$

As integration is not quite relevant to most optimization techniques, we will not discuss integration any further. We will introduce more whenever needed in later chapters. In the rest of the chapter, we will review the vector algebra and eigenvalues of matrices before we move onto the introduction of optimality conditions.

1.3 Vectors

Loosely speaking, a vector is a quantity with a magnitude and a direction in practice. However, mathematically speaking, a vector can be represented by a set of ordered scalars or numbers. For example, a three-dimensional vector in the Cartesian coordinates can be written as

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{1.65}$$

where a_1, a_2, a_3 (or x, y, z) are its three components, along x-, y-, and z-axes, respectively. A vector is usually denoted in a lowercase boldface. Here, we write the component as a column vector. Alternatively, a vector can be equally represented by a row vector in the form:

$$\boldsymbol{a} = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix} = \begin{pmatrix} x & y & z \end{pmatrix}. \tag{1.66}$$

A column vector can be converted into a row vector by a simple transpose (using notation T as a superscript to denote this operation) or vice versa. We have

$$\begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} x & y & z \end{pmatrix}. \tag{1.67}$$

The magnitude or length of a three-dimensional vector is its Cartesian norm

$$|\boldsymbol{a}| = \sqrt{a_1^2 + a_2^2 + a_3^2} = \sqrt{x^2 + y^2 + z^2}. \tag{1.68}$$

1.3.1 Vector Algebra

In general, a vector in an n-dimensional space ($n \geq 1$) can be written as a column vector

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{1.69}$$

or a row vector

$$\boldsymbol{x} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}. \tag{1.70}$$

Its length can be written as

$$||\boldsymbol{x}|| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}, \tag{1.71}$$

which is the Euclidean norm.

The addition or substraction of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are the addition or substraction of their corresponding components, that is

$$\boldsymbol{u} \pm \boldsymbol{v} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \pm \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} u_1 \pm v_1 \\ u_2 \pm v_2 \\ \vdots \\ u_n \pm v_n \end{pmatrix}. \tag{1.72}$$

The dot product, also called the inner product, of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined as

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} \equiv \boldsymbol{u} \cdot \boldsymbol{v} = \sum_{i=1}^{n} u_i v_i = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n. \tag{1.73}$$

1.3.2 Norms

For an n-dimensional vector $\boldsymbol{x}$, we can define a p-norm or L_p-norm (also L^p-norm) as

$$||\boldsymbol{x}||_p \equiv \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p} = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \qquad (p > 0). \tag{1.74}$$

Obviously, the Cartesian norm or length is an L_2-norm

$$||\boldsymbol{x}||_2 = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \tag{1.75}$$

Three most widely used norms are $p = 1, 2$, and ∞. When $p = 2$, it becomes the Cartesian L_2-norm as discussed above. When $p = 1$, the L_1-norm is given by

$$||\boldsymbol{x}||_1 = |x_1| + |x_2| + \cdots + |x_n|. \tag{1.76}$$

For $p = \infty$, it becomes

$$||\boldsymbol{x}||_\infty = \max\{|x_1|, |x_2|, \ldots, |x_n|\} = x_{\max}, \tag{1.77}$$

which is the largest absolute component of $\boldsymbol{x}$. This is because

$$\begin{aligned} ||\boldsymbol{x}||_\infty &= \lim_{p\to\infty} \left(\sum_{i=1}^{p} |x_i|^p \right)^{1/p} = \lim_{p\to\infty} \left(|x_{\max}|^p \sum_{i=1}^{n} \left| \frac{x_i}{x_{\max}} \right|^p \right)^{1/p} \\ &= x_{\max} \lim_{p\to\infty} \left(\sum_{i=1}^{n} \left| \frac{x_i}{x_{\max}} \right| \right)^{1/p} = x_{\max}, \end{aligned} \tag{1.78}$$

where we have used the fact that $|x_i/x_{\max}| < 1$ (except for one component, say, $|x_k| = x_{\max}$). Thus, $\lim_{p\to\infty} |x_i/x_{\max}|^p \to 0$ for all $i \neq k$. Thus, the sum of all ratio terms is 1. That is

$$\left(\lim_{p\to\infty} \left| \frac{x_i}{x_{\max}} \right|^p \right)^{1/p} = 1. \tag{1.79}$$

In general, for any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in the same space, we have the following equality:

$$||\boldsymbol{u}||_p + ||\boldsymbol{v}||_p \geq ||\boldsymbol{u} + \boldsymbol{v}||_p \quad (p \geq 0). \tag{1.80}$$

Example 1.11 For two vectors $\boldsymbol{u} = [1 \quad 2 \quad 3]^{\mathrm{T}}$ and $\boldsymbol{v} = [1 \quad -2 \quad -1]^{\mathrm{T}}$, we have

$$\begin{aligned} &\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = 1 \times 1 + 2 \times (-2) + 3 \times (-1) = -6, \\ &||\boldsymbol{u}||_1 = |1| + |2| + |3| = 6, \quad ||\boldsymbol{v}||_1 = |1| + |-2| + |-1| = 4, \\ &||\boldsymbol{u}||_2 = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}, \quad ||\boldsymbol{v}||_2 = \sqrt{1^2 + (-2)^2 + (-1)^2} = \sqrt{6}, \\ &||\boldsymbol{u}||_\infty = \max\{|1|, |2|, |3|\} = 3, \quad ||\boldsymbol{v}||_\infty = \max\{|1|, |-2|, |-1|\} = 2, \end{aligned}$$

and

$$\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v} = \begin{bmatrix} 1+1 & 2+(-2) & 3+(-1) \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 2 & 0 & 2 \end{bmatrix}^{\mathrm{T}}$$

whose norms are

$$||\boldsymbol{w}||_1 = |2| + |0| + |2| = 2, \quad ||\boldsymbol{w}||_\infty = \max\{|2|, |0|, |2|\} = 2,$$

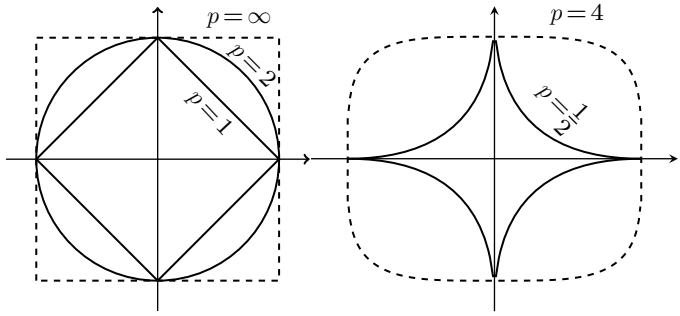

Figure 1.6 Different p-norms for $p = 1, 2$, and ∞ (left) as well as $p = 1/2$ and $p = 4$ (right).

$$||\boldsymbol{w}||_2 = \sqrt{2^2 + 0^2 + 2^2} = \sqrt{8}.$$

Using the above values, it is straightforward to verify that

$$||\boldsymbol{u}||_p + ||\boldsymbol{v}||_p \geq ||\boldsymbol{u} + \boldsymbol{v}||_p \quad (p = 1, 2, \infty).$$

1.3.3 2D Norms

To get a clearer picture about the differences between different norms, we now focus on the vectors in the two-dimensional (2D) Cartesian coordinates (x, y). For $\boldsymbol{u} = (x, y)^{\mathrm{T}}$, we have

$$||\boldsymbol{u}||_p = (|x|^p + |y|^p)^{1/p} \quad (p > 0). \tag{1.81}$$

Obviously, in special cases of $p = 1, 2, \infty$, we have

$$||\boldsymbol{u}||_1 = |x| + |y|, \quad ||\boldsymbol{u}||_2 = \sqrt{x^2 + y^2}, \quad ||\boldsymbol{u}||_\infty = \max\{|x|, |y|\}. \tag{1.82}$$

In order to show the main differences, we can let $-1 \leq x, y \leq 1$ and thus the 2-norm becomes a unit circle $x^2 + y^2 = 1$. All the other norms with different p values can be plotted around this unit circle, as shown in Figure 1.6.

1.4 Matrix Algebra

1.4.1 Matrices

A matrix is a rectangular array of numbers such as

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 7 & 8 \\ -1 & 3.7 \end{pmatrix}. \tag{1.83}$$

Matrix $\boldsymbol{A}$ has two rows and three columns, thus its size is said to be 2×3 or 2 by 3. The element in the first row and first column is $a_{11} = 1$, and the element

in the second row and third column is $a_{23} = 6$. Similarly, matrix $\boldsymbol{B}$ has a size of 2×2.

It is customary to use a boldface uppercase letter to represent a matrix, while its element is written in a corresponding lowercase letter. Thus, a matrix $\boldsymbol{A}$ of size $m \times n$ can, in general, be written as

$$\boldsymbol{A} = [a_{ij}] = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad [a_{ij}] \in \mathbb{R}^{m \times n}, \tag{1.84}$$

where $1 \leq i \leq m, 1 \leq j \leq n$. It is worth pointing out that a vector can be considered as a special case of matrices, and thus matrices are the natural extension of vectors. Here, we assume that all the numbers are real numbers, that is $a_{ij} \in \mathbb{R}$, which is most relevant to the contents in this book. In general, the entries in a matrix can be complex numbers. Here, we have used $\mathbb{R}^{m \times n}$ to denote the fact that all the elements in $\boldsymbol{A}$ span a space with a dimensionality of $m \times n$.

The transpose or transposition of an $m \times n$ matrix $\boldsymbol{A} = [a_{ij}]$ is obtained by turning columns into rows and vice versa. This operation is denoted by T or $\boldsymbol{A}^{\mathrm{T}}$. That is

$$\boldsymbol{A}^{\mathrm{T}} = [a_{ij}]^{\mathrm{T}} = [a_{ji}] = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}, \quad [a_{ji}] \in \mathbb{R}^{n \times m}. \tag{1.85}$$

In a special case when $m = n$, the matrix becomes a square matrix. If a transpose of a square matrix $\boldsymbol{A}$ is equal to itself, the matrix is said to be symmetric. That is

$$\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A} \quad \text{or} \quad a_{ij} = a_{ji}, \tag{1.86}$$

for all $1 \leq i, j \leq n$.

For a square matrix $\boldsymbol{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$, its diagonal elements are $a_{ii} (i = 1, 2, \ldots, n)$. The trace of the matrix is the sum of all its diagonal elements

$$\mathrm{tr}(\boldsymbol{A}) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}. \tag{1.87}$$

A special and useful square matrix $[a_{ij}]$ is the identity matrix where the diagonal elements are one ($a_{ii} = 1$) and all other elements are zero ($a_{ij} = 0$ if $i \neq j$). That is

$$\boldsymbol{I} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}. \tag{1.88}$$

Addition or subtraction of two matrices is possible only if they are of the same size. For example, if $\boldsymbol{A} = [a_{ij}]$ and $\boldsymbol{B} = [b_{ij}]$ where $1 \le i \le m$ and $1 \le j \le n$, we have

$$\begin{aligned}\boldsymbol{A} \pm \boldsymbol{B} &= [a_{ij}] \pm [b_{ij}] = [a_{ij} \pm b_{ij}] \\ &= \begin{pmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \cdots & a_{1n} \pm b_{1n} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \cdots & a_{2n} \pm b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} \pm b_{m1} & a_{m2} \pm b_{m2} & \cdots & a_{mn} \pm b_{mn} \end{pmatrix}. \end{aligned} \tag{1.89}$$

Example 1.12 For matrices $\boldsymbol{A} = \begin{pmatrix} 2 & 3 & -1 \\ 1 & 2 & 5 \end{pmatrix}$, $\boldsymbol{B} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$, and $\boldsymbol{D} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 0 \end{pmatrix}$ we have

$$\begin{aligned}\boldsymbol{A} + \boldsymbol{D} &= \begin{pmatrix} 2 & 3 & -1 \\ 1 & 2 & 5 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 2+1 & 3+1 & -1+1 \\ 1+2 & 2+(-1) & 5+0 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 0 \\ 3 & 1 & 5 \end{pmatrix}. \end{aligned}$$

The transpose of these matrices are

$$\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} 2 & 1 \\ 3 & 2 \\ -1 & 5 \end{pmatrix}, \quad \boldsymbol{D}^{\mathrm{T}} = \begin{pmatrix} 1 & 2 \\ 1 & -1 \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{B}^{\mathrm{T}} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} = \boldsymbol{B},$$

which means that the square matrix $\boldsymbol{B}$ is symmetric. In addition, the trace of $\boldsymbol{B}$ is $\mathrm{tr}(\boldsymbol{B}) = 1 + 3 = 4$.

The multiplication of two matrices requires a special condition that the number of columns of the first matrix must be equal to the number of rows of the second matrix. If $\boldsymbol{A}$ has an $m \times n$ matrix and $\boldsymbol{B}$ is an $n \times p$ matrix, then the production $\boldsymbol{C} = \boldsymbol{AB}$ is an $m \times p$ matrix. The element c_{ij} is obtained by the dot product of the ith row of $\boldsymbol{A}$ and the jth column of $\boldsymbol{B}$. That is

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}. \tag{1.90}$$

Let us look at an example.

Example 1.13 For $\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & -1 \\ -2 & 5 \end{pmatrix}$, we have $\boldsymbol{C} = \boldsymbol{AB}$ so that

$$c_{11} = \begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = 1 \times 1 + 2 \times (-2) = -3,$$

$$c_{12} = \begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 5 \end{pmatrix} = 1 \times (-1) + 2 \times 5 = 9,$$

$$c_{21} = \begin{pmatrix} 3 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = -5, \quad c_{22} = \begin{pmatrix} 3 & 4 \end{pmatrix} \begin{pmatrix} -1 \\ 5 \end{pmatrix} = 17,$$

$$c_{31} = \begin{pmatrix} 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = -7, \quad c_{32} = \begin{pmatrix} 5 & 6 \end{pmatrix} \begin{pmatrix} -1 \\ 5 \end{pmatrix} = 25.$$

Thus, we have

$$\boldsymbol{C} = \boldsymbol{AB} = \begin{pmatrix} -3 & 9 \\ -5 & 17 \\ -7 & 25 \end{pmatrix}.$$

However, $\boldsymbol{BA}$ does not exist because it is not possible to carry out the multiplication. In general, $\boldsymbol{AB} \neq \boldsymbol{BA}$ even if they exist, which means that matrix multiplication is not commutative.

It is straightforward to show that

$$(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}} \boldsymbol{A}^{\mathrm{T}}. \tag{1.91}$$

Example 1.14 Let us revisit the previous example. We know that

$$\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}, \quad \boldsymbol{B}^{\mathrm{T}} = \begin{pmatrix} 1 & -2 \\ -1 & 5 \end{pmatrix}.$$

Thus, we have

$$\begin{aligned} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{A}^{\mathrm{T}} &= \begin{pmatrix} 1 & -2 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \\ &= \begin{pmatrix} -3 & -5 & -7 \\ 9 & 17 & 25 \end{pmatrix} = \begin{pmatrix} -3 & 9 \\ -5 & 17 \\ -7 & 25 \end{pmatrix} = (\boldsymbol{AB})^{\mathrm{T}}. \end{aligned}$$

For a square matrix $\boldsymbol{A}$ and an identity matrix $\boldsymbol{I}$ of the same size, it is straightforward to check that

$$\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}. \tag{1.92}$$

For a square matrix of size $n \times n$, if there exists another unique matrix $\boldsymbol{B}$ of the same size satisfying

$$\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}, \tag{1.93}$$

then $\boldsymbol{B}$ is called the inverse matrix of $\boldsymbol{A}$. Here, $\boldsymbol{I}$ is an $n \times n$ identity matrix. In this case, we often denote $\boldsymbol{B} = \boldsymbol{A}^{-1}$, which means

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}. \tag{1.94}$$

In general, $\boldsymbol{A}^{-1}$ may not exist or be unique. One useful test condition is the determinant to be introduced next.

In addition, in a special case when the inverse of $\boldsymbol{A}$ is the same as its transpose $\boldsymbol{A}^{\mathrm{T}}$ (i.e. $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}$), then $\boldsymbol{A}$ is said to be orthogonal, which means

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}, \quad \boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}. \tag{1.95}$$

It is easy to check that the rotation matrix

$$\boldsymbol{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{1.96}$$

is orthogonal because $\boldsymbol{R}^{-1} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \boldsymbol{R}^{\mathrm{T}}$.

1.4.2 Determinant

The determinant of a square matrix $\boldsymbol{A}$ is just a number, denoted by $\det(\boldsymbol{A})$. In the simplest case, for a 2×2 matrix, we have

$$\det(\boldsymbol{A}) = \det \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc. \tag{1.97}$$

For a 3×3 matrix, we have

$$\begin{aligned} \det(\boldsymbol{A}) &= \det \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= a_{11} \det \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \det \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \det \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22}). \end{aligned} \tag{1.98}$$

In general, the determinant of a matrix can be calculated using a recursive formula such as the Leibniz formula or the Laplace expansion with the adjugate matrices. Interested readers can refer to more advanced literature on this topic.

A square matrix $\boldsymbol{A}$ can have a unique inverse matrix if $\det(\boldsymbol{A}) \neq 0$. Otherwise, the matrix is called singular and not invertible. For an invertible 2×2 matrix

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{1.99}$$

its inverse can be conveniently calculated by

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \quad ad - bc \neq 0. \tag{1.100}$$

As an exercise, we leave the reader to show that this is true.

1.4.3 Rank of a Matrix

The rank of a matrix $\boldsymbol{A}$ is a useful concept, and it is the maximum number of linearly independent columns or rows. For example, the rank of matrix

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 0 \\ 2 & 3 & 3 \end{pmatrix} \tag{1.101}$$

is 2 because there first two rows are independent, and the third row is the sum of the first two rows. We can write it as

$$\text{rank}(\boldsymbol{A}) = 2. \tag{1.102}$$

In some specialized literature, the number of linearly independent rows of $\boldsymbol{A}$ is called row rank, while the number of linearly independent columns if called column rank. It can be proved that the row rank is always equal to the column rank for the same matrix. In general, for an $m \times n$ matrix $\boldsymbol{A}$, we have

$$\text{rank}(\boldsymbol{A}) \leq \min\{m, n\}. \tag{1.103}$$

When the equality holds (i.e. $\text{rank}(\boldsymbol{A}) = \min\{m, m\}$), the matrix is said to be full rank. Thus, the following matrices

$$\boldsymbol{C} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 2 & 2 \\ 2 & 3 & 4 & 5 \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 3 & 7 \end{pmatrix} \tag{1.104}$$

are all full rank matrices because their ranks are 3 and 2, respectively.

For an $n \times n$ matrix, it becomes a full rank matrix if its rank is n. A useful full rank test is that the determinant of $\boldsymbol{A}$ is not zero. That is $\det(\boldsymbol{A}) \neq 0$. For example, matrix

$$\boldsymbol{B} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 0 \\ 2 & 3 & 7 \end{pmatrix} \tag{1.105}$$

is a full rank matrix because $\text{rank}(\boldsymbol{B}) = 3$ and $\det(\boldsymbol{B}) = -4$.

There are a few methods such as the Gauss elimination can be used to compute the rank of a matrix. Readers can refer to more advanced literature on this topic.

1.4.4 Frobenius Norm

Similar to the norms for vectors, there are also various ways to define norms for a matrix. For a matrix of size $m \times n$, the Frobenius norm is defined by

$$||\boldsymbol{A}||_{\mathrm{F}} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}, \tag{1.106}$$

which is equivalent to

$$||\boldsymbol{A}||_{\mathrm{F}} = \sqrt{\mathrm{tr}(A^{\mathrm{T}}A)} = \sqrt{\mathrm{diag}(A^{\mathrm{T}}A)}. \tag{1.107}$$

The maximum absolute column sum norm is defined by

$$||\boldsymbol{A}||_1 = \max_{1\le j\le n}\sum_{i=1}^{m}|a_{ij}|. \tag{1.108}$$

Similarly, the maximum absolute row sum norm is defined by

$$||\boldsymbol{A}||_\infty = \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|. \tag{1.109}$$

Example 1.15 For $\boldsymbol{A} = \begin{pmatrix} 1 & -2 & 3 \\ -5 & 0 & 7 \end{pmatrix}$, we have

$$||\boldsymbol{A}||_1 = \max\{|1| + |-5|, |-2| + |0|, |3| + |7|\} = \max\{6, 2, 10\} = 10,$$

$$||\boldsymbol{A}||_\infty = \max\{|1| + |-2| + |3|, |-5| + |0| + |7|\} = \max\{6, 12\} = 12,$$

and

$$||\boldsymbol{A}||_{\mathrm{F}} = \sqrt{|1|^2 + |-2|^2 + |3|^2 + |-5|^2 + |0|^2 + |7|^2} = \sqrt{88}.$$

Other norms can also be defined for different applications.

1.5 Eigenvalues and Eigenvectors

An eigenvalue λ of a square matrix $\boldsymbol{A}$ is defined by

$$\boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{u}, \tag{1.110}$$

where a nonzero eigenvector $\boldsymbol{u}$ exists for a corresponding λ. An $n \times n$ matrix can have at most n different eigenvalues and thus n corresponding eigenvectors.

Multiplying the above Eq. (1.110) by an identity $n \times n$ matrix $\boldsymbol{I}$, we have

$$\boldsymbol{IAu} = \boldsymbol{I}\lambda\boldsymbol{u} = (\lambda\boldsymbol{I})\boldsymbol{u}, \tag{1.111}$$

which becomes

$$\boldsymbol{IAu} - (\lambda\boldsymbol{I})\boldsymbol{u} = (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{u} = 0, \tag{1.112}$$

where we have used $\boldsymbol{IA} = \boldsymbol{A}$. In order to obtain a non-trial solution $\boldsymbol{u} \neq 0$, it is thus required that the matrix $\boldsymbol{A} - \lambda\boldsymbol{I}$ is not invertible. In other words, its determinant should be zero. That is

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0, \tag{1.113}$$

which is equivalent to a polynomial of order n. Such a polynomial is called the characteristic polynomial of $\boldsymbol{A}$. All the eigenvalues form a set, called the spectrum of matrix $\boldsymbol{A}$.

Example 1.16 The eigenvalue of $\boldsymbol{A} = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$ can be calculated by

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \det \begin{vmatrix} a - \lambda & b \\ b & a - \lambda \end{vmatrix} = 0,$$

which leads to

$$(a - \lambda)^2 - b^2 = 0,$$

or

$$a - \lambda = \pm b.$$

Thus, we have

$$\lambda_1 = a + b, \quad \lambda_2 = a - b.$$

For example, if $\boldsymbol{A} = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}$, we have

$$\lambda_1 = 5, \quad \lambda_2 = -1.$$

The trace of $\boldsymbol{A}$ is $\mathrm{tr}(\boldsymbol{A}) = 2 + 2 = 4$. Here, the sum of both eigenvalues are $\lambda_1 + \lambda_2 = 5 + (-1) = 4$, which means that $\mathrm{tr}(\boldsymbol{A}) = \lambda_1 + \lambda_2$.

In general, if a square matrix has n different eigenvalues $\lambda_i (i = 1, 2, \ldots, n)$, we have

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i. \tag{1.114}$$

Let us look at another example.

Example 1.17 For $\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ -3 & 8 \end{pmatrix}$, its eigenvalues can be obtained by

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \det \begin{vmatrix} 1-\lambda & 2 \\ -3 & 8-\lambda \end{vmatrix} = 0,$$

which gives

$$(1-\lambda)(8-\lambda) - 2 \times (-3) = 0,$$

or

$$\lambda^2 - 9\lambda + 14 = (\lambda - 2)(\lambda - 7) = 0.$$

Thus, the two eigenvalues are

$$\lambda_1 = 2, \quad \lambda_2 = 7.$$

The eigenvector $\boldsymbol{u} = (a \;\; b)^{\mathrm{T}}$ corresponding to $\lambda_1 = 2$ can be obtained by the original definition

$$\boldsymbol{A}\boldsymbol{u} = \lambda_1 \boldsymbol{u},$$

or

$$\begin{pmatrix} 1 & 2 \\ -3 & 8 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 2 \begin{pmatrix} a \\ b \end{pmatrix}.$$

This is equivalent to two equations

$$\begin{cases} a + 2b = 2a, \\ -3a + 8b = 2b. \end{cases}$$

Both equations give $a = 2b$, which means that they are not linear independent because one can be obtained by the other via some minor algebraic manipulations. This means that we can determine the direction of the eigenvector, not the magnitude uniquely. In some textbooks, it is assumed that the magnitude of an eigenvector should be one. That is $||\boldsymbol{u}||_2 = 1 = \sqrt{a^2 + b^2}$. However, in many textbooks and many software packages, they usually assume that the first component is 1, which makes subsequent calculations much easier. Thus, we can set $a = 1$, thus $b = 1/2$. The eigenvector for eigenvalue 2 becomes

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}.$$

Similarly, the eigenvector $\boldsymbol{u}_2 = [c \;\; d]^{\mathrm{T}}$ for $\lambda_2 = 7$ can be obtained by

$$\begin{pmatrix} 1 & 2 \\ -3 & 8 \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = 7 \begin{pmatrix} c \\ d \end{pmatrix}.$$

As both equations lead to $d = 3c$, we impose $c = 1$, which means $d = 3$. Thus, the eigenvector $\boldsymbol{u}_2$ for $\lambda_2 = 7$ is

$$\boldsymbol{u}_2 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

It is worth pointing out that the eigenvectors $-\boldsymbol{u}_1$ and $-\boldsymbol{u}_2$ for λ_1 and λ_2, respectively, are equally valid.

In addition, in some textbooks and many software packages, the unity of the eigenvectors is used, instead of setting the first component as 1. In this case, the above vectors are multiplied by a normalization or scaling constant (usually the length or magnitude). Therefore, in the above example, the eigenvectors can also be written equivalently as

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \boldsymbol{u}_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 3 \end{pmatrix}. \tag{1.115}$$

A useful theorem is that if a square matrix is real and symmetric $\boldsymbol{A}^T = \boldsymbol{A} \in \mathbb{R}^{n\times n}$, all its eigenvalues are real, and eigenvectors corresponding to distinct eigenvalues are orthogonal. That is, if $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are two eigenvectors for $\lambda_1 \neq \lambda_2$, respectively, we have $\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_2 = \boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = 0$.

However, eigenvalues for real symmetric matrices may not be distinct. For example, both eigenvalues of $\boldsymbol{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ are 1, and they are not distinct.

1.5.1 Definiteness

A square symmetric matrix $\boldsymbol{A}$ (i.e. $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}$) is said to be positive definite if all its eigenvalues are strictly positive ($\lambda_i > 0$, where $i = 1, 2, \ldots, n$). By multiplying both sides of $\boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{u}$ by $\boldsymbol{u}^{\mathrm{T}}$, we have

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u} = \boldsymbol{u}^{\mathrm{T}}\lambda\boldsymbol{u} = \lambda\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}, \tag{1.116}$$

which leads to

$$\lambda = \frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}}. \tag{1.117}$$

Since $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = ||\boldsymbol{u}||_2^2 > 0$, this means that

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u} > 0, \qquad \text{if } \lambda > 0. \tag{1.118}$$

In fact, for any vector $\boldsymbol{v}$, the following relationship holds:

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v} > 0. \tag{1.119}$$

For $\boldsymbol{v}$ can be a unit vector, all the diagonal elements of $\boldsymbol{A}$ should be strictly positive as well. If the equal sign is included in the definition, we have semidefiniteness. That is, $\boldsymbol{A}$ is called positive semidefinite if $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u} \geq 0$, and negative semidefinite if $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u} \leq 0$ for all $\boldsymbol{u}$.

If all the eigenvalues are nonnegative or $\lambda_i \geq 0$, then the matrix is positive semi-definite. If all the eigenvalues are nonpositive or $\lambda_i \leq 0$, then the matrix is negative semidefinite. In general, an indefinite matrix can have both positive and negative eigenvalues. Furthermore, the inverse of a positive definite matrix is also positive definite. For a linear system $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{f}$, if $\boldsymbol{A}$ is positive definite, the system can be solved more efficiently by matrix decomposition methods.

Let us look at an example.

Example 1.18 For matrices

$$\boldsymbol{A} = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 7 & 3 \\ 3 & 7 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix},$$

the eigenvalues of $\boldsymbol{A}$ are 1 and 5, which are both positive. Thus, $\boldsymbol{A}$ is positive definite. Similarly, $\boldsymbol{B}$ is also positive definite because its two eigenvalues are 4 and 10. However, $\boldsymbol{C}$ is indefinite because one of its eigenvalues is negative (-1) and the other eigenvalue is positive (5).

The definiteness of matrices can be useful to determine if a multivariate function has a local maximum or minimum. It is also useful to see if an expression can be written as a quadratic form.

1.5.2 Quadratic Form

Quadratic forms are widely used in optimization, especially in convex optimization and quadratic programming. Loosely speaking, a quadratic form is a homogenous polynomial of degree 2 of n variables. For example, $3x^2 + 10xy + 7y^2$ is a binary quadratic form, while $x^2 + 2xy + y^2 - y$ is not.

For a real $n \times n$ symmetric matrix $\boldsymbol{A}$ and a vector $\boldsymbol{u}$ of n elements, their combination

$$Q = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u} \tag{1.120}$$

is called a quadratic form. Since $\boldsymbol{A} = [a_{ij}]$, we have

$$\begin{aligned} Q = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u} &= \sum_{i=1}^{n}\sum_{j=1}^{n} u_i a_{ij} u_j = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} u_i u_j \\ &= \sum_{i=1}^{n} a_{ii} u_i^2 + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} a_{ij} u_i u_j. \end{aligned} \tag{1.121}$$

Example 1.19 For the symmetric matrix $\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$ and $\boldsymbol{u} = \begin{pmatrix} u_1 & u_2 \end{pmatrix}^{\mathrm{T}}$, we have

$$\begin{aligned} \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u} &= \begin{pmatrix} u_1 & u_2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \\ &= \begin{pmatrix} u_1 & u_2 \end{pmatrix} \begin{pmatrix} u_1 + 2u_2 \\ 2u_1 + 5u_2 \end{pmatrix} = 3u_1^2 + 10u_1u_2 + 7u_2^2. \end{aligned}$$

In fact, for a binary quadratic form $Q(x, y) = ax^2 + bxy + cy^2$, we have

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = (\alpha + \beta)x^2 + (\alpha + 2\beta + \gamma)xy + (\beta + \gamma)y^2.$$

If this is equivalent to $Q(x, y)$, it requires that

$$\alpha + \beta = a, \qquad \alpha + 2\beta + \gamma = b, \qquad \beta + \gamma = c,$$

which leads to $a + c = b$. This means that not all arbitrary quadratic functions $Q(x, y)$ are quadratic form.

If $\boldsymbol{A}$ is real symmetric, its eigenvalues λ_i are real, and the eigenvectors $\boldsymbol{v}_i$ of distinct eigenvalues λ_i are orthogonal to each other. Therefore, we can write $\boldsymbol{u}$ using the eigenvector basis and we have

$$\boldsymbol{u} = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i. \tag{1.122}$$

In addition, $\boldsymbol{A}$ becomes diagonal in this basis. That is

$$\boldsymbol{A} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}. \tag{1.123}$$

Subsequently, we have

$$\boldsymbol{A}\boldsymbol{u} = \sum_{i=1}^{n} \alpha_i \boldsymbol{A}\boldsymbol{v}_i = \sum_{i=1}^{n} \lambda_i \alpha_i \boldsymbol{v}_i, \tag{1.124}$$

which means that

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u} = \sum_{j=1}^{n} \sum_{i=1}^{n} \lambda_i \alpha_j \alpha_i \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{v}_i = \sum_{i=1}^{n} \lambda_i \alpha_i^2, \tag{1.125}$$

where we have used the fact that $\boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{v}_i = 1$ (by normalizing the eigenvectors so as to have a magnitude of unity).

1.6 Optimization and Optimality

Optimization is everywhere, from engineering design and business planning to artificial intelligence and industries. After all, time and resources are limited, and optimal use of such valuable resources is crucial. In addition, designs of products have to maximize the performance, sustainability, energy efficiency, and to minimize the costs and wastage. Therefore, optimization is specially important for engineering applications, business planning, and industries.

1.6.1 Minimum and Maximum

One of the simplest optimization problems is to find the minimum of a function such as $f(x) = x^2$ in the real domain. As x^2 is always nonnegative, it is easy to guess that the minimum occurs at $x = 0$.

From basic calculus, we know that, for a given curve described by $f(x)$, its gradient $f'(x)$ describes the rate of change. When $f'(x) = 0$, the curve has a horizontal tangent at that particular point. This means that it becomes a point of special interest. In fact, the maximum or minimum of a curve can only occur at

$$f'(x_*) = 0, \tag{1.126}$$

which is a critical condition or stationary condition. The solution x_* to this equation corresponds to a stationary point and there may be multiple stationary points for a given curve.

In order to see if it is a maximum or minimum at $x = x_*$, we have to use the information of its second derivative $f''(x)$. In fact, $f''(x_*) > 0$ corresponds to a minimum, while $f''(x_*) < 0$ corresponds to a maximum. Let us see a concrete example.

Example 1.20 To find the minimum of $f(x) = x^2 e^{-x^2}$, we have the stationary condition $f'(x) = 0$ or

$$f'(x) = 2x \times e^{-x^2} + x^2 \times (-2x)e^{-x^2} = 2(x - x^3)e^{-x^2} = 0.$$

As $e^{-x^2} > 0$, we have

$$x(1 - x^2) = 0, \quad \text{or} \quad x = 0, \quad \text{and} \quad x = \pm 1.$$

The second derivative is given by

$$f''(x) = 2e^{-x^2}(1 - 5x^2 + 2x^4),$$

which is an even function with respect to x.

So at $x = \pm 1$, $f''(\pm 1) = 2[1 - 5(\pm 1)^2 + 2(\pm 1)^4]e^{-(\pm 1)^2} = -4e^{-1} < 0$. Thus, there are two maxima that occur at $x_* = \pm 1$ with $f_{\max} = e^{-1}$. At $x = 0$, we have $f''(0) = 2 > 0$, thus the minimum of $f(x)$ occurs at $x_* = 0$ with $f_{\min}(0) = 0$.

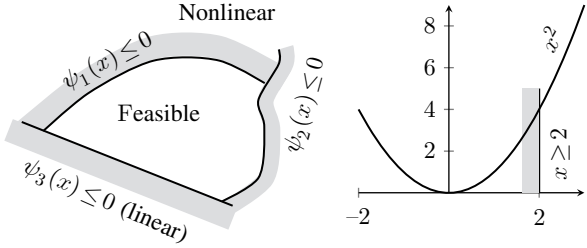

Figure 1.7 Feasible domain with nonlinear inequality constraints $\psi_1(x)$ and $\psi_2(x)$ (left) as well as a linear inequality constraint $\psi_3(x)$. An example with an objective of $f(x) = x^2$ subject $x \geq 2$ (right).

In mathematical programming, there are many important concepts, and we will first introduce the concepts of feasible solutions, optimality criteria, strong local optima, and weak local optima.

1.6.2 Feasible Solution

A point $\boldsymbol{x}$ which satisfies all the constraints is called a feasible point and thus it is a feasible solution to the problem. The set of all feasible points is called the feasible region (see Figure 1.7).

For example, we know that the domain $f(x) = x^2$ consists of all the real numbers. If we want to minimize $f(x)$ without any constraint, all solutions such as $x = -1$, $x = 1$, and $x = 0$ are feasible. In fact, the feasible region is the whole real axis. Obviously, $x = 0$ corresponds to $f(0) = 0$ as the true minimum.

However, if we want to find the minimum of $f(x) = x^2$ subject to $x \geq 2$, it becomes a constrained optimization problem. The points such as $x = 1$ and $x = 0$ are no longer feasible because they do not satisfy $x \geq 2$. In this case, the feasible solutions are all the points that satisfy $x \geq 2$. So $x = 2$, $x = 100$, and $x = 10^8$ are all feasible. It is obvious that the minimum occurs at $x = 2$ with $f(2) = 2^2 = 4$. That is, the optimal solution for this problem occurs at the boundary point $x = 2$ (see Figure 1.7).

1.6.3 Gradient and Hessian Matrix

We can extend the optimization procedure for univariate functions to multivariate functions using partial derivatives and relevant conditions. Let us start with an example

$$\text{Minimize} \quad f(x, y) = x^2 + y^2 \quad (x, y \in \mathbb{R}). \tag{1.127}$$

It is obvious that $x = 0$ and $y = 0$ is the minimum solution because $f(0, 0) = 0$. The question is how to solve this problem formally. We can extend the stationary condition to partial derivatives, and we have $\partial f / \partial x = 0$ and $\partial f / \partial y = 0$. In this case, we have

$$\frac{\partial f}{\partial x} = 2x + 0 = 0, \quad \frac{\partial f}{\partial y} = 0 + 2y = 0. \tag{1.128}$$

The solution is obviously $x_* = 0$ and $y_* = 0$.

Now how do we know that it corresponds to a maximum or minimum? If we try to use the second derivatives, we have four different partial derivatives such as f_{xx} and f_{yy} and which one should we use? In fact, we need to define a Hessian matrix from these second partial derivatives and we have

$$\boldsymbol{H} = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}. \tag{1.129}$$

Since $\partial x \partial y = \partial y \partial x$ or

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}, \tag{1.130}$$

we can conclude that the Hessian matrix is always symmetric. In the case of $f = x^2 + y^2$, it is easy to check that the Hessian matrix is

$$\boldsymbol{H} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}. \tag{1.131}$$

Mathematically speaking, if $\boldsymbol{H}$ is positive definite, then the stationary point (x_*, y_*) corresponds to a local minimum. Similarly, if $\boldsymbol{H}$ is negative definite, the stationary point corresponds to a maximum. Since the Hessian matrix here does not involve any x or y, it is always positive definite in the whole search domain $(x, y) \in \mathbb{R}^2$, so we can conclude that the solution at point $(0, 0)$ is the global minimum.

Obviously, this is a special case. In general, the Hessian matrix will depend on the independent variables, but the definiteness test conditions still apply. That is, positive definiteness of a stationary point means a local minimum. Alternatively, for bivariate functions, we can define the determinant of the Hessian matrix in Eq. (1.129) as

$$\Delta = \det(\boldsymbol{H}) = f_{xx} f_{yy} - (f_{xy})^2. \tag{1.132}$$

At the stationary point (x_*, y_*), if $\Delta > 0$ and $f_{xx} > 0$, then (x_*, y_*) is a local minimum. If $\Delta > 0$ but $f_{xx} < 0$, it is a local maximum. If $\Delta = 0$, it is inconclusive and we have to use other information such as higher-order derivatives. However, if $\Delta < 0$, it is a saddle point. A saddle point is a special point where a local minimum occurs along one direction while the maximum occurs along another (orthogonal) direction.

In fact, for a multivariate function $f(x_1, x_2, \ldots, x_n)$ in an n-dimensional space, the stationary condition can be extended to

$$\boldsymbol{G} = \nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)^{\mathrm{T}} = 0, \tag{1.133}$$

where $\boldsymbol{G}$ is called the gradient vector. The second derivative test becomes the definiteness of the Hessian matrix

$$\boldsymbol{H} = \begin{pmatrix} \frac{\partial^2 f}{\partial {x_1}^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial {x_2}^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial {x_n}^2} \end{pmatrix}. \tag{1.134}$$

At the stationary point defined by $\boldsymbol{G} = \nabla f = 0$, the positive definiteness of $\boldsymbol{H}$ gives a local minimum, while the negative definiteness corresponds to a local maximum. In essence, the eigenvalues of the Hessian matrix $\boldsymbol{H}$ determine the local behavior of the function. As we mentioned before, if $\boldsymbol{H}$ is positive semidefinite, it corresponds to a local minimum.

1.6.4 Optimality Conditions

A point $\boldsymbol{x}_*$ is called a strong local maximum of the nonlinearly constrained optimization problem if $f(\boldsymbol{x})$ is defined in a δ-neigborhood $N(\boldsymbol{x}_*, \delta)$ and satisfies $f(\boldsymbol{x}_*) > f(\boldsymbol{u})$ for $\forall \boldsymbol{u} \in N(\boldsymbol{x}_*, \delta)$, where $\delta > 0$ and $\boldsymbol{u} \neq \boldsymbol{x}_*$. If $\boldsymbol{x}_*$ is not a strong local maximum, the inclusion of equality in the condition $f(\boldsymbol{x}_*) \geq f(\boldsymbol{u})$ for $\forall \boldsymbol{u} \in N(\boldsymbol{x}_*, \delta)$ defines the point $\boldsymbol{x}_*$ as a weak local maximum (see Figure 1.8). The local minima can be defined in a similar manner when $>$ and $\geq$ are replaced by $<$ and $\leq$, respectively.

Figure 1.8 shows various local maxima and minima. Point A is a strong local maximum, while point B is a weak local maximum because there are many (in fact infinite) different values of $\boldsymbol{x}$ which will lead to the same value of $f(\boldsymbol{x}_*)$. Point D is the global maximum, and point E is the global minimum. In addition, point F is a strong local minimum.

However, point C is a strong local minimum, but it has a discontinuity in $f'(\boldsymbol{x}_*)$; the stationary condition for this point $f'(\boldsymbol{x}_*) = 0$ is not valid. We will not deal with this type of minima or maxima in detail, though the subgradient method should work well if the function is convex.

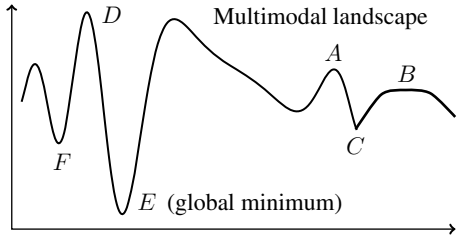

Figure 1.8 Local optima, weak optima, and global optimality.

As we briefly mentioned before, for a smooth curve $f(x)$, optimal solutions usually occur at stationary points where $f'(x) = 0$. This is not always the case because optimal solutions can also occur at the boundary, as we have seen in the previous example of minimizing $f(x) = x^2$ subject to $x \geq 2$. In our present discussion, we will assume that both $f(\boldsymbol{x})$ and $f'(\boldsymbol{x})$ are always continuous, or $f(\boldsymbol{x})$ is everywhere twice-continuously differentiable. Obviously, the information of $f'(x)$ is not sufficient to determine whether a stationary point is a local maximum or minimum. Thus, higher-order derivatives such as $f''(x)$ are needed, but we do not make any assumption at this stage. We will discuss this further in detail in later chapters.

1.7 General Formulation of Optimization Problems

Whatever the real-world applications may be, it is usually possible to formulate an optimization problem in a general mathematical form. All optimization problems with an explicit objective $f(\boldsymbol{x})$ can, in general, be expressed as a nonlinearly constrained optimization problem

$$\text{Maximize/minimize} \quad f(\boldsymbol{x}), \quad \boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n,$$

$$\text{Subject to } \phi_j(\boldsymbol{x}) = 0 \ \ (j = 1, 2, \ldots, M),$$

$$\psi_k(\boldsymbol{x}) \leq 0 \ \ (k = 1, \ldots, N), \tag{1.135}$$

where $f(\boldsymbol{x})$, $\phi_j(\boldsymbol{x})$ and $\psi_k(\boldsymbol{x})$, are scalar functions of the design vector $\boldsymbol{x}$. Here, the components x_i of $\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$ are called design or decision variables, and they can be either continuous, discrete, or a mixture of these two. The vector $\boldsymbol{x}$ is often called the decision vector, which varies in an n-dimensional space $\mathbb{R}^n$. It is worth pointing out that we use a column vector here for $\boldsymbol{x}$ (thus with a transpose T). We can also use a row vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ and the results will be the same, though some formulations may be slightly different. Different textbooks may use slightly different formulations. Once we are aware of such minor variations, this causes no difficulty or confusion.

It is worth pointing out that the objectives are explicitly known in all the optimization problems to be discussed in this book. However, in reality, it is often difficult to quantify what we want to achieve, but we still try to optimize certain things such as the degree of enjoyment or the quality of service on holiday. In other cases, it might be impossible to write the objective function in any explicit form mathematically. In any case, we always assume that the values of an objective function are always computable.

Exercises

1.1 Find the first and second derivatives of $f(x) = \sin(x)/x$.

1.2 Find the gradient and Hessian matrix of $f(x, y, z) = x^2 + y^2 + 2xy + 3yz + z^3$. Is the Hessian matrix symmetric?

1.3 Show that $\int_0^\infty x^3 e^{-x} dx = 6$.

1.4 Find the eigenvalues and eigenvectors of $\boldsymbol{A} = \begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix}$.

1.5 In the second exercise, the Hessian matrix at $z = 1$ becomes

$$\boldsymbol{H} = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 2 & 3 \\ 0 & 3 & 6 \end{pmatrix},$$

what are its eigenvalues? Is this matrix positive definite?

1.6 Show that $f(x, y) = (x - 1)^2 + x^2 y^2$ has a minimum at $(1, 0)$.

Further Reading

Boyd, S.P. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge, UK: Cambridge University Press.

Eriksson, K., Estep, D., and Johnson, C. (2004). *Applied Mathematics: Body and Soul, Volume 1: Derivatives and Geometry in IR3*. Berlin: Springer-Verlag.

Gill, P.E., Murray, W., and Wright, M.H. (1982). *Practical Optimization*. Bingley: Emerald Publishing.

Kreyszig, E. (2010). *Advanced Engineering Mathematics*, 10e. Hoboken, NJ: Wiley.

Nocedal, J. and Wright, S.J. (2006). *Numerical Optimization*, 2e. New York: Springer.

Yang, X.S. (2010). *Engineering Optimization: An Introduction with Metaheurisitic Applications*. Hoboken, NJ: Wiley.

Yang, X.S. (2014). *Nature-Inspired Optimization Algorithms*. London: Elsevier.

Yang, X.S. (2017). *Engineering Mathematics with Examples and Applications*. London: Academic Press.