Lu Sun, and many more.

A Notebook on Calculus

To all family members, friends and communities members who have been dedicating to the presentation of this notebook, and to all students, researchers and faculty members who might find this notebook helpful.

Contents

Foreword **vii**

Preface **xi**

List of Figures **xiii**

List of Tables **xv**

I Limit, Derivative and Integral 1

1 Limit **3**
- 1.1 Limit of a Sequence 3
  - 1.1.1 A Motivating Example 3
  - 1.1.2 Limit of a Sequence 6
  - 1.1.3 Calculation of the Limit of a Sequence 8
- 1.2 Limit of a Function 10
  - 1.2.1 A Motivating Example 10
  - 1.2.2 Limit of a Function 11
  - 1.2.3 Calculation of the Limit of a Function 13

2 Derivative **17**
- 2.1 A Motivating Example 17
- 2.2 Derivative of a Function 21
- 2.3 Calculation of the Derivative of a Function 21

3 Integral **23**
- 3.1 A Motivating Example 23
- 3.2 Integral of a Function 28
- 3.3 Calculation of the Integral of a Function 30

4 Applications **33**
- 4.1 Newton's Method 33
- 4.2 Taylor Series 34

II Multivariable Function, Partial Derivative and Multiple Integral 37

5 Multivariable Function 39
5.1 Brief Introduction to Vector and Matrix 39
5.1.1 Basic Concepts 40
5.1.2 Matrix Multiplication 41
5.1.3 Block Matrix 42
5.1.4 Identity Matrix and Square Matrix Inverse 43
5.2 Multivariable Function 44

6 Partial derivative 45
6.1 A Motivating Example 45
6.2 Partial Derivative 48
6.3 Gradient 51
6.3.1 Motivating Example 51
6.3.2 Gradient of a Multivariable Function 53
6.4 Jacobian Matrix 53

7 Multiple Integral 55
7.1 A Motivating Example 55
7.2 Multiple Integral 55

8 Applications 57
8.1 Neural Network Back-propagation 57
8.2 Bayesian Inference 57

III Differential Equation 59

IV Functional and Calculus of Variations 61

Bibliography 63

Foreword

If a piece of software or an e-book can be made completely open source, why not a notebook?

This brings me back to the summer of year 2009, when I just started my third year as a high school student in Harbin No. 3 High School. In around August and September of every year, that is, when the results of Gaokao (National College Entrance Examination of P. R. China, annually held in July) are released, you would find people selling notebooks photocopies claimed to be collected from the top scorers of the exam. Much as I was interested in what these notebooks look like, I myself was not expecting to actually learn anything from them, mainly for the following three reasons.

First of all, some (in fact many) of these notebooks were more difficult to understand than the textbooks. I guess we cannot blame the top scorers for being too smart and make things sometimes extremely brief, or otherwise overwhelmingly complicated.

Secondly, why would I want to adapt to notebooks of others when I have my own? And by the way, I was positive that mine would be as good as theirs, given that I had been putting the same time (three years of high school, only for 6 modules!) and effort learning the courses and preparing the notebooks.

And lastly, as a student in Harbin No. 3 High School, I knew that the top scorers of the coming year would probably be a schoolmate next door, perhaps even a good friend of mine. Why would I want to pay a great amount of penny to a complete stranger in a photocopy shop for his or her notebook, rather than ask from him or her directly?

However, things have changed later on after entering a university as an undergraduate student. I think the main cause of the change is that, since in the university there are so many modules and materials to learn, students are often distracted from digging into one book or module very deeply. (For those who still can concentrate, you have my highest respect.) The situation becomes even worse as I become a Ph.D. student, this time due to that I have to concentrate on one subject entirely, and can hardly split much time on other irrelevant but still important and interesting contents.

This motivates me to start reading and taking notebooks for selected books and articles such as journal papers and magazines. I have a bunch of notebooks with me, most of them are physical. My very first notebook is on *Numerical Analysis*, an entrance level module for engineering background students. Till today I have on my hand dozens of notebooks. One day it suddenly came to me: why not digitalize them, and make them accessible online and open source, and let everyone read and edit it?

—

As majority of open source software, this notebook (and it applies to the other notebooks in this series) does not come with any "warranty" of any kind, meaning that there is no guarantee for the statement and knowledge in this notebook to be exactly correct as it is not peer reviewed. **Do NOT cite this notebook in your academic research paper or book!** Of course, if you find anything here useful with your research, please trace back to the origin of the citation, and read it yourself, and on top of that determine whether or not to use it in your research.

This notebook is suitable as:

- a quick reference guide;
- a brief introduction to the subject;
- a "cheat sheet" for students to prepare for the exam (Don't bring it to the exam unless it is allowed by your lecture!) or for lectures to prepare the teaching materials.

This notebook is NOT suitable as:

- a direct research reference;
- a replacement to the textbook;

because as explained the notebook is NOT peer reviewed and it is meant to be simple and easy to read. It is not necessary brief, but all the tedious explanation and derivation, if any, shall be "fold into appendix" and a reader can easily skip those things without any interruption to the reading.

—

Although this notebook is open source, the reference materials of this notebook, including many textbooks, journal papers, conference proceedings, etc., may not be open source. Very likely many of these reference materials are licensed or copyrighted. Please legitimately access these materials and properly use them if necessary.

Some of the figures in this notebook is drawn using Excalidraw, a very interesting tool for machine to emulate hand-writing. The Excalidraw project can be found in GitHub, *excalidraw/excalidraw*.

Preface

This notebook is on *Calculus*, a very important mathematical tool that was invented back in Newton's time or even earlier. It has now become entrance level module for mathematics and engineering background students in year one in the university.

Initially, the invention of calculus, including the introduction of differentiation and integration, is of course used to explain things such as the concept of "speed" as a differential of distance over time. You might easily come up with some common use cases of calculus, for example calculating the tangent of a curve, and calculating the volume of an arbitrarily shaped container. Other applications which may not make too much sense for beginners, for example the derivation of cycloid, are also obtained from calculus. Many advanced mathematical tools themselves are built on top of calculus, for example fourier transform, which is widely used in signal processing. Without a solid understanding of calculus, it is hardly possible for one to use these tools confidently and effectively.

The key reference of this notebook is listed below. During the development of the notebook, this list may become longer and longer.

Book *Calculus Metric Version Eighth Edition* by James Stewart, published by Cengage Learning [1].

Book *Calculus* by Gibert Strang (Massachusetts Institute of Technology), published by Wellesley-Cambridge Press [2]. This book is available at MIT Open Courseware (*ocw.mit.edu*). There are countless number of great learning materials there.

List of Figures

1.1 Plot of a_n as a function of n in the motivating example. The readings of a_n are given in Table 1.2. 6
1.2 Plot of y as a function of x in the motivating example. 11
1.3 Plot of $y = \sin\left(\frac{1}{x}\right)$. 15

2.1 Plot of y as a function of x in the motivating example. 18
2.2 Slope of secant and tangent of $y = f(x)$ at $x = 2$. 19

3.1 Plot of (3.1) and $N = 3$ trapezoids. 24
3.2 Use $N = 20$ trapezoids to approximate the red area. 26
3.3 Calculation of the area of trapezoids. Variables $N = 3$ and $a = 0.5$ are used in the plot for demonstration. 27
3.4 Plot of $f(x) = \sin(x) + 0.5$ from 0 to 2π. The area above $y = 0$ is surrounded by the red dashed line, and the area below $y = 0$ by the blue dashed line. The definite integral $\int_0^{2\pi} f(x)dx$ in this case is the red area subtracting the blue area. 32

4.1 Plot of function $y = 2^x$ in red solid line, and its approximations using first-order, 5th-order and 10th-order Taylor series in blue solid line, blue dashed line and blue dot line respectively. . . 35

6.1 Plot of function $y = f(x_1, x_2)$ in 3-D. 46
6.2 The calculation of z using x and y. 49
6.3 Plot of $y = f(x_1, x_2)$ in 3-D. 52
6.4 Contour line of $y = f(x_1, x_2)$. 53

List of Tables

1.1 Calculate a_n for any arbitrary n in the motivating example. . 4
1.2 Calculate a_n for any arbitrary n in the motivating example, but longer table. 5
1.3 Convergence of commonly seen $\{a_n\}$ and $\{s_n\}$. Variable c, r are constant real numbers. 8
1.4 Limit of commonly seen elementary functions. 14

2.1 Derivative of commonly seen functions. 22

3.1 Indefinite integral of commonly seen functions. 30

Part I

Limit, Derivative and Integral

1

Limit

CONTENTS

1.1 Limit of a Sequence 3
  1.1.1 A Motivating Example 3
  1.1.2 Limit of a Sequence 6
  1.1.3 Calculation of the Limit of a Sequence 8
1.2 Limit of a Function 9
  1.2.1 A Motivating Example 10
  1.2.2 Limit of a Function 11
  1.2.3 Calculation of the Limit of a Function 13

Consider a sequence $\{a_n\}$ or a function $f(x)$. In this chapter, the values of a_n or $f(x)$ when n or x grows towards infinity are discussed. The value of $f(x)$ when x approaches a constant value is discussed.

1.1 Limit of a Sequence

A motivating example of is given in Section 1.1.1. The definition of the limit of a sequence is given in Section 1.1.2. The calculation of the limit of a sequence is discussed in 1.1.3, mainly on the proof of convergence of a sequence.

1.1.1 A Motivating Example

We use $\{a_n\}$ to denote a sequence. In $\{a_n\}$, the positive integer n is the index of the elements in the sequence, where a_1 represents the first element of $\{a_n\}$, and a_2 the second element, and so on. A sequence has at least one element, i.e. the first element a_1. It may have finite elements, in which case it is called a *finite sequence*. Or, it may have infinite elements, in which case it is called an *infinite sequence*. In this notebook, we are mostly interested in the infinite sequence.

A motivating example is given in below to illustrate the limit of an infinite sequence.

A Motivating Example

Consider an infinite sequence $\{a_n\}$ whose elements are recursively calculated by

$$a_1 = 1, \tag{1.1}$$

$$a_n = a_{n-1} + \left(\frac{1}{2}\right)^{n-1}. \tag{1.2}$$

Q1: Calculate a_n for any arbitrary n.

Q2: Calculate the feasible domain of n such that a_n reaches/exceeds 1.95.

Q3: Calculate the feasible domain of n such that a_n reaches/exceeds 2.

The old school way of solving Q1 is rather simple: use a table to list down different n and its associated a_n. The value of a_n can be manually calculated for small n, as shown in Table 1.1. In theory, this table can continue on and on, thus calculating a_n for any arbitrary n. From Table 1.1, for any $n \geq 6$, $a_n \geq 1.95$.

TABLE 1.1
Calculate a_n for any arbitrary n in the motivating example.

n	$a_n - a_{n-1} = \left(\frac{1}{2}\right)^{n-1}$	a_n
1	—	1
2	0.5	1.5
3	0.25	1.75
4	0.125	1.875
5	0.0625	1.9375
6	0.03125	1.96875
7	0.015625	1.984375
⋮	⋮	⋮

To find out the feasible range of n such that $a_n \geq 2$, intuitively thinking, we might simply require a larger table, say Table 1.2. From Table 1.2, it can been seen that as n grows larger and larger, the increment $a_n - a_{n-1} = \left(\frac{1}{2}\right)^{n-1}$ becomes smaller and smaller, and the increment is just barely enough to top a_n to 2.

An alternative method to find the feasible range is to derive an analytical equation of a_n as a function of n. Then we might be able to solve $a_n \geq 2$ for n. Notice that Recursively using (1.2) for $t-1$ times and substituting (1.1)

TABLE 1.2
Calculate a_n for any arbitrary n in the motivating example, but longer table.

n	$a_n - a_{n-1} = \left(\frac{1}{2}\right)^{n-1}$	a_n
1	1	1
2	0.5	1.5
3	0.25	1.75
4	0.125	1.875
5	0.0625	1.9375
6	0.03125	1.96875
7	0.015625	1.984375
8	0.0078125	1.9921875
9	0.00390625	1.99609375
10	0.001953125	1.998046875
11	0.0009765625	1.9990234375
12	0.00048828125	1.99951171875
13	0.000244140625	1.999755859375
14	0.0001220703125	1.9998779296875
15	0.00006103515625	1.99993896484375
$\vdots$	$\vdots$	$\vdots$

into (1.2) gives

$$\begin{aligned} a_n &= \sum_{i=1}^{n}\left(\frac{1}{2}\right)^{i-1} && (1.3)\\ &= 1+\sum_{i=2}^{n}\left(\frac{1}{2}\right)^{i-1}, && (1.4)\end{aligned}$$

and multiplying $\frac{1}{2}$ on (1.3) gives

$$\begin{aligned} \frac{1}{2}a_n &= \sum_{i=1}^{n}\left(\frac{1}{2}\right)^{i}\\ &= \sum_{i=2}^{n}\left(\frac{1}{2}\right)^{i-1}+\left(\frac{1}{2}\right)^{n}. && (1.5)\end{aligned}$$

Subtracting (1.5) from (1.4) gives

$$\begin{aligned} \frac{1}{2}a_n &= 1-\left(\frac{1}{2}\right)^{n},\\ a_n &= 2-\left(\frac{1}{2}\right)^{n-1}. && (1.6)\end{aligned}$$

Equation (1.6) can be verified using the results in Table 1.2. It suggests

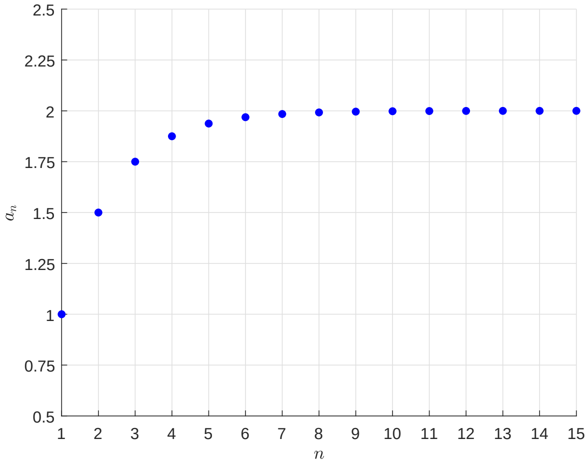

FIGURE 1.1
Plot of a_n as a function of n in the motivating example. The readings of a_n are given in Table 1.2.

that the elements in sequence $\{a_n\}$ will never reach 2 for any n, although a_n can be very close to 2 as n increases. This can also be shown from the plot of (1.6) as given in Fig. 1.1.

The following features can be observed for sequence $\{a_n\}$ from both (1.6) and Fig. 1.1.

- The sequence is monotonically increasing, as $a_{n-1} < a_n$ for any n.
- The sequence is bounded, as $a_n < 2$ for any n.
- The sequence can get "as close to 2 as we like", in the sense that for any value smaller than 2, however close to 2 it is (say, 1.9999), a_n will at some point exceeds that value and get even closer to 2, for large enough n.

This existence of sequence $\{a_n\}$ reveals an important yet not intuitive fact that it is possible to find a monotonically increasing yet bounded sequence. Another way to look at it is that it is possible to add infinite number of positive values together, yet the result is a finite value. From the fact that $\{a_n\}$ gets as close to a certain value as possible when n gets large enough, the limit of sequence is defined, as introduced in the next Section 1.1.2.

1.1.2 Limit of a Sequence

Given an infinite sequence $\{a_n\}$, if $\{a_n\}$ is bounded and it gets close to a value L "as close as we like" for large n, then L is called the limit of sequence $\{a_n\}$. The formal definition of the limit of a sequence is given as follows.

Definition of the limit of sequence:

A sequence $\{a_n\}$ has the limit L if for any $\varepsilon > 0$, there is always a corresponding integer N, such that if $n > N$, $|a_n - L| < \varepsilon$. This is denoted by

$$\lim_{n\to\infty} a_n = L,$$

or

$$a_n \to L \quad \text{as} \quad n \to \infty,$$

and in this case we say "sequence $\{a_n\}$ is convergent" and "sequence $\{a_n\}$ converges to L as n approaches infinity".

The sequence $\{a_n\}$ in (1.6) is an example of a convergent sequence that converges to 2. We can easily prove this using the definition of the limit of a sequence as follows.

Given any $\varepsilon > 0$, solving

$$|V(N) - 2| < \varepsilon$$

gives

$$\begin{aligned} \left|2 - \left(\frac{1}{2}\right)^{N-1} - 2\right| &< \varepsilon, \\ N &> 1 - \log_2 \varepsilon. \end{aligned} \tag{1.7}$$

For example, if specifying $\varepsilon = 0.05$, using (1.7) gives $N > 5.32$. This implies that from $n \geq 6$ onwards, $|a_n - 2| < 0.05$. This matches the observation given in Table 1.1.

If there is no such limit L for an infinite sequence $\{a_n\}$, we say "sequence $\{a_n\}$ is divergent" or "sequence $\{a_n\}$ diverges".

As a special case of divergent sequences, if $\{a_n\}$ becomes unbounded as n approaches infinity, we say "sequence $\{a_n\}$ diverges to infinity". The definition is as follows.

Definition of sequence diverging to infinity:

For a sequence $\{a_n\}$, if for any arbitrary positive value M, there is always a corresponding integer N, such that if $n > N$,

$|a_n| > M$, we say "$\{a_n\}$ diverges to infinity". This is denote it by

$$\lim_{n\to\infty} a_n = \infty.$$

Another sequence $\{s_n\}$ where $s_n = \sum_{i=1}^{n} a_n$ is used to denote the sum of the first n elements in the infinite sequence $\{a_n\}$. Apparently, $\{s_n\}$ itself is also an infinite sequence which may or may not converge depending on $\{a_n\}$. As n approaches infinity, $\{s_n\}$ approaches $\sum_{i=1}^{\infty} a_n$, which is called the infinite sum of the series $\{a_n\}$.

1.1.3 Calculation of the Limit of a Sequence

As the first step in the calculation of the limit of a sequence, the convergence of the sequence needs to be proved. In other words, we need to prove the existence of the limit before calculating it. This section mainly focuses on the proof of convergence of a sequence. Usually, after proving the convergence, the limit can be obtained by either using numerical methods, or calculating the limit of the function a_n as a function of n. More details of calculating the limit of a function are given in later Section 1.2.

Generally speaking, there is no systematic way of proving the convergence/divergence of an infinite sequence, and sometimes the proof can be very difficult or even impossible. For some commonly seen sequence, the limits and infinite sum are concluded in the following Table 1.3.

TABLE 1.3
Convergence of commonly seen $\{a_n\}$ and $\{s_n\}$. Variable c, r are constant real numbers.

Category	a_n	s_n	$\lim_{n\to\infty} a_n$	$\lim_{n\to\infty} s_n$
Polynomial	c	nc	c	No(∞)
	n	$\frac{n(n+1)}{2}$	No(∞)	No(∞)
	n^2	$\frac{n(n+1)(2n+1)}{6}$	No(∞)	No(∞)
	n^3	$\frac{n^2(n+1)^2}{4}$	No(∞)	No(∞)
Power Series	$cr^{n-1}, \|r\| < 1$	$\frac{c(1-r^n)}{1-r}$	0	$\frac{c}{1-r}$
Others	n^{-1}	—	0	No(∞)
	n^{-2}	—	0	$\frac{\pi}{6}$

"No(∞)" stands for "diverges to infinity".

If an interested sequence falls into one of the above categories, or somewhat similar to the above categories, or can be expressed as a sum/multiplication/division of sequences from the above categories, then its convergence/divergence might be proved a bit easier. For example, if $c_n = a_n + b_n$, and both a_n and b_n converge, then

$$\lim_{n\to\infty} c_n \quad = \quad \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n.$$

This can be proved using the definition of the limitation of the sequence.

Some other interesting features regarding sequence convergence are given as follows. Given two sequences $\{a_n\}$ and $\{b_n\}$, if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = L \neq 0,$$

then $\{a_n\}$ and $\{b_n\}$ must behave the same in terms of convergence, meaning that both of them must converge or diverge at the same time.

It is intuitive and not difficult to prove that for $\{s_n\}$ to converge, i.e. $\lim_{n\to\infty} s_n = s$, it is necessary (but not sufficient) for its associated $\{a_n\}$ to converge to zero, i.e. $\lim_{n\to\infty} a_n = 0$. Do notice that it is possible to have a divergent $\{s_n\}$ even if a_n converges to zero. An example is the harmonic series given in Table 1.3 where $a_n = \frac{1}{n}$. It is obvious that $\lim_{n\to\infty} a_n = 0$, but in fact $\lim_{n\to\infty} s_n = \infty$.

The famous monotone convergence theorem states that *if a sequence* $\{a_n\}$ *is monotonically increasing or decreasing, and it is at the same time bounded, i.e.* $|a_n| < M$ *for all* n*, then* $\{a_n\}$ *must be convergent.* The proof of this theorem is more difficult than it appears to be, thus is not included in the notebook.

From the monotone convergence theorem, we know that for a sequence $\{a_n\}$, if $\sum_{i=1}^{\infty} |a_n|$ exists, then the infinite sum $\sum_{i=1}^{\infty} a_n$ must also exist. This can be illustrated simply by splitting $\{a_n\}$ into $\{a_n^+\}$ and $\{a_n^-\}$, where

$$\begin{aligned} a_n^+ &= \left\{ \begin{array}{cc} a_n & a_n \geq= 0 \\ 0 & a_n < 0 \end{array} \right. , \\ a_n^- &= \left\{ \begin{array}{cc} -a_n & a_n < 0 \\ 0 & a_n \geq 0 \end{array} \right. . \end{aligned}$$

Apparently, $|a_n| = a_n^+ + a_n^-$ and $a_n = a_n^+ - a_n^-$, and both $\sum_n a_n^+$ and $\sum_n a_n^-$ are monotonically increasing positive sequences. Since $\sum_{i=1}^{\infty} |a_n| = \sum_{i=1}^{\infty} a_n^+ + \sum_{i=1}^{\infty} a_n^-$ is finite, both $\sum_{i=1}^{\infty} a_n^+$ and $\sum_{i=1}^{\infty} a_n^-$ must also be finite. Therefore, both $\sum_n a_n^+$ and $\sum_n a_n^-$ must be convergent according to the monotone convergence theorem. This implies that $\sum_n a_n = \sum_n a_n^+ - \sum_n a_n^-$ must also be a convergent sequence as it is the sum of two convergent sequences. In this case, a_n is called “absolutely convergent”. Absolutely convergent sequence must have a bounded infinite sum, but it might be not true wise versa.

There are some famous sequence widely used in both mathematics and applied mathematics, such as the famous Taylor serious. The application of Taylor serious, and many more important sequence, can be found in varieties of textbooks related to numerical analysis, signal processing, control engineering, etc., and they will not be covered in details in this notebook.

1.2 Limit of a Function

A motivating example is given in Section 1.2.1. The definition of the limit of a function is given in Section 1.2.2. The calculation of the limit of a function is discussed in 1.2.3.

1.2.1 A Motivating Example

A motivating example is given below to illustrate the limit of a function.

A Motivating Example

Consider function

$$y = f(x) = \begin{cases} (x-1)^2 & x \neq 1 \\ 1 & x = 1 \end{cases} . \tag{1.8}$$

Q1: Obtain the domain for x and range for y.
Q2: Calculate y at $x = 1$.
Q3: Calculate y when x is in a small "neighbourhood" of 1, but $x \neq 1$.

The plot of y as a function of x is given in Fig 1.2. It is clear from the figure that the domain and range of the function are $x \in \mathbb{R}$ and $y \in \mathbb{R}, y > 0$ respectively. Substituting $x = 1$ into (1.8) gives $y = 1$, as it is also shown in Fig 1.2.

For Q3, we need to rewrite "x is in a small "neighbourhood" of 1, but $x \neq 1$" in a bit more precise manner. An intuitive way is to define a small "threshold area" near $x = 1$, say, $-\delta < x - 1 < \delta, x \neq 1$ with δ being a very small positive value. Notice that $x = 1$ is out of our concern and the value of y near $x = 1$ has nothing to do with y at $x = 1$.

Next, we can calculate y subject to $-\delta < x - 1 < \delta, x \neq 1$. Clearly, the range of y relates to the choice of δ. Substituting $-\delta < x - 1 < \delta, x \neq 1$ into (1.8) gives $0 < y < \delta^2$. With δ being chosen smaller and smaller, the range of y would become smaller and smaller, and y will eventually approach 0 (although y cannot be precisely 0).

This discussion is essentially a study of y when x is "as close as we like but not equal" to 1. We have learned that in this motivating example:

- The value of y when x is "as close as we like to 1" does not rely on the value of y at $x = 1$.
- The value of y when x is "as close as we like to 1" floats in a small range depending on the size of the "neighbourhood" of 1 where x residents. The size of the "neighbourhood" is quantified by δ in this example. Thus, the range of y is related to the choice of δ.

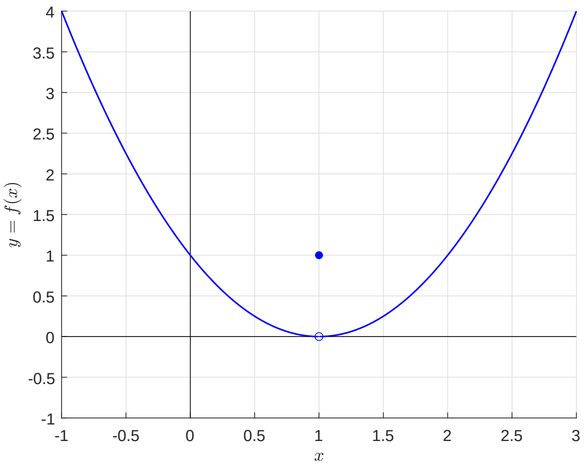

FIGURE 1.2
Plot of y as a function of x in the motivating example.

- With a proper choice of δ, the value of y can be as close as we like to 0 since $0 < y < \delta^2$. For example, to achieve $0 < y < 10^{-6}$, simply choose $\delta = 10^{-3}$ for the neighbourhood of x around 1.

The above features give a brief idea of limit of a function. In the example, y can be made as close as we like to 0 simply by making x close enough to 1 by choosing a small δ.

1.2.2 Limit of a Function

The formal definition of the limit of a function follows the similar idea shown in Section 1.2.1 as follows. Notice that there are a few different but equivalent ways to define the limit of a function. Here the "ε-δ definition" is introduced.

Definition of the limit of a function at $x \to a$:

A function $f(x)$ of x has the limit L at $x = a$ if for any $\varepsilon > 0$, there is always a corresponding $\delta > 0$, such that if $|x - a| < \delta$, $|f(x) - L| < \varepsilon$, with the prerequisite that $|x - a| < \delta$ is defined for $f(x)$. This is denoted by

$$\lim_{x \to a} f(x) \quad = \quad L,$$

or

$$f(x) \to L \quad \text{as} \quad x \to a.$$

Using the definition above, it can be proved easily that for the motivating example in Section 1.2.1, the function $y = f(x)$ has a limit of $\lim_{x\to 1} f(x) = 0$. Notice that $\lim_{x\to a} f(x) = L$ does not necessarily require $f(a) = L$. As a matter of fact, $f(x)$ does not even need to be defined at $x = a$, as long as it is defined at the neighbour of $x = a$.

Similar to the definition of the limit of a function, the definition of one-sided limit of a function is given below. The one-sided limit is similar but weaker than the definition of the limit of a function in the sense that it only concerns one side of the neighbour of $x = a$.

Definition of the one-sided limit of a function:

A function $f(x)$ of x has the one-side left limit L^- at $x = a$ if for any $\varepsilon > 0$, there is always a corresponding $\delta > 0$, such that if $a - \delta < x < a$, $|f(x) - L^-| < \varepsilon$, with prerequisite that $a - \delta < x < a$ is defined for $f(x)$. This is denoted by

$$\lim_{x\to a^-} f(x) = L^-,$$

or

$$f(x) \to L^- \quad \text{as} \quad x \to a^-.$$

A function $f(x)$ of x has the one-side right limit L^+ at $x = a$ if for any $\varepsilon > 0$, there is always a corresponding $\delta > 0$, such that if $a < x < a + \delta$, $|f(x) - L^+| < \varepsilon$, with prerequisite that $a < x < a + \delta$ is defined for $f(x)$. This is denoted by

$$\lim_{x\to a^+} f(x) = L^+,$$

or

$$f(x) \to L^+ \quad \text{as} \quad x \to a^+.$$

It is clear from the definition that a function $f(x)$ has a limit of L at $x = a$ if and only if it has both one-sided left limit L^- and one-sided right limit L^+ at $x = a$ and $L^- = L^+ = L$, i.e.

$$\lim_{x\to a} f(x) = L \quad \Leftrightarrow \quad \lim_{x\to a^-} f(x) = \lim_{x\to a^+} f(x) = L.$$

Furthermore, if function $f(x)$ has a limit L at $x = a$, and also $f(x) = L$, the function $f(x)$ is called *continuous at* $x = a$. The example given in the motivating example in Section 1.2.1 is not continuous at $x = 1$ as $\lim_{x \to 1} = 0$ while $f(x)|_{x=1} = 1$, which can be seen from Fig. 1.2. However, it is continuous everywhere else. For instance, at $x = 0$, $\lim_{x \to 0} = 1$ and $f(x)|_{x=0} = 1$.

The definition of the limit of a function $f(x)$ when x approaches infinity is given below. It is quite similar to the definition of the limit of an infinite sequence.

Definition of the limit of a function at $x \to \pm\infty$:

A function $f(x)$ of x has the limit L at $x \to +\infty$ (sometimes denoted as $x \to \infty$ for simplicity) if for any $\varepsilon > 0$, there is always a corresponding δ, such that if $x > \delta$, $|f(x) - L| < \varepsilon$, with prerequisite that $x > \delta$ is defined for $f(x)$. This is denoted by

$$\lim_{x \to +\infty} f(x) \quad = \quad L,$$

or

$$f(x) \to L \quad \text{as} \quad x \to +\infty.$$

A function $f(x)$ of x has the limit L at $x \to -\infty$ if for any $\varepsilon > 0$, there is always a corresponding δ, such that if $x < \delta$, $|f(x) - L| < \varepsilon$, with prerequisite that $x < \delta$ is defined for $f(x)$. This is denoted by

$$\lim_{x \to -\infty} f(x) \quad = \quad L,$$

or

$$f(x) \to L \quad \text{as} \quad x \to -\infty.$$

In special cases where the function is unbounded, the limit of the function does not exist and we can use $\lim_{x \to a} f(x) = \pm\infty$ and $\lim_{x \to \pm\infty} f(x) = \pm\infty$ to represent the cases.

1.2.3 Calculation of the Limit of a Function

The calculation of the limit of many commonly seen elementary functions are often obvious and easy. This is because these functions are often continuous in the domain, and for a continuous function the limit $\lim_{x\to a} f(x)$ can be obtained by simply substituting $x = a$ into the function. The limit $\lim_{x\to\infty} f(x)$ might be slightly difficult but mostly can be obtained from the definition.

Some examples are given below in Table 1.4.

TABLE 1.4

Limit of commonly seen elementary functions.

Category	$f(x)$	$\lim_{x\to a} f(x)$	$\lim_{x\to\infty} f(x)$
Polynomial	$p(x)$	$p(a)$	No(∞)
Root	$\sqrt{x}$	$\sqrt{a}$ for $a > 0$	No(∞)
Rational	$\dfrac{q(x)}{p(x)}$	depends	depends
Trigonometric	$\sin(x)$, $\cos(x)$	$\sin(a)$, $\cos(a)$	No
Exponential	e^{-x}	e^{-a}	0 as $x \to +\infty$ No(∞) as $x \to -\infty$
Logarithm	$\log_e(x)$	$\log_e(a)$ for $a > 0$	No(∞)

"No(∞)" stands for "unbounded";

For the case of rational function, if $p(a) \neq 0$, $\lim_{x\to a} \frac{q(x)}{p(x)} = \frac{q(a)}{p(a)}$. If $p(a) = 0$, $\lim_{x\to a} \frac{q(x)}{p(x)}$ depends on the order and coefficients of $q(x)$ and $p(x)$ and may not exist. The limit $\lim_{x\to\infty} \frac{q(x)}{p(x)}$ depends on the order and coefficients of $q(x)$ and $p(x)$ and may not exist.

It is worth mentioning a few typical cases where a given function $f(x)$ does not have a limit and/or is not continuous.

Case 1: function $f(x)$ does not converge at a neighbourhood of $x = a$, thus does not have a one-sided limit. For example, consider

$$f(x) = \sin\left(\frac{1}{x}\right).$$

The function is defined at $x \in \mathbb{R}, x \neq 0$, but it does not have a one-sided limit $\lim_{x\to 0^-} f(x)$ or $\lim_{x\to 0^+} f(x)$, as when $x \to 0$, function $f(x)$ oscillates, as shown in Fig. 1.3.

Case 2: function $f(x)$ is unbounded at a neighbourhood of $x = a$, therefore does not have a one-sided limit. For example, consider

$$f(x) = \left|\frac{1}{x}\right|.$$

Apparently, $\lim_{x\to 0^-} f(x)$ or $\lim_{x\to 0^+} f(x)$ does not exist.

Case 3: function $f(x)$ has one-sided limits $\lim_{x\to 0^-} f(x)$ and $\lim_{x\to 0^+} f(x)$, but $\lim_{x\to 0^-} f(x) \neq \lim_{x\to 0^+} f(x)$. For example, consider

$$f(x) = \text{sign} = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}.$$

From the definition, $\lim_{x\to 0^-} f(x) = -1$ and $\lim_{x\to 0^+} f(x) = 1$, therefore, $\lim_{x\to 0} f(x)$ does not exist.

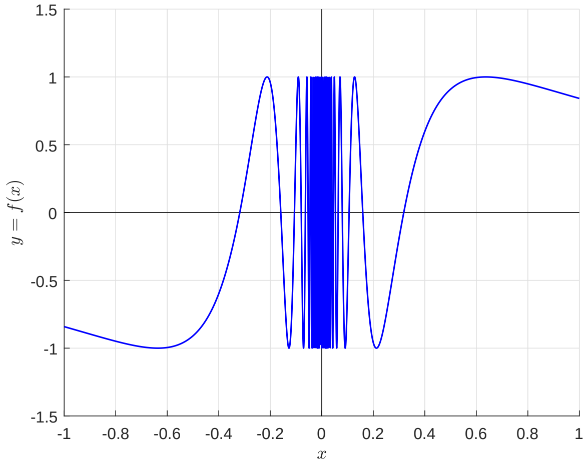

FIGURE 1.3
Plot of $y = \sin\left(\frac{1}{x}\right)$.

Some commonly used tricks are as follows. For $g(x) = f_1(x) + f_2(x)$, if the limits exist for $f_1(x)$ and $f_2(x)$ at $x \to a$, then

$$\lim_{x\to a} g(x) \quad = \quad \lim_{x\to a} f_1(x) + \lim_{x\to a} f_2(x).$$

The same is true for the cases $g(x) = f_1(x)f_2(x)$ and $g(x) = \frac{f_1(x)}{f_2(x)}$, subject to $\lim_{x\to a} f_2(x) \neq 0$ when it is the denominator. And it holds true for $x \to \pm\infty$ as well.

In the case of $g(x) = \frac{f_1(x)}{f_2(x)}$ and $\lim_{x\to a} f_2(x) = 0$, the discussion is more complicated. Sometimes, *L′Hôpital's rule* can become handy. The proof of L′Hôpital's rule requires solid calculus foundation and is out of the scope of this notebook. Generally speaking, for $g(x) = \frac{f_1(x)}{f_2(x)}$ with $\lim_{x\to a} f_1(x) = 0$, $\lim_{x\to a} f_2(x) = 0$, and $f_1(x)$, $f_2(x)$ being both continuous and differentiable at $x = a$, and $\lim_{x\to a} f_2'(x) \neq 0$, it is possible to calculate $\lim_{x\to a} g(x)$ using (1.9) if its right side exists.

$$\lim_{x\to a} g(x) = \lim_{x\to a} \frac{f_1'(x)}{f_2'(x)}. \tag{1.9}$$

where $f'(x)$ is the derivative of $f(x)$. The explanation to derivative and differential is given in a later chapter.

The limit of an infinite sequence is linked to the limit of the associated function at $x \to \infty$. For example, for sequence $\{a_n\}$, if $a_n = f(n)$ and $\lim_{n\to\infty} f(n) = L$, $\lim_{n\to\infty} a_n = L$.

2

Derivative

CONTENTS

2.1 A Motivating Example 17
2.2 Derivative of a Function 20
2.3 Calculation of the Derivative of a Function 21

Consider a continuous function $f(x)$ defined in $[a, b]$. In this chapter the ratio of change of $f(x)$ versus x is quantitively derived and studied.

2.1 A Motivating Example

Consider the following motivating example.

A Motivating Example

Consider

$$y = f(x) = (|x|-1)^2 . \tag{2.1}$$

Obviously the above function (2.1) is continuous in $x \in \mathbb{R}$. We want to study the change $\Delta y = f(x + \Delta x) - f(x)$ given a small deviation Δx at different values of x.

Q1: Describe Δy given a small deviation Δx at $x = 2$, and describe the ratio $\frac{\Delta y}{\Delta x}$ when $\Delta x \to 0$.

Q2: Describe Δy given a small deviation Δx at $x = 0$, and describe the ratio $\frac{\Delta y}{\Delta x}$ when $\Delta x \to 0$.

Q3: Describe the ratio $\frac{\Delta y}{\Delta x}$ at any x when $\Delta x \to 0$.

As a first step, plot y as a function of x from (2.1) in Fig.2.1 for convenient analysis.

Consider $x = 2$. Variable $\Delta y = f(x + \Delta x) - f(x)$ is given in (2.2). Notice that the deviation Δx is supposed to be small, i.e. $|\Delta x| \approx 0$. Therefore, we can

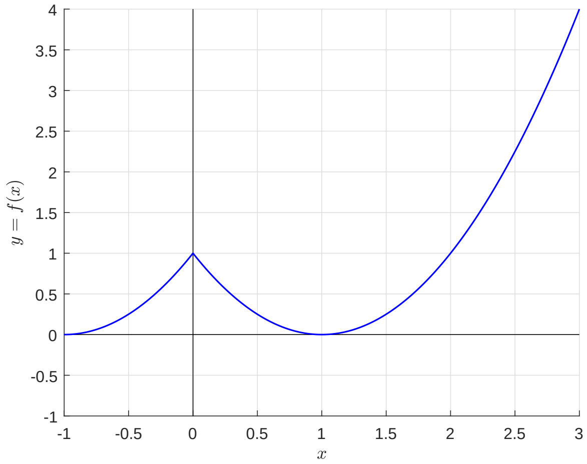

FIGURE 2.1
Plot of y as a function of x in the motivating example.

safely assume $2 + \Delta x \geq 0$ for simplicity.

$$\Delta y = f(2 + \Delta x) - f(2) = (2 + \Delta x - 1)^2 - (2 - 1)^2 = 2\Delta x + \Delta x^2. \quad (2.2)$$

With (2.2), it is possible to calculate y at $x = 2+\Delta x$ by using $y = f(2)+\Delta y$ for small Δx. For example, to calculate $f(1.9)$, substituting $\Delta x = -0.1$ into (2.2) gives $\Delta y = -0.19$, thus $f(1.9) = f(2) - 0.19 = 0.81$.

From (2.2), the ratio $\frac{\Delta y}{\Delta x}$ can be calculated as

$$\left.\frac{\Delta y}{\Delta x}\right|_{x=2} = 2 + \Delta x, \quad (2.3)$$

where $(\cdot)|_{x=a}$ represents substituting $x = a$ into $(\cdot)$.

Notice that (2.3) can be interpreted geometrically as the slope of secant of $(2, 1)$ and $(x + \Delta x, f(x + \Delta x))$. For example, when $\Delta x = 0.5$, the slope of secant $(2, 1)$ and $(2.5, 2.25)$ can be obtained using (2.3) as 2.5, which is shown as the red solid line in Fig. 2.2. Other two examples when $\Delta x = 0.9$ and $\Delta x = -0.5$ are given by red dashed line and red dot-dashed line in Fig. 2.2 respectively, and their slopes can be calculated as 2.9 and 1.5 respectively using (2.3).

Apparently, the slope of the secant depends on Δx, which can be seen from both equation (2.3) and Fig. 2.2. From the figure, when $\Delta x \to 0$, the slope of the tangent at $x = 2$ can be obtained, as given by the blue dashed line in Fig. 2.2. From (2.3),

$$\lim_{\Delta x \to 0} \left.\frac{\Delta y}{\Delta x}\right|_{x=2} = 2 \quad (2.4)$$

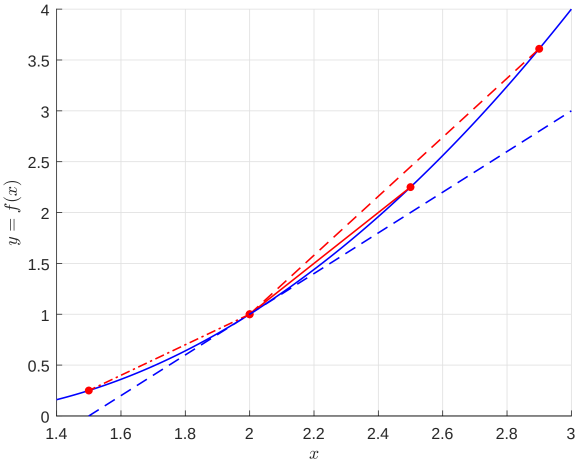

FIGURE 2.2
Slope of secant and tangent of $y = f(x)$ at $x = 2$.

which is the slope of the tangent of $y = f(x)$ at $x = 2$.

Consider $x = 0$. Variable $\Delta y = f(x + \Delta x) - f(x)$ can be obtained as given in (2.5). Different from (2.2), this time the analytical equation has two forms, depending on the sign of Δx.

$$\Delta y = f(\Delta x) - f(0) = \begin{cases} \Delta x^2 - 2\Delta x & \Delta x > 0 \\ \Delta x^2 + 2\Delta x & \Delta x < 0 \end{cases} . \tag{2.5}$$

From (2.5), the ratio $\frac{\Delta y}{\Delta x}$ can be calculated as

$$\left.\frac{\Delta y}{\Delta x}\right|_{x=0} = \begin{cases} -2 + \Delta x & \Delta x > 0 \\ 2 + \Delta x & \Delta x < 0 \end{cases} . \tag{2.6}$$

Equation (2.6) implies that

$$\begin{aligned} \lim_{\Delta x \to 0^-} \left.\frac{\Delta y}{\Delta x}\right|_{x=0} &= 2, \\ \lim_{\Delta x \to 0^+} \left.\frac{\Delta y}{\Delta x}\right|_{x=0} &= -2, \end{aligned}$$

and $\lim_{\Delta x \to 0} \left.\frac{\Delta y}{\Delta x}\right|_{x=0}$ does not exist. This can be intuitively comprehended from Fig. 2.1 as there is no tangent for the curve at $x = 0$.

Consider Q3. From what have been achieved in Q1 and Q2, we know that at different value of x, the ratio $\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$ can be different. In this example, we need to discuss three sub-cases $x < 0$, $x = 0$ and $x > 0$. Notice that since

Δx is small ($\Delta x \to 0$ is studied), we assume $x + \Delta x$ has the same sign with x when $x \neq 0$.

When $x < 0$,

$$\begin{aligned}\Delta y &= f(x+\Delta x) - f(x)\\ &= (-x-\Delta x-1)^2 - (-x-1)^2\\ &= \Delta x^2 + 2x\Delta x + 2\Delta x,\end{aligned}$$

and the ratio is

$$\frac{\Delta y}{\Delta x} = \Delta x + 2x + 2.$$

Thus,

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = 2x + 2 \tag{2.7}$$

When $x = 0$, as discussed previously, $\lim_{x\to 0} \frac{\Delta y}{\Delta x}$ does not exist.

When $x > 0$,

$$\begin{aligned}\Delta y &= f(x+\Delta x) - f(x)\\ &= (x+\Delta x-1)^2 - (x-1)^2\\ &= \Delta x^2 + 2x\Delta x - 2\Delta x,\end{aligned}$$

and the ratio is

$$\frac{\Delta y}{\Delta x} = \Delta x + 2x - 2.$$

Thus,

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = 2x - 2 \tag{2.8}$$

To summarize (2.7) and (2.8)

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \left\{\begin{array}{ll} 2x+2 & x<0 \\ 2x-2 & x>0 \end{array}\right., \tag{2.9}$$

which can be interpreted as the slope of tangent of curve $y = f(x)$ at different x in Fig. 2.1.

This motivating example shows that for a function $y = f(x)$, the ratio $\lim_{\Delta x\to 0} \frac{f(x+\Delta x)}{\Delta x}$ may change for different x, and sometimes may not even exist. The limit $\lim_{\Delta x\to 0} \frac{f(x+\Delta x)}{\Delta x}$ itself is also a function of x, which we call *the derivative of* $f(x)$.

2.2 Derivative of a Function

The formal definition of the derivative of a scalar function $f(x)$ with respect to x is given as follows.

Definition of the derivative of a function:

The derivative of $f(x)$ at $x = a$, denoted by $f'(a)$, is given as follows

$$f'(a) = \lim_{\Delta x \to 0} \frac{f(a + \Delta x) - f(a)}{\Delta x} \tag{2.10}$$

if such limit in (2.10) exists. In this case, $f(x)$ is called differentiable at $x = a$.

If $y = f(x)$, equation (2.10) can also be written as

$$\left.\frac{dy}{dx}\right|_{x=a} = f'(a),$$

where $\frac{dy}{dx}$ can be taken as an alternative denotation of $f'(x)$, and $\frac{d}{dx}$ is called the *differentiation operator*.

It is easy to prove that a necessary condition for a function $f(x)$ to be differentiable at $x = a$ is that $f(x)$ is continuous at $x = a$.

2.3 Calculation of the Derivative of a Function

The derivative of commonly seen functions are concluded in the following Table 2.1.

As two very commonly seen special cases of a^x and $\log_a x$ in Table 2.1,

$$\begin{aligned} \frac{d}{dx}e^x &= e^x, \\ \frac{d}{dx}\ln x &= \frac{1}{x}. \end{aligned}$$

TABLE 2.1
Derivative of commonly seen functions.

$f(d)$	$f'(x)$	Comments
c	0	
x^n	nx^{n-1}	$n \neq 0$
$\sin(x)$	$\cos(x)$	
$\cos(x)$	$-\sin(x)$	
a^x	$a^x \ln a$	$a > 0$
$\log_a x$	$\frac{1}{\ln a}\frac{1}{x}$	$a, x > 0$

With both $f_1(x)$ and $f_2(x)$ differentiable,

$$\begin{aligned}
\frac{d}{dx}\left(af_1(x) + bf_2(x)\right) &= a\frac{d}{dx}f_1(x) + b\frac{d}{dx}f_2(x), \\
\frac{d}{dx}\left(f_1(x)f_2(x)\right) &= f_2(x)\frac{d}{dx}f_1(x) + f_1(x)\frac{d}{dx}f_2(x), \\
\frac{d}{dx}\left(\frac{f_1(x)}{f_2(x)}\right) &= \frac{f_2(x)\frac{d}{dx}f_1(x) - f_1(x)\frac{d}{dx}f_2(x)}{\left(\frac{d}{dx}f_2(x)\right)^2}.
\end{aligned}$$

If $f = f(x)$ and $g = g(f)$, $\frac{dg}{dx}$ can be calculated using *chain rule for composite function* as

$$\frac{dg}{dx} = \frac{d}{df}g(f)\frac{d}{dx}f(x).$$

For example, if $f = 3x^2 - 2$ and $g = f^2$,

$$\frac{dg}{dx} = \frac{d}{df}g(f)\frac{d}{dx}f(x) = (2f)(6x) = 36x^3 - 24x.$$

There are a bunch of theorems not covered with much details in this notebook. For example, the famous *mean value theorem* says if a function $f(x)$ is continuous in $[a, b]$ and has a derivative everywhere in (a, b), there must be such $a < c < b$ that

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

Some of these theorems are widely used in the derivation and proof of other theorems.

3

Integral

CONTENTS

3.1 A Motivating Example 23
3.2 Integral of a Function 28
3.3 Calculation of the Integral of a Function 30

Consider a continuous function $f(x)$ defined in $[a, b]$. We want to find the "antiderivative" of $f(x)$, i.e. find a function $F(x)$ whose derivative is $f(x)$. In Section 3.1, a motivating example is given to illustrate a use case for such "antiderivative". The formal definition of integral is given in Section 3.2, and the calculation of integral for common functions are given in Section 3.3.

3.1 A Motivating Example

Consider the following motivating example where we would like to calculate the area between $y = x^2$ and $y = 0$ for $0 \le x \le 1$.

A Motivating Example

Consider

$$y = x^2, 0 \le x \le 1. \tag{3.1}$$

A plot of (3.1) is given in Fig. 3.1. To obtain the area of the red shape, approximate the red area with the sum of N trapezoids, as shown by the blue dashed curves in Fig. 3.1.

Q1: Calculate the sum of the area of the N trapezoids.

Q2: On top of Q1, let $N \to \infty$ to obtain the area of the red shape.

Q3: Given $0 \le a \le 1$, calculate the area of the trapezoids in between two vertical lines $x = 0$ and $x = a$.

Q4: Given $0 \le a \le 1$, calculate the area of the red in between two vertical lines $x = 0$ and $x = a$.

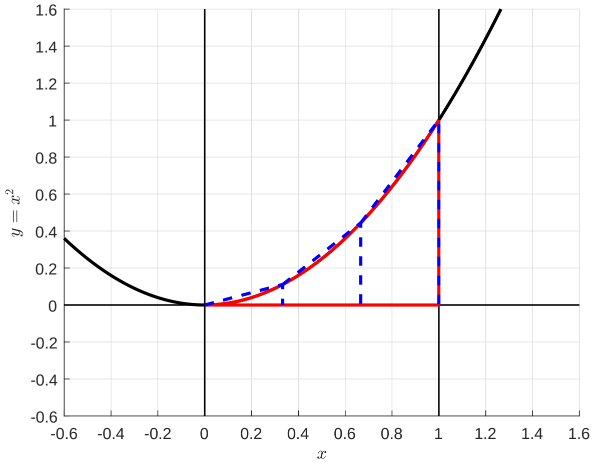

FIGURE 3.1
Plot of (3.1) and $N = 3$ trapezoids.

Notice that in Fig. 3.1, $N = 3$ trapezoids is used in the plot for clearer demonstration. In practice, more trapezoids shall be considered to get a better approximation. We can all agree on that with a larger choice of N, a better approximation can be obtained. When $N \to \infty$, the blue dashed trapezoids approaches the red shape geometrically.

The area of the i-th trapezoid can be calculated as follows. (Let the left side smallest triangle be the first trapezoid, and the right side largest trapezoid be the N-th trapezoid.)

$$\begin{aligned} \Delta x &= \frac{1}{N}, \\ S_i &= \frac{((i-1)\Delta x)^2 + (i\Delta x)^2}{2}\Delta x, \end{aligned}$$

where Δx, $((i-1)\Delta x)^2$ and $(i\Delta x)^2$ are the altitude, shorter base and longer base of the i-th trapezoid, respectively.

Therefore, the total area of the trapezoids is

$$\begin{aligned} s_T &= \sum_{i=1}^{N} S_i \\ &= \sum_{i=1}^{N} \frac{((i-1)\Delta x)^2 + (i\Delta x)^2}{2}\Delta x \\ &= \sum_{i=1}^{N} \frac{\left((i-1)\frac{1}{N}\right)^2 + \left(i\frac{1}{N}\right)^2}{2}\frac{1}{N} \\ &= \frac{\sum_{i=1}^{N}\left(2i^2 - 2i + 1\right)}{2N^3}. \end{aligned} \tag{3.2}$$

Using Table 1.3, equation (3.2) becomes

$$\begin{aligned} s_T &= \frac{\frac{N(N+1)(2N+1)}{3} - N(N+1) + N}{2N^3} \\ &= \frac{2N^3 + N}{6N^3}. \end{aligned} \tag{3.3}$$

Equation (3.3) gives the sum of the area of the trapezoids. For example, substituting $N = 3$ into (3.3) gives $s_T = 0.3519$. This illustrates the case where 3 trapezoids are used to approximate the area of the red as shown in Fig. 3.1, and the approximated area is 0.3519. With larger N, a better estimation can be obtained. For example, $N = 5$ gives $s_T = 0.3400$ and $N = 10$ gives $s_T = 0.3350$ and $N = 20$ gives $s_T = 0.3337$. A plot of using 20 trapezoids to approximate the red area is given in Fig. 3.2. Comparing the two figures 3.1 and 3.2, we can see that $N = 20$ is a fairly good approximation and it gives more accurate result than $N = 3$.

To calculate the precise red area, let N approaches infinity in (3.3) to get

$$\begin{aligned} S &= \lim_{N\to\infty} \frac{2N^3 + N}{6N^3} \\ &= \lim_{N\to\infty} \frac{2 + \frac{1}{N^2}}{6} \\ &= \frac{1}{3}. \end{aligned}$$

The calculation of the trapezoids area in between $x = 0$ and $x = a$, $0 \le a \le 1$ is tedious but not theoretically difficult. The basic idea is to discuss a in each single trapezoid individually, and finally use a piece-wise function to represent the area. The larger N is, the more segments it will require in the result.

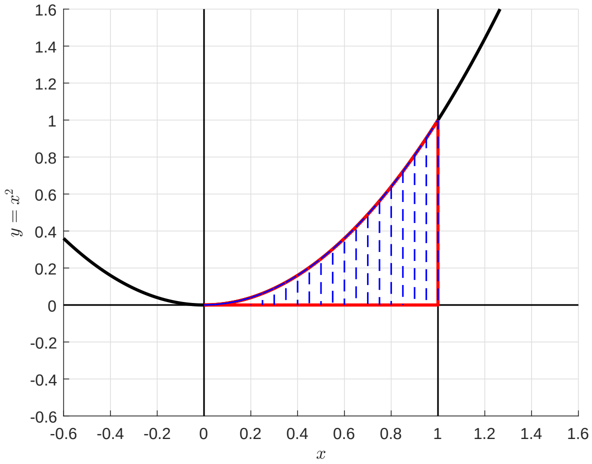

FIGURE 3.2
Use $N = 20$ trapezoids to approximate the red area.

Here we simply use $N = 3$ as an example. The result for any arbitrarily large N can be obtained similarly, just with more calculation and probably even support from a computer.

The upper edge of the trapezoids for $N = 3$ is given by 3.4. It is plotted as the solid blue line in Fig. 3.3.

$$z_3 = \begin{cases} \frac{1}{3}x & 0 \leq x \leq \frac{1}{3} \\ x - \frac{2}{9} & \frac{1}{3} < x \leq \frac{2}{3} \\ \frac{5}{3}x - \frac{2}{3} & \frac{2}{3} < x \leq 1 \end{cases} . \tag{3.4}$$

The area of the trapezoids in between $x = 0$ and $x = a$ for $N = 3$ is given by the red dashed area in Fig. 3.3, where $a = 0.5$ is used in the plot just for demonstration. The area as a function of a is given as

$$s_T^a = \begin{cases} \frac{1}{6}a^2 & 0 \leq a \leq \frac{1}{3} \\ \frac{1}{2}a^2 - \frac{2}{9}a + \frac{1}{27} & \frac{1}{3} < a \leq \frac{2}{3} \\ \frac{5}{6}a^2 - \frac{2}{3}a + \frac{5}{27} & \frac{2}{3} < a \leq 1 \end{cases} . \tag{3.5}$$

The following can be observed from (3.4) and (3.5).

- Equation (3.5) is only an approximation to the area in $x = 0$, $x = a$, $y = 0$ and $y = x^2$. This is because the blue dashed trapezoids are only an approximation to $y = x^2$ given by the black bashed line in Fig. 3.3.

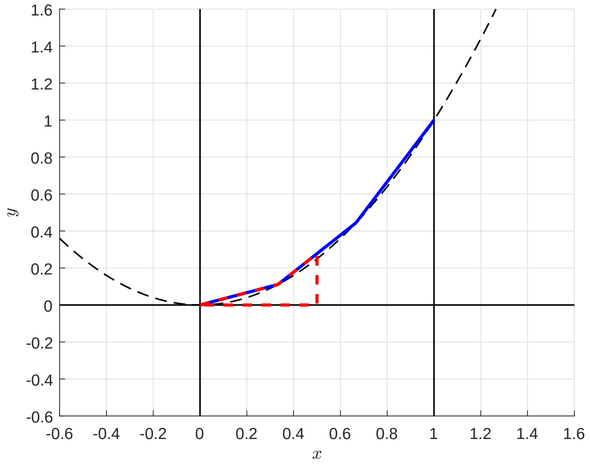

FIGURE 3.3
Calculation of the area of trapezoids. Variables $N = 3$ and $a = 0.5$ are used in the plot for demonstration.

- By increasing N, equation (3.5) gives a better and better approximation. When $N \to \infty$, the accurate area in $x = 0$, $x = a$, $y = 0$ and $y = x^2$ can be obtained.
- VERY IMPORTANT: Equation (3.4) is the derivative of (3.5) for $0 < x < 1$, if replacing the notation "a" in (3.5) with "x".

The last statement above might be difficult to comprehend and explain, but it can be verified rather easily using (3.4) and (3.5).

Notice that in this motivating example, $y = x^2$, $0 < a < 1$ and $N = 3$ are chosen for demonstration purpose without losing generality, thus, this statement shall hold true for any continuous $y = f(x)$ in (3.1) and any arbitrarily large N. We could have chosen $N \to \infty$ for any $y = f(x)$ to calculate the area in $x = a$, $x = b$, $y = 0$ and $y = f(x)$ to get precise (not approximated, since $N \to \infty$) result.

Furthermore, when $N \to \infty$, we know that the trapezoids edge given by (3.4) will perfectly overlap $y = f(x)$. And we know that (3.4) will always be the derivative of the area equation (3.5). Thus, we can derive (3.5) for the $N \to \infty$ case by simply looking for a function $F(x)$ whose derivative is $f(x)$. After that, substituting a and b into $F(b) - F(a)$ gives the precise area surrounded by $x = a$, $x = b$, $y = 0$ and $y = f(x)$.

An intuitive explanation to this statement is given as follows.

The increment of the area function $F(x)$ from any value $x = a$ to $x = a + \Delta x$ is by definition the area in between $x = a$, $x = a + \Delta x$, $y = 0$ and $y = f(x)$. Since Δx is very small (in fact, when $N \to \infty$, $\Delta x \to 0$),

$f(a+\Delta x) \to f(a)$ and this small area is $f(a) \times \Delta x$. On the other hand, the derivative of $F(x)$ at $x = a$ is defined as this increment $f(a) \times \Delta x$ divided by Δx. Therefore, the derivative of $F(x)$ at $x = a$ is $f(a)$. Alternatively, we can say *the area function $F(x)$ for $y = f(x)$ is the "antiderivative" of $f(x)$*, and in the next section we will meet its official name, the "integral".

3.2 Integral of a Function

In this section, the definitions of *definite integral* and *indefinite integral* are given. As a first step, Riemann integral is introduced. Then, Riemann integral is "translated" into the definition of definite integral.

Definition of Riemann integral:

Consider a function $f(x)$ defined on interval $[a,b]$. Let $[a,b]$ be split into N consecutive segments whose length are given by $\lambda_1, ..., \lambda_N$, with the longest segment's length being $\lambda = \max\{\lambda_1, ..., \lambda_N\}$. Let x_i be a sample randomly taken inside the i-th segment.

If for any arbitrarily small ε, there is always such δ that as long as $\lambda < \delta$,

$$\left| \sum_N f(x_i)\lambda_i - S \right| < \varepsilon,$$

for a constant S, then S is called the Riemann integral of function $f(x)$ on interval $[a,b]$.

The segment splitting and sampling used in Riemann integral is more general than what Section 3.1 has been using. In Section 3.1, we have been assuming even splitting segment $\lambda_i = \lambda_j$ for any two segments, and also determinant sampling in the segment $x_i = \underset{x}{arg f(x)} = \frac{1}{2}\left(f(x_l) + f(x_r)\right)$ where x_l, x_r are the left and right boundary of the segment respectively. By assuming continuous function $f(x)$ in Section 3.1, it is guaranteed that such x_i exists. The approach used in Section 3.1 is only a special case of Riemann integral, but it would work just fine for most of the cases.

A more intuitive definition of definite integral is given below. The idea is inherited from the motivating example in Section 3.1. It is not as "strong" as the Riemann's definition, but should be sufficient for most of the use cases.

Definition of definite and indefinite integrals in an intuitive manner:

Consider a continuous function $f(x)$ defined on interval $[a, b]$. Let $[a, b]$ be split into N consecutive segments with equal length Δx. Let x_i be a sample in the i-th segment. The *definite integral* of $f(x)$ on interval $[a, b]$ is given by the following equation

$$S = \lim_{\Delta x \to 0} \sum_{N} f(x_i)\Delta x, \tag{3.6}$$

if the right side limit exists.

In such case, we can rewrite (3.6) with the following denotations.

$$S = \int_a^b f(x)dx. \tag{3.7}$$

where a and b are called the lower and upper bound of the integral, respectively. The "dx" in (3.7) can be taken as $\Delta x \to 0$.

Equation (3.7) can be solved by finding such $F(x)$ that $\frac{d}{dx}F(x) = f(x)$, and

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

Function $F(x)$ is called the *indefinite integral* of function $f(x)$, and it is denoted by

$$F(x) = \int f(x)dx,$$

which does not come with the lower and upper bound, and it is often not unique.

The integral may not exist for some $f(x)$, or at the very least it is impossible to derive an analytical equation $F(x)$ associated with that $f(x)$. It is often easier to derive the derivative of a function, i.e. from $F(x)$ to $f(x)$, than the other way around.

Notice that in the definitions of definite and indefinite integral, continuous $f(x)$ is assumed. For those piece-wise functions that is not continuous at certain values in its range, particular caution is required, for example, to analyze it piece-by-piece considering boundary conditions.

TABLE 3.1

Indefinite integral of commonly seen functions.

$f(x)$	$F(x) = \int f(x)dx$	Comments
a	$ax + c$	
x^n	$\frac{1}{n+1}x^{n+1} + \mathrm{c}$	$n \neq -1$
x^{-1}	$\ln\lvert x\rvert + c$	$x \neq 0$
$\frac{1}{ax+b}$	$\frac{1}{a}\ln\lvert ax+b\rvert + c$	$a \neq 0$, $ax + b \neq 0$
$\sin(x)$	$-\cos(x) + c$	
$\cos(x)$	$\sin(x) + c$	
e^x	$e^x + c$	
$\ln x$	$x\ln x - x + c$	
$\frac{1}{\sqrt{1-x^2}}$	$\frac{1}{\sin(x)} + c$	$\lvert x\rvert < 1$
$\frac{1}{1+x^2}$	$\frac{\cos(x)}{\sin(x)} + c$	

In the case of Riemann integral, the assumption is more relaxed as continuity of $f(x)$ is not required. However, it is still possible that for some $f(x)$, Riemann integral does not exist. For example, Dirichlet function,

$$D(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & \text{otherwise} \end{cases}, \tag{3.8}$$

is not continuous or differentiable at any x, and it is not Riemann integrable on any interval.

3.3 Calculation of the Integral of a Function

Some commonly seen indefinite integral is given in Table 3.1. When calculating indefinite integral, there is always an arbitrary constant c in the result, as the constant in $F(x)$ does not affect the derivative $f(x)$.

If a function is similar to the functions presented in Table 3.1, its indefinite integral might be achievable.

The following rules are commonly used when calculating the integral.

$$\begin{aligned} \int f(x) + g(x) &= \int f(x)dx + \int g(x)dx + c, \\ \int af(x)dx &= a\int f(x)dx. \end{aligned}$$

The integral can be calculated by parts as follows.

$$\int u(x)v'(x)dx = uv - \int u'(x)v(x)dx,$$

where $u'(x)$, $v'(x)$ are the derivative of $u(x)$ and $v(x)$ respectively. For example, to solve the integral of $f(x) = x\sin(x)$, let $u(x) = x$, $v'(x) = \sin(x)$. From Table 3.1, we know that $u'(x) = 1$, and $v(x) = -\cos(x) + c$.

$$\begin{aligned}
\int f(x)dx &= \int u(x)v'(x)dx \\
&= u(x)v(x) - \int u'(x)v(x)dx \\
&= x\left(-\cos(x) + c_1\right) - \int \left(-\cos(x) + c_1\right) dx \\
&= -x\cos(x) + \int \cos(x)dx \\
&= -x\cos(x) + \sin(x) + c.
\end{aligned}$$

Inspired by the chain rule of calculation of derivative, the integral can be calculated by substitution as follows.

$$\int f\left(g(x)\right) dg(x) = \int f\left(g(x)\right) g'(x)dx = F\left(g(x)\right) + c,$$

where $F(x) = \int f(x)dx$.

In general, the definite integral of a continuous function can be calculated by substituting the upper and lower bound into the indefinite integral and do the following subtraction

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

Notice that sometimes $F(b) - F(a)$ is also denoted by $F(x)|_a^b$.

As illustrated in the motivating example in Section 3.1, for $a < b$, $\int_a^b f(x)dx$ can be interpreted as the accumulated area between $y = f(x)$ and $y = 0$, within boundary $x = a$ and $x = b$.

When $f(x) \geq 0$, $\int_a^b f(x)dx \geq 0$ is the area surrounded by $y = f(x)$, $y = 0$, $x = a$ and $x = b$. When $f(x) \leq 0$, $\int_a^b f(x)dx \leq 0$ is the negative of the area. Otherwise, $\int_a^b f(x)dx$ is the difference of area above and below $y = 0$ as shown in Fig. 3.4.

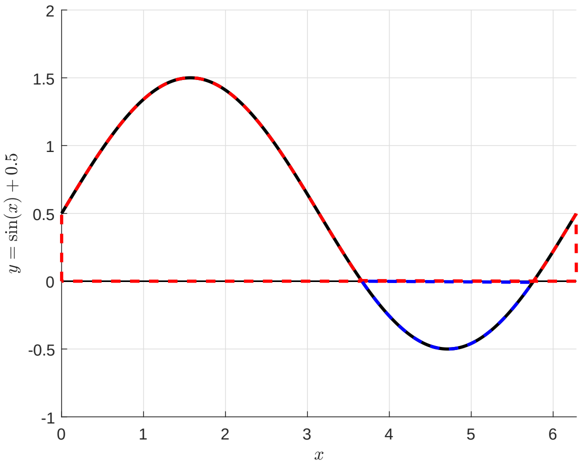

FIGURE 3.4
Plot of $f(x) = \sin(x) + 0.5$ from 0 to 2π. The area above $y = 0$ is surrounded by the red dashed line, and the area below $y = 0$ by the blue dashed line. The definite integral $\int_0^{2\pi} f(x)dx$ in this case is the red area subtracting the blue area.

4

Applications

CONTENTS

4.1 Newton's Method 33
4.2 Taylor Series 34

This chapter introduces some interesting and widely known use cases of derivatives and integrals. Section 4.1 introduces Newton's method, a widely used numerical method to solve equation $f(x) = 0$ for some continuous function $f(x)$. Section 4.2 introduces Taylor serious, a widely used numerical method to approximate the value of $f(x)$ near a specific x_0.

4.1 Newton's Method

Consider solving an equation $f(x) = 0$ for continuous function $f(x)$. Sometimes it is possible to construct a function $g(x)$, such that $x = x + g(x)$ has the same solution with $f(x) = 0$. Equation $f(x) = 0$ might then be solved recursively using $x^{k+1} = x^k + g(x^k)$ where k is the recursive index. In this notebook, we are not going to discuss how $g(x)$ can be constructed and what limitations of this method may have.

In Newton's method, $g(x)$ is constructed as $g(x) = -\frac{f(x)}{f'(x)}$. For Newton's method to converge to the correct solution, there is some restrictions to $f(x)$ and also the choice of x^0 as the initial guess. The detailed discussion to these restrictions are not covered in this notebook.

The basic procedures for Newton's method are given below.

Step 1: Determine a feasible range $[a, b]$ **where the solution to** $f(x) = 0$ **must lie inside.** This can be done by having a rough guess of an range $[a, b]$ near the solution, and make sure $f(a)f(b) < 0$. It can be proved that for a continuous function $f(x)$, $f(a)f(b) < 0$ guarantees a solution in $[a, b]$.

Step 2: Have a initial guess $x^0 \in [a, b]$**. Initialize** $k = 0$**.**

Step 3: Calculate $f'(x^k)$, then calculate x^{k+1} as follows.

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}.$$

Step 3 is iterated to for recursive calculation of x^k. The iterations stops when either of the following happens: (a) $|f(x^k)|< \varepsilon$; or (b) $|x^{k+1} - x^k|< \varepsilon$, where ε is a pre-defined threshold parameter. The finally calculated x^k is the numerical solution of the original equation $f(x) = 0$ using Newton's method.

4.2 Taylor Series

Consider a continuous and $(n+1)$-th order differentiable function $f(x)$ defined on interval $[a, b]$. Taylor series introduces a way to approach a function $f(x)$ at any $x \in [a, b]$ using a sum of sequence consisting $f(x)$ and its derivatives at a particular $x_0 \in [a, b]$.

Taylor series claims that such $f(x)$ can be expressed by the following equation

$$f(x) = P(x, x_0), +R(x, x_0), \tag{4.1}$$

where

$$\begin{aligned} P(x, x_0) &= \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k, \\ R(x, x_0) &= \frac{f^{(n+1)}\left(x_0 + \theta(x - x_0)\right)}{(n+1)!}(x - x_0)^{n+1}, \end{aligned} \tag{4.2}$$

with $\theta \in (0, 1)$, and $f^{(n)}(x)$ the n-th order derivative of $f(x)$, i.e. "the derivative of derivative of ... derivative of $f(x)$", where there are n "derivative" in the sentence.

Equation (4.1) uses $P(x, x_0)$ to approximate $f(x)$. It can be seen from (4.2) that with $n \to \infty$ or $x \to x_0$, the remainder $R(x, x_0) = f(x) - P(x, x_0)$ approaches zero (notice that $f^{(n+1)}$ is assumed bounded). This implies that the approximation performs better when higher order Taylor series is used, or when the target point x is close to the evaluated point x_0.

The following Fig. 4.1 gives an example of using Taylor series to approximate function $y = 2^x$ at $x_0 = 0$. In this example, it can be seen that the approximation gets better when higher order Taylor series is used.

For polynomial functions $f(x)$, $f^{(n)}(x) = 0$ for large enough n, depending on the degree of the polynomial. For a N-th-order polynomial, $f^{(n)}(x) = 0$ for $n > N$, thus its n-th-order (or higher) Taylor series $P(x, x_0) = f(x)$ for any choice of x_0.

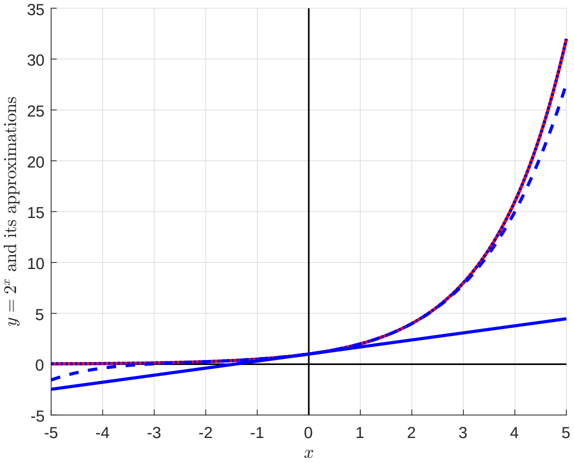

FIGURE 4.1
Plot of function $y = 2^x$ in red solid line, and its approximations using first-order, 5th-order and 10th-order Taylor series in blue solid line, blue dashed line and blue dot line respectively.

Part II

Multivariable Function, Partial Derivative and Multiple Integral

5

Multivariable Function

CONTENTS

5.1 Brief Introduction to Vector and Matrix 39
  5.1.1 Basic Concepts 40
  5.1.2 Matrix Multiplication 41
  5.1.3 Block Matrix 42
  5.1.4 Identity Matrix and Square Matrix Inverse 43
5.2 Multivariable Function 43

This chapter introduces functions with multiple inputs and/or outputs. Usually, these inputs and outputs are put into vectors for computation and presentation convenience. Section 5.1 gives a very brief introduction to the basics of vector and matrix operations. Section 5.2 introduces the concept of multivariable functions, including the multiple input function and the vector function.

From calculus perspective, this chapter clears the preliminary knowledge required for later Chapters 6 and 7.

5.1 Brief Introduction to Vector and Matrix

Detailed introduction to vector and matrix can be found in any *linear algebra* textbook, where there are the geometric interpretation of vector, then linear equation represented by product of matrix and vector, then column and null space of matrix, then rank of matrix and determinant of square matrix, then inverse of matrix, then linear transformation of matrix, then eigenvalue of matrix, then norm of vector and matrix, then algebraic Riccati equation, and so on. This list can go almost eternally. To make things even more complicated, depending on the application, the vector and matrix may have different physical meanings. For example, the speed of motion and the bus voltage phasors of a microgrid can all be represented by vectors, but might be completely two different things.

In the context of this notebook, however, most of the above are out of the scope. We will take vector and matrix as a way of organizing scalar data. Basically, a vector is a 1-D chain of scalar variables, and a matrix is a 2-D rectangular mesh of scalar variables. In many cases such as calculating matrix product, a vector can be taken as a special case of matrix, and a scalar is a special case of a vector. Of course, the very basics such as product of matrices are still required.

The preliminary vector and matrix knowledge used in this notebook is summarized in the rest of this section as follows.

5.1.1 Basic Concepts

A vector x (sometimes denoted as bold text $\mathbf{x}$ in textbooks) is a finite sequence of scalar variables organized in a 1-D chain. The length of the vector, or the dimension of the vector, is the number of elements in the vector. For vector x with n elements, the elements are denoted by $x_1, x_2, \cdots, x_n$, and x_i is called the i-th element of the vector.

Most of the vectors in this notebook are by default *column vectors*, which means that the n elements are put into n-row-1-column, as follows. In this case, we say *"x is a n dimensional column vector"* or *"x is a $n \times 1$ vector"*.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \tag{5.1}$$

A row vector, on the other hand, puts the n elements into 1-row-n column. A row vector is like a column vector flipped diagonally, as shown below in (5.2). We call the flipping operation *"transpose"*, denoted by $(\cdot)^T$ (used in this textbook) or $(\cdot)'$.

$$x^T = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}, \tag{5.2}$$

where x^T is a row vector and it is a transpose of the column vector x previously given in (5.1). The transpose of a column vector is a row vector, and vise versa. The transpose of a scalar is itself.

A matrix A (sometimes denoted as bold $\mathbf{A}$ in textbooks) is a finite number of scalar variables organized in 2-D mesh, with each scalar taking a particular position. The number of rows and columns are the dimension of the matrix. For example, if A has m rows and n columns, we say *"A has a dimension of $m \times n$"* or *"A is a $m \times n$ matrix"*, shown as follows.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \ldots & a_{m,n} \end{bmatrix}, \tag{5.3}$$

where $a_{i,j}$ is the i-th row, j-th column element of A.

The transpose operation is also defined on matrix. By applying transpose on A in (5.3), a $n \times m$ matrix A^T can be obtained, where the i-th row, j-th column element in the transpose matrix A^T is the j-th row, i-th column element in the original matrix A.

The vectors or matrices with the same dimension can be added together by adding each associated pair of elements together.

5.1.2 Matrix Multiplication

The product of two matrices is defined as follows.

Matrix Multiplication:

Consider matrices A and B. As a prerequisite of calculating AB, the number of column in the first matrix A must equal to the number of row in the second matrix B.

Let A and B be $m \times p$ matrix and $p \times n$ matrix respectively. The matrix product $C = AB$ is a $m \times n$ matrix with each element calculated by

$$\begin{aligned} c_{i,j} &= a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \ldots + a_{i,p}b_{p,j} \\ &= \sum_{k=1}^{p} a_{i,k}b_{k,j}, \end{aligned}$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

It is clearly from the definition that AB does not equal to BA. In the example above, if $m \neq n$, BA does not exist in the first place.

It can be proved by definition that if $C = AB$, $C^T = (AB)^T = B^T A^T$.

In terms of matrix multiplication, a vector is just a special case of matrix. Therefore, the product of a matrix with a vector $y = Ax$ where A is a $m \times n$

matrix and x a $n \times 1$ matrix, is a $m \times 1$ vector given by

$$\begin{aligned} y &= Ax \\ &= \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \begin{bmatrix} \sum_{k=1}^{n} a_{1,k}x_k \\ \sum_{k=1}^{n} a_{2,k}x_k \\ \vdots \\ \sum_{k=1}^{n} a_{m,k}x_k \end{bmatrix}. \end{aligned}$$

The product of row vector u^T and column v, both n dimensional vectors, is

$$\begin{aligned} u^T v &= \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\ &= \sum_{k=1}^{n} u_k v_k \end{aligned}$$

which is a scalar. As a special case, $x^T x = \sum_{k=1}^{n} x_i^2$ for a n dimensional vector x.

5.1.3 Block Matrix

For the convenience of calculation and interpretation, sometimes a large dimension matrix is split into a combination of smaller dimension sub matrices.

For example, consider matrix A with dimension $m \times p$ and matrix B with dimension $p \times n$. Matrix A can be split into two sub matrix A_1 and A_2, where A_1 consists of the first m_1 rows of A thus $m_1 \times p$ dimension, and A_2 consists of the rest $m_2 = m - m_1$ rows of A thus $m_2 \times p$ dimension, i.e.

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}.$$

The calculation of $C = AB$ can be done as follows

$$C = AB = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} B = \begin{bmatrix} A_1 B \\ A_2 B \end{bmatrix}.$$

Similarly, if split matrix B into two sub matrices, with B_1 the first n_1 columns and B_2 the rest $n_2 = n - n_1$ columns of B, then

$$C = AB = A \begin{bmatrix} B_1 & B_2 \end{bmatrix} = \begin{bmatrix} AB_1 & AB_2 \end{bmatrix}$$

Furthermore, splitting both A and B simultaneously gives

$$C = AB = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \begin{bmatrix} B_1 & B_2 \end{bmatrix} = \begin{bmatrix} A_1B_1 & A_1B_2 \\ A_2B_1 & A_2B_2 \end{bmatrix} \tag{5.4}$$

Equation (5.4) sometimes helps to speed up the matrix product as it allows to split the calculation into independent pieces. But more importantly, equation (5.4) gives a lot of insights into matrix operations and linear transformation. The details are not covered in this notebook.

5.1.4 Identity Matrix and Square Matrix Inverse

A matrix is called a *square matrix* if it has the same number of rows and columns. For example, if matrix A has a dimension of $n \times n$, then it is a square matrix of dimensional n. The elements with the same row and column index, i.e. $a_{i,i}, i = 1, \ldots, n$, are called the diagonal elements.

The *identity matrix*, denoted by I, is a special type of square matrix as given in (5.5). Its diagonal elements are 1, with the rest elements 0.

$$I = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} \tag{5.5}$$

A n dimensional identity matrix is denoted by I_n. The multiplication of the identity matrix with any matrix from either left or right side does not change the matrix, i.e. for any matrix A with dimension $m \times n$, $I_mA = AI_n = A$.

A square matrix may have an associated inverse matrix. The definition of inverse matrix is given as follows.

Definition of inverse of a matrix:

Consider a square matrix A with dimension $n \times n$. Matrix A is called *invertible* if such matrix A^{-1} with dimension $n \times n$ exists that

$$AA^{-1} = A^{-1}A = I_n,$$

and A^{-1} is called the *inverse* of matrix A.

Notice that matrix A^{-1} may not exist for some A, depending on the determinant of A. Details can be found in linear algebra textbooks and are not given in this notebook.

5.2 Multivariable Function

In Chapters 2 and 3, we have been considering single-input-single-output functions only, i.e. for function $y = f(x)$, we have been discussing only the cases where y and x are scalars so far.

In many other cases, however, a function may have multiple input and/or output variables. For example, consider the following function used to calculate the mechanical energy of an object

$$e = f(v, h) = \frac{1}{2}mv^2 + mgh.$$

where m, v and h are the mass, motion speed and height (related to ground) of the object, and g is the free-fall acceleration, $g = 9.8m/s^2$ on the earth. This is a typical multivariable function, where the function $z = f(x, y)$ depends on multiple independent variables x and y.

When there are multiple outputs for a function, the outputs are often out into a column vector and the function is called a *vector function*. Two examples of vector functions are given below.

$$\begin{aligned} f(x) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 3 & -2 \end{bmatrix} x, & (5.6) \\ g(x) &= \begin{bmatrix} g_1(x) \\ g_2(x) \end{bmatrix} = \begin{bmatrix} x_1^2 + x_2^2 \\ 0.5x_1 + e^{x_2} \end{bmatrix}, & (5.7) \end{aligned}$$

where $f(x)$ and $g(x)$ are both vectors with dimension 3×1 and 2×1 respectively. Equation (5.6) is a linear function with 2 inputs $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$ and 3 outputs $y = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}^T$, while (5.7) is a nonlinear function with 2 inputs $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$ and 2 outputs $g = \begin{bmatrix} g_1 & g_2 \end{bmatrix}^T$.

6

Partial derivative

CONTENTS

6.1 A Motivating Example 45
6.2 Partial Derivative 48
6.3 Gradient 51
  6.3.1 Motivating Example 51
  6.3.2 Gradient of a Multivariable Function 53
6.4 Jacobian Matrix 53

Partial derivative studies the effect of a small deviation of one particular independent variable on the multivariable function. It is similar with the normal derivative in many ways, but also has its unique characteristics.

In Section 6.1, a motivating example is given to illustrate the motivation of introducing partial derivative. In Section 6.2, the definition of partial derivative is given. In Sections 6.3 and 6.4, two very important and commonly used tool derived from partial derivative, namely gradient and Jacobian matrix, are introduced respectively.

6.1 A Motivating Example

Consider the following motivating example where $y = f(x_1, x_2)$ is a multivariable function with 2 inputs.

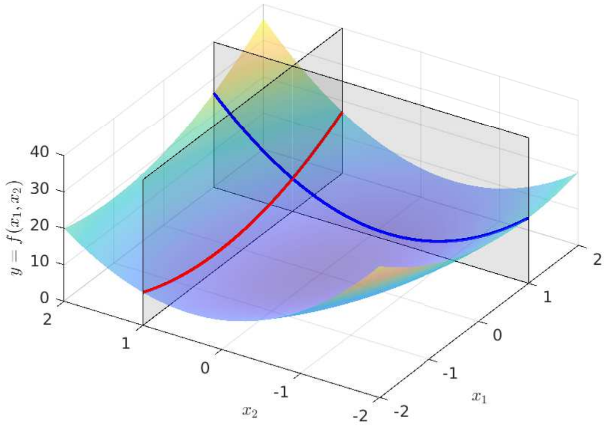

FIGURE 6.1
Plot of function $y = f(x_1, x_2)$ in 3-D.

A Motivating Example

Consider

$$y = 2x_1^2 + x_2^2 + 2x_1x_2. \tag{6.1}$$

Q1: Let $x_2 = 1$ be a constant. Derive y as a function of x_1, and calculate its derivative with respect to x_1. Similarly, let $x_1 = 1$ be a constant and derive y as a function of x_2, and calculate its derivative with respect to x_2.

Q2: At $(x_1, x_2) = (1, 1)$, consider small vibrations Δx_1 and Δx_2. Approximate Δy as a function of Δx_1 and Δx_2 using differentiation.

Q3: Find such x_1 and x_2 that y is minimized.

Equation (6.1) can be plot in 3-D as Fig. 6.1.

Let $x_2 = 1$ be constant. Equation (6.1) becomes

$$y = f(x_1, 1) = 2x_1^2 + 2x_1 + 1,$$

which is given by the intersection of curved surface (equation (6.1)) and the vertical flat ($x_2 = 1$) given by the red solid line in Fig. 6.1. Its derivative with respect to x_1 can be easily obtained as

$$\frac{d}{dx_1} f(x_1, 1) = 4x_1 + 2. \tag{6.2}$$

Similarly, at $x_1 = 1$, (6.1) becomes

$$y = f(1, x_2) = x_2^2 + 2x_2 + 2,$$

which is given by the blue solid line in Fig. 6.1. Its derivative with respect to x_2 is

$$\frac{d}{dx_2}f(1,x_2) \quad = \quad 2x_2+2. \tag{6.3}$$

Consider small vibrations Δx_1 and Δx_2. Firstly, let x_2 remain constant and only apply Δx_1 on $y=f(1,1)$. In this case,

$$\Delta y = f(1+\Delta x_1,1)-f(1,1) \approx \frac{d}{dx_1}f(x_1,1)\Delta x_1, \tag{6.4}$$

where $\frac{d}{dx_1}f(x_1,1)$ is given in (6.2).

On top of (6.4), consider vibration Δx_2. The derivative of function $f(1+\Delta x_1,x_2)$ with respect to x_2 depends on Δx_1 and it is generally unknown without specifying Δx_1. However, since (6.1) is continuous and "smooth" (Notice that "smooth" has specific definition in mathematics. Here, just take its literal meaning: it is perfectly polished without any edges or spikes.), we can intuitively understand that when Δx_1 is small, $\frac{d}{dx_2}f(1+\Delta x_1,x_2) \approx \frac{d}{dx_2}f(1,x_2)$, and $\frac{d}{dx_2}f(1+\Delta x_1,x_2) \to \frac{d}{dx_2}f(1,x_2)$ as $\Delta x_1 \to 0$. Thus, we have

$$\begin{aligned}\Delta y &= (f(1+\Delta x_1,1+\Delta x_2)-f(1+\Delta x_1,1))+(f(1+\Delta x_1,1)-f(1,1)) \\ &\approx \frac{d}{dx_2}f(1,x_2)\Delta x_2+\frac{d}{dx_1}f(x_1,1)\Delta x_1. \end{aligned}\tag{6.5}$$

In (6.5), consider $\Delta x_1 \to 0$ and $\Delta x_2 \to 0$. Denote such Δx_1 and Δx_2 as dx_1 and dx_2 respectively, and the associated Δy as dy. Here "$d(\cdot)$" represents the *infinitesimal change*. Equation (6.5) then becomes

$$dy \quad = \quad \frac{d}{dx_2}f(1,x_2)dx_2+\frac{d}{dx_1}f(x_1,1)dx_1. \tag{6.6}$$

Equation (6.6) can be extended to more general cases where $(x_1,x_2)=(1,1)$ is not specified. The terms $\frac{d}{dx_1}f(x_1,1)$ and $\frac{d}{dx_2}f(1,x_2)$ needs to be changed accordingly. For example, $\frac{d}{dx_1}f(x_1,1)$ given by (6.2) shall be changed to

$$\left.\frac{d}{dx_1}f(x_1,x_2)\right|_{x_2 \text{ is constant}} \quad = \quad 4x_1+2x_2$$

with x_2 being any arbitrary constant.

The denotation $\left.\frac{d}{dx_1}f(x_1,x_2)\right|_{x_2 \text{ is constant}}$ can sometimes become ambiguous if not handled carefully. One of the reasons could be that "$d(\cdot)$" is used as infinitesimal change as shown before. In this case both dx_1 and dx_2 contribute to $df(x_1, x_2)$, but we would want $\frac{d}{dx_1}f(x_1, x_2)$ here to reflect the deviation of $f(x_1, x_2)$ caused by dx_1 only, and it is not convenient to list down all the other variables and put "is/are constant" everywhere in the equation. Therefore, instead of saying $\left.\frac{d}{dx_1}f(x_1, x_2)\right|_{x_2 \text{ is constant}}$, we denote

$$\frac{\partial}{\partial x_1}f(x_1, x_2) = 4x_1 + 2x_2.$$

The operator $\frac{\partial}{\partial x}$ is called *partial derivative* (with respect to x), often followed by a multivariable function $f(x, y)$ where x is one of its inputs. It is similar to the derivative, but emphasizing that the interested function is multivariable, and the derivative of this function with respect to ONLY ONE variable is being studied (and the rest variables assumed constant). Since $\partial f(x)$ is rarely or never used as the infinitesimal of $f(x)$, it will hopefully be less ambiguous.

The formal definition of partial derivative is given in next Section 6.2.

Equation (6.6) for general (x_1, x_2), therefore, becomes

$$dy = \frac{\partial}{\partial x_2}f(x_1, x_2)dx_2 + \frac{\partial}{\partial x_1}f(x_1, x_2)dx_1. \tag{6.7}$$

The minimum $y = f(x_1, x_2)$ and its associated x_1 and x_2 can be found by rewriting (6.1) as

$$y = x_1^2 + (x_1 + x_2)^2 .$$

Therefore, the minimum y is $y = 0$, at $x_1 = 0$ and $x_1+x_2 = 0$, i.e., $x_1 = x_2 = 0$.

This can also be solved using (6.7). At the minimum point of y, $\frac{\partial}{\partial x_2}f(x_1, x_2)$ and $\frac{\partial}{\partial x_1}f(x_1, x_2)$ must be both zero. Otherwise, it is always possible to add/subtract a small Δx_1 or Δx_2 to further decrease y. Thus, from (6.1)

$$\begin{aligned} \frac{\partial}{\partial x_1}f(x_1, x_2) &= 4x_1 + 2x_2 = 0 \\ \frac{\partial}{\partial x_2}f(x_1, x_2) &= 2x_1 + 2x_1 = 0 \end{aligned}$$

which yields $x_1 = x_2 = 0$.

6.2 Partial Derivative

In the case of single input scalar function $f(x), x \in \mathbb{R}$, there is no point to define partial and total derivative, as the derivative with respect to that single variable by itself is the total derivative.

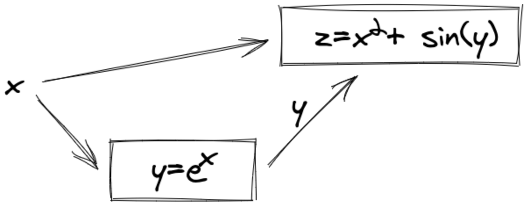

FIGURE 6.2
The calculation of z using x and y.

In the case of multivariable function with multiple inputs, $f(x), x \in \mathbb{R}^{R\times 1}$, the definition partial derivative is given below.

Definition of Partial Integral:

Consider function $y = f(x_1, ..., x_n)$ where y is the scalar output and $x = [x_1, ..., x_n]^T$ is a $n \times 1$ vector input. The partial derivative of $y = f(x)$ with respect to x_i is given by

$$\frac{\partial}{\partial x_i} f(x) = \lim_{\Delta x_i \to 0} \frac{f(x_1, ..., x_i + \Delta x_i, ..., x_n) - f(x_1, ..., x_n)}{\Delta x_i}, \tag{6.8}$$

with $x_1, ..., x_{i-1}, x_{i+1}, ..., x_n$ remaining constant, and only x_i is allowed to vary.

Here is another example to help better understand partial derivative. Consider $z = f(x, y)$ a multivariable function of both inputs x, y as follows

$$z = f(x, y) = x^2 + \sin(y), \tag{6.9}$$

where $y = g(x)$ is a single input function of x as follows

$$y = g(x) = e^x. \tag{6.10}$$

Apparently, the value of z ultimately depends on x alone, but in the intermediate calculation, z depends on both x and y. The calculation flow is visualized in Fig. 6.2.

In this case, the total derivative $\frac{d}{dx} f(x, y)$ can be expressed as follows

$$\frac{d}{dx} f(x, y) \quad = \quad \frac{\partial}{\partial x} f(x, y) + \frac{\partial}{\partial y} f(x, y) \frac{d}{dx} y. \tag{6.11}$$

where in this example,

$$\begin{aligned}\frac{\partial}{\partial x}f(x,y) &= 2x \\ \frac{\partial}{\partial y}f(x,y) &= \cos(y) \\ \frac{d}{dx}y &= e^x\end{aligned}$$

and finally

$$\frac{d}{dx}f(x,y) = 2x + \cos(e^x)e^x$$

Sometimes (6.11) is rewritten as follows

$$\begin{aligned}df(x,y) &= \frac{\partial}{\partial x}f(x,y)dx + \frac{\partial}{\partial y}f(x,y)dy \\ dy &= \left(\frac{d}{dx}y\right)dx\end{aligned}$$

where $d(\cdot)$ represents the infinitesimal change of a variable.

Equation (6.11) serves as a good example to show the relationship and difference between the total derivative $\frac{d}{dx}f(x,y)$ and the partial derivative $\frac{\partial}{\partial x}f(x,y)$. When calculating total derivative, all variables that would affect the value of the function must be taken into consideration, while when calculating partial derivative, only one input variable is studied, with the rest variables remaining constant.

For a function with 1 output and n inputs, sometimes it is convenient to put the partial derivative to each input in a vector. For example, for $y = f(x)$ where $x = [x_1, ..., x_n]^T$, denote

$$\frac{\partial}{\partial x}f(x) = \left[\begin{array}{ccc}\frac{\partial}{\partial x_1}f(x) & \ldots & \frac{\partial}{\partial x_n}f(x)\end{array}\right] \tag{6.12}$$

as the *scalar-by-vector derivative*. Notice that (6.12) is usually given as a row vector. In some research papers the result is given as a column vector, i.e. the transpose of (6.12).

For a function with m outputs and 1 input $y = f(x)$ where $y = [y_1, ..., y_m]^T = [f_1(x), ..., f_m(x)]^T$, its *vector-by-scalar derivative* is given by

$$\frac{\partial}{\partial x}f(x) = \left[\begin{array}{c}\frac{\partial}{\partial x}f_1(x) \\ \vdots \\ \frac{\partial}{\partial x}f_m(x)\end{array}\right]. \tag{6.13}$$

And finally for a function with m outputs and n inputs $y = f(x)$ where $y = [y_1, ..., y_m]^T = [f_1(x), ..., f_m(x)]^T$ and $x = [x_1, ..., x_n]^T$, the *vector-by-vector* derivative is given by

$$\frac{\partial}{\partial x} f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f_1(x) & \dots & \frac{\partial}{\partial x_n} f_1(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} f_m(x) & \dots & \frac{\partial}{\partial x_n} f_m(x) \end{bmatrix}. \tag{6.14}$$

Partial derivative and the above equations (6.12), (6.13) and (6.14) have many applications. Two of those most popular use cases are *gradient* and *Jacobian matrix*. Since they are so important and widely used, it is worth especially introducing them in specific Sections 6.3 and 6.4.

6.3 Gradient

For a scalar-valued multivariable function $y = f(x)$ where y is a scalar and x is a vector $x = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \in \mathbb{R}^n$, the gradient studies the "direction" of Δx in $\mathbb{R}^n$ space that causes y to increase/decrease the fastest.

The gradient of such function $f(x)$ is a vector function of x, denoted by $\nabla f(x)$. Notice that $\nabla f(x) \in \mathbb{R}^n$ is a vector in the same space with x, since it indicates a "direction" of x.

Section 6.3.1 gives a motivating example of gradient, and Section 6.3.2 gives its formal definition.

6.3.1 Motivating Example

The following motivating example helps to illustrate the calculation and use case of gradient.

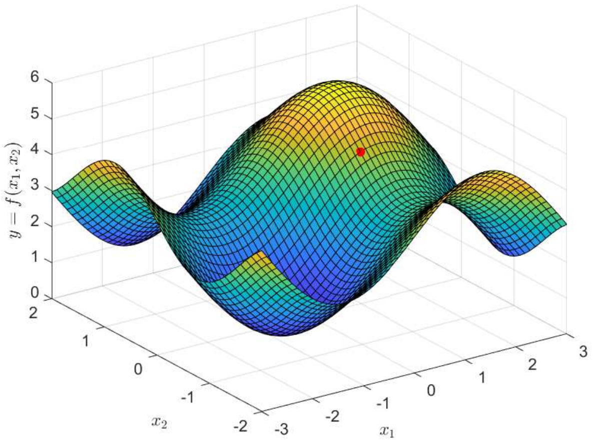

FIGURE 6.3
Plot of $y = f(x_1, x_2)$ in 3-D.

A Motivating Example

Consider the following function $y = f(x_1, x_2)$

$$y = f(x_1, x_2) = 2\sin(x_1) + \sin\left(\frac{x_1}{2} + \pi\right) + \sin(2x_2) \tag{6.15}$$

where $x_1 \in [-3, 3]$ and $x_2 \in [-2, 2]$. The 3-D plot and the contour line of the function are given in Figs 6.3 and 6.4 respectively.

The objective is to obtain such x_1 and x_2 to maximize y. By applying a glomal optimization on (6.15), $x_1 = 1.3764$, $x_2 = 0.7854$ can be obtained. However, in this example we are assuming that the global optimization is not applicable due to the lack of information or computational power. Try to obtain the target coordinate x_1 and x_2 without using global optimization.

We are using iterations to find the maximum point of y and associated x_1 and x_2 as follows.

Step 1: Start at initial point $x^0 = [1, 0]$, which is marked by a red dot on both Figs. 6.3 and 6.4.

Step 2: Determine a small change $\Delta x^k = [\Delta x_1^k, \Delta x_2^k]$. The length of this small change $|\Delta x^k| = \sqrt{\left(\Delta x_1^k\right)^2 + \left(\Delta x_2^k\right)^2} = \lambda$ is a constant. Determine such $[\Delta x_1^k, \Delta x_2^k]$ that maximize $f(x_1^k + \Delta x_1^k, x_2^k + \Delta x_2^k)$.

Step 3: Let $x^{k+1} = x^k + \Delta x^k$.

Steps 2 and 3 are executed recursively until $f(x_1^{k+1}, x_2^{k+1}) \approx f(x_1^k +$

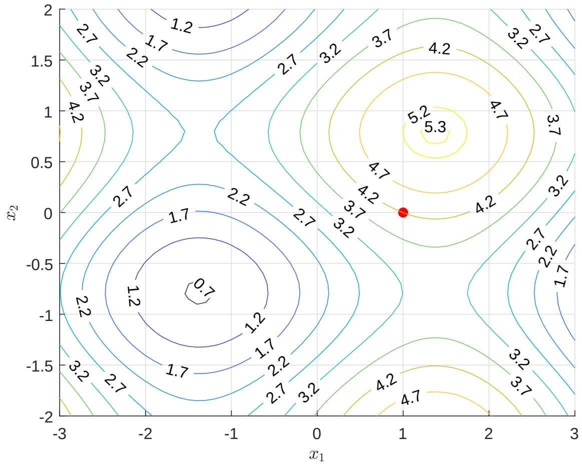

FIGURE 6.4
Contour line of $y = f(x_1, x_2)$.

$\Delta x_1^k, x_2^k)$. In this way, hopefully x^k can slowly "climb up" to the maximum point of y in Fig. 6.3.

6.3.2 Gradient of a Multivariable Function

"nobreak

6.4 Jacobian Matrix

7

Multiple Integral

CONTENTS

7.1 A Motivating Example 55
7.2 Multiple Integral 55

“nobreak

7.1 A Motivating Example

“nobreak

7.2 Multiple Integral

8

Applications

CONTENTS

8.1 Neural Network Back-propagation 57
8.2 Bayesian Inference 57

“nobreak

8.1 Neural Network Back-propagation

“nobreak

8.2 Bayesian Inference

Part III

Differential Equation

Part IV

Functional and Calculus of Variations

Bibliography

[1] James Stewart. *Calculus Metric Version Eighth Edition*. Cengage Learning, 2015.

[2] Gibert Strang. *Calculus*. Wellesley-Cambridge Press, 1991.