

3

MATHEMATICAL REVIEW

3.1 INTRODUCTION

The study of numerical algorithms requires a certain knowledge of Arithmetic, Algebra, and Analysis. In Chapter 2 we covered some of the Arithmetic needed to analyze computer numbers systems. This chapter is a very brief review of some of the Algebra and Mathematical Analysis we will need in subsequent chapters. Linear Algebra will be reviewed in Chapter 5.

The first four sections of this chapter should be familiar to all students. The 5th and 6th sections, on Sequences in Normed Linear Spaces, and Contraction Mappings and Successive Approximation, may not be so familiar. We will spend most of the time on these.

3.2 SEQUENCES, SERIES, AND LIMITS

Definition 3.1. *Geometric Series.* A series of the form

$$a + ar + ar^2 + \dots + ar^k + \dots \quad (3.1)$$

is called a *geometric series*. The ratio of successive terms is the constant r .

We want to derive an expression for the sum of the first n terms S_n and then use this to find the expression for the infinite series. Both expressions are useful.

$$\begin{aligned} S_n &= a + ar + ar^2 + ar^3 \dots + ar^{n-1}. && \text{Multiplying both sides by } r \text{ we get} \\ rS_n &= ar + ar^2 + ar^3 \dots + ar^{n-1} + ar^n. \end{aligned}$$

Subtracting the last expression from the previous one gives

$$(1 - r)S_n = a - ar^n = a(1 - r^n).$$

If $r \neq 1$ we may divide both sides by $(1 - r)$, giving

$$S_n = \sum_{k=0}^n ar^k = \begin{cases} \frac{a(1-r^{n+1})}{1-r}, & r \neq 1 \\ na, & r = 1 \end{cases} \quad (3.2)$$

It is obvious that if $|r| < 1$ then $r^n \rightarrow 0$ as $n \rightarrow \infty$. Therefore,

$$\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} \frac{a(1-r^{n+1})}{1-r} = \frac{a}{1-r}, \text{ if } |r| < 1. \quad (3.3)$$

Example 3.1.

$$S_n(x) = x^2 + \frac{x^2}{1+x^2} + \cdots + \frac{x^2}{(1+x^2)^n} \quad (3.4)$$

This is a geometric series with ratio $r = 1/(1+x^2)$. Hence

$$S(x) = \lim_{n \rightarrow \infty} S_n(x) = \frac{x^2}{1 - 1/(1+x^2)} = 1 + x^2, \text{ if } \left| \frac{1}{1+x^2} \right| < 1, \text{ or } x \neq 0.$$

The function $S(x)$ seems well-behaved. We have $\lim_{x \rightarrow 0} S(x) = 1$, but if we use (3.4) we get $S(0) = 0$. Hence this function has a discontinuity at $x = 0$.

Theorem 3.1. CONVERGENT SERIES. *If the series*

$$\sum_{k=1}^{\infty} u_k = u_1 + u_2 + \cdots + u_k + \cdots$$

converges, then

$$\lim_{k \rightarrow \infty} u_k = 0.$$

The converse of this theorem is not true :

$$\text{The Harmonic Series } \sum_{k=1}^{\infty} \frac{1}{k} \text{ diverges although } \lim_{k \rightarrow \infty} \frac{1}{k} = 0.$$

Theorem 3.2. THE p SERIES . *The series*

$$\sum_{k=1}^{\infty} \frac{1}{k^p} \quad (3.5)$$

converges if $p > 1$ and diverges if $p \leq 1$. It is not necessary that p be an integer.

3.2.1 The Binomial Theorem

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}, \text{ integer } n \geq 0. \quad (3.6)$$

A more useful form is found if we divide both sides by x^n and set $z = y/x$. This gives

$$(1+z)^n = \sum_{k=0}^n \binom{n}{k} z^{n-k}, \text{ integer } n \geq 0. \quad (3.7)$$

Problem 3.1. Use the Binomial theorem to show that

$$\sum_{k=1}^{n-1} k^n < n^n. \quad (3.8)$$

Problem 3.2. Using (3.8), show that

$$\sum_{k=1}^{n-1} \left(\frac{k}{n}\right)^n \quad (3.9)$$

converges to a limit $l \in (0, 1)$. Calculate the first 3 decimal digits of this limit.

3.2.2 Harmonic Numbers

Harmonic numbers, although rarely used in ‘standard’ mathematics, occur quite frequently in the study of algorithms and combinatorics.

The series

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{k} + \cdots$$

is called the Harmonic series.

The Harmonic Number H_n is

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = \sum_{k=1}^n \frac{1}{k}, \quad \text{integer } n \geq 1. \quad (3.10)$$

The Harmonic series diverges to ∞ . This was first proved by Nicole d’Oresme, Bishop of Lisieux, (c. 1323–1382), as follows :

$$\begin{aligned} \lim_{n \rightarrow \infty} H_n &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \left(\frac{1}{9} + \frac{1}{10} + \frac{1}{11} + \frac{1}{12} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15} + \frac{1}{16}\right) + \cdots \\ &> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \left(\frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16}\right) + \cdots \\ &= 1 + \frac{1}{2} + \frac{2}{4} + \frac{4}{8} + \frac{8}{16} + \cdots \\ &= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots, \end{aligned}$$

which obviously diverges to ∞ . There are many other proofs but this is still the simplest.

We can obtain bounds on H_n by noting that the staircase function $1/k$ is bounded below by the continuous function $1/x$ and above by $1/(x-1)$, as shown in Figure 3.1. Thus the area under $1/x$ is a lower bound on H_n , the area under $1/(x-1)$ is an upper bound, and we get

$$\int_1^n \frac{1}{x} dx = \ln n < H_n < \ln n + 1 = 1 + \int_2^n \frac{1}{x} dx. \quad (3.11)$$

Now $\lim_{n \rightarrow \infty} \ln n = +\infty$. Therefore $\lim_{n \rightarrow \infty} H_n = +\infty$, another proof that H_n diverges.

Using asymptotic expansions ?? get

$$H_n = \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{\epsilon_n}{120n^4}, \quad 0 < \epsilon_n < 1$$

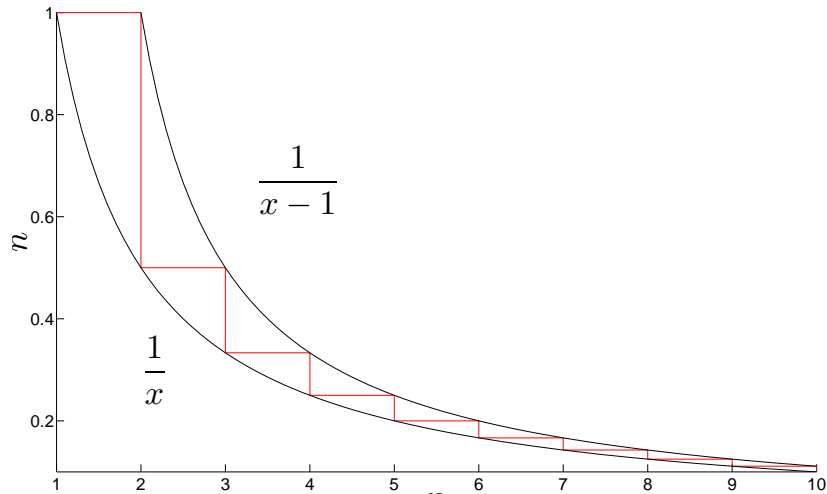


Figure 3.1 : Bounds on Harmonic Series

which is very accurate for large n . Using this formula, with $\epsilon_n = 1$, we get

$$H_{1,000,000} = 14.3927267228657236313811275$$

We can see that the Harmonic series grows very slowly towards ∞ .

Euler's Constant γ

Euler, among many other things, showed that

$$\lim_{n \rightarrow \infty} (H_n - \ln n) = \gamma = 0.5772156649\dots \tag{3.12}$$

which shows that H_n lies at about 0.58 of the distance between $\ln n$ and $\ln n + 1$. The constant γ is called *Euler's Constant*.

This constant γ is something of a mystery because it is not known whether it is rational, irrational, or transcendental. Indeed, if it is rational, i.e., $\gamma = p/q$, then it has been shown that, of necessity, $q > 10^{242,080}$, i.e, an integer with at least 242,080 digits. To add to the mystery, it seems to be harder to compute than π , a transcendental number.

Here are the first 218 digits of γ ,

$\gamma = 0.57721566$
 4901532860 6065120900 8240243104 2159335939 9235988057 6723488486 7726777664
 6709369470 6329174674 9514631447 2498070824 8096050401 4486542836 2241739976
 4492353625 3500333742 9373377376 7394279259 5258247094 9160087352 0394816567

3.2.3 Bounding a Series by Integrals

We saw an example of such a bound in equation(3.11). Here is a generalization of the bounds :

If $f(x)$ is a bounded, real *decreasing* function for $x \geq 1$ then the series

$$S_n = \sum_{k=1}^n f(k) \quad (3.13)$$

is bounded above and below as follows :

$$\int_1^{n+1} f(x) dx \leq \sum_{k=1}^n f(k) \leq f(1) + \int_1^n f(x) dx \quad (3.14)$$

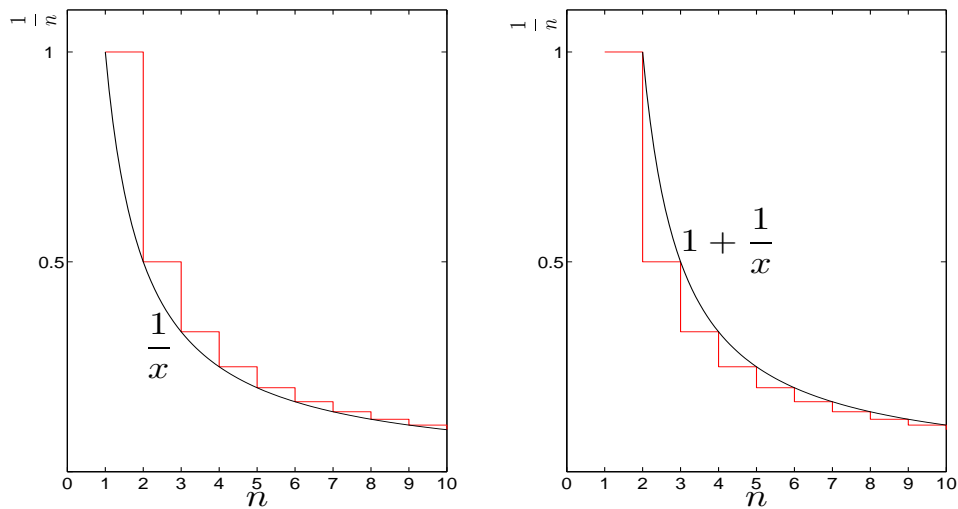


Figure 3.2 : Integral Bounds on a Series

3.3 FUNCTIONS IN \mathbb{R}^1

Let E and F be sets. With each element x of E , let there be associated a unique element $f(x)$ of F . Then f is called a function from E into F , and f is said to map E into F . We write $f: E \rightarrow F$ to indicate it. Let f be a function from E into F . For x in E , the point $f(x)$ in F is called the image of x or the value of f at x . Similarly, for $A \subseteq E$, the set $\{y \in F : y = f(x) \text{ for some } x \in A\}$ is called the image of A . In particular, the image of E is called the range of f . Moving in the opposite direction, for $B \subseteq F$, $f^{-1}(B) = \{x \in E : f(x) \in B\}$ is called the inverse image of B under f . Obviously, the inverse of F is E . Terms like mapping, operator, transformation are synonyms for the term 'function' with varying shades of meaning depending on the context and on the sets E and F . We shall become familiar with them in time. Sometimes, we write $x \mapsto f(x)$ to indicate the mapping f ; for instance, the mapping $x \mapsto x^3 + 5$ from \mathbb{R} into \mathbb{R} is the function $f: \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(x) = x^3 + 5$.

3.3.1 Continuous Functions

see any Calculus textbook

3.4 CALCULUS IN \mathbb{R}^1

3.4.1 Derivatives

The derivative of a function $f(x)$ at $x = x_0$ is defined as

$$\boxed{\frac{df(x_0)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta f(x_0)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.} \quad (3.15)$$

The **tangent to $f(x)$ at the point x_0** is a linear function $l(x) = a_0 + a_1x$ which must satisfy

$$l(x_0) = f(x_0) \quad \text{and} \quad l'(x_0) = f'(x_0).$$

These conditions give $a_1 = f'(x_0)$ and $a_0 + f'(x_0)x_0 = f(x_0)$, or $a_0 = f(x_0) - f'(x_0)x_0$. Hence the equation of the **tangent line** is

$$\boxed{l(x) = f(x_0) + f'(x_0)(x - x_0).} \quad (3.16)$$

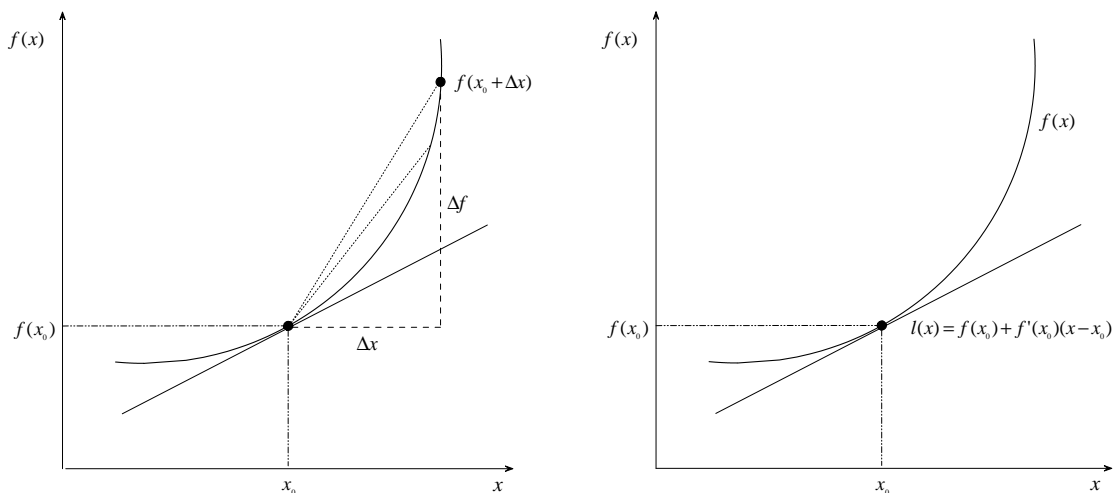


Figure 3.3 : Derivative and Tangent Line

The importance of the derivative is not that it gives us a number — the slope of $f(x)$ at x_0 — but that it allows us to construct a linear function $l(x)$ that is tangent to $f(x)$ at x_0 . This view is important when we move to higher dimensions where the ‘derivative-as-number’ no longer works. Furthermore, this linear function is a **local approximation to $f(x)$ at x_0** .

The linear local approximation concept is fundamentally important in the construction and analysis of numerical algorithms. Without this idea, most important problems could not be solved numerically and they certainly could not be solved symbolically. Analog computation would be the only alternative.

We will see linear local approximation used in many of the algorithms discussed in the chapters to follow.

Theorem 3.3. THE INTERMEDIATE VALUE THEOREM. *Given that $f(x)$ is continuous on $[a, b]$ and that $f(a) = A$, $f(b) = B$. If C is any number between A and B then there exists a real number c between a and b such that $f(c) = C$.*

Theorem 3.4. THE MEAN VALUE THEOREM FOR DERIVATIVES. *If $f(x)$ is continuous and differentiable on $[a, b]$ then there exists a real number $c \in [a, b]$ such that*

$$\boxed{\frac{f(b) - f(a)}{b - a} = f'(c).} \quad (3.17)$$

A specialization of this theorem, when $f(a) = f(b)$, is called *Rolle's Theorem*. This gives

$$f'(c) = 0,$$

for some $c \in [a, b]$. This implies that $f(x)$ has a maximum or minimum for some $c \in [a, b]$, if $f(a) = f(b)$.

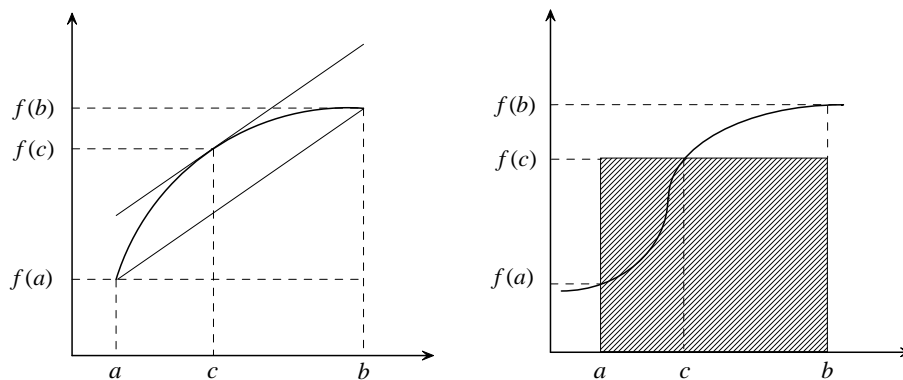


Figure 3.4 : Mean Value Theorem for Derivatives & Integrals

3.4.2 Integrals

see any Calculus textbook

Theorem 3.5. THE MEAN VALUE THEOREM FOR INTEGRALS. *If $f(x)$ is continuous on $[a, b]$ then there exists a real number c in (a, b) such that*

$$\boxed{\int_a^b f(x) dx = f(c)(b - a).} \quad (3.18)$$

This theorem says that the area under the curve $f(x)$ between a and b is equal to the area of the rectangle whose width is $b - a$ and height is $f(c)$, for some $c \in [a, b]$.

Example 3.2. *Monte Carlo Sampling.* An interesting application of the MVT for integrals is finding the area under a curve (evaluating the integral) by *Monte Carlo Sampling*. If we knew the value of c then we could calculate the area by evaluating $f(c)(b - a)$.

We do not know c and so we will guess the value of c and use it to calculate an estimate of the integral. In fact we will make a sequence of guesses by randomly picking values in $[a, b]$. Thus we will generate a sequence of random values $\{x_1, x_2, \dots, x_n\}$ and for each x_i evaluate $f(x_i)(b - a)$. This is the area

of the rectangle whose base is $b - a$ and whose height is $f(x_i)$ (see Fig. 3.5). We then average these guesses to get an estimate of the integral. That is,

$$\int_a^b f(x) dx \approx I(n) = \frac{1}{n} \sum_{i=1}^n f(x_i)(b-a) = \frac{(b-a)}{n} \sum_{i=1}^n f(x_i).$$

Now, the x_i 's are (pseudo-) random variables and therefore $I(n)$ is a random variable. It can be shown that the expected value of $I(n)$ converges to the value of the integral. That is,

$$\lim_{n \rightarrow \infty} E(I(n)) = \int_a^b f(x) dx,$$

with variance $\text{Var}(I(n)) = \sigma_f^2/n$, where σ_f is a constant that depends on the shape of the function. This means that the error in the estimate $\sqrt{\text{Var}(I(n))} = \sigma_f/\sqrt{n} \rightarrow 0$ as $n \rightarrow \infty$. If we define the *accuracy* of the estimate as $A(n) = \sqrt{n}/\sigma_f$, we see that to double the accuracy we must *quadruple* the number of guesses n . Although the Monte Carlo method may seem a rather frivolous method for evaluating an integral, it is still the only method for evaluating multiple integrals over complicated domains.

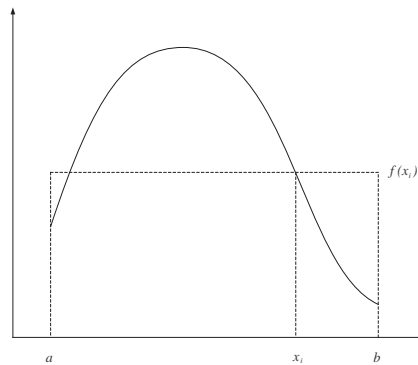


Figure 3.5 : Monte Carlo Estimation of Integral $\int_a^b f(x) dx$

3.4.3 Higher Mean Value Theorems

We showed above that

$$\frac{f(b) - f(a)}{(b-a)} = f'(c) \quad \text{or} \quad f(b) - f(a) = (b-a)f'(c), \quad a < c < b.$$

An alternative statement is

$$f(a+h) - f(a) = hf'(a + \theta'h), \quad 0 < \theta' < 1.$$

We can extend this, by imposing further restrictions on $f(x)$, to get,

$$f(b) - f(a) = (b-a)f'(a) + \frac{1}{2}(b-a)^2 f''(c), \quad a < c < b.$$

If we put $b = a + h$, we get

$$f(a + h) = f(a) + hf'(a) + \frac{1}{2}h^2 f''(a + \theta''h), \quad 0 < \theta'' < 1.$$

This is the *Second Order Mean Value Theorem*.

Note in the above that any number c such that $a < c < b$, is representable as $a + \theta(b - a)$, where $0 < \theta < 1$.

3.4.4 Power Series Expansion of Functions

A **power series** is a series of the form

$$a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots + a_k(x - x_0)^k + \cdots = \sum_{k=0}^{\infty} a_k(x - x_0)^k,$$

where x_0 and $a_k, k = 0, 1, 2, \dots$, are constants.

If the power series converges for certain values of x then we may define a function of x

$$f(x) = \sum_{k=0}^{\infty} a_k(x - x_0)^k, \quad (3.19)$$

for those values of x . This is called a **power series expansion of $f(x)$ about the point x_0** .

It can be shown that such a series can be differentiated and integrated, term-by-term, to give $f'(x)$ and $\int f(x) dx$.

Theorem 3.6. POWER SERIES UNIQUENESS. *If a function $f(x)$ can be expanded about x_0 in a power series, then this expansion is unique, i.e., the constants a_k , for $k = 1, 2, \dots$, are unique.*

It is obvious that the constants a_k , for $k = 1, 2, \dots$, must depend on $f(x)$ for any x and x_0 . Following in the footsteps of Colin MacLaurin, let us simplify matters and determine the relationship between them when $x_0 = 0$, that is,

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_kx^k + \cdots = \sum_{k=0}^{\infty} a_kx^k$$

If $x = 0$ then $f(0) = a_0$.

Having obtained a_0 , get rid of it by differentiating both sides :

$$f'(x) = a_1 + 2a_2x + 3a_3x^2 + \cdots + ka_kx^{k-1} + \cdots = \sum_{k=1}^{\infty} ka_kx^{k-1}$$

If $x = 0$ then $f'(0) = a_1$. Again, get rid of a_1 by differentiating both sides :

$$f''(x) = 2 \cdot 1a_2 + 3 \cdot 2a_3x + \cdots + k \cdot (k-1)a_kx^{k-2} + \cdots = \sum_{k=2}^{\infty} k(k-1)a_kx^{k-2}$$

If $x = 0$ then $f''(0) = 2 \cdot 1a_2$.

$$f'''(x) = 3 \cdot 2 \cdot 1a_3 + 4 \cdot 3 \cdot 2a_4x + \cdots + k \cdot (k-1) \cdot (k-2)a_kx^{k-3} + \cdots = \sum_{k=3}^{\infty} k(k-1)(k-2)a_kx^{k-3}$$

If $x = 0$ then $f'''(0) = 3 \cdot 2 \cdot 1 a_3$.

We can now see the general pattern :

$$f^{(n)}(x) = n \cdot (n-1) \cdot (n-2) \cdots 1 a_n + \cdots = \sum_{k=n}^{\infty} k(k-1)(k-2) \cdots (k-n+1) a_k x^{k-n}$$

If $x = 0$ then $f^{(n)}(0) = n \cdot (n-1) \cdot (n-2) \cdots 1 a_n$, and we get the general expression

$$a_n = \frac{f^{(n)}(0)}{n \cdot (n-1) \cdot (n-2) \cdots 1} = \frac{f^{(n)}(0)}{n!}$$

We have just developed

Theorem 3.7. MACLAURIN'S THEOREM. *Given a function $f(x)$ such that $f'(x), f''(x), \dots, f^{(n)}(x)$ exist on $[a, b]$, then*

$$f(x) = f(0) + f'(0)x + \frac{1}{2!}f''(0)x^2 + \cdots + \frac{1}{(n-1)!}f^{(n-1)}(0)x^{n-1} + R_n(x), \quad (3.20)$$

where the remainder is

$$R_n(x) = \frac{f^{(n)}(c)}{n!}x^n, \quad \text{for some } c \in [0, x].$$

Taylor's Theorem.

Taylor's theorem is a generalization of the two mean value theorems above. It is one of the most useful theorems in mathematics and is used extensively in the theory and practice of Numerical Methods. It allows us to replace a complicated function with a simpler (polynomial) approximation, under certain restrictions on $f(x)$.

Theorem 3.8. TAYLOR'S THEOREM. *Given a function $f(x)$ such that $f'(x), f''(x), \dots, f^{(n)}(x)$ exist on $[a, b]$, then*

$$f(x) = f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \cdots + \frac{1}{(n-1)!}f^{(n-1)}(a)(x-a)^{n-1} + R_n(x) \quad (3.21)$$

where the remainder is

$$R_n(x) = \frac{f^{(n)}(c)}{n!}(x-a)^n, \quad \text{for some } c \in [a, x].$$

This is called the Taylor series expansion of $f(x)$ about the point a . We may think of Taylor's expansion as a polynomial approximation to $f(x)$ plus an error term. That is,

$$f(x) = P_{n-1}(x) + R_n(x),$$

where P_{n-1} is an $(n-1)$ -degree polynomial and $R_n(x)$ is an error term. This theorem is used again and again in devising and analysing numerical methods. In general we replace a complicated function $f(x)$ with a polynomial approximation that is easy to compute.

We get the alternative form by letting $b = a + h$:

$$f(a+h) = f(a) + hf'(a) + \frac{1}{2}h^2f''(a) + \cdots +, \quad 0 < \theta'' < 1.$$

Taylor's Series.

Functions that have derivatives of all orders may be expanded in an infinite Taylor series. We then have

$$f(x) = f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \cdots + \frac{1}{(k-1)!}f^{(k-1)}(a)(x-a)^{k-1} + \cdots$$

When $a = 0$ we get the *MacLaurin Series*

$$f(x) = f(0) + f'(0)x + \frac{1}{2!}f''(0)x^2 + \cdots + \frac{1}{(n-1)!}f^{(n-1)}(0)x^{n-1} + \cdots$$

Example 3.3 (*Well-known Taylor-Maclaurin Series*).

1. The Taylor series for e^x about a is

$$\begin{aligned} e^x &= e^a + e^a(x-a) + \frac{1}{2}e^a(x-a)^2 + \cdots + \frac{1}{n!}e^a(x-a)^k + \cdots \\ &= e^a \sum_{k=0}^{\infty} \frac{(x-a)^k}{k!} \end{aligned}$$

If we let $a = 0$ we get

$$e^x = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^k}{k!} + \cdots$$

2. The Taylor series for $\sin x$ about a is

$$\sin x = \sin a + \cos a(x-a) - \frac{\sin a(x-a)^2}{2!} - \frac{\cos a(x-a)^3}{3!} + \frac{\sin a(x-a)^4}{4!} + \frac{\cos a(x-a)^5}{5!} + \cdots$$

If we let $a = 0$ all the $\sin()$ terms are 0 and the $\cos()$ terms are 1, so we get

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} + \cdots$$

3.5 LINEAR SPACES

Most iterative algorithms may be thought of as generating a sequence of points in an abstract space that converges to a limit which is the solution of some problem. The space must have two features :

1. *Algebraic Properties* which allow one point to be mapped into another.
2. *Metric Properties* which allow us to measure the closeness of one point to another.

3.5.1 Linear Vector Spaces

Definition 3.2. *Linear Vector Space.* A linear vector space X is a set of elements x, y, \dots along with two operations on these elements :

Vector Addition : for any two elements $x, y \in X$ there is a unique element $x + y \in X$, and

Scalar Multiplication : for each element $x \in X$ and each scalar $\alpha \in F$ there is a unique element $\alpha x \in X$.

Vector addition and scalar multiplication have the following properties :

1. $x + y = y + x$.
2. $x + (y + z) = (x + y) + z$.
3. There exists a unique element $0 \in X$ such that $x + 0 = x$, for all $x \in X$.
4. For each element $x \in X$ there exists a unique inverse $-x$ such that $x + (-x) = 0$.
5. $\alpha(\beta x) = (\alpha\beta)x$ for all $\alpha, \beta \in F, x \in X$.
6. $\alpha(x + y) = \alpha x + \alpha y$.
7. $(\alpha + \beta)x = \alpha x + \beta y$.
8. $\mathbf{1}(x) = x, \mathbf{1} \in F$.

These axioms or properties define the algebraic properties of the vector space X . The field F is either \mathbb{R}^1 or \mathbb{C}^1 .

The linear vector spaces that we use are \mathbb{R}^n , the space of real n -dimensional vectors; \mathbb{P}^n , the space of polynomials of degree n with real coefficients; and $\mathcal{C}[a, b]$, the space of functions continuous on $[a, b]$.

Definition 3.3. *Linear Combination.*

An expression of the form

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n, \text{ where } \alpha_k \in F, \text{ and } x_k \in X, \quad (3.22)$$

is a **linear combination** of the set $\{x_1, x_2, \dots, x_n\}$.

Definition 3.4. *Linear Dependence.*

A set of vectors $\{x_1, x_2, \dots, x_n\}$ is **linearly dependent** if there are constants $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, not all zero, such that

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n = 0. \quad (3.23)$$

If there are no such constants, then the set of vectors is **linearly independent**.

If we can find a set of n vectors $\{x_1, x_2, \dots, x_n\} \in X$ which are independent, but every set of $n+1$ vectors is dependent, then n is the **dimension of the space X** .

3.5.2 Normed Linear Vector Spaces

These are linear vector spaces on which the notions of *distance* and *angle* between two vectors are defined. These concepts will be used in all subsequent chapters.

Definition 3.5. *Norm.* A norm on a linear vector space X is a real number, denoted by $\|x\|$, for each $x \in X$, such that

- N1.** $\|x\| \geq 0$
- N2.** $\|x\| = 0$, if and only if $x = 0$.
- N3.** $\|ax\| = |a|\|x\|$, for all $a \in \mathbb{R}^1$, Homogeneity.
- N4.** $\|x + y\| \leq \|x\| + \|y\|$, Triangle Inequality.

We define the **distance** between two vectors $x, y \in X$ as

$$d(x, y) = \|x - y\|.$$

The norm $\|x\|$ may be viewed as the **length** of a vector x or the distance between the origin 0 and x because $\|x - 0\| = \|x\|$.

Standard Norms.

The norms used most often on \mathbb{R}^n are as follows, where $x = [x_1 \ x_2 \ \dots \ x_n]$:

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad (3.24)$$

$$\|x\|_2 = \left[\sum_{i=1}^n |x_i|^2 \right]^{1/2} \quad (3.25)$$

$$\|x\|_p = \left[\sum_{i=1}^n |x_i|^p \right]^{1/p} \quad (3.26)$$

$$\|x\|_\infty = \lim_{p \rightarrow \infty} \|x\|_p = \max_i |x_i| \quad (3.27)$$

These are called the **1-, 2-, p-, and infinity norms**, respectively.

The shape of a geometrical figure in \mathbb{R}^n changes as the norm changes. This can be seen in Figure 3.6, where the set $\|x\|_p = 1$, $x \in \mathbb{R}^2$, is plotted for values of $p = 1, 2, \dots, 10$. Obviously, $\lim_{p \rightarrow \infty} \{\|x\|_p = 1\} = \{\|x\|_\infty = 1\} = \{\max_i \|x_i\| = 1\}$ is the square $\{(-1, 1), (1, 1), (1, -1), (-1, -1)\}$.

Exercise 3.1. Show that $\|x\|_1$ and $\|x\|_\infty$ are indeed norms, according to the definition. That is, show that N1, N2, and N3 hold.

Exercise 3.2. Define $d(x, y) = \|x - y\|$ for each of the standard norms above. What is the distance between $x = (1, 2)$ and $y = (5, 10)$ for the $\|\cdot\|_2$ norm.

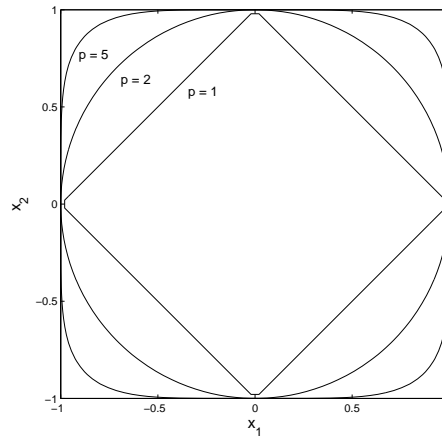


Figure 3.6 : The set $\{\|x\|_p = 1\}$ for different values of p

3.5.3 Inner Product Spaces

Definition 3.6. Inner Product. An inner product of two vectors x and y in a vector space X , is a real number, denoted by $\langle x, y \rangle$, which obeys the following axioms :

IP1. $\langle x, x \rangle \geq 0$, and $\langle x, x \rangle = 0$ iff $x = 0$.

IP2. $\langle ax, y \rangle = a \langle x, y \rangle$, for all scalars $a \in \mathbb{R}^1$.

IP3. $\langle x, y \rangle = \langle y, x \rangle$.

IP4. $\langle x + z, z \rangle = \langle x, z \rangle + \langle y, z \rangle$.

Any vector space X with an inner product defined on it, is an **Inner Product Space**.

The space \mathbb{R}^n is an inner product space with the standard **inner or dot product** of two vectors $x, y \in \mathbb{R}^n$ defined as

$$\langle x, y \rangle = x \cdot y = x^T y = \sum_{i=1}^n x_i y_i. \quad (3.28)$$

We define an **inner product norm** as

$$\|x\|_{ip} = \sqrt{\langle x, x \rangle} = \sqrt{x^T x} = \sqrt{\sum_{i=1}^n x_i x_i} = \sqrt{\sum_{i=1}^n x_i^2} \quad (3.29)$$

We define the **angle between two vectors** as

$$\theta(x, y) = \cos^{-1} \frac{\langle x, y \rangle}{\sqrt{\langle x, x \rangle \langle y, y \rangle}} = \cos^{-1} \frac{\langle x, y \rangle}{\|x\|_{ip} \|y\|_{ip}} \quad (3.30)$$

This gives us an alternative definition of inner product :

$$\langle x, y \rangle = \|x\| \|y\| \cos \theta.$$

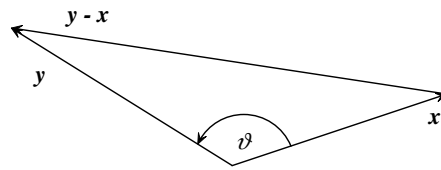


Figure 3.7 : The angle between two vectors

This definition of angle is a generalization from 2–dimensional Euclidean space.

Two vectors x and y are **orthogonal** if the angle θ between them is 90° or $\pi/2$ radians. If $x \neq 0$ and $y \neq 0$ and $\theta = \pi/2$, then we have

$$\cos \theta = \frac{\langle x, y \rangle}{\sqrt{\langle x, x \rangle \langle y, y \rangle}} = 0 \Rightarrow \langle x, y \rangle = 0, \text{ or,}$$

$$\boxed{\langle x, y \rangle = 0 \Leftrightarrow x \perp y.} \quad (3.31)$$

Given two vectors x and y , the **orthogonal projection** of y onto x is the vector $y_p = ax$, where a is a unique number such that $y - y_p = y - ax$ is perpendicular to x . To find a we have $(y - ax) \perp x$ and so $\langle y - ax, x \rangle = 0$. Using IP2 and IP4 above we get $\langle y - ax, x \rangle = \langle y, x \rangle - a \langle x, x \rangle = 0$, or

$$a = \frac{\langle x, y \rangle}{\langle x, x \rangle} = \frac{\langle x, y \rangle}{\|x\|^2}, \text{ or } y_p = \frac{\langle x, y \rangle}{\|x\|^2} x.$$

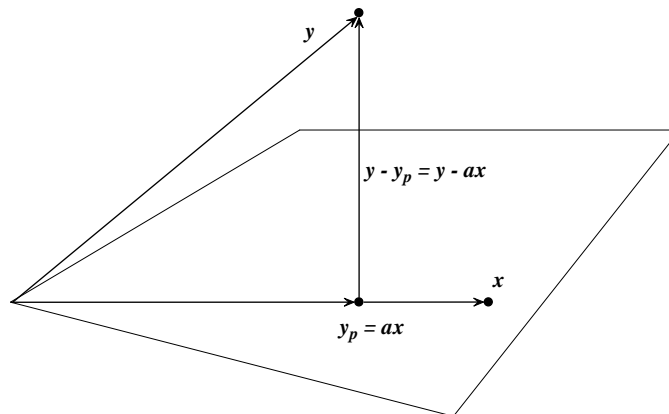


Figure 3.8 : Orthogonal Projection of y onto x

Theorem 3.9 (PYTHAGORAS). *Two vectors $x, y \in \mathbb{R}^n$ are orthogonal if and only if*

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Proof. Applying the definition of inner-product norm, IP4, and $\langle x, y \rangle = 0$ we get¹

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + 2 \langle x, y \rangle + \langle y, y \rangle = \langle x, x \rangle + \langle y, y \rangle = \|x\|^2 + \|y\|^2.$$

□

¹Why is Euclid's proof in the standard school geometry books so long?

Orthogonal Sets of Vectors.

Definition 3.7 (Orthogonal Sets). A set of vectors $S = \{x_1, x_2, \dots, x_k\} \in \mathbb{R}^n$ is an *orthogonal set* if each pair of distinct vectors of the set is orthogonal. That is, $\langle x_i, x_j \rangle = 0$ whenever $i \neq j$.

Theorem 3.10 (ORTHOGONAL BASIS). If $S = \{x_1, x_2, \dots, x_k\} \in \mathbb{R}^n$ is an orthogonal set of vectors, then S is linearly independent and hence is a basis for the subspace of \mathbb{R}^n spanned by S .

3.5.4 Examples of Normed Linear Vector Spaces

3.5.5 Mappings

Mappings, also called *transformations* or *functions* are fundamental to mathematics.

Definition 3.8. *Mapping.* Let X and Y be linear spaces and let D be a subset of X . A *mapping* or *transformation* T is a ‘rule’ which associates with every element (vector) $x \in D$ a *single* element (vector) $y \in Y$. Symbolically, a mapping is written as either $T : X \rightarrow Y$, or $y = T(x)$, or $y = Tx$.

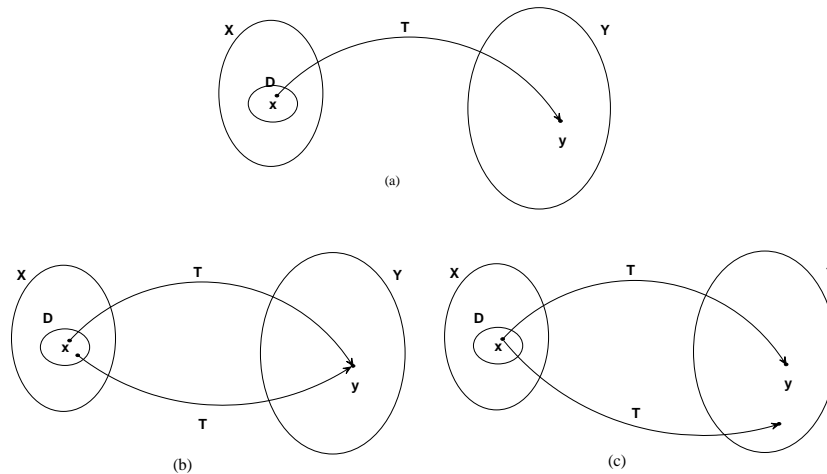


Figure 3.9 : Mappings — (c) is not a mapping

Examples of Mappings

1. $X = Y = \mathbb{R}^1$, $y = T(x) = 4 \sin x$, $D = [-\pi, \pi]$, $R = [-4, 4]$.
2. $X = Y = \mathbb{R}^2$, $y = T(x) = Ax$, where

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

This transforms the point $x = (x_1, x_2)$ into the point $y = (x_1 + x_2, 2x_1 + 5x_2)$.

3. Any norm $\|x\|$ is a mapping $T : X \rightarrow \mathbb{R}^1$.
4. Integration is a mapping $T : \mathcal{C}[a, b] \rightarrow \mathcal{C}[a, b]$, $g = T(f) : g(t) = \int f(t) dt$.

Linear Transformations

These are an important class of mappings which include matrices, derivatives, and integrals.

Definition 3.9 (Linear Transformation). These are mappings $T : X \rightarrow Y$ that have the property

$$T(a_1x_1 + a_2x_2) = a_1T(x_1) + a_2T(x_2), \text{ where } x_1, x_2 \in X \text{ and } a_1, a_2 \in \mathbb{R}^1. \quad (3.32)$$

Example 3.4 (Matrices). Any matrix $A_{m \times n}$ may be viewed as a mapping of vectors x in \mathbb{R}^n to vectors y in \mathbb{R}^m . The mapping is performed by matrix-vector multiplication, i.e., $y = Ax$. It is obvious that this mapping is linear because for any two vectors x_1 and x_2 in \mathbb{R}^n we have

$$A(x_1 + x_2) = Ax_1 + Ax_2 = y_1 + y_2.$$

Example 3.5 (Differentiation and Integration). Given any two functions f and g in $\mathcal{C}[a, b]$, we have

$$\frac{d(f+g)}{dt} = \frac{df}{dt} + \frac{dg}{dt} \quad \text{and} \quad \int (f+g) dt = \int f dt + \int g dt.$$

Thus differentiation and integration are linear transformations from $\mathcal{C}[a, b]$ to $\mathcal{C}[a, b]$.

Example 3.6 (Non-Linear Transformations). The norm of a vector is not linear because, in general, $\|x+y\| \neq \|x\| + \|y\|$. Likewise, the square of a vector x , defined as $f(x) = x^2 = (x, x) = \sum x_i^2$ is not linear because

$$(x+y)^2 = \sum_{i=1}^n (x_i + y_i)^2 = \sum_{i=1}^n (x_i^2 + 2x_i y_i + y_i^2) \neq \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2.$$

3.5.6 The Hierarchy of Functions

Functions are usually ranked according to their ‘niceness’ or degree of smoothness. Below is a list of function sets, in increasing order of smoothness. Naturally the last on this list is the set of constant functions — nothing is smoother than a straight line whose slope(derivative) is 0.

- | | |
|---|---|
| 1. L^p Functions | 7. Infinitely Differentiable functions. |
| 2. Bounded functions. | 8. Analytic functions. |
| 3. Continuous functions. | 9. Entire functions. |
| 4. Lipschitz-continuous functions. | 10. Polynomials of restricted degree. |
| 5. Differentiable functions. | 11. Constants. |
| 6. n -times Differentiable functions. | |

3.6 SEQUENCES IN NORMED LINEAR SPACES

Most of the algorithms discussed in this course generate sequences of vectors in some space X . We must be able to analyse these sequences so that we can determine the behavior of these algorithms. The most interesting feature of sequences is their *convergence* or *divergence*.

Definition 3.10. *Convergence to a Limit.* Let $\{x_1, x_2, \dots, x_k, \dots\}$ be a sequence of vectors in a normed linear space X . The sequence is *convergent to a limit* $x \in X$ if for each $\epsilon > 0$ there exists a positive integer n_ϵ such that if $k \geq n_\epsilon$ then $d(x_k, x) < \epsilon$.

Convergence to a limit is usually written as either $\lim_{k \rightarrow \infty} \{x_k\} = x$, or $x_k \rightarrow x$, or $d(x_k, x) \rightarrow 0$, or $\|x_k - x\| \rightarrow 0$.

The problem with this definition of convergence is that it provides no *constructive* way to test a sequence for convergence because we must know the limit x of the sequence to test it. From an algorithmic point of view this begs the question — if we know the limit of the sequence there is no point in constructing an algorithm to find it. The following theorem gives a constructive way of testing the convergence of a sequence.

Theorem 3.11. CAUCHY SEQUENCE & CONVERGENCE.² *A sequence $\{x_1, x_2, \dots, x_k, \dots\}$ in a normed linear space X is a Cauchy Sequence if for each $\epsilon > 0$ there exists an integer k such that*

$$d(x_{k+p}, x_k) = \|x_{k+p} - x_k\| < \epsilon, \text{ for } p = 1, 2, \dots \quad (3.33)$$

Furthermore, every Cauchy sequence converges to some vector x .

This is sometimes called the *Cauchy Convergence Criterion*. We note that this is a theorem rather than a definition. It is proved in Whittaker & Watson, who say that

“This result is one of the most important and fundamental theorems of analysis.”³

Theorem 3.12.

1. *Every convergent sequence is Cauchy.*
2. *Every Cauchy sequence is bounded.*
3. *Every subsequence of a Cauchy sequence is Cauchy.*

Although every Cauchy sequence converges to some vector x , this vector may not be in the space X , i.e., each element x_k of the sequence is in X but the limit x is not. For example, the sequence $\{x_k\} = \{1/k\}$ is a Cauchy sequence in the half-open interval $X = (0, 1] \subset \mathbb{R}^1$, but the limit of the sequence $x = 0$ is not in X . In this sense the space X is not *complete*.

Definition 3.11. *Complete Normed Linear Space.* A normed linear space X in which every Cauchy sequence has its limit in X is a *Complete Normed Linear Space*. Such spaces are usually called *Banach Spaces*.

The spaces that we use, such as \mathbb{R}^n , \mathbb{P}^n , $\mathcal{C}[0, 1]$, are complete.

²A. Cauchy, *Analyse Algèbrique*, 1821

³Whittaker & Watson, *A Course of Modern Analysis, 4th Ed.*, Cambridge University Press, 1927.

J. Dieudonné, in his *Foundations of Modern Analysis*, Academic Press, 1960, says on page 142, “... what is probably the most useful theorem in Analysis, the mean value theorem, ...”

3.7 SUCCESSIVE APPROXIMATIONS AND CONTRACTION MAPPINGS

Contraction mappings and successive approximations form the theoretical base on which many numerical algorithms are built. Successive approximation is a general prototype for algorithms that generate a sequence of approximate solutions that tend towards a solution of some problem. The theory of contraction mappings defines the conditions under which a successive approximation algorithm converges to a solution.

3.7.1 Successive Approximations

We saw in Chapter 1, Section 1.3.1, that the heart of an iterative algorithm is the the step

$$s_{new} \leftarrow T(s_{old}),$$

where s_{old} is the current guess at the solution and s_{new} is the refined guess. We may view this algorithm as generating a sequence of refined guesses $\{x_0, x_1, \dots, x_k, \dots\}$ in some linear space X . Given x_0 , the sequence is generated as follows:

$$x_1 = T(x_0), x_2 = T(x_1), \dots, x_k = T(x_{k-1}).$$

If this sequence converges to some limit $x \in X$ then we say that this limit is approached by the process of *Successive Approximation*. We sometimes write this sequence as

$$\boxed{\{x_{k+1} = T(x_k)\} \text{ or } \{x_{k+1} = T^k(x_0)\}.} \quad (3.34)$$

Notice that if the sequence $\{x_{k+1} = T(x_k)\}$ converges to some point x then we must have $x = T(x)$. This point x is called a **fixed point** of the mapping T .

Example 3.7. *Successive Approximations.* Here are some successive approximation sequences that converge to a fixed point. For simplicity we have chosen $X = \mathbb{R}^1$.

1. $T(x) = \sqrt{x}$, with $x_0 = 2$, gives the sequence

$$\{2.0, 1.4142, 1.1892, 1.0905, \dots\} \rightarrow 1.0.$$

Note, that at the limit $x = 1$, we have $x = T(x)$, a fixed point.

2. $T(x) = x + x(1 - ax)$, with $a = 3.0$ and $x_0 = 0.5$, gives the sequence

$$\{.500000, .250000, .312500, .3320313, .333328, .333333\} \rightarrow 1/3.$$

Note again that at the limit $x = 1/3$, we have $x = T(x)$. Indeed, for any a the mapping T has a *fixed point* $x = 1/a$ such that $x = T(x)$. Note also that this gives a method for calculating $1/a$ using addition and multiplication only.

3. $T(x) = x + x(1 - ax)$, with $a = 3.0$ and $x_0 = 0.7$, gives the sequence

$$\{0.7000, -0.007000 - 0.1547, -0.3812, -1.198, \dots, -5.199 \times 10^{20}, \dots\}$$

What has gone wrong? We have the same fixed point $x = 1/3$, but we have not converged to it. Why? Because we have not started close enough to the fixed point.

4. $T(x) = 1 + x$, with $x_0 = 0$ gives the sequence

$$\{0, 1, 2, 3, 4, \dots\} \rightarrow \infty.$$

This sequence diverges for all x_0 .

Exercise 3.3. For which values of x_0 does the sequence in example 1 above converge? Consider the generalization $T(x) = \sqrt[n]{x}$. For which values of x_0 and n does the sequence converge?

Example 3.8. What is the value of the following expression as $n \rightarrow \infty$:

$$S_n = \sqrt{\underbrace{2 + \sqrt{2 + \dots + \sqrt{2}}}_{n \text{ times}}}$$

If we let $x_0 = 2$, $T(x_0) = \sqrt{2}$, and $T(x_1) = \sqrt{2 + \sqrt{2}} = \sqrt{2 + T(x_0)}$, then in general we have

$$T(x_n) = \sqrt{2 + T(x_{n-1})}$$

Suppose $\{x_n\}$ converges to a fixed point x , then we have

$$T(x) = \sqrt{2 + T(x)} = \sqrt{2 + x} = x,$$

or

$$x^2 - x - 2 = 0$$

This is a quadratic in x which has the solutions 2 and -1 . Now $x (= S)$ must be positive and so we have the curious result

$$\lim_{n \rightarrow \infty} S_n = x = \sqrt{2 + \sqrt{2 + \dots + \sqrt{2 + \dots}}} = 2$$

It should be obvious that the convergence of a sequence generated by successive approximation depends on the transformation T . We now define a new type of transformation and show that under certain conditions all such transformations generate successive approximation sequences that converge to a fixed point $x = T(x)$.

3.7.2 Contraction Mappings

Definition 3.12 (Contraction Mapping). A mapping $T : X \rightarrow X$, of a complete normed linear space X into itself is a *contraction mapping* if

$$d(T(x_1), T(x_2)) \leq c d(x_1, x_2), \text{ for all } x_1, x_2 \in X \text{ and } 0 \leq c < 1. \quad (3.35)$$

Stated in norms we get

$$\|T(x_1) - T(x_2)\| \leq c \|x_1 - x_2\|, \text{ for all } x_1, x_2 \in X. \quad (3.36)$$

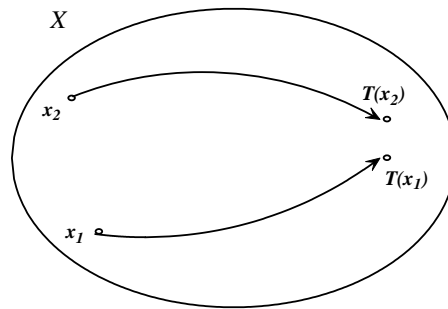


Figure 3.10 : A Contraction Mapping

In other words, a contraction mapping is such that the distance between a pair of transformed points $T(x_1)$ and $T(x_2)$ is less than the distance between the points themselves.

Now, consider the successive approximation sequence $\{x_{k+1} = T(x_k)\}$, where T is a contraction mapping. Let x_k, x_{k+1}, x_{k+2} be three successive points in this sequence and let $d_k = d(x_{k+1}, x_k)$ and $d_{k+1} = d(x_{k+2}, x_{k+1})$ be the distances between these points. Then we have

$$d_{k+1} = d(x_{k+2}, x_{k+1}) = d(T(x_{k+1}), T(x_k)) \leq c d(x_{k+1}, x_k) < d_k.$$

This means that the distance between successive pairs of points becomes smaller and $d_k \rightarrow 0$ as $k \rightarrow \infty$. Thus a point x is approached such that $x = T(x)$. This is called a *Fixed Point* of T .

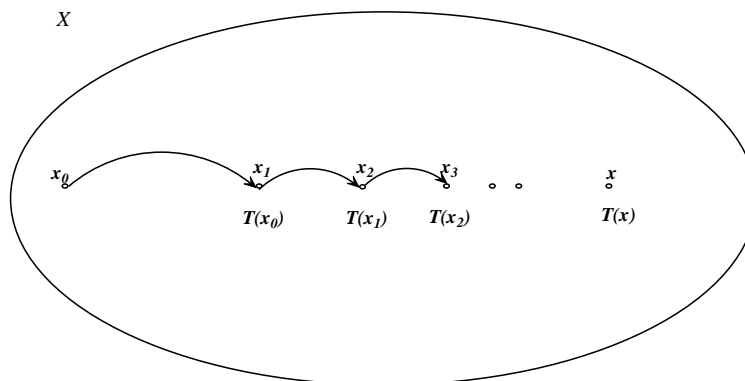


Figure 3.11 : Convergence to a Fixed Point

We now put all this on a rigorous footing by stating and proving Banach's famous fixed point theorem.

Theorem 3.13 (Banach's Fixed Point Theorem). *If T is a contraction mapping $T : X \rightarrow X$, where X is a complete normed linear space, then T has a unique fixed point $x \in X$.*

Proof. We prove this theorem in three stages :

1. Generate the sequence $\{x_k = T(x_{k-1})\}$ and show that it is a Cauchy sequence.
2. Show that the limit of the Cauchy sequence is a fixed point of T , i.e., $x = T(x)$.
3. Show that the fixed point x is unique.

1. We show that the sequence is Cauchy as follows :

$$\begin{aligned}
 d(x_{k+1}, x_k) &= d(T(x_k), T(x_{k-1})) \\
 &\leq cd(x_k, x_{k-1}) \\
 &= cd(T(x_{k-1}), T(x_{k-2})) \\
 &\leq c^2d(x_{k-1}, x_{k-2}) \\
 &\vdots \\
 &\leq c^k d(x_1, x_0).
 \end{aligned}$$

Hence $d(x_{k+1}, x_k) \leq c^k d(x_1, x_0)$. Now, by the triangle inequality we have, for $p > k$

$$d(x_k, x_p) \leq d(x_k, x_{k+1}) + d(x_{k+1}, x_{k+2}) + \cdots + d(x_{p-1}, x_p).$$

Combining this with the previous result that $d(x_{k+1}, x_k) \leq c^k d(x_1, x_0)$ we get

$$\begin{aligned}
 d(x_k, x_p) &\leq (c^k + c^{k+1} + \cdots + c^{p-1})d(x_1, x_0) \\
 &= c^k(1 + c + c^2 + \cdots + c^{p-k-1})d(x_1, x_0) \\
 &< c^k \left(\sum_{i=0}^{\infty} c^i \right) d(x_1, x_0) \\
 &= \frac{c^k}{1-c} d(x_1, x_0)
 \end{aligned}$$

Hence $d(x_k, x_p) \rightarrow 0$ as $k, p \rightarrow \infty$ and so the sequence is a Cauchy sequence. This sequence has a limit $x \in X$ because X is a complete normed linear space.

2. We show that the limit x is a fixed point of T as follows :

$$\begin{aligned}
 d(x, T(x)) &\leq d(x, x_k) + d(x_k, T(x)) \\
 &\leq d(x, x_k) + cd(x_{k-1}, x)
 \end{aligned}$$

Now, the right-hand side goes to zero as $x_k \rightarrow x$. Hence, $d(x, T(x)) \rightarrow 0$ as $x_k \rightarrow x$, which means $x = T(x)$.

3. This fixed point is unique : assume that it is not unique and that there are two fixed points $x = T(x)$ and $y = T(y)$. Then we have

$$d(x, y) = d(T(x), T(y)) \leq cd(x, y) < d(x, y).$$

This is clearly impossible and so $x = y$. □

We have proved the *existence* and *uniqueness* of the fixed point of any contraction mapping $T : X \rightarrow X$. We note that this proof is *constructive* in that we generated a sequence, starting at an arbitrary $x_0 \in X$, that converged to the fixed point $x \in X$.

Lemma 3.1. *A contraction T on a normed linear space X is a continuous mapping.*

3.7.3 Numeric Examples of Contraction Mappings

Example 3.9 (Zeno's Paradox). The Hare (or Achilles) and the Tortoise are to have a race. Because the Hare runs 10 times faster than the Tortoise, the Tortoise is given a 1 league head start. It is obvious that the Hare will always beat the Tortoise over a sufficiently long distance. For example, if the Hare covers a distance of 2 leagues in 2 time units then the Tortoise will have covered $1 + 2/10 = 1.2$ leagues. However, we could argue as follows : by the time the Hare reaches the 1-league mark, the Tortoise is at $1 + 1/10 = 1.1$ leagues; by the time the Hare reaches the 1.1-league mark, the Tortoise is at $1.1 +$

which has the fixed point at $x = \frac{1}{a}$. [VERIFY] This can be used to find the reciprocal of a using subtraction and multiplication only. Thus we may implement the division operation b/a as $\frac{1}{a} \times b$, reciprocation followed by multiplication. Some older Cray supercomputers used this method, with sometimes disastrous consequences.

If we start with $x_0 = 0.5$, then we get the sequence

$$\begin{aligned} x_1 &= x_0(2 - 7x_0) = 0.5(2 - 3.5) \\ &= -0.750000000000000000000000000000 \\ x_2 &= -5.437500000000000000000000000000 \\ x_3 &= -217.8398437500000000000000000000 \\ x_4 &= -332615.0623626708984375000000 \\ x_5 &= -774430123203.7888814338948578 \\ x_6 &= -4198194110079598242745487.49 \end{aligned}$$

This sequence is diverging to $-\infty$. What has gone wrong? The problem is that $T(x)$ is not a contraction mapping everywhere in \mathbb{R}^1 . We can show that we must start with x_0 in the interval $(0, 2/a)$ for the sequence $\{x_{k+1} := T(x_k)\}$ to converge to $1/a$. In the example above we started with $x_0 = 0.5$, which is outside the interval $(0, 2/7)$. Note that this interval is open and does not include either 0 or $2/7$. If we start with $x_0 = 0$ or $x_0 = 2/7$ then the iterations converge to 0 in one step. Thus the mapping $T(x) = x(2 - ax)$ has 2 fixed points: $1/a$ and 0. The second fixed point is meaningless in this context. See Figure 3.14.

Graphical Interpretation of Successive Approximations

The sequence generated by $\{x_k = T(x_{k-1})\}$ can be visualized by plotting the functions $y = x$ and $y = T(x)$. These curves meet at a fixed point $y = x = T(x)$, as shown in Figure 3.13(a). We interpret the vertical movement as $x_{new} \leftarrow T(x_{old})$ and the horizontal movement as $x_{old} \leftarrow x_{new}$.

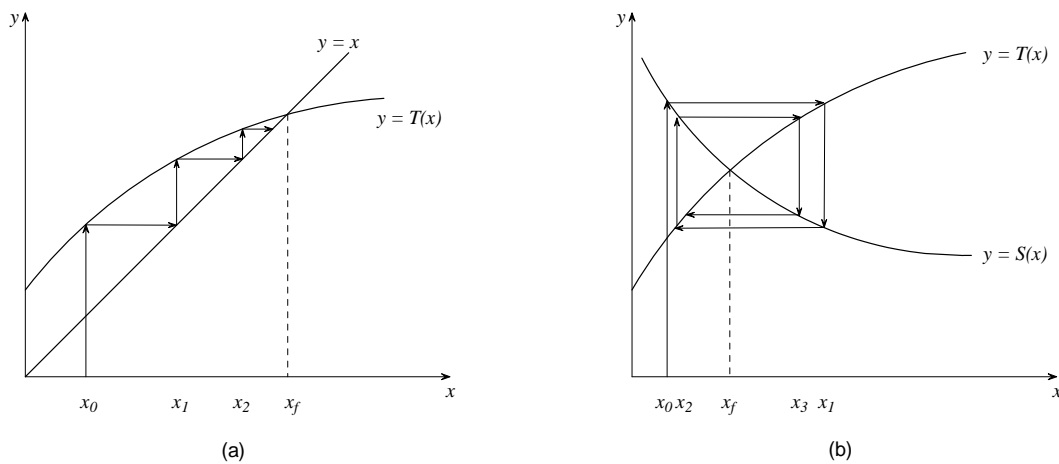


Figure 3.13 : Successive Approximation to a Fixed Point

Figure 3.13(b) shows that we can generalize $x_{k+1} := T(x_k)$ to $S(x_{k+1}) := T(x_k)$. What does this generalization mean? If S^{-1} exists then we have $x_{k+1} := S^{-1}(T(x_k))$ and we are back to the original fixed point iteration formula. But if we do not have the inverse of S then we must use the following interpretation of $S(x_{k+1}) := T(x_k)$: given x_k , calculate $T(x_k)$ (vertical move), and then find x_{k+1} such that

$S(x_{k+1}) = T(x_k)$. Set $x_{k+1} := x_k$ (horizontal move). For example, if S is defined by a table of (S_t, x_t) pairs, then we would search this table until we found $S_t = T(x_k)$. Set $x_{k+1} := x_t$.

Contraction on a Ball

Banach's contraction mapping theorem is too restrictive to be generally useful: it requires that T be a contraction on the entire space X . This requirement is rarely met in practice, as we have just seen in Example 3.10. In most applications the mapping T is a contraction on a subset B of X and we need to modify Banach's theorem appropriately.

It is obvious that we must carefully choose the starting point x_0 if the sequence $\{x_{k+1} := T(x_k)\}$ is to converge. We must choose $x_0 \in D \subseteq X$ so that the iteration sequence remains in D , where T is a contraction.

Theorem 3.14 (BANACH'S FIXED POINT THEOREM ON A BALL). *Let T be a mapping $T : X \rightarrow X$, where X is a complete normed linear space, and let it be a contraction on a subset of X defined as the ball $B = \{x | d(x, x_0) \leq r\}$. If we choose x_0 such that*

$$d(x_0, T(x_0)) < (1 - c)r, \quad [\star]$$

where $c < 1$ is the contraction parameter, then the sequence $\{x_{k+1} := T(x_k)\}$ converges to a limit $x \in B$. This limit x is a fixed point of T and is unique.

Proof. We must show that the sequence $\{x_{k+1} := T(x_k)\}$ remains in B . Then Banach's theorem is applied to this sequence and the result follows.

We have, from the part (1) of the proof of Banach's theorem 3.13

$$d(x_k, x_p) \leq \frac{c^k}{1 - c} d(x_1, x_0), \quad p > k.$$

By assumption $[\star]$ $d(x_0, T(x_0)) = d(x_0, x_1) < (1 - c)r$, and so

$$d(x_k, x_p) \leq \frac{c^k}{1 - c} d(x_1, x_0) < \frac{c^k}{1 - c} (1 - c)r = c^k r.$$

Replacing k by 0 we get $d(x_0, x_p) < r$. Thus x_p is in B for all p . Apply Banach's theorem to this sequence and the result follows. \square

Theorem 3.15 (GENERAL FIXED POINT THEOREM). *Let X be a complete metric space and let T be a continuous transformation of $X \rightarrow X$. If T^k is a contraction for some $k \geq 1$, then T has a unique fixed point.*

Example 3.11 (Contraction on a Ball). Newton's method for calculating the reciprocal $\frac{1}{a}$ is

$$x_{k+1} := x_k (2 - ax_k) \triangleq T(x_k). \quad (3.37)$$

Notice that this iteration function uses multiplication and subtraction only. The steps of Newton's method are shown graphically in Figure 3.14

The sequence in Figure 3.14 is converging to $1/3$, i.e., $a = 3$. The sequence starts with $x_0 = 0.01$. It can be shown?? that we must choose x_0 in the interval (1-dimensional ball) $0 < x_0 < 2/a$ for the sequence to converge.

If we look back at the second part of Example 3.10 we can see the cause of the divergence. In that example $a = 7$ and we started with $x_0 = 1/2$, which is outside the interval $(0, 2/a) = (0, 2/7)$ because $1/2 > 2/7$.

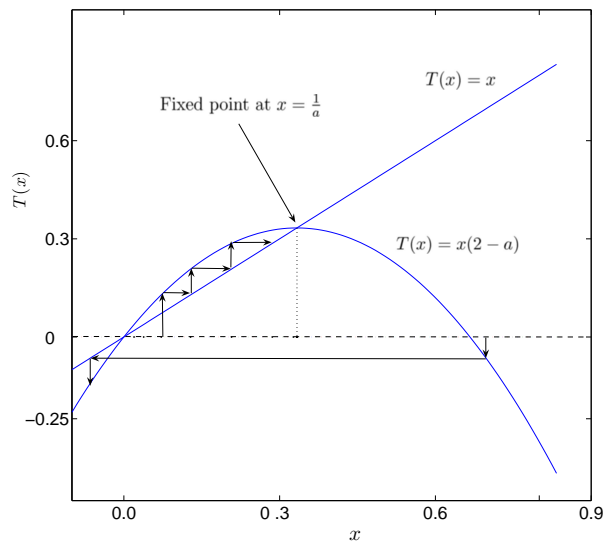


Figure 3.14 : Converging to a fixed point of $T(x) = x(2 - 3x)$

Symbolic Examples of Contraction Mappings

The method of successive approximation with contraction mappings is not limited to numerical calculation. It is a very general method that can be applied symbolically (with due care) to a great many problems.

The examples that follow are purely symbolic and involve neither a computer nor numerical computation as such. However in <http://www.derekroconnor.net/NA/Notes/RecSqRoot.pdf> we show how various iterative algorithms can be performed *semi-symbolically* using Computer Algebra Systems, such as MAXIMA, MATHEMATICA, MAPLE, etc.

Example 3.12 (*Linear Contraction Mapping in \mathbb{R}^1*). $T(x) = a + bx$ with $|b| < 1$ is a contraction mapping on \mathbb{R}^1 because, for any $x, y \in \mathbb{R}^1$ we have

$$\begin{aligned} d(T(x), T(y)) &= |T(x) - T(y)| = |a + bx - a - by| \\ &= |b|(x - y)| < |x - y| = d(x, y). \end{aligned}$$

Let $x_0 = 0$. Then we have the sequence

$$\begin{aligned} x_1 &= a \\ x_2 &= a + ab \\ x_3 &= a + ab + ab^2 \\ &\vdots \\ x_k &= a + ab + ab^2 + \dots + ab^{k-1} \end{aligned}$$

In the limit we have

$$\lim_{k \rightarrow \infty} x_k = a \sum_{k=0}^{\infty} b^k = \frac{a}{1-b}, \text{ for all } |b| < 1.$$

This is, of course, the solution of $x = T(x) = a + bx$.

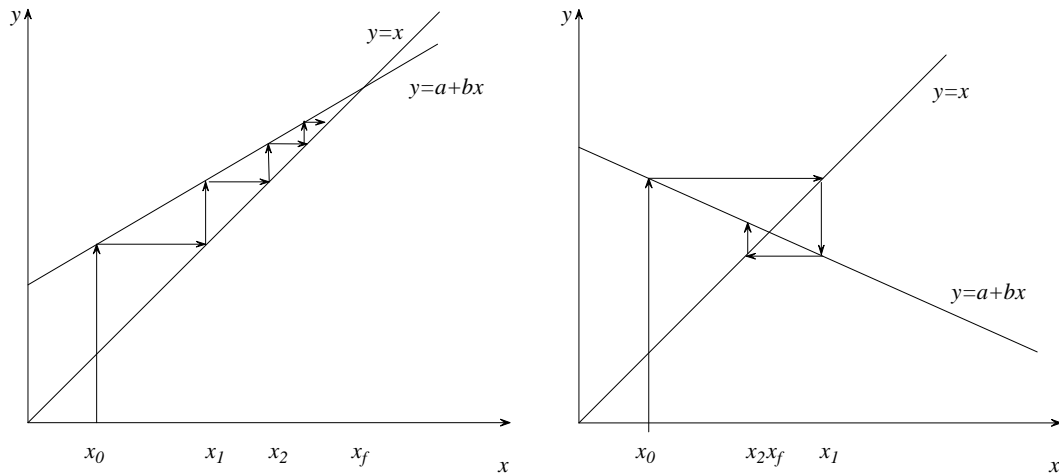


Figure 3.15 : Successive Approximation for Example 3.12

Example 3.13 (*Linear Contraction Mapping in \mathbb{R}^n*). This example shows that the result of Example 1 can be generalized to $X = \mathbb{R}^n$. Let $T(x) = a + Bx$, where $x, a \in \mathbb{R}^n$, and $B \in \mathbb{R}^{n \times n}$, i.e., a $n \times n$ real matrix. Assume that some suitable norm $\|\cdot\|$ has been defined for matrices and that $\|B\| \leq c < 1$. $T(x)$ is a contraction mapping on \mathbb{R}^n because, for any $x, y \in \mathbb{R}^n$ we have

$$\begin{aligned} d(T(x), T(y)) &= \|T(x) - T(y)\| = \|a + Bx - a - By\| \\ &\leq \|B\| \|(x - y)\| \leq c\|x - y\| < d(x, y). \end{aligned}$$

Let $x_0 = 0$, the zero vector. Then we have the sequence

$$x_1 = a, x_2 = a + Ba, x_3 = a + Ba + B^2a = \dots \text{ which gives}$$

$$x_k = Ia + Ba + B^2a + \dots + B^{k-1}a = \left(\sum_{i=0}^{k-1} B^i \right) a.$$

In the limit we have

$$\lim_{k \rightarrow \infty} x_k = \left(\sum_{i=0}^{\infty} B^i \right) a = (I - B)^{-1}a, \text{ for all } \|B\| < 1.$$

This is, of course, the solution of $x = T(x) = a + Bx$. Note that

$$\sum_{i=0}^{\infty} B^i = (I - B)^{-1}, \text{ for all } \|B\| < 1,$$

is the n -dimensional equivalent of the geometric series, $\sum_{i=0}^{\infty} b^i = 1/(1 - b)$, for all $|b| < 1$. The matrix geometric series is called a *Neumann Series*.

As we said before, successive approximation and contraction mappings are very general concepts which can be applied to a wide variety of problems. The next example shows how an integral equation is solved in the Banach space \mathcal{C} , the space of continuous functions.

Example 3.14 (*Contraction Mapping in \mathcal{C}*). Solve the integral equation

$$f(x) = 1 - \int_0^x f(t) dt.$$

In this problem we are looking for a *function* $f \in \mathcal{C}$, the space of continuous functions, that satisfies the equation. We will use the iteration

$$f_{k+1}(x) = 1 - \int_0^x f_k(t) dt, \quad k = 0, 1, 2, \dots$$

This will generate a sequence of approximations f_0, f_1, \dots, f_k that approaches f in the limit, we hope. How do we start? Choose $f_0(x) = \tan^{-1}(x)$? Or $f(x) = \log(x)$?

This is one of the hardest questions in the study of numerical algorithms. Of necessity, no (general) algorithm can tell us where to start. It simply carries out a sequence of instructions *after* it has been given a starting point.

Let us choose $f_0(x) \equiv 0$ for all x . Then we get the sequence

$$\begin{aligned} f_1(x) &= 1 - \int_0^x 0 dt = 1 \\ f_2(x) &= 1 - \int_0^x 1 dt = 1 - x \\ f_3(x) &= 1 - \int_0^x (1 - x) dt = 1 - x + \frac{x^2}{2} \\ f_4(x) &= 1 - \int_0^x (1 - x + x^2/2) dt = 1 - x + \frac{x^2}{2} - \frac{x^3}{6} \end{aligned}$$

As we said above, an algorithm cannot tell us what function to use as the starting approximation. Neither can it tell us name of the function (in the form of a power series) that it has generated. We must do that.

It is obvious that the sequence of polynomials generated above is tending to the power series expansion of e^{-x} . Is this a fixed point of $T(f) = 1 - \int_0^x f(t) dt$? Plug it into $f = T(f)$ and we get

$$e^{-x} = 1 - \int_0^x e^{-t} dt = 1 - [-e^{-t}]_0^x = 1 + e^{-x} - e^0 = e^{-x}.$$

Thus e^{-x} is a fixed point.

This example is a particular case of the more general integral equation

$$f(x) = g(x) + \int_a^b K(x, y) f(y) dy,$$

where $g(x)$ and $K(x, y)$ are given.

Exercise 3.4 (Functional Equations). What functions satisfy these functional equations :

1. $f(x + y) = f(x) + f(y)$.
2. $f(x + y) = f(x)f(y)$.

Example 3.15 (Bellman's Shortest Path Equations). Let $G = (N, A)$ be a directed, weighted graph, where N is a set of *nodes*, and A is a set of ordered pairs of nodes ($u \rightarrow v$) called *arcs*. Let d_{uv} be the weight (length, cost, etc.) of arc $(u, v) \in A$.

We wish to find the lengths of the shortest paths from some node r to all other nodes in G . Let $D(v)$ be the length of the shortest path from node r to node $v \in N - \{r\}$.

We apply Bellman's *Principle of Optimality*?? to the length of the shortest path to any node v . This path must have a final arc (u, v) , for some node u . Thus the shortest path from r to v is a path from r to u followed by an arc from u to v . Bellman's principle says that the path from r to u must also be a shortest path from r to u . This means that the shortest path distances must satisfy the following equations :

$$\begin{aligned} D(r) &= 0, \\ D(v) &= \min_{u \neq v} \{D(u) + d_{uv}\}, \quad v \in N - \{r\}. \end{aligned}$$

It is obvious that $D(\cdot)$ is a non-linear function. Bellman proposed that these equations be solved by successive approximations as follows : Initialize the D 's

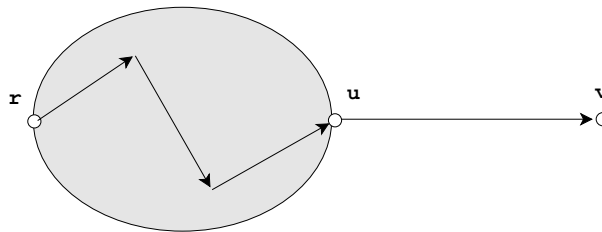


Figure 3.16 : Bellman's Principle for a Shortest Path

$$\begin{aligned} D^0(r) &= 0, \\ D^0(u) &= \infty, \quad u \neq r, \end{aligned}$$

and then compute the $k + 1$ st approximation

$$D^{k+1}(v) = \min\{D^k(v), \min_{u \neq v} (D^k(u) + d_{uv})\}. \quad (3.38)$$

It is obvious that, for each v ,

$$D^1(v) \geq D^2(v) \dots \geq D^{k+1}(v).$$

Bellman showed that the successive approximations $D^k(v)$ converge to their optimum shortest path values after $n - 1$ iterations, i.e., $D^{n-1}(v) = D(v)$, the length of the shortest path from r to v , if the graph G contains no *cycles of negative length*.

If the weighted graph has a negative cycle then the iterations diverge to ∞ , because every time the negative cycle is traversed the length of the path decreases. Disaster? No. An obvious corollary to Bellman's theorem is : if the iteration count k in equation (3.38) reaches n then the graph has a negative cycle. Thus Bellman's shortest path algorithm can be used to detect negative cycles in directed weighted graphs.

Attractive and Repulsive Fixed Points.

Not all fixed points can be found by successive approximation. The mapping $T(x)$ shown in Figure 3.17 has two fixed points but only the lower one can be approached by successive approximation. The upper fixed point cannot be approached, no matter how close to it we start.

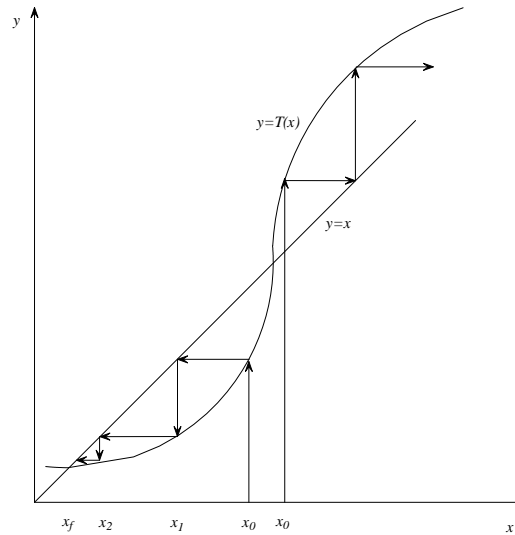


Figure 3.17 : Attractive & Repulsive Fixed Points.

TO BE COMPLETED

3.8 NOTES

3.8.1 Convergence of H_n in Finite Precision

In MATLAB `fzero('eps*harmonic(n)-1/n',1e15)` finds a zero at $n = 135,975,317,602,139 \approx 10^{14}$. That is, the harmonic series converges after 10^{14} terms in double precision.

Using `fzero('sqrt(eps)*harmonic(n)-1/n',1e7)` gives the single precision value $n = 4,237,549$.

Assume that each addition requires 5 clock cycles on a Pentium IV. Then convergence occurs after 5×10^{14} clock ticks. If the speed of the processor is S GHz = $S \times 10^9$ clock ticks per second then convergence occurs after $T_c = 5 \times 10^{14} / S \times 10^9$ secs. On a 3GHz machine this gives $T_c = 226625$ secs \approx 63 hours. This is a gross under-estimate because we are excluding all sorts of overheads, such as loads, store, and looping.

On an 800MHz Pentium III Xeon which does 2×10^6 terms per second we get $T_c = 10^{14} / 2 \times 10^6 = 0.5 \times 10^8$ secs \approx 13889 hours \approx 578 days \approx 1.58 years. In single precision the time to convergence is $T_c = 2$ secs.

3.8.2 MATLAB and MAXIMA Notes