Chapter 3

The Contraction Mapping Theorem

In this chapter we state and prove the *contraction mapping theorem*, which is one of the simplest and most useful methods for the construction of solutions of linear and nonlinear equations. We also present a number of applications of the theorem.

3.1 Contractions

Definition 3.1 Let (X, d) be a metric space. A mapping $T : X \to X$ is a *contraction mapping*, or *contraction*, if there exists a constant c, with $0 \leq c < 1$, such that

$$d(T(x), T(y)) \leq c\, d(x, y) \tag{3.1}$$

for all $x, y \in X$.

Thus, a contraction maps points closer together. In particular, for every $x \in X$, and any $r > 0$, all points y in the ball $B_r(x)$, are mapped into a ball $B_s(Tx)$, with

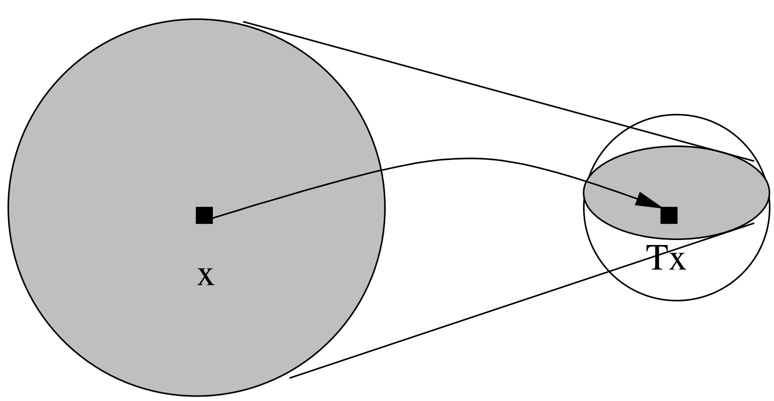

Fig. 3.1 T is a contraction.

$s < r$. This is illustrated in Figure 3.1. Sometimes a map satisfying (3.1) with $c = 1$ is also called a contraction, and then a map satisfying (3.1) with $c < 1$ is called a *strict contraction*. It follows from (3.1) that a contraction mapping is uniformly continuous.

If $T : X \to X$, then a point $x \in X$ such that

$$T(x) = x \tag{3.2}$$

is called a *fixed point* of T. The contraction mapping theorem states that a strict contraction on a complete metric space has a unique fixed point. The contraction mapping theorem is only one example of what are more generally called fixed-point theorems. There are fixed-point theorems for maps satisfying (3.1) with $c = 1$, and even for arbitrary continuous maps on certain metric spaces. For example, the *Schauder fixed point theorem* states that a continuous mapping on a convex, compact subset of a Banach space has a fixed point. The proof is topological in nature (see Kantorovich and Akilov [27]), and we will not discuss such fixed point theorems in this book.

In general, the condition that c is strictly less than one is needed for the uniqueness and the existence of a fixed point. For example, if $X = \{0, 1\}$ is the discrete metric space with metric determined by $d(0, 1) = 1$, then the map T defined by $T(0) = 1$, $T(1) = 0$ satisfies (3.1) with $c = 1$, but T does not have any fixed points. On the other hand, the identity map on any metric space satisfies (3.1) with $c = 1$, and every point is a fixed point.

It is worth noting that (3.2), and hence its solutions, do not depend on the metric d. Thus, if we can find any metric on X such that X is complete and T is a contraction on X, then we obtain the existence and uniqueness of a fixed point. It may happen that X is not complete in any of the metrics for which one can prove that T is a contraction. This can be an indication that the solution of the fixed point problem does not belong to X, but to a larger space, namely the completion of X with respect to a suitable metric d.

Theorem 3.2 (Contraction mapping) If $T : X \to X$ is a contraction mapping on a complete metric space (X, d), then there is exactly one solution $x \in X$ of (3.2).

Proof. The proof is constructive, meaning that we will explicitly construct a sequence converging to the fixed point. Let x_0 be any point in X. We define a sequence (x_n) in X by

$$x_{n+1} = Tx_n \qquad \text{for } n \geq 0.$$

To simplify the notation, we often omit the parentheses around the argument of a map. We denote the nth iterate of T by T^n, so that $x_n = T^n x_0$.

First, we show that (x_n) is a Cauchy sequence. If $n \geq m \geq 1$, then from (3.1) and the triangle inequality, we have

$$\begin{aligned}
d(x_n, x_m) &= d(T^n x_0, T^m x_0) \\
&\leq c^m d(T^{n-m} x_0, x_0) \\
&\leq c^m \left[d(T^{n-m} x_0, T^{n-m-1} x_0) + d(T^{n-m-1} x_0, T^{n-m-2} x_0) \right. \\
&\qquad \left. + \cdots + d(T x_0, x_0) \right] \\
&\leq c^m \left[\sum_{k=0}^{n-m-1} c^k \right] d(x_1, x_0) \\
&\leq c^m \left[\sum_{k=0}^{\infty} c^k \right] d(x_1, x_0) \\
&\leq \left(\frac{c^m}{1-c} \right) d(x_1, x_0),
\end{aligned}$$

which implies that (x_n) is Cauchy. Since X is complete, (x_n) converges to a limit $x \in X$. The fact that the limit x is a fixed point of T follows from the continuity of T:

$$Tx = T \lim_{n\to\infty} x_n = \lim_{n\to\infty} T x_n = \lim_{n\to\infty} x_{n+1} = x.$$

Finally, if x and y are two fixed points, then

$$0 \leq d(x, y) = d(Tx, Ty) \leq c\, d(x, y).$$

Since $c < 1$, we have $d(x, y) = 0$, so $x = y$ and the fixed point is unique. □

3.2 Fixed points of dynamical systems

A dynamical system describes the evolution in time of the state of some system. Dynamical systems arise as models in many different disciplines, including physics, chemistry, engineering, biology, and economics. They also arise as an auxiliary tool for solving other problems in mathematics, and the properties of dynamical systems are of intrinsic mathematical interest.

A dynamical system is defined by a state space X, whose elements describe the different states the system can be in, and a prescription that relates the state $x_t \in X$ at time t to the state at a previous time. We call a dynamical system continuous or discrete, depending on whether the time variable is continuous or discrete. For a continuous dynamical system, the time t belongs to an interval in $\mathbb{R}$, and the dynamics of the system is typically described by an ODE of the form

$$\dot{x} = f(x), \tag{3.3}$$

5 Contraction mapping principle

5.1 Prelimenaries

Now it is time to put aside our computers and try to look at the methods to solve numerically scalar equations from a somewhat deeper theoretical point of view.

Recall that our goal is to analyze the equation of the form

$$f(x) = 0, \quad x \in \mathbf{R}, \tag{5.1}$$

where f is a given function. For the following I will assume that this function is continuous (and if necessary is differential as many times as needed for my reasonings). I write "analyze" and not "solve," because I remind you again that the very first question for any mathematical problem is not "How to solve it?" but "Does a solution exist?" We already had one sufficient condition for f to have root on $[a, b]$, namely, if f is continuous and $f(a)f(b) < 0$ then there must be a point $c \in (a, b)$ such that $f(c) = 0$. Here is another useful similar statement (which can be even called a "baby version of Brouwer's fixed point theorem").

Theorem 5.1. *Let $g\colon [a, b] \longrightarrow [a, b]$ be a continuous function. Then there exists $\hat{x} \in [a, b]$ such that*

$$\hat{x} = g(\hat{x}). \tag{5.2}$$

Quite naturally, such $\hat{x}$ is called a *fixed point* of g. The key thing here is to note that g maps closed interval into itself.

Proof. Consider

$$f(x) = x - g(x),$$

which is a continuous function as the difference of two continuous functions.

I have

$$f(a) = a - g(a) \leq 0, \quad f(b) = b - g(b) \geq 0,$$

and hence $f(a)f(g) \leq 0$. Therefore f is either zero at a or b, or, by the intermediate value theorem, inside $[a, b]$. In other words, there is $\hat{x} \in [a, b]$ such that $f(\hat{x}) = 0$ or

$$\hat{x} = g(\hat{x})$$

as required. ■

Why do we care about equations of the form $x = g(x)$? The short answer goes as follows: I start with (5.1) and *transform* it into (5.2). Now my goal will be to find a fixed point of g. My hope is that if I start with some good x_0, I can find $x_1 = g(x_0)$, then $x_2 = g(x_1)$, i.e., use (5.2) as a recurrent formula, the examples of which we already saw before, and produce the sequence of *simple iterates* $x_0, x_1, x_2, \ldots$ If this sequence converges to a fixed point of g, then this fixed point will be a root of f and I am done.

First I remark that there are always different ways to go from (5.1) to (5.2) including the obvious one

$$f(x) = 0 \implies x + f(x) = x \implies g(x) = x,\ g(x) = x + f(x).$$

Some of them work and some do not as the following example shows.

Math 488/688: Numerical Analysis

e-mail: artem.novozhilov@ndsu.edu. Fall 2021

Example 5.2. Let

$$f(x) = e^x - 2x - 1.$$

If you make a graph of this function, you can see that there must be a root somewhere in $[1, 2]$. Indeed

$$f(1) = e - 3 < 0, \quad f(2) = e^2 - 5 > 0,$$

and f is continuous. Let me try first to transform $f(x) = 0$ into $g(x) = x$ naively:

$$x = \frac{e^x - 1}{2} = g(x).$$

I need g to map $[1, 2]$ into $[1, 2]$, but the first check $g(1) = (e - 1)/2 < 1$ fails. So this will not work.

Now I try the following:

$$e^x = 2x + 1 \implies x = \log(2x + 1) = g(x).$$

Now $g(1) = \log 3 > 1$, $g(2) = \log 5 < 2$, and log is a monotone function, there are no hills and valleys. Hence

$$g\colon [1, 2] \longrightarrow [1, 2],$$

and I know (thanks to Theorem 5.1) that there must be a fixed point of g in $[1, 2]$. So I can try to build my sequence of iterates $x_0, x_1, \ldots$, but first I would like to find a sufficient condition that this sequences converges *prior* to performing actual calculations.

5.2 Contraction mapping principle

Definition 5.3. *Function $g\colon [a, b] \longrightarrow [a, b]$ is called a contraction on $[a, b]$ if there exists a constant q, $0 \le q < 1$ such that*

$$|g(x) - g(y)| \le q|x - y| \tag{5.3}$$

for any $x, y \in [a, b]$.

Since $|x - y|$ measures the distance between x and y then *contraction* means geometrically that the distance between the images of x, y under function g is strictly less than the distance between x and y. If we iterate g then at each step we will *contract* the distance from my current iterate to the unknown root, so intuitively it is expected that at the end we'll end up at a point, which must be our fixed point! Now let me convince you that this intuition is indeed correct.

Theorem 5.4 (Contraction mapping principle)**.** *Let $g\colon [a, b] \longrightarrow [a, b]$ and g be a contraction. Then g has a unique fixed point $\hat{x} \in [a, b]$, and, moreover, for any $x_0 \in [a, b]$ the sequence of simple iterates $x_0, x_1, \ldots$ obtained through the recurrence $x_{n+1} = g(x_n)$, $n = 0, 1, 2, \ldots$ converges to $\hat{x}$. Additionally, the following estimate holds*

$$|x_k - \hat{x}| \le \frac{q^k}{1 - q}|x_0 - x_1|\,. \tag{5.4}$$

In the proof I will use a few basic mathematical facts, which you saw in other classes. Let me recall them here. First, sequence (x_n) is said to converge to $\hat{x}$ (is said to have the limit $\hat{x}$) if for any $\varepsilon > 0$ there is natural number N such that for all $n \ge N$ I have $|x_n - \hat{x}| < \varepsilon$. For instance sequence $x_n = 1$ converges to 1, sequence $x_n = \frac{1}{n}$ converges to 0, and sequence $x_n = (-1)^n$ has no limit (does not converge, this sequence diverges). In particular, I will need the fact that the sequence $x_n = Aq^n$ for any constant A and $0 \leq q < 1$ converges to 0. Second, the absolute value satisfies the triangle inequality

$$|x + y| \leq |x| + |y|.$$

Now I am ready for a proof.

Proof. First, I will prove uniqueness. Assume that I have two fixed points $\hat{x} = g(x)$, $\hat{y} = g(\hat{y})$ in $[a, b]$. Then

$$|\hat{x} - \hat{y}| = |g(\hat{x}) - g(\hat{y})| \leq q|\hat{x} - \hat{y}|$$

since g is a contraction. Or,

$$(1 - q)|\hat{x} - \hat{y}| \leq 0.$$

Since $1 - q > 0$ by assumption I must have $|\hat{x} - \hat{y}| = 0$ (absolute value cannot be negative) and hence $\hat{x} = \hat{y}$.

Now to the existence. By Theorem 5.1 I know that a fixed point $\hat{x}$ must exist (this is a pure existence statement without indication how to find this fixed point). Now I will present a constructive approach to the fixed point. I consider a sequence of iterates

$$x_{n+1} = g(x_n), \quad n = 0, 1, 2, 3, \ldots,$$

starting with some arbitrary $x_0 \in [a, b]$. Let me estimate

$$\begin{aligned} |x_n - \hat{x}| = |g(x_{n-1}) - g(\hat{x})| \leq q|x_{n-1} - \hat{x}| \leq \\ \leq q|g(x_{n-2}) - g(\hat{x})| \leq q^2|x_{n-2} - \hat{x}| \leq \ldots \\ \leq q^{n-1}|x_1 - \hat{x}| \leq q^n|x_0 - \hat{x}|. \end{aligned}$$

Since $q^n \to 0$ I conclude that $x_n \to \hat{x}$ as required.

Moreover,

$$|x_1 - \hat{x}| \leq q|x_0 - \hat{x}| = q|x_0 - x_1 + x_1 - \hat{x}| \leq q|x_0 - x_1| + q|x_1 - \hat{x}|$$

implies that

$$|x_1 - \hat{x}| \leq \frac{q}{1 - q}|x_1 - x_0|,$$

which, together with the previous, finishes the proof:

$$|x_n - \hat{x}| \leq \frac{q^n}{1 - q}|x_1 - x_0|.$$

∎

Remark 5.5. The contraction mapping principle is important for two main reasons (in addition that it guarantees the existence and uniqueness of a fixed point). The first reason is that it provides a constructive way to approach to the fixed point: start with some x_0 and produce a sequence. It must converge to $\hat{x}$. The second reason is that it allows estimating the error we make. For instance, we want that

$$|x_n - \hat{x}| \leq \varepsilon.$$

(Say, if I need k correct digits, I would request that $\varepsilon = 0.5 \cdot 10^{-k}$). The I can calculate

$$\frac{q^n}{1-q}|x_1 - x_0| \leq \varepsilon \implies n \geq \frac{\log \frac{\varepsilon(1-q)}{|x_1-x_0|}}{\log q}\,.$$

Finally, I need a relatively simple procedure to see if my function actually a contraction. For this I will need the mean value theorem (MVT) from Calculus I. I will state it without proof, you can consult any calculus book for this.

Theorem 5.6 (MVT)**.** *Let $g\colon [a,b] \longrightarrow \mathbf{R}$ be a continuously differentiable function. Then for any $a \leq x < y \leq b$ there is point $c \in [x,y]$ such that*

$$\frac{g(y) - g(x)}{y - x} = fg'(c).$$

This theorem has a simple geometric meaning: No matter which points x and y you will take and draw a straight line connecting these points, you can always find a point between x and y such that the tangent line to the graph of g at this point has exactly the same slope as the secant line through x and y (make a drawing).

Now to the contraction. From the mean value theorem it follows that

$$|g(y) - g(x)| = |g'(c)||y - x|.$$

Now let $q = \max_{c\in[a,b]} |g'(c)|$, which must exist since g' is continuous function on a closed interval. If I assume that $0 \leq q < 1$ I conclude that

$$|g(y) - g(x)| \leq q|y - x|,$$

for any $x, y \in [a,b]$, i.e., g is a contraction on $[a,b]$.

Example 5.7. Let me continue Example 5.2. I have that

$$x = g(x) = \log(2x + 1),$$

and $g\colon [1,2] \longrightarrow [1,2]$. Now I want to check that g is a contraction.

Indeed, I have

$$g'(x) = \frac{2}{2x+1}\,, \quad g''(x) = -\frac{4}{(2x+1)^2}\,,$$

which implies that g' monotonously decreasing on $[1,2]$ and hence its maximum at the point $x = 1$:

$$\max_{x\in[1,2]} g'(x) = \frac{2}{3} < 1,$$

and hence for any $x, y \in [1,2]$

$$|g(x) - g(y)| \leq \frac{2}{3}|x - y|,$$

i.e., g is a contraction. This implies, according to the theorem I proved, that if I take any x_0 from $[1,2]$ my sequence (x_n) will converge to the unique root. Moreover, if I want to make sure that my absolute error does not exceed $0.5 \cdot 10^{-6}$ (I hope to have approximately 6 correct digits after the decimal point), then, using my formula above,

$$k \geq \frac{\log \frac{\varepsilon(1-q)}{|x_1 - x_0|}}{\log q},$$

which gives $k \geq 32.77$ for $x_0 = 1, x_1 = 1.09$. This is, of course, a pessimistic estimate (since I use inequalities in all my reasonings, this is the worst possible case). You can numerically check that $k = 25$ is sufficient to find the fixed point $\hat{x}$, and hence the root of $f(x) = 0$ with the required precision.

Finally, I want a somewhat simpler criterion to guarantee that my iterates converge. As usual, I have to pay a price for this. Now it will be a local statement (which means it is true only for some neighborhood of $\hat{x}$), which will follow directly from the general contraction mapping principle.

Corollary 5.8. *Let $g\colon [a, b] \longrightarrow [a, b]$ be a continuously differentiable and have a fixed point $\hat{x} \in [a, b]$. Then if $|g'(\hat{x})| < 1$ then there is a neighborhood $I = [\hat{x} - \delta, \hat{x} + \delta]$ with some $\delta > 0$ such that the sequence (x_n) converges to $\hat{x}$ if $x_0 \in I$.*

Proof. Since $|g'(\hat{x})| < 1$, and g' is continuous then there must be an interval $I = [\hat{x} - \delta, \hat{x} + \delta]$, $\delta > 0$ where $|g'(x)| \leq q < 1$. Since $g\colon I \longrightarrow I$ is a contraction on I, there must be a unique fixed point $\hat{x} \in I$ to which any sequence $x_{n+1} = g(x_n)$ starting in I will converge. ∎

Remark 5.9. Very naturally, any fixed point of g for which $|g'(\hat{x})| < 1$ holds is called *stable* fixed point or *sink*. Analogously, I can define an *unstable* fixed point $\hat{x}$ of g such that $|g'(\hat{x})| > 1$. I will leave all the details for you to work through.

5.3 Rate of convergence

How fast my simple iterates approach the fixed point? Recall that the formal mathematical notion for this is the order and rate of convergence.

Let

$$\varepsilon_n = |x_n - \hat{x}|$$

be the absolute error at the n-th step. I have

$$\lim_{n\to\infty} \frac{\varepsilon_{n+1}}{\varepsilon_n} = \lim_{n\to\infty} \frac{|x_{n+1} - \hat{x}|}{|x_n - \hat{x}|} = \lim_{n\to\infty} \frac{|g(x_n) - g(\hat{x})|}{|x_n - \hat{x}|} = |g'(\hat{x})|,$$

where I again used the mean value theorem. Hence I conclude that simple iterates have the linear order of convergence with the rate of convergence

$$\mu = |g'(\hat{x})|.$$

This value has a direct interpretation. Namely,

$$\rho = -\log_{10} \mu$$

gives the number of correct digits gained in one iteration. Or, equivalently, the quantity

$$\left\lfloor \frac{1}{\rho} \right\rfloor + 1$$

is the number of iterations to gain at least one more correct digit.

Recall, that the bisection method has the rate of convergence $\mu = 1/2$, hence

$$\rho = -\log_{10}\frac{1}{2} \approx 0.3, \quad \lfloor 1/\rho \rfloor + 1 = 4,$$

i.e., for the bisection method one needs to perform 4 steps to gain another correct digit.

Remark 5.10. It is possible that $\mu = 1$ or $\mu = 0$. In the former case we say that convergence *sublinear* in the latter case — *superlinear*. Ideally I want a method that would have μ as close as possible to zero.