Chapter 2

Spaces of continuous functions

In this chapter we shall apply the theory we developed in the previous chapter to spaces where the elements are continuous functions. We shall study completeness and compactness of such spaces and take a look at some applications.

2.1 Modes of continuity

If (X, d_X) and (Y, d_Y) are two metric spaces, the function $f : X \to Y$ is continuous at a point a if for each $\epsilon > 0$ there is a $\delta > 0$ such that $d_Y(f(x), f(a)) < \epsilon$ whenever $d_X(x, a) < \delta$. If f is also continuous at another point b, we may need a different δ to match the same ϵ. A question that often comes up is when we can use the *same* δ for *all* points x in the space X. The function is then said to be *uniformly continuous* in X. Here is the precise definition:

Definition 2.1.1 *Let $f : X \to Y$ be a function between two metric spaces. We say that f is* uniformly continuous *if for each $\epsilon > 0$ there is a $\delta > 0$ such that $d_Y(f(x), f(y)) < \epsilon$ for all points $x, y \in X$ such that $d_X(x, y) < \delta$.*

A function which is continuous at all points in X, but not uniformly continuous, is often called *pointwise continuous* when we want to emphasize the distinction.

Example 1 The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is pointwise continuous, but not uniformly continuous. The reason is that the curve becomes steeper and steeper as $|x|$ goes to infinity, and that we hence need increasingly smaller δ's to match the same ϵ (make a sketch!) See Exercise 1 for a more detailed discussion.

If the underlying space X is compact, pointwise continuity and uniform continuity is the same. This means that a continuous function defined on a closed and bounded subset of $\mathbb{R}^n$ is always uniformly continuous.

Proposition 2.1.2 *Assume that X and Y are metric spaces. If X is compact, all continuous functions $f : X \to Y$ are uniformly continuous.*

Proof: We argue contrapositively: Assume that f is *not* uniformly continuous; we shall show that f is not continuous.

Since f fails to be uniformly continuous, there is an $\epsilon > 0$ we cannot match; i.e. for each $\delta > 0$ there are points $x, y \in X$ such that $d_X(x, y) < \delta$, but $d_Y(f(x), f(y)) \geq \epsilon$. Choosing $\delta = \frac{1}{n}$, there are thus points $x_n, y_n \in X$ such that $d_X(x_n, y_n) < \frac{1}{n}$ and $d_Y(f(x_n), f(y_n)) \geq \epsilon$. Since X is compact, the sequence $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ converging to a point a. Since $d_X(x_{n_k}, y_{n_k}) < \frac{1}{n_k}$, the corresponding sequence $\{y_{n_k}\}$ of y's also converge to a. We are now ready to show that f is not continuous at a. Had it been, the two sequences $\{f(x_{n_k})\}$ and $\{f(y_{n_k})\}$ would both have converged to $f(a)$, something they clearly can not since $d_Y(f(x_n), f(y_n)) \geq \epsilon$ for all $n \in \mathbb{N}$. $\Box$

There is an even more abstract form of continuity that will be important later. This time we are not considering a single function, but a whole collection of functions:

Definition 2.1.3 *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $\mathcal{F}$ be a collection of functions $f : X \to Y$. We say that $\mathcal{F}$ is* equicontinuous *if for all $\epsilon > 0$, there is a $\delta > 0$ such that for all $f \in \mathcal{F}$ and all $x, y \in X$ with $d_X(x, y) < \delta$, we have $d_Y(f(x), f(y)) < \epsilon$.*

Note that in the case, the same δ should not only hold at all points $x, y \in X$, but also for all functions $f \in \mathcal{F}$.

Example 2 Let $\mathcal{F}$ be the set of all contractions $f : X \to X$. Then $\mathcal{F}$ is equicontinuous, since we can can choose $\delta = \epsilon$. To see this, just note that if $d_X(x, y) < \delta = \epsilon$, then $d_X(f(x), f(y)) \leq d_X(x, y) < \epsilon$ for all $x, y \in X$ and all $f \in \mathcal{F}$.

Equicontinuous families will be important when we study compact sets of continuous functions in Section 1.5.

Exercises for Section 2.1

1. In this problem we shall prove that $f(x) = x^2$ is not uniformly continuous on $\mathbb{R}$.

   a) Show that if ϵ is a positive, real number, then $\lim_{a\to\infty} \left(\sqrt{a^2 + \epsilon} - a\right) = 0$.

b) Assume that a and ϵ are positive, real numbers. Show that the largest number δ such that $|f(x) - f(a)| < \epsilon$ whenever $|x - a| < \delta$, is $\delta = \sqrt{a^2 + \epsilon} - a$.

c) Show that f is not uniformly continuous on $\mathbb{R}$.

2. Prove that the function $f : (0, 1) \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$ is not uniformly continuous.

3. A function $f :\to X \to Y$ between metric spaces is said to be *Lipschitz-continuous with Lipschitz constant* K if $d_Y(f(x), f(y)) \leq K d_X(x, y)$ for all $x, y \in X$. Asume that $\mathcal{F}$ is a collection of functions $f : X \to Y$ with Lipschitz constant K. Show that $\mathcal{F}$ is equicontinuous.

4. Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function and assume that the derivative f' is bounded. Show that f is uniformly continuous.

2.2 Modes of convergence

In this section we shall study two ways in which a sequence $\{f_n\}$ of continuous functions can converge to a limit function f: pointwise convergence and uniform convergence. The distinction is rather simililar to the distinction between pointwise and uniform continuity in the previous section — in the pointwise case, a condition can be satisfied in different ways for different x's, in the uniform, case it must be satisfied in the same way for all x. We begin with pointwise convergence:

Definition 2.2.1 *Let* (X, d_X) *and* (Y, d_Y) *be two metric space, and let* $\{f_n\}$ *be a sequence of functions* $f_n : X \to Y$. *We say that* $\{f_n\}$ converges pointwise *to a function* $f : X \to Y$ *if* $f_n(x) \to f(x)$ *for all* $x \in X$. *This means that for each* x *and each* $\epsilon > 0$, *there is an* $N \in \mathbb{N}$ *such that* $d_Y(f_n(x), f(x)) < \epsilon$ *when* $n \geq N$.

Note that the N in the last sentence of the definition depends on x — we may need a much larger N for some x's than for others. If we can use the *same* N for all $x \in X$, we have uniform convergence. Here is the precise definition:

Definition 2.2.2 *Let* (X, d_X) *and* (Y, d_Y) *be two metric space, and let* $\{f_n\}$ *be a sequence of functions* $f_n : X \to Y$. *We say that* $\{f_n\}$ converges uniformly *to a function* $f : X \to Y$ *if for each* $\epsilon > 0$, *there is an* $N \in \mathbb{N}$ *such that if* $n \geq N$, *then* $d_Y(f_n(x), f(x)) < \epsilon$ *for all* $x \in X$.

At first glance, the two definitions may seem confusingly similar, but the difference is that in the last one, the *same* N should work simultaneously for all x, while in the first we can adapt N to each individual x. Hence uniform convergence implies pointwise convergence, but a sequence may converge pointwise but not uniformly. Before we look at an example, it will be useful to reformulate the definition of uniform convergence.

Proposition 2.2.3 *Let (X, d_X) and (Y, d_Y) be two metric space, and let $\{f_n\}$ be a sequence of functions $f_n : X \to Y$. For any function $f : X \to Y$ the following are equivalent.*

(i) $\{f_n\}$ *converges uniformly to* f.

(ii) $\sup\{d_Y(f_n(x), f(x)) \,|\, x \in X\} \to 0$ *as* $n \to \infty$.

Hence uniform convergence means that the "maximal" distance between f and f_n goes to zero.

Bevis: $(i) \implies (ii)$ Assume that $\{f_n\}$ converges uniformly to f. For any $\epsilon > 0$, we can find an $N \in \mathbb{N}$ such that $d_Y(f_n(x), f(x)) < \epsilon$ for all $x \in X$ and all $n \geq N$. This means that $\sup\{d_Y(f_n(x), f(x)) \,|\, x \in X\} \leq \epsilon$ for all $n \geq N$, and since ϵ is arbitrary, this means that $\sup\{d_Y(f_n(x), f(x)) \,|\, x \in X\} \to 0$.

$(ii) \implies (i)$ Assume that $\sup\{d_Y(f_n(x), f(x)) \,|\, x \in X\} \to 0$ as $n \to \infty$. Given an $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that $\sup\{d_Y(f_n(x), f(x)) \,|\, x \in X\} < \epsilon$ for all $n \geq N$. But then we have $d_Y(f_n(x), f(x) < \epsilon$ for all $x \in X$ and all $n \geq N$, which means that $\{f_n\}$ converges uniformly to f. □

Here is an example which shows clearly the distinction between pointwise and uniform convergence:

Example 1 Let $f_n : [0, 1] \to \mathbb{R}$ be the function in Figure 1. It is constant zero except on the interval $[0, \frac{1}{n}]$ where it looks like a tent of height 1.

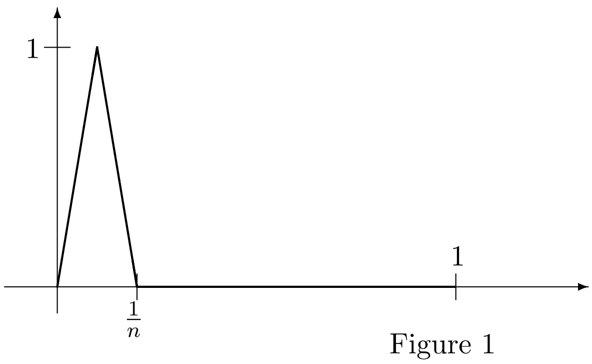

Figure 1

If you insist, the function is defined by

$$f_n(x) = \begin{cases} 2nx & \text{if } 0 \leq x < \frac{1}{2n} \\ -2nx + 2 & \text{if } \frac{1}{2n} \leq x < \frac{1}{n} \\ 0 & \text{if } \frac{1}{n} \leq x \leq 1 \end{cases}$$

but it is much easier just to work from the picture.

The sequence $\{f_n\}$ converges pointwise to 0, because at every point $x \in [0,1]$ the value of $f_n(x)$ eventually becomes 0 (for $x = 0$, the value is always 0, and for $x > 0$ the "tent" will eventually pass to the left of x.) However, since the maximum value of all f_n is 1, $\sup\{d_Y(f_n(x), f(x)) \,|\, x \in [0,1]\} = 1$ for all n, and hence $\{f_n\}$ does not converge uniformly to 0.

When we are working with convergent sequences, we would often like the limit to inherit properties from the elements in the sequence. If, e.g., $\{f_n\}$ is a sequence of *continuous* functions converging to a limit f, we are often interested in showing that f is also continuous. The next example shows that this is not always the case when we are dealing with pointwise convergence.

Example 2: Let $f_n : \mathbb{R} \to \mathbb{R}$ be the function in Figure 2.

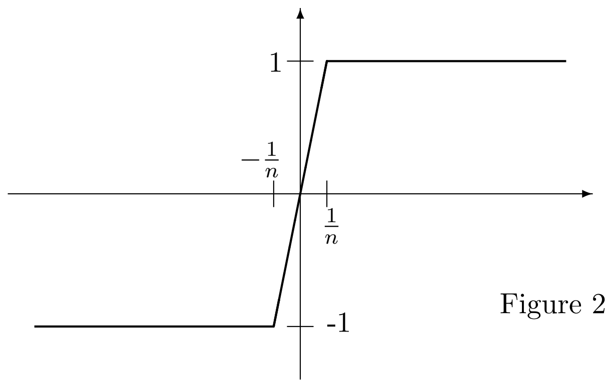

Figure 2

It is defined by

$$f_n(x) = \begin{cases} -1 & \text{if } x \leq -\frac{1}{n} \\ \\ nx & \text{if } -\frac{1}{n} < x < \frac{1}{n} \\ \\ 1 & \text{if } \frac{1}{n} \leq x \end{cases}$$

The sequence $\{f_n\}$ converges pointwise to the function, f defined by

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

but although all the functions $\{f_n\}$ are continuous, the limit function f is not.

If we strengthen the convergence from pointwise to uniform, the limit of a sequence of continuous functions is always continuous.

Proposition 2.2.4 *Let (X, d_X) and (Y, d_Y) be two metric spaces, and assume that $\{f_n\}$ is a sequence of continuous functions $f_n : X \to Y$ converging uniformly to a function f. Then f is continuous.*

Proof: Let $a \in X$. Given an $\epsilon > 0$, we must find a $\delta > 0$ such that $d_Y(f(x), f(a)) < \epsilon$ whenever $d_X(x, a) < \delta$. Since $\{f_n\}$ converges uniformly to f, there is an $N \in \mathbb{N}$ such that when $n \geq N$, $d_Y(f(x), f_n(x)) < \frac{\epsilon}{3}$ for *all* $x \in X$. Since f_N is continuous at a, there is a $\delta > 0$ such that $d_Y(f_N(x), f_N(a)) < \frac{\epsilon}{3}$ whenever $d_X(x, a) < \delta$. If $d_X(x, a) < \delta$, we then have

$$d_Y(f(x), f(a)) \leq d_Y(f(x), f_N(x)) + d_Y(f_N(x), f_N(a)) + d_Y(f_N(a), f(a)) <$$

$$< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

and hence f is continuous at a. □

The technique in the proof above is quite common, and arguments of this kind are often referred to as $\frac{\epsilon}{3}$-arguments.

Exercises for Section 2.2

1. Let $f_n : \mathbb{R} \to \mathbb{R}$ be defined by $f_n(x) = \frac{x}{n}$. Show that $\{f_n\}$ converges pointwise, but not uniformly to 0.

2. Let $f_n : (0, 1) \to \mathbb{R}$ be defined by $f_n(x) = x^n$. Show that $\{f_n\}$ converges pointwise, but not uniformly to 0.

3. The function $f_n : [0, \infty) \to \mathbb{R}$ is defined by $f_n(x) = e^{-x} \left(\frac{x}{n}\right)^{ne}$.

   a) Show that $\{f_n\}$ converges pointwise.

   b) Find the maximum value of f_n. Does $\{f_n\}$ converge uniformly?

4. The function $f_n : (0, \infty) \to \mathbb{R}$ is defined by

   $$f_n(x) = n(x^{1/n} - 1)$$

   Show that $\{f_n\}$ converges pointwise to $f(x) = \ln x$. Show that the convergence is uniform on each interval $(\frac{1}{n}, n)$, $n \in \mathbb{N}$, but not on $(0, \infty)$.

5. Let $f_n : \mathbb{R} \to \mathbb{R}$ and assume that the sequence $\{f_n\}$ of continuous functions converges uniformly to $f : \mathbb{R} \to \mathbb{R}$ on all intervals $[-k, k]$, $k \in \mathbb{N}$. Show that f is continuous.

6. Assume that X is a metric space and that f_n, g_n are functions from X to $\mathbb{R}$. Show that if $\{f_n\}$ and $\{g_n\}$ converge uniformly to f and g, respectively, then $\{f_n + g_n\}$ converge uniformly to $f + g$.

7. Assume that $f_n : [a,b] \to \mathbb{R}$ are continuous functions converging uniformly to f. Show that
$$\int_a^b f_n(x)\,dx \to \int_a^b f(x)\,dx$$
Find an example which shows that this is not necessarily the case if $\{f_n\}$ only coverges pointwise to f.

8. Let $f_n : \mathbb{R} \to \mathbb{R}$ be given by $f_n(x) = \frac{1}{n}\sin(nx)$. Show that $\{f_n\}$ converges uniformly to 0, but that the sequence $\{f_n'\}$ of derivates does not converge. Sketch the graphs of f_n to see what is happening.

9. Let (X,d) be a metric space and assume that the sequence $\{f_n\}$ of continuous functions converges uniformly to f. Show that if $\{x_n\}$ is a sequence in X converging to x, then $f_n(x_n) \to f(x)$. Find an example which shows that this is not necessarily the case if $\{f_n\}$ only converges pointwise to f.

10. Assume that the functions $f_n : X \to Y$ converges uniformly to f, and that $g : Y \to Z$ is uniformly continuous. Show that the sequence $\{g \circ f_n\}$ converges uniformly. Find an example which shows that the conclusion does not necessarily hold if g is only pointwise continuous.

11. Assume that $\sum_{n=0}^{\infty} M_n$ is a convergent series of positive numbers. Assume that $f_n : X \to \mathbb{R}$ is a sequence of continuous functions defined on a metric space (X,d). Show that if $|f_n(x)| \leq M_n$ for all $x \in X$ and all $n \in N$, then the partial sums $s_N(x) = \sum_{n=0}^{N} f_n(x)$ converge uniformly to a continuous function $s : X \to \mathbb{R}$ as $N \to \infty$. (This is called *Weierstrass' M-test*).

12. Assume that (X,d) is a compact space and that $\{f_n\}$ is a decreasing sequence of continuous functions converging pointwise to a continuous function f. Show that the convergence is uniform (this is called *Dini's theorem*).

2.3 The spaces $C(X,Y)$

If (X,d_X) and (Y,d_Y) are metric spaces, we let
$$C(X,Y) = \{f : X \to Y \mid f \text{ is continuous}\}$$
be the collection of all continuous functions from X to Y. In this section we shall see how we can turn $C(X,Y)$ into a metric space. To avoid certain technicalities, we shall restrict ourselves to the case where X is compact as this is sufficient to cover most interesting applications (see Exercise 4 for one possible way of extending the theory to the non-compact case).

The basic idea is to measure the distance between two functions by looking at the point they are the furthest apart; i.e. by
$$\rho(f,g) = \sup\{d_Y(f(x),g(x)) \mid x \in X\}$$
Our first task is to show that ρ is a metric on $C(X,Y)$. But first we need a lemma:

Lemma 2.3.1 *Let* (X, d_X) *and* (Y, d_Y) *be metric spaces, and assume that* X *is compact. If* $f.g : X \to Y$ *are continuous functions, then*

$$\rho(f, g) = \sup\{d_Y(f(x), g(x)) \,|\, x \in X\}$$

is finite, and there is a point $x \in X$ *such that* $d_Y(f(x), g(x)) = \rho(f, g)$.

Proof: The result will follow from the Extreme Value Theorem (Theorem 1.5.9) if we can only show that the function

$$h(x) = d_Y(f(x), g(x))$$

is continiuous. By the triangle inequality

$$d_Y(f(x), g(x)) \leq d_Y(f(x), f(y)) + d_Y(f(y), g(y)) + d_Y(g(y), g(x))$$

and hence

$$d_Y(f(x), g(x)) - d_Y(f(y), g(y)) \leq d_Y(f(x), f(y)) + d_Y(g(y), g(x))$$

By symmetry, we also have

$$d_Y(f(y), g(y)) - d_Y(f(x), g(x)) \leq d_Y(f(x), f(y)) + d_Y(g(y), g(x))$$

and thus

$$|h(x) - h(y)| = |d_Y(f(x), g(x)) - d_Y(f(y), g(y))| \leq$$

$$\leq d_Y(f(x), f(y)) + d_Y(g(y), g(x))$$

To prove that f is continuous at x, observe that since f and g are continuous at x, there is for any given $\epsilon > 0$ a $\delta > 0$ such that $d_Y(f(x), f(y)) < \frac{\epsilon}{2}$ and $d_Y(g(x), g(y)) < \frac{\epsilon}{2}$ when $d_X(x, y) < \delta$. But then

$$|h(x) - h(y)| \leq d_Y(f(x), f(y)) + d_Y(g(y), g(x)) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

whenever $d_X(x, y) < \delta$, and hence h is continuous. □

We are now ready to prove that ρ is a metric on $C(X, Y)$:

Proposition 2.3.2 *Let* (X, d_X) *and* (Y, d_Y) *be metric spaces, and assume that* X *is compact. Then*

$$\rho(f, g) = \sup\{d_Y(f(x), g(x)) \,|\, x \in X\}$$

defines a metric on $C(X, Y)$.

Proof: By the lemma, $\rho(f,g)$ is always finite, and we only have to prove that ρ satisfies the three properties of a metric: positivity, symmetry, and the triangle inequality. The first two are more or less obvious, and we concentrate on the triangle inequality:

Assume that f,g,h are three functions in $C(X,Y)$; we must show that

$$\rho(f,g) \leq \rho(f,h) + \rho(h,g)$$

According to the lemma, there is a point $x \in X$ such that $\rho(f,g) = d_Y f(x), g(x))$. But then

$$\rho(f,g) = d_Y(f(x),g(x)) \leq d_Y(f(x),h(x)) + d_Y(h(x),g(x)) \leq \rho(f,h) + \rho(h,g)$$

where we have used the triangle inequality in Y and the definition of ρ. □

Not surprisingly, convergence in $C(X,Y)$ is exactly the same as uniform convergence.

Proposition 2.3.3 *A sequence $\{f_n\}$ converges to f in $(C(X,Y),\rho)$ if and only if it converges uniformly to f.*

Proof: According to Proposition 2.2.3, $\{f_n\}$ converges uniformly to f if and only if

$$\sup\{d_Y(f_n(x), f(x)) \,|\, x \in X\} \to 0$$

This just means that $\rho(f_n, f) \to 0$, which is to say that $\{f_n\}$ converges to f in $(C(X,Y),\rho)$. □

The next result is the starting point for nany applications; it shows that $C(X,Y)$ is complete if Y is.

Theorem 2.3.4 *Assume that (X, d_X) is a compact and (Y, d_Y) a complete metric space. Then $C(X,Y),\rho)$ is complete.*

Proof: Assume that $\{f_n\}$ is a Cauchy sequence in $C(X,Y)$. We must prove that f_n converges to a function $f \in C(X,Y)$.

Fix an element $x \in X$. Since $d_Y(f_n(x), f_m(x)) \leq \rho(f_n, f_m)$ and $\{f_n\}$ is a Cauchy sequence in $(C(X,Y),\rho)$, the function values $\{f_n(x)\}$ form a Cauchy sequence in Y. Since Y is complete, $\{f_n(x)\}$ converges to a point $f(x)$ in Y. This means that $\{f_n\}$ converges *pointwise* to a function $f : X \to Y$. We must prove that $f \in C(X,Y)$ and that $\{f_n\}$ converges to f in the ρ-metric.

Since $\{f_n\}$ is a Cauchy sequence, we can for any $\epsilon > 0$ find an $N \in \mathbb{N}$ such that $\rho(f_n, f_m) < \frac{\epsilon}{2}$ when $n,m \geq N$. This means that all $x \in X$ and all $n,m \geq N$, $d_Y(f_n(x), f_m(x)) < \frac{\epsilon}{2}$. If we let $m \to \infty$, we see that for all $x \in X$ and all $n \geq N$

$$d_Y(f_n(x), f(x)) = \lim_{m\to\infty} d_Y(f_n(x), f_m(x)) \leq \frac{\epsilon}{2} < \epsilon$$

This means that $\{f_n\}$ converges uniformly to f. According to Proposition 2,2.4, f is continuos and belongs to $C(X,Y)$, and according to the propositio above, $\{f_n\}$ converges to f in $(C(X,Y),\rho)$. □

In the next section we shall combine the result above with Banach's Fixed Point Theorem to prove our first real application.

Exercises to Section 2.3

1. Let $f,g : [0,1] \to \mathbb{R}$ be given by $f(x) = x$, $g(x) = x^2$. Find $\rho(f,g)$.

2. Let $f,g : [0,2\pi] \to \mathbb{R}$ be given by $f(x) = \sin x$, $g(x) = \cos x$. Find $\rho(f,g)$.

3. Complete the proof of Proposition 2.3.2 by showing that ρ satisfies the first two conditions on a metric (positivity and symmetry)

4. The main reason why we have restricted the theory above to the case where X is compact, is that if not,

$$\rho(f,g) = \sup\{d_Y(f(x),g(x)) \mid x \in X\}$$

may be infinite, and then ρ is not a metric. In this problem we shall sketch a way to avoid this problem.

A function $f : X \to Y$ is called *bounded* if there is a point $a \in Y$ and a constant $K \in \mathbb{R}$ such that $d_Y(a,f(x)) \leq K$ for all $x \in X$ (it doesn't matter which point a we use in this definition). Let $C_0(X,Y)$ be the set of all bounded, continuous functions $f : X \to Y$, and define

$$\rho(f,g) = \sup\{d_Y(f(x),g(x)) \mid x \in X\}$$

a) Show that $\rho(f,g) < \infty$ for all $f,g \in C_0(X,Y)$.

b) Show by an example that there need not be a point x in X such that $\rho(f,g) = d_Y(f(x),g(x))$.

c) Show that ρ is a metric on $C_0(X,Y)$.

d) Show that if a sequence $\{f_n\}$ of functions in $C_0(X,Y)$ converges uniformly to a function f, then $f \in C_0(X,Y)$.

e) Assume that (Y,d_Y) is complete. Show that $(C_0(X,Y),\rho)$ is complete.

f) Let c_0 be the set of all bounded sequences in $\mathbb{R}$. If $\{x_n\}$, $\{y_n\}$ are in c_0, define

$$\rho(\{x_n\},\{y_n\}) = \sup(|x_n - y_n| \mid n \in \mathbb{N}\}$$

Prove that (c_0,ρ) is a complete metric space. (*Hint:* You may think of c_0 as $C_0(\mathbb{N},\mathbb{R})$ where $\mathbb{N}$ has the discrete metric).

2.4 Applications to differential equations

Consider a system of differential equations

$$\begin{array}{ccc} y_1'(t) & = & f_1(t, y_1(t), y_2(t), \dots, y_n(t)) \\ y_2'(t) & = & f_2(t, y_1(t), y_2(t), \dots, y_n(t)) \\ \vdots & \vdots & \vdots \quad\quad \vdots \\ y_n'(t) & = & f_n(t, y_1(t), y_2(t), \dots, y_n(t)) \end{array}$$

with initial conditions $y_1(0) = Y_1$, $y_2(0) = Y_2$, $\dots$, $y_n(0) = Y_n$. In this section we shall use Banach's Fixed Point Theorem 1.4.5 and the completeness of $C([0,a], \mathbb{R}^n)$ to prove that under reasonable conditions such systems have a unique solution.

We begin by introducing vector notation to make the formulas easier to read:

$$\mathbf{y}(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix}$$

$$\mathbf{y}_0 = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

and

$$\mathbf{f}(t, \mathbf{y}(t)) = \begin{pmatrix} f_1(t, y_1(t), y_2(t), \dots, y_n(t)) \\ f_2(t, y_1(t), y_2(t), \dots, y_n(t)) \\ \vdots \\ f_n(t, y_1(t), y_2(t), \dots, y_n(t)) \end{pmatrix}$$

In this notation, the system becomes

$$\mathbf{y}'(t) = \mathbf{f}(t, \mathbf{y}(t)), \qquad \mathbf{y}(0) = \mathbf{y}_0 \tag{2.4.1}$$

The next step is to rewrite the differential equation as an integral equation. If we integrate on both sides of (2.4.1), we get

$$\mathbf{y}(t) - \mathbf{y}(0) = \int_0^t \mathbf{f}(s, \mathbf{y}(s))\, ds$$

i.e.

$$\mathbf{y}(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}(s, \mathbf{y}(s))\, ds \tag{2.4.2}$$

On the other hand, if we start with a solution of (2.4.2) and differentiate, we arrive at (2.4.1). Hence solving (2.4.1) and (2.4.2) amounts to exactly the same thing, and for us it will be convenient to concentrate on (2.4.2).

Let us begin by putting an arbitrary, continuous function $\mathbf{z}$ into the right hand side of (2.4.2). What we get out is another function $\mathbf{u}$ defined by

$$\mathbf{u}(t) = y_0 + \int_0^t \mathbf{f}(s, \mathbf{z}(s))\, ds$$

We can think of this as a function F mapping continuous functions $\mathbf{z}$ to continuous functions $\mathbf{u} = F(\mathbf{z})$. From this point of view, a solution $\mathbf{y}$ of the integral equation (2.4.2) is just a fixed point for the function F — we are looking for a $\mathbf{y}$ such that $\mathbf{y} = F(\mathbf{y})$. (Don't worry if you feel a little dizzy; that's just normal at this stage! Note that F is a function acting on a function $\mathbf{z}$ to produce a new function $\mathbf{u} = F(\mathbf{z})$ — it takes some time to get used to such creatures!)

Our plan is to use Banach's Fixed Point Theorem to prove that F has a unique fixed point, but first we have to introduce a crucial condition. We say that the function $\mathbf{f} : [a, b] \times \mathbb{R}^n \to \mathbb{R}^n$ is *uniformly Lipschitz with Lipschitz constant* K *on the interval* $[a, b]$ if K is a real number such that

$$|\mathbf{f}(t, \mathbf{y}) - \mathbf{f}(t, \mathbf{z})| \leq K|\mathbf{y} - \mathbf{z}|$$

for all $t \in [a, b]$ and all $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$. Here is the key observation in our argument.

Lemma 2.4.1 *Assume that* $\mathbf{y}_0 \in \mathbb{R}^n$ *and that* $\mathbf{f} : [0, \infty] \times \mathbb{R}^n \to \mathbb{R}^n$ *is continuous and uniformly Lipschitz with Lipschitz constant* K *on* $[0, \infty)$*. If* $a < \frac{1}{K}$*, the map*

$$F : C([0, a], \mathbb{R}^n) \to C([0, a], \mathbb{R}^n)$$

defined by

$$F(\mathbf{z})(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}(t, \mathbf{z}(t))\, dt$$

is a contraction.

Remark: The notation here is rather messy. Remember that $F(\mathbf{z})$ is a function from $[0, a]$ to $\mathbb{R}^n$. The expression $F(\mathbf{z})(t)$ denotes the value of this function at point $t \in [o, a]$.

Proof: Let $\mathbf{v}, \mathbf{w}$ be two elements in $C([0, a], \mathbb{R}^n)$, and note that for any $t \in [0, a]$

$$|F(\mathbf{v})(t) - F(\mathbf{w})(t)| = |\int_0^t \big(\mathbf{f}(s, \mathbf{v}(s)) - \mathbf{f}(s, \mathbf{w}(s))\big)\, ds| \leq$$

$$\leq \int_0^t |\mathbf{f}(s,\mathbf{v}(s)) - \mathbf{f}(s,\mathbf{w}(s))|\, ds \leq \int_0^t K|\mathbf{v}(s) - \mathbf{w}(s)|\, ds \leq$$

$$\leq K \int_0^t \rho(\mathbf{v},\mathbf{w})\, ds \leq K \int_0^a \rho(\mathbf{v},\mathbf{w})\, ds = Ka\, \rho(\mathbf{v},\mathbf{w})$$

Taking the supremum over all $t \in [0,a]$, we get

$$\rho(F(\mathbf{v}), F(\mathbf{w})) \leq Ka\, \rho(\mathbf{v},\mathbf{w}).$$

Since $Ka < 1$, this means that F is a contraction. □

We are now ready for the main theorem.

Theorem 2.4.2 *Assume that* $\mathbf{y}_0 \in \mathbb{R}^n$ *and that* $\mathbf{f} : [0,t] \times \mathbb{R}^n \to \mathbb{R}^n$ *is continuous and uniformly Lipschitz on* $[0,\infty)$*. Then the initial value problem*

$$\mathbf{y}'(t) = \mathbf{f}(t,\mathbf{y}(t)), \qquad \mathbf{y}(0) = \mathbf{y}_0 \tag{2.4.3}$$

has a unique solution $\mathbf{y}$ *on* $[0,\infty)$*.*

Proof: Let K be the uniform Lipschitz constant, and choose a number $a < 1/K$. According to the lemma, the function

$$F : C([0,a],\mathbb{R}^n) \to C([0,a],\mathbb{R}^n)$$

defined by

$$F(\mathbf{z})(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}(t,\mathbf{z}(t))\, dt$$

is a contraction. Since $C([0,a],\mathbb{R}^n)$ is complete by Theorem 2.3.4, Banach's Fixed Point Theorem tells us that F has a unique fixed point $\mathbf{y}$. This means that the integral equation

$$\mathbf{y}(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}(s,\mathbf{y}(s))\, ds \tag{2.4.4}$$

has a unique solution on the interval $[0,a]$. To extend the solution to a longer interval, we just repeat the argument on the interval $[a,2a]$, using $\mathbf{y}(a)$ as initial value. The function we then get, is a solution of the integral equation (2.4.4) on the extended interval $[0,2a]$ as we for $t \in [a,2a]$ have

$$\mathbf{y}(t) = \mathbf{y}(a) + \int_a^t \mathbf{f}(s,\mathbf{y}(s))\, ds =$$

$$= \mathbf{y}_0 + \int_0^a \mathbf{f}(s,\mathbf{y}(s))\, ds + \int_a^t \mathbf{f}(s,\mathbf{y}(s))\, ds = \mathbf{y}_0 + \int_0^t \mathbf{f}(s,\mathbf{y}(s))\, ds$$

Continuing this procedure to new intervals $[2a,3a]$, $[3a,4a]$, we see that the integral equation (2.4.3) has a unique solution on all of $[0,\infty)$. As we have already observed that equation (2.4.3) has exactly the same solutions as equation (2.4.4), the theorem is proved. □

In the exercises you will see that the conditions in the theorem are important. If they fail, the equation may have more than one solution, or a solution defined only on a bounded interval.

Exercises to Section 2.4

1. Solve the initial value problem

$$y' = 1 + y^2, \qquad y(0) = 0$$

and show that the solution is only defined on the interval $[0, \pi/2)$.

2. Show that the functions

$$y(t) = \begin{cases} 0 & \text{if } 0 \leq t \leq a \\ (t-a)^{\frac{3}{2}} & \text{if } t > a \end{cases}$$

where $a \geq 0$ are all solutions of the initial value problem

$$y' = \frac{3}{2} y^{\frac{1}{3}}, \qquad y(0) = 0$$

Remember to check that the differential equation is satisfied at $t = a$.

3. In this problem we shall sketch how the theorem in this section can be used to study higher order systems. Assume we have a second order initial value problem

$$u''(t) = g(t, u(t), u'(t)) \qquad u(0) = a, u'(0) = b \qquad (*)$$

where $g : [0, \infty) \times \mathbb{R}^2 \to \mathbb{R}$ is a given function. Define a function $\mathbf{f} : [0, \infty) \times \mathbb{R}^2 \to \mathbb{R}^2$ by

$$\mathbf{f}(t, u, v) = \begin{pmatrix} v \\ g(t, u, v) \end{pmatrix}$$

Show that if

$$\mathbf{y}(t) = \begin{pmatrix} u(t) \\ v(t) \end{pmatrix}$$

is a solution of the initial value problem

$$\mathbf{y}'(t) = \mathbf{f}(t, \mathbf{y}(t)). \qquad \mathbf{y}(0) = \begin{pmatrix} a \\ b \end{pmatrix},$$

then u is a solution of the original problem $(*)$.

2.5 Compact subsets of $C(X, \mathbb{R}^m)$

The compact subsets of $\mathbb{R}^m$ are easy to describe — they are just the closed and bounded sets. This characterization is extremely useful as it is much easier to check that a set is closed and bounded than to check that it satisfies the definition of compactness. In the present section we shall prove a similar kind of characterization of compact sets in $C(X, \mathbb{R}^m)$ — we shall show that a subset of $C(X, \mathbb{R}^m)$ is compact if and only if it it closed, bounded and equicontinuous. This is known as the Arzelà-Ascoli Theorem. But before we turn to it, we have a question of independent interest to deal with.

Definition 2.5.1 *Let (X, d) be a metric space and assume that A is a subset of X. We say that A is* dense *in X if for each $x \in X$ there is a sequence from A converging to x.*

We know that $\mathbb{Q}$ is dense in $\mathbb{R}$ — we may, e.g., approximate a real number by longer and longer parts of its decimal expansion. For $x = \sqrt{2}$ this would mean the approximating sequence

$$a_1 = 1.4 = \frac{14}{10}, \ \ a_2 = 1.41 = \frac{141}{100}, \ \ a_3 = 1.414 = \frac{1414}{1000}, \ \ a_4 = 1.4142 = \frac{14142}{10000}, \ldots$$

Recall that $\mathbb{Q}$ is countable, but that $\mathbb{R}$ is not. Still every element in the uncountable set $\mathbb{R}$ can be approximated arbitrarily well by elements in the much smaller set $\mathbb{Q}$. This property turns out to be so useful that it deserves a name.

Definition 2.5.2 *A metric set (X, d) is called* separable *if it has a countable, dense subset A.*

Our first result is a simple, but rather surprising connection between separability and compactness.

Proposition 2.5.3 *All compact metric (X, d) spaces are separable. We can choose the countable dense set A in such a way that for any $\delta > 0$, there is a finite subset A_δ of A such that all elements of X are within distance less than δ of A_δ, i.e. for all $x \in X$ there is an $a \in A_\delta$ such that $d(x, a) < \delta$.*

Proof: We use that a compact space X is totally bounded (recall Theorem 1.5.11). This mean that for all $n \in \mathbb{N}$, there is a finite number of balls of radius $\frac{1}{n}$ that cover X. The centers of all these balls form a countable subset A of X (to get a listing of A, first list the centers of the balls of radius 1, then the centers of the balls of radius $\frac{1}{2}$ etc.). We shall prove that A is dense in X.

Let x be an element of X. To find a sequence $\{a_n\}$ from A converging to x, we first pick the center a_1 of (one of) the balls of radius 1 that x belongs to, then we pick the center a_2 of (one of) the balls of radius $\frac{1}{2}$ that x belong to, etc. Since $d(x, a_n) < \frac{1}{n}$, $\{a_n\}$ is a sequence from A converging to x.

To find the set A_δ, just choose $m \in \mathbb{N}$ so big that $\frac{1}{m} < \delta$, and let A_δ consist of the centers of the balls of radius $\frac{1}{m}$. □

We are now ready to turn to $C(X, \mathbb{R}^m)$. First we recall the definition of equicontinuous sets of functions from Section 2.1.

Definition 2.5.4 *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $\mathcal{F}$ be a collection of functions $f : X \to Y$. We say that $\mathcal{F}$ is* equicontinuous *if for all $\epsilon > 0$, there is a $\delta > 0$ such that for all $f \in \mathcal{F}$ and all $x, y \in X$ with $d_X(x, y) < \delta$, we have $d_Y(f(x), f(y)) < \epsilon$.*

We dive straight in by first proving the most difficult and important ingredient in the Arzelà-Ascoli Theorem.

Proposition 2.5.5 *Assume that (X, d) is a compact metric space, and let $\{f_n\}$ be a bounded and equicontinuous sequence in $C(X, \mathbb{R}^m)$. Then $\{f_n\}$ has a subsequence converging in $C(X, \mathbb{R}^m)$.*

Proof: Since X is compact, there is a countable, dense subset

$$A = \{a_1, a_2 \ldots, a_n, \ldots\}$$

as in Proposition 2.5.3. The hard part of the proof is to find a subsequence $\{g_k\}$ of $\{f_n\}$ which is such that $\{g_k(a)\}$ converges for all $a \in A$. Let us first check that such a sequence $\{g_k\}$ will necessarily converge in $C(X, \mathbb{R}^m)$ (it is here we need the equicontinuity), and hence do the job for us.

Since $C(X, \mathbb{R}^m)$ is complete, it suffices to prove that $\{g_k\}$ is a Cauchy sequence in $C(X, \mathbb{R}^m)$. Given an $\epsilon > 0$, we must thus find an $N \in \mathbb{N}$ such that $\rho(g_n, g_m) < \epsilon$ when $n, m \geq N$. Since the original sequence is equicontinuous, there exists a $\delta > 0$ such that if $d_X(x, y) < \delta$, then $d_{\mathbb{R}}(f_n(x), f_n(y)) < \frac{\epsilon}{4}$ for all n. Since $\{g_k\}$ is a subsequence of $\{f_n\}$, we clearly have $d_{\mathbb{R}}(g_k(x), g_k(y)) < \frac{\epsilon}{4}$ for all k. Choose a finite subset A_δ of A such that any element in X is within less than δ of an element in A_δ. Since the sequences $\{g_k(a)\}$, $a \in A_\delta$, converge, they are all Cauchy sequences, and we can find an $N \in \mathbb{N}$ such that when $n, m \geq N$, $d_{\mathbb{R}^m}(g_n(a), g_m(a)) < \frac{\epsilon}{4}$ for all $a \in A_\delta$ (here we are using that A_δ is finite).

For any $x \in X$, we can find an $a \in A_\delta$ such that $d_X(x, a) < \delta$. But then for all $n, m \geq N$,

$$d_{\mathbb{R}^m}(g_n(x), g_m(x)) \leq d_{\mathbb{R}^m}(g_n(x), g_n(a)) + d_{\mathbb{R}^m}(g_n(a), g_m(a)) + d_{\mathbb{R}^m}(g_m(a), g_m(x))$$

$$< \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \frac{3\epsilon}{4}$$

Since this holds for any $x \in X$, we must have $\rho(g_n, g_m) \leq \frac{3\epsilon}{4} < \epsilon$ for all $n, m \geq N$, and hence $\{g_k\}$ is a Cauchy sequence and converges.

It remains to prove that the original sequence $\{f_n\}$ really has a subsequence $\{g_k\}$ such that $\{g_k(a)\}$ converges for all $a \in A$. We begin a little less ambitiously by showing that $\{f_n\}$ has a subsequence $\{f_n^{(1)}\}$ such that $\{f_n^{(1)}(a_1)\}$ converges (recall that a_1 is the first element in our listing of the countable set A). Next we show that $\{f_n^{(1)}\}$ has a subsequence $\{f_n^{(2)}\}$ such that both $\{f_n^{(2)}(a_1)\}$ and $\{f_n^{(2)}(a_2)\}$ converge. Continuing taking subsequences in this way, we shall for each $j \in \mathbb{N}$ find a sequence $\{f_n^{(j)}\}$ such that $\{f_n^{(j)}(a)\}$ converges for $a = a_1, a_2, \ldots, a_j$. Finally, we shall construct the sequence $\{g_k\}$ by combining all the sequences $\{f_n^{(j)}\}$ in a clever way.

Let us start by constructing $\{f_n^{(1)}\}$. Since the sequence $\{f_n\}$ is bounded, $\{f_n(a_1)\}$ is a bounded sequence in $\mathbb{R}^m$, and by Bolzano-Weierstrass' Theorem, it has a convergent subsequence $\{f_{n_k}(a_1)\}$. We let $\{f_n^{(1)}\}$ consist of the functions appearing in this subsequence. If we now apply $\{f_n^{(1)}\}$ to a_2, we get a new bounded sequence $\{f_n^{(1)}(a_2)\}$ in $\mathbb{R}^m$ with a convergent subsequence. We let $\{f_n^{(2)}\}$ be the functions appearing in this subsequence. Note that $\{f_n^{(2)}(a_1)\}$ still converges as $\{f_n^{(2)}\}$ is a subsequence of $\{f_n^{(1)}\}$. Continuing in this way, we see that we for each $j \in \mathbb{N}$ have a sequence $\{f_n^{(j)}\}$ such that $\{f_n^{(j)}(a)\}$ converges for $a = a_1, a_2, \ldots, a_j$. In addition, each sequence $\{f_n^{(j)}\}$ is a subsequence of the previous ones.

We are now ready to construct a sequence $\{g_k\}$ such that $\{g_k(a)\}$ converges for *all* $a \in A$. We do it by a diagonal argument, putting g_1 equal to the first element in the first sequence $\{f_n^{(1)}\}$, g_2 equal to the second element in the second sequence $\{f_n^{(2)}\}$ etc. In general, the k-th term in the g-sequence equals the k-th term in the k-th f-sequence $\{f_n^k\}$, i.e. $g_k = f_k^{(k)}$. Note that except for the first few elements, $\{g_k\}$ is a subsequence of *any* sequence $\{f_n^{(j)}\}$. This means that $\{g_k(a)\}$ converges for all $a \in A$, and the proof is complete. □

As a simple consequence of this result we get:

Corollary 2.5.6 *If (X, d) is a compact metric space, all bounded, closed and equicontinuous sets $\mathcal{K}$ in $C(X, \mathbb{R}^m)$ are compact.*

Proof: According to the proposition, any sequence in $\mathcal{K}$ has a convergent subsequence. Since $\mathcal{K}$ is closed, the limit must be in $\mathcal{K}$, and hence $\mathcal{K}$ is compact. □

As already mentioned, the converse of this result is also true, but before we prove it, we need a technical lemma that is quite useful also in other situations:

Lemma 2.5.7 *Assume that* (X, d_X) *and* (Y, d_Y) *are metric spaces and that* $\{f_n\}$ *is a sequence of continuous function from* X *to* Y *which converges uniformly to* f*. If* $\{x_n\}$ *is a sequence in* X *xonverging to* a*, then* $\{f_n(x_n)\}$ *converges to* $f(a)$*.*

Remark This lemma is not as obvious as it may seem — it is not true if we replace uniform convergence by pointwise!

Proof of Lemma 2.5.7: Given $\epsilon > 0$, we must show how to find an $N \in \mathbb{N}$ such that $d_Y(f_n(x_n), f(a)) < \epsilon$ for all $n \geq N$. Since we know from Proposition 2.2.4 that f is continuous, there is a $\delta > 0$ such that $d_Y(f(x), f(a)) < \frac{\epsilon}{2}$ when $d_X(x, a) < \delta$. Since $\{x_n\}$ converges to x, there is an $N_1 \in \mathbb{N}$ such that $d_X(x_n, a) < \delta$ when $n \geq N_1$. Also, since $\{f_n\}$ converges uniformly to f, there is an $N_2 \in \mathbb{N}$ such that if $n \geq N_2$, then $d_Y(f_n(x), f(x)) < \frac{\epsilon}{2}$ for all $x \in X$. If we choose $N = \max\{N_1, N_2\}$, we see that if $n \geq N$,

$$d_Y(f_n(x_n), f(a)) \leq d_Y(f_n(x_n), f(x_n)) + d_Y(f(x_n), f(a)) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

and the lemma is proved. □

We are finally ready to prove the main theorem:

Theorem 2.5.8 (Arzelà-Ascoli's Theorem) *Let* (X, d_X) *be a compact metric space. A subset* $\mathcal{K}$ *of* $C(X, \mathbb{R}^m)$ *is compact if and only if it is closed, bounded and equicontinuous.*

Proof: It remains to prove that a compact set $\mathcal{K}$ in $C(X, \mathbb{R}^m)$ is closed, bounded and equicontinuous. Since compact sets are always closed and bounded according to Proposition 1.5.4, if suffices to prove that $\mathcal{K}$ is equicontinuous. We argue by contradiction: We assume that the compact set $\mathcal{K}$ is *not* equicontinuous and show that this leads to a contradiction.

Since $\mathcal{K}$ is not equicontinuous, there must be an $\epsilon > 0$ which can not be matched by any δ; i.e. for any $\delta > 0$, there is a function $f \in \mathcal{K}$ and points $x, y \in X$ such that $d_X(x, y) < \delta$, but $d_{\mathbb{R}^m}(f(x), f(y)) \geq \epsilon$. If we put $\delta = \frac{1}{n}$, we get at function $f_n \in \mathcal{K}$ and points $x_n, y_n \in X$ such that $d_X(x_n, y_n) < \frac{1}{n}$, but $d_{\mathbb{R}^m}(f_n(x_n), f(y_n)) \geq \epsilon$. Since $\mathcal{K}$ is compact, there is a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ which converges (uniformly) to a function $f \in \mathcal{K}$. Since X is compact, the corresponding subsequence $\{x_{n_k}\}$ of $\{x_n\}$, has a subsequence $\{x_{n_{k_j}}\}$ converging to a point $a \in X$. Since $d_X(x_{n_{k_j}}, y_{n_{k_j}}) < \frac{1}{n_{k_j}}$, the corresponding sequence $\{y_{n_{k_j}}\}$ of y's also converges to a.

Since $\{f_{n_{k_j}}\}$ converges uniformly to f, and $\{x_{n_{k_j}}\}$, $\{y_{n_{k_j}}\}$ both converge to a, the lemma tells us that

$$f_{n_{k_j}}(x_{n_{k_j}}) \to f(a) \quad \text{and} \quad f_{n_{k_j}}(y_{n_{k_j}}) \to f(a)$$

But this is impossible since $d_{\mathbb{R}^m}(f(x_{n_{k_j}}), f(y_{n_{k_j}})) \geq \epsilon$ for all j. Hence we have our contradiction, and the theorem is proved. □

Exercises for Section 2.5

1. Show that $\mathbb{R}^n$ is separable for all n.

2. Show that a subset A of a metric space (X, d) is dense if and only if all open balls $\mathrm{B}(a, r)$, $a \in X$, $r > 0$, contain elements from X.

3. Assume that (X, d) is a complete metric space, and that A is a dense subset of X. We let A have the subset metric d_A.

   a) Assume that $f : A \to \mathbb{R}$ is uniformly continuous. Show that if $\{a_n\}$ is a sequence from A converging to a point $x \in X$, then $\{f(a_n)\}$ converges. Show that the limit is the same for all such sequences $\{a_n\}$ converging to the same point x.

   b) Define $\bar{f} : X \to \mathbb{R}$ by putting $\bar{f}(x) = \lim_{n\to\infty} f(a_n)$ where $\{a_n\}$ is a sequence from a converging to x. We call f the *continuous extension of f to X*. Show that $\bar{f}$ is uniformly continuous.

   c) Let $f : \mathbb{Q} \to \mathbb{R}$ be defined by

   $$f(q) = \begin{cases} 0 & \text{if } q < \sqrt{2} \\ \\ 1 & \text{if } q > \sqrt{2} \end{cases}$$

   Show that f is continuous on $\mathbb{Q}$ (we are using the usual metric $d_{\mathbb{Q}}(q, r) = |q - r|$). Is f uniformly continuous?

   d) Show that f does not have a continuous extension to $\mathbb{R}$.

4. Let K be a compact subset of $\mathbb{R}^n$. Let $\{f_n\}$ be a sequence of contractions of K. Show that $\{f_n\}$ has uniformly convergent subsequence.

5. A function $f : [-1, 1] \to \mathbb{R}$ is called *Lipschitz continuous with Lipschitz constant* $K \in \mathbb{R}$ if

   $$|f(x) - f(y)| \leq K|x–y|$$

   for all $x, y \in [-1, 1]$. Let $\mathcal{K}$ be the set of all Lipschitz continuous functions with Lipschitz constant K such that $f(0) = 0$. Show that $\mathcal{K}$ is a compact subset of $C([-1, 1], \mathbb{R})$.

6. Assume that (X, d_X) and (Y, d_Y) are two metric spaces, and let $\sigma : [0, \infty) \to [0, \infty)$ be a nondecreasing, continuous function such that $\sigma(0) = 0$. We say that σ is a *modulus of continuity* for a function $f : X \to Y$ if

   $$d_Y(f(u), f(v)) \leq \sigma(d_X(u, v))$$

   for all $u, v \in X$.

   a) Show that a family of functions with the same modulus of continuity is equicontinuous.

b) Assume that (X, d_X) is compact, and let $x_0 \in X$. Show that if σ is a modulus of continuity, then the set

$$\mathcal{K} = \{f : X \to \mathbb{R}^n \ : \ f(x_0) = \mathbf{0} \text{ and } \sigma \text{ is modulus of continuity for } f\}$$

is compact.

c) Show that all functions in $C([a,b], \mathbb{R}^m)$ has a modulus of continuity.

7. A metric space (X, d) is called *locally compact* if for each point $a \in X$, there is a *closed* ball $\overline{\mathrm{B}}(a; r)$ centered at a that is compact. (Recall that $\overline{\mathrm{B}}(a; r) = \{x \in X \ : \ d(a, x) \leq r\}$). Show that $\mathbb{R}^m$ is locally compact, but that $C([0,1], \mathbb{R})$ is not.

2.6 Differential equations revisited

In Section 2.4, we used Banach's Fixed Point Theorem to study initial value problems of the form

$$\mathbf{y}'(t) = \mathbf{f}(t, \mathbf{y}(t)), \quad \mathbf{y}(0) = \mathbf{y}_0 \tag{2.6.1}$$

or equivalently

$$\mathbf{y}(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}(s, \mathbf{y}(s))\, ds \tag{2.6.2}$$

In this section we shall see how Arzelà-Ascoli's Theorem can be used to prove existence of solutions under weaker conditions than before. But in the new approach we shall also lose something — we can only prove that the solutions exist in small intervals, and we can no longer guarantee uniqueness.

The starting point is Euler's method for finding approximate solutions to differential equations. If we want to approximate the solution starting at $\mathbf{y}_0$ at time $t = 0$, we begin by partitioning time into discrete steps of length Δt; hence we work with the time line

$$T = \{t_0, t_1, t_2, t_3 \ldots\}$$

where $t_0 = 0$ and $t_{i+1} - t_i = \Delta t$. We start the approximate solution $\hat{\mathbf{y}}$ at $\mathbf{y}_0$ and move in the direction of the derivative $\mathbf{f}(t_0, \mathbf{y}_0)$, i.e. we put

$$\hat{\mathbf{y}}(t) = \mathbf{y}_0 + \mathbf{f}(t_0, \mathbf{y}_0)(t - t_0)$$

for $t \in [t_0, t_1]$. Once we reach t_1, we change directions and move in the direction of the new derivative $\mathbf{f}(t_1, \hat{\mathbf{y}}(t_1))$ so that we have

$$\hat{\mathbf{y}}(t) = \hat{\mathbf{y}}(t_1) + \mathbf{f}(t_0, \hat{\mathbf{y}}(t_1))(t - t_1)$$

for $t \in [t_1, t_2]$. If we insert the expression for $\hat{\mathbf{y}}(t_1)$, we get:

$$\hat{\mathbf{y}}(t) = \mathbf{y}_0 + \mathbf{f}(t_0, \mathbf{y}_0)(t_1 - t_0) + \mathbf{f}(t_1, \hat{\mathbf{y}}(t_1))(t - t_1)$$

If we continue in this way, changing directions at each point in T, we get

$$\hat{\mathbf{y}}(t) = \mathbf{y}_0 + \sum_{i=0}^{k-1} \mathbf{f}(t_i, \hat{\mathbf{y}}(t_i))(t_{i+1} - t_i) + \mathbf{f}(t_k, \hat{\mathbf{y}}(t_k))(t - t_k)$$

for $t \in [t_k, t_{k+1}]$. If we observe that

$$\mathbf{f}(t_i, \hat{\mathbf{y}}(t_i))(t_{i+1} - t_i) = \int_{t_i}^{t_{i+1}} \mathbf{f}(t_i, \hat{\mathbf{y}}(t_i)\, ds\,,$$

we can rewrite this expression as

$$\hat{\mathbf{y}}(t) = \mathbf{y}_0 + \sum_{i=0}^{k-1} \int_{t_i}^{t_{i+1}} \mathbf{f}(t_i, \hat{\mathbf{y}}(t_i)\, ds + \int_{t_k}^{t} \mathbf{f}(t_k, \hat{\mathbf{y}}(t_k)\, ds$$

If we also introduce the notation

$$\underline{s} = \text{the largest } t_i \in T \text{ such that } t_i \leq s,$$

we may express this more compactly as

$$\hat{\mathbf{y}}(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}(\underline{s}, \hat{\mathbf{y}}(\underline{s}))\, ds$$

We can also write this as

$$\hat{\mathbf{y}}(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}(s, \hat{\mathbf{y}}(s))\, ds + \int_0^t \big(\mathbf{f}(\underline{s}, \hat{\mathbf{y}}(\underline{s})) - \mathbf{f}(s, \hat{\mathbf{y}}(s))\big)\, ds$$

where the last term measures how much $\hat{\mathbf{y}}$ "deviates" from being a solution of equation (2.6.2).

Intuitively, one would think that the approximate solution $\hat{\mathbf{y}}$ will converge to a real solution $\mathbf{y}$ when the step size Δt goes to zero. To be more specific, if we let $\hat{\mathbf{y}}_n$ be the approximate solution we get when we choose $\Delta t = \frac{1}{n}$, we would expext the squence $\{\hat{\mathbf{y}}_n\}$ to converge to a solution of (2). It turns out that in the most general case we can not quite prove this, but we can instead use the Arzelà-Ascoli Theorem to find a *subsequence* converging to a solution.

Before we turn to the proof, it will useful to see how intergals of the form

$$I_k(t) = \int_0^t \mathbf{f}(s, \hat{\mathbf{y}}_k(s))\, ds$$

behave when the functions $\hat{\mathbf{y}}_k$ converge uniformly to a limit $\mathbf{y}$.

Lemma 2.6.1 *Let* $\mathbf{f} : [0,\infty) \times \mathbb{R}^m$ *be a continuous function, and assume that* $\{\hat{\mathbf{y}}_k\}$ *is a sequence of continuous functions* $\hat{\mathbf{y}}_k : [0,a] \to \mathbb{R}^m$ *converging uniformly to a function* $\mathbf{y}$. *Then the integral functions*

$$I_k(t) = \int_0^t \mathbf{f}(s, \hat{\mathbf{y}}_k(s))\, ds$$

converge uniformly to

$$I(t) = \int_0^t \mathbf{f}(s, \mathbf{y}(s))\, ds$$

on $[0,a]$.

Proof: Since the sequence $\{\hat{\mathbf{y}}_k\}$ converges unifomly, it is bounded, and hence there is a constant K such that $|\hat{\mathbf{y}}_k(t)| \leq K$ for all $k \in \mathbb{N}$ and all $t \in [0,a]$ (prove this!). The continuous function $\mathbf{f}$ is uniformly continuous on the compact set $[0,a] \times [-K,K]^m$, and hence for every $\epsilon > 0$, there is a $\delta > 0$ such that if $|\mathbf{y} - \mathbf{y}'| < \delta$, then $|\mathbf{f}(s,\mathbf{y}) - \mathbf{f}(s,\mathbf{y}')| < \frac{\epsilon}{a}$ for all $s \in [0,a]$. Since $\{\hat{\mathbf{y}}_k\}$ converges uniformly to $\mathbf{y}$, there is an $N \in \mathbb{N}$ such that if $n \geq N$, $|\hat{\mathbf{y}}_n(s) - \mathbf{y}(s)| < \delta$ for all $s \in [0,a]$. But then

$$|I_n(t) - I(t)| = |\int_0^t \big(\mathbf{f}(s, \hat{\mathbf{y}}_n(s)) - \mathbf{f}(s, \mathbf{y}(s))\big)\, ds| \leq$$

$$\leq \int_0^t \big|\mathbf{f}(s, \hat{\mathbf{y}}_n(s)) - \mathbf{f}(s, \mathbf{y}(s))\big|\, ds < \int_0^a \frac{\epsilon}{a}\, ds = \epsilon$$

for all $t \in [0,a]$, and hence $\{I_k\}$ converges uniformly to I. □

We are now ready for the main result.

Theorem 2.6.2 *Assume that* $\mathbf{f} : [0,\infty) \times \mathbb{R}^m \to \mathbb{R}^m$ *is a continuous function and that* $\mathbf{y}_0 \in \mathbb{R}^m$. *Then there exists a positive real number* a *and a function* $\mathbf{y} : [0,a] \to \mathbb{R}^m$ *such that* $\mathbf{y}(0) = \mathbf{y}_0$ *and*

$$y'(t) = \mathbf{f}(t, \mathbf{y}(t)) \quad \textit{for all } t \in [0,a]$$

Remark Note that there is no uniqueness statement (the problem may have more than one solution), and that the solution is only guaranteed to exist on a bounded intervall (it may disappear to infinity after finite time).

Proof of Theorem 2.6.2: Choose a big, compact subset $C = [0,R] \times [-R,R]^m$ of $[0,\infty) \times \mathbb{R}^m$ containing $(0, \mathbf{y}_0)$ in its interior. By the Extreme Value Theorem, the components of $\mathbf{f}$ have a maximum value on C, and hence there exists a number $M \in \mathbb{R}$ such that $|f_i(t, \mathbf{y})| \leq M$ for all $(t, \mathbf{y}) \in C$ and all $i = 1, 2, \ldots, m$. If the initial value has components

$$\mathbf{y}_0 = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix}$$

we choose $a \in \mathbb{R}$ so small that the set

$$A = [0, a] \times [Y_1 - Ma, Y_1 + Ma] \times [Y_2 - Ma, Y_2 + Ma] \times \cdots \times [Y_m - Ma, Y_m + ma]$$

is contained in C. This may seem mysterious, put the point is that our approximate solutions of the differential equation can never leave the area

$$[Y_1 - Ma, Y_1 + Ma] \times [Y_2 - Ma, Y_2 + Ma] \times \cdots \times [Y_m - Ma, Y + ma]$$

while $t \in [0, a]$ since all the derivatives are bounded by M.

Let $\hat{\mathbf{y}}_n$ be the approximate solution obtained by using Euler's method on the interval $[0, a]$ with time step $\frac{a}{n}$. The sequence $\{\hat{\mathbf{y}}_n\}$ is bounded since $(t, \hat{\mathbf{y}}_n(t)) \in A$, and it is equicontinuous since the components of $\mathbf{f}$ are bounded by M. By Proposition 2.5.5, $\hat{\mathbf{y}}_n$ has a subsequece $\{\hat{\mathbf{y}}_{n_k}\}$ converging uniformly to a function $\mathbf{y}$. If we can prove that $\mathbf{y}$ solves the integral equation

$$\mathbf{y}(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}(s, \mathbf{y}(s))\, ds$$

for all $t \in [0, a]$, we shall have proved the theorem.

From the calculations at the beginning of the section, we know that

$$\hat{\mathbf{y}}_{n_k}(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}(s, \hat{\mathbf{y}}_{n_k}(s))\, ds + \int_0^t \big(\mathbf{f}(\underline{s}, \hat{\mathbf{y}}_{n_k}(\underline{s})) - \mathbf{f}(s, \hat{\mathbf{y}}_{n_k}(s))\big)\, ds \quad (2.6.3)$$

and according to the lemma

$$\int_0^t \mathbf{f}(s, \hat{\mathbf{y}}_{n_k}(s))\, ds \to \int_0^t \mathbf{f}(s, \mathbf{y}(s))\, ds \quad \text{uniformly for } t \in |0, a]$$

If we can only prove that

$$\int_0^t \big(\mathbf{f}(\underline{s}, \hat{\mathbf{y}}_{n_k}(\underline{s})) - \mathbf{f}(s, \hat{\mathbf{y}}_{n_k}(s))\big)\, ds \to 0 \quad (2.6.4)$$

we will get

$$\mathbf{y}(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}(s, \mathbf{y}(s))\, ds$$

as $k \to \infty$ in (2.6.3), and the theorem will be proved

To prove (2.6.4), observe that since A is a compact set, $\mathbf{f}$ is uniformly continuous on A. Given an $\epsilon > 0$, we thus find a $\delta > 0$ such that $|\mathbf{f}(s,\mathbf{y}) - \mathbf{f}(s',\mathbf{y}')| < \frac{\epsilon}{a}$ when $|(s,\mathbf{y}) - (s',\mathbf{y})| < \delta$ (we are measuring the distance in the ordinary $\mathbb{R}^{m+1}$-metric). Since

$$|(\underline{s}, \hat{\mathbf{y}}_{n_k}(\underline{s})) - (s, \hat{\mathbf{y}}_{n_k}(s))| \leq |(\Delta t, M\Delta t, \ldots, M\Delta t)| = \sqrt{1 + nM^2}\, \Delta t\,,$$

we can clearly get $|(\underline{s}, \hat{\mathbf{y}}_{n_k}(\underline{s})) - (s, \hat{\mathbf{y}}_{n_k}(s))| < \delta$ by choosing k large enough (and hence Δt small enough). For such k we then have

$$|\int_0^t \left(\mathbf{f}(\underline{s}, \hat{\mathbf{y}}_{n_k}(\underline{s})) - \mathbf{f}(s, \hat{\mathbf{y}}_{n_k}(s))\right| < \int_0^a \frac{\epsilon}{a}\, ds = \epsilon$$

and hence

$$\int_0^t \left(\mathbf{f}(\underline{s}, \hat{\mathbf{y}}_{n_k}(\underline{s})) - \mathbf{f}(s, \hat{\mathbf{y}}_{n_k}(s))\right) ds \to 0$$

as $k \to \infty$. As already observed, this completes the proof. □

Remark An obvious question at this stage is why didn't we extend our solution beyond the interval $[0, a]$ as we did in the proof of Theorem 2.4.2? The reason is that in the present case we do not have control over the length of our intervals, and hence the second interval may be very small compared to the first one, the third one even smaller, and so one. Even if we add an infinite number of intervals, we may still only cover a finite part of the real line. There are good reasons for this: the differential equation may only have solutions that survive for a finite amount of time. A typical example is the equation

$$y' = (1 + y^2), \quad y(0) = 0$$

where the (unique) solution $y(t) = \tan t$ goes to infinity when $t \to \frac{\pi}{2}^-$.

The proof above is a simple, but typical example of a wide class of compactness arguments in the theory of differential equations. In such arguments one usually starts with a sequence of approximate solutions and then uses compactness to extract a subsequence converging to a solution. Compactness methods are strong in the sense that they can often prove local existence of solutions under very general conditions, but they are weak in the sense that they give very little information about the nature of the solution. But just knowing that a solution exists, is often a good starting point for further explorations.

Exercises for Section 2.6

1. Prove that if $\mathbf{f}_n : [a, b] \to \mathbb{R}^m$ are continuous functions converging uniformly to a function $\mathbf{f}$, then the sequence $\{\mathbf{f}_n\}$ is bounded in the sense that there is a constant $K \in \mathbb{R}$ such that $|\mathbf{f}_n(t)| \leq K$ for all $n \in \mathbb{N}$ and all $t \in [a, b]$ (this property is used in the proof of Lemma 2.6.1).

2. Go back to exercises 1 and 2 in Section 2.4. Show that the differential equations satisfy the conditions of Theorem 2.6.2. Comment.

3. It is occasionally useful to have a slightly more general version of Theorem 2.6.2 where the solution doesn't just start a given point, but passes through it:

   Teorem *Assume that* $\mathbf{f} : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ *is a continuous function. For any* $t_0 \in \mathbb{R}$ *and* $\mathbf{y}_0 \in \mathbb{R}^m$*, there exists a positive real number* a *and a function* $\mathbf{y} : [t_0 - a, t_0 + a] \to \mathbb{R}^m$ *such that* $\mathbf{y}(t_0) = \mathbf{y}_0$ *and*

   $$y'(t) = \mathbf{f}(t, \mathbf{y}(t)) \quad \text{for all } t \in [t_0 - a, t_0 + a]$$

   Prove this theorem by modifying the proof of Theorem 2.6.2 (run Euler's method "backwards" on the interval $[t_0 - a, t_0]$).

2.7 Polynomials are dense in $C([a,b],\mathbb{R})$

From calculus we know that many continuous functions can be approximated by their Taylor polynomials, but to have Taylor polynomials of all orders, a function f has to be infinitely differentiable, i.e. the higher order derivatives $f^{(k)}$ have to exist for all k. Most continuous functions are not differentiable at all, and the question is whether they still can be approximated by polynomials. In this section we shall prove:

Theorem 2.7.1 (Weierstrass' Theorem) *The polynomials are dense in* $C([a,b],\mathbb{R})$ *for all* $a, b \in \mathbb{R}$*,* $a < b$*. In other words, for each continuous function* $f : [a,b] \to \mathbb{R}$*, there is a sequence of polynomials* $\{p_n\}$ *converging uniformly to* f*.*

The proof I shall give (due to the Russian mathematician Sergei Bernstein (1880-1968)) is quite surprising; it uses probability theory to establish the result for the interval $[0,1]$, and then a straight forward scaling argument to extend it to all closed and bounded intervals.

The idea is simple: Assume that you are tossing a biased coin which has probability x of coming up "heads". If you toss it more and more times, you expect the proportion of times it comes up "heads" to stabilize around x. If somebody has promised you an award of $f(X)$ dollars, where X is the actually proportion of "heads" you have had during your (say) 1000 first tosses, you would expect your average award to be close to $f(x)$. If the number of tosses was increased to 10 000, you would feel even more certain.

Let us fomalize this: Let Y_i be the outcome of the i-th toss in the sense that Y_i has the value 0 if the coin comes up "tails" and 1 if it comes up "heads". The proportion of "heads" in the first N tosses is then given by

$$X_N = \frac{1}{N}(Y_1 + Y_2 + \cdots + Y_N)$$

Each Y_i is binomially distributed with mean $\mathrm{E}(Y_i) = x$ and variance $\mathrm{Var}(Y_i) = x(1-x)$, and since they are independent, we see that

$$\mathrm{E}(X_N) = \frac{1}{N}(\mathrm{E}(Y_1) + \mathrm{E}(Y_2) + \cdots \mathrm{E}(Y_N)) = x$$

and

$$\mathrm{Var}(X_N) = \frac{1}{N^2}(\mathrm{Var}(Y_1) + \mathrm{Var}(y_2) + \cdots + \mathrm{Var}(Y_N)) = \frac{1}{N}x(1-x)$$

(if you don't remember these formulas from probability theory, we shall derive them by analytic methods in the exercises). As N goes to infinity, we would expect X_N to converge to x with probability 1. If the "award function" f is continuous, we would also expect our average award $\mathrm{E}(f(X_N))$ to converge to $f(x)$.

To see what this has to do with polynomials, let us compute the average award $\mathrm{E}(f(X_N))$. Since the probability of getting exactly k heads in N tosses is $\binom{N}{k}x^k(1-x)^{n-k}$, we get

$$\mathrm{E}(f(X_N)) = \sum_{k=0}^{N} f(\frac{k}{N})\binom{N}{k}x^k(1-x)^{N-k}$$

Our expectation that $\mathrm{E}(f(X_N)) \to f(x)$ as $N \to \infty$, can therefore be rephrased as

$$\sum_{k=0}^{N} f(\frac{k}{N})\binom{N}{k}x^k(1-x)^{N-k} \to f(x) \quad N \to \infty$$

If we expand the parentheses $(1-x)^{N-k}$, we see that the expressions on the right hand side are just polynomials in x, and hence we have arrived at the hypothesis that the polynomials

$$p_N(x) = \sum_{k=0}^{N} f(\frac{k}{N})\binom{N}{k}x^k(1-x)^{N-k}$$

converge to $f(x)$. We shall prove that this is indeed the case, and that the convergence is uniform.

Proposition 2.7.2 *If $f : [0,1] \to \mathbb{R}$ is a continuous function, the* Bernstein polynomials

$$p_N(x) = \sum_{k=0}^{N} f(\frac{k}{N})\binom{N}{k}x^k(1-x)^{N-k}$$

converge uniformly to f on $[0,1]$.

Proof: Given $\epsilon > 0$, we must show how to find an N such that $|f(x) - p_n(x)| < \epsilon$ for all $n \geq N$ and all $x \in [0,1]$. Since f is continuous on the compact set $[0,1]$, it has to be uniformly continuous, and hence we can find a $\delta > 0$ such that $|f(u) - f(v)[< \frac{\epsilon}{2}$ whenever $|u - v| < \delta$. Since $p_n(x) = \mathrm{E}(f(X_n))$, we have

$$|f(x)-p_n(x)| = |f(x)-\mathrm{E}(f(X_n))| = |\mathrm{E}(f(x)-f(X_n))| \leq \mathrm{E}(|f(x)-f(X_n)|)$$

We split the last expectation into two parts: the cases where $|x - X_n| < \delta$ and the rest:

$$\mathrm{E}(|f(x)-f(X_n)|) = \mathrm{E}(\mathbf{1}_{\{|x-X_n|<\delta\}}|f(x)-f(X_n)|)+\mathrm{E}(\mathbf{1}_{\{|x-X_n|\geq\delta\}}|f(x)-f(X_n)|)$$

where the indicator function $\mathbf{1}_{\{|x-X_n|<\delta\}}$ is defined by

$$\mathbf{1}_{\{|x-X_n|<\delta\}} = \begin{cases} 1 & \text{if } |x - X_n| < \delta \\ 0 & \text{otherwise} \end{cases}$$

and oppositely for $\mathbf{1}_{\{|x-X_n|\geq\delta\}}$. By choice of δ, we have for the first term

$$\mathrm{E}(\mathbf{1}_{\{|x-X_n|<\delta\}}|f(x) - f(X_n)|) \leq \mathrm{E}\left(\mathbf{1}_{\{|x-X_n|<\delta\}}\frac{\epsilon}{2}\right) \leq \frac{\epsilon}{2}$$

For the second term, we first note that since f is a continuous function on a compact interval, it must be bounded by a constant M. Hence

$$\mathrm{E}(\mathbf{1}_{\{|x-X_n|\geq\delta\}}|f(x)-f(X_n)|) \leq 2M\mathrm{E}(\mathbf{1}_{\{|x-X_n|\geq\delta\}}) \leq 2M\mathrm{E}\left(\left(\frac{|x-X_n|}{\delta}\right)^2\right) =$$

$$= \frac{2M}{\delta^2}\mathrm{E}(|x - X_n|^2) = \frac{2M}{\delta^2}\mathrm{Var}(X_n) = \frac{2Mx(1-x)}{\delta^2 n} \leq \frac{M}{2\delta^2 n}$$

where we in the last step used that $\frac{1}{4}$ is the maximal value of $x(1-x)$ on $[0,1]$. If we now choose $N \geq \frac{M}{\delta^2\epsilon}$, we see that we get

$$\mathrm{E}(\mathbf{1}_{\{|x-X_n|\geq\delta\}}|f(x) - f(X_n)|) < \frac{\epsilon}{2}$$

for all $n \geq N$. Combining all the inequalities above, we see that if $n \geq N$, we have for all $x \in [0,1]$

$$|f(x) - p_n(x)| \leq \mathrm{E}(|f(x) - f(X_n)|) =$$

$$= \mathrm{E}(\mathbf{1}_{\{|x-X_n|<\delta\}}|f(x) - f(X_n)|) + \mathrm{E}(\mathbf{1}_{\{|x-X_n|\geq\delta\}}|f(x) - f(X_n)|) <$$

$$< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

and hence the Bernstein polynomials p_n converge uniformly to f. □

To get Weierstrass' result, we just have to move functions from arbitrary intervals $[a, b]$ to $[0, 1]$ and back. The function

$$T(x) = \frac{x-a}{b-a}$$

maps $[a, b]$ bijectively to $[0, 1]$, and the inverse function

$$T^{-1}(y) = a + (b-a)y$$

maps $[0, 1]$ back to $[a, b]$. If f is a continuous function on $[a, b]$, the function $\hat{f} = f \circ T^{-1}$ is a continuous function on $[0, 1]$ taking exactly the same values in the same order. If $\{q_n\}$ is a sequence of pynomials converging uniformly to $\hat{f}$ on $[0, 1]$, then the functions $p_n = q_n \circ T$ converge uniformly to f on $[a, b]$. Since

$$p_n(x) = q_n(\frac{x-a}{b-a})$$

the p_n's are polynomials, and hence Weierstrass' theorem is proved.

Remark Weierstrass' theorem is important because many mathematical arguments are easier to perform on polynomials than on continuous functions in general. If the property we study is preserved under uniform limits (i.e. if the if the limit function f of a uniformly convergent sequence of functions $\{f_n\}$ always inherits the property from the f_n's), we can use Weierstrass' Theorem to extend the argument from polynomials to all continuous functions. There is an extension of the result called the Stone-Weierstrass Theorem which generalizes the result to much more general settings.

Exercises for Section 2.7

1. Show that there is no sequence of polynomials that converges uniformly to the continuous function $f(x) = \frac{1}{x}$ on $(0, 1)$.
2. Show that there is no sequence of poynomials that converges uniformly to the function $f(x) = e^x$ on $\mathbb{R}$.
3. In this problem
$$f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$
   a) Show that if $x \neq 0$, then the n-th derivative has the form
   $$f^{(n)}(x) = e^{-1/x^2} \frac{P_n(x)}{x^{N_n}}$$
   where P_n is a polynomial and $N_n \in \mathbb{N}$.
   b) Show that $f^{(n)}(0) = 0$ for all n.

c) Show that the Taylor polynomials of f at 0 do not converge to f except in the point 0.

4. Assume that $f : [a, b] \to \mathbb{R}$ is a continuous function such that $\int_a^b f(x)x^n\, dx = 0$ for all $n = 0, 1, 2, 3, \ldots$.

   a) Show that $\int_a^b f(x)p(x)\, dx = 0$ for all polynomials p.

   b) Use Weierstrass' theorem to show that $\int_a^b f(x)^2\, dx = 0$. Conclude that $f(x) = 0$ for all $x \in [a, b]$.

5. In this exercise we shall show that $C([a, b], \mathbb{R})$ is a separable metric space.

   a) Assume that (X, d) is a metric space, and that $S \subset T$ are subsets of X. Show that if S is dense in (T, d_T) and T is dense in (X, d), then S is dense in (X, d).

   b) Show that for any polynomial p, there is a sequence $\{q_n\}$ of polynomials with rational coefficients that converges uniformly to p on $[a, b]$.

   c) Show that the polynomials with rational coefficients are dense in $C([a, b], \mathbb{R})$.

   d) Show that $C([a, b], \mathbb{R})$ is separable.

6. In this problem we shall reformulate Bernstein's proof in purely analytic terms, avoiding concepts and notation from probability theory. You should keep the Binomial Formula

$$(a + b)^N = \sum_{k=0}^{N} \binom{n}{k} a^k b^{N-k}$$

and the definition $\binom{N}{k} = \frac{N(N-1)(N-2)\cdot\ldots\cdot(N-k+1)}{1\cdot2\cdot3\cdot\ldots\cdot k}$ in mind.

   a) Show that $\sum_{k=0}^{N} \binom{N}{k} x^k(1-x)^{N-k} = 1$.

   b) Show that $\sum_{k=0}^{N} \frac{k}{N}\binom{N}{k} x^k(1-x)^{N-k} = x$ (this is the analytic version of $\mathrm{E}(X_N) = \frac{1}{N}(\mathrm{E}(Y_1) + \mathrm{E}(Y_2) + \cdots \mathrm{E}(Y_N)) = x$)

   c) Show that $\sum_{k=0}^{N} \left(\frac{k}{N} - x\right)^2 \binom{N}{k} x^k(1-x)^{N-k} = \frac{1}{N}x(1-x)$ (this is the analytic version of $\mathrm{Var}(X_N) = \frac{1}{N}x(1-x)$). *Hint:* Write $(\frac{k}{N} - x)^2 = \frac{1}{N^2}\left(k(k-1) + (1-2xN)k + N^2x^2\right)$ and use points b) and a) on the second and third term in the sum.

   d) Show that if p_n is the n-th Bernstein polynomial, then

$$|f(x) - p_n(x)| \leq \sum_{k=0}^{n} |f(x) - f(k/n)| \binom{n}{k} x^n(1-x)^{n-k}$$

   e) Given $\epsilon > 0$, explain why there is a $\delta > 0$ such that $|f(u) - f(v)| < \epsilon/2$ for all $u, v \in [0, 1]$ such that $|u - v| < \delta$. Explain why

$$|f(x) - p_n(x)| \leq \sum_{\{k : |\frac{k}{n} - x| < \delta\}} |f(x) - f(k/n)| \binom{n}{k} x^n(1-x)^{n-k} +$$

$$+ \sum_{\{k:|\frac{k}{n}-x|\geq\delta\}} |f(x) - f(k/n)| \binom{n}{k} x^n (1-x)^{n-k} \leq$$

$$< \frac{\epsilon}{2} + \sum_{\{k:|\frac{k}{n}-x|\geq\delta\}} |f(x) - f(k/n)| \binom{n}{k} x^n (1-x)^{n-k}$$

f) Show that there is a constant M such that $|f(x)| \leq M$ for all $x \in [0,1]$. Explain all the steps in the calculation:

$$\sum_{\{k:|\frac{k}{n}-x|\geq\delta\}} |f(x) - f(k/n)| \binom{n}{k} x^n (1-x)^{n-k} \leq$$

$$\leq 2M \sum_{\{k:|\frac{k}{n}-x|\geq\delta\}} \binom{n}{k} x^n (1-x)^{n-k} \leq$$

$$\leq 2M \sum_{k=0}^{n} \left(\frac{\frac{k}{n} - x}{\delta}\right)^2 \binom{n}{k} x^n (1-x)^{n-k} \leq \frac{2M}{n\delta^2} x(1-x) \leq \frac{M}{2n\delta^2}$$

g) Explain why we can get $|f(x) - p_n(x)| < \epsilon$ by chosing n large enough, and explain why this proves Proposition 2.7.2.

2.8 Baire's Category Theorem

Recall that a subset A of a metric space (X,d) is *dense* if for all $x \in X$ there is a sequence from A converging to x. An equivalent definition is that all balls in X contain elements from A. To show that a set S is *not* dense, we thus have to find an open ball that does not intersect S. Obviously, a set can fail to be dense in parts of X, and still be dense in other parts. The following definition catches our intiution of a set set that is not dense anywhere.

Definition 2.8.1 *A subset S of a metric space (X,d) is said to be* nowhere dense *if for all nonempty, open sets $G \subset X$, there is a ball* $\mathrm{B}(x;r) \subset G$ *that does not intersect S.*

This definition simply says that no matter how much we restrict our attention, we shall never find an area in X where S is dense.

Example 1. $\mathbb{N}$ is nowhere dense in $\mathbb{R}$.

Nowhere dense sets are sparse in an obvious way. The following definition indicates that even countable unions of nowhere dense sets are unlikely to be very large.

Definition 2.8.2 *A set is called* meager *if it is a countable union of nowhere dense sets. The complement of a meager set is called* comeager.[1]

Example 2. $\mathbb{Q}$ is a meager set in $\mathbb{R}$ as it can be written as a countable union $\mathbb{Q} = \bigcup_{a\in\mathbb{Q}}\{a\}$ of the nowhere dense singletons $\{a\}$. By the same argument, $\mathbb{Q}$ is also meager in $\mathbb{Q}$.

The last part of the example shows that a meager set can fill up a metric space. However, in *complete* spaces the meager sets are always "meager" in the following sense:

Theorem 2.8.3 (Baire's Category Theorem) *Assume that M is a meager subset of a complete metric space (X, d). Then M does not contain any open balls, i.e. M^c is dense in X.*

Proof: Since M is meager, it can be written as a union $M = \bigcup_{k\in\mathbb{N}} N_k$ of nowhere dense sets N_k. Given a ball $\mathrm{B}(a;r)$, our task is to find an element $x \in \mathrm{B}(a;r)$ which does not belong to M.

We first observe that since N_1 is nowhere dense, there is a ball $\mathrm{B}(a_1;r_1)$ inside $\mathrm{B}(a;r)$ which does not intersect N_1. By shrinking the radius r_1 slightly if necessary, we may assume that the *closed* ball $\overline{\mathrm{B}}(a_1;r_1)$ is contained in $\mathrm{B}(a;r)$, does not intersect N_1, and has radius less than 1. Since N_2 is nowhere dense, there is a ball $\mathrm{B}(a_2;r_2)$ inside $\mathrm{B}(a_1;r_1)$ which does not intersect N_2. By shrinking the radius r_2 if necessary, we may assume that the closed ball $\overline{\mathrm{B}}(a_2;r_2)$ does not intersect N_2 and has radius less than $\frac{1}{2}$. Continuing in this way, we get a sequence $\{\overline{\mathrm{B}}(a_k;r_k)\}$ of closed balls, each contained in the previous, such that $\overline{\mathrm{B}}(a_k;r_k)$ has radius less than $\frac{1}{k}$ and does not intersect N_k.

Since the balls are nested and the radii shrink to zero, the centers a_k form a Cauchy sequence. Since X is complete, the sequence converges to a point x. Since each ball $\overline{\mathrm{B}}(a_k;r_k)$ is closed, and the "tail" $\{a_n\}_{n=k}^{\infty}$ of the sequence belongs to $\overline{\mathrm{B}}(a_k;r_k)$, the limit x also belongs to $\overline{\mathrm{B}}(a_k;r_k)$. This means that $x \notin N_k$ for all k, and hence $x \notin M$. Since $\overline{\mathrm{B}}(a_1;r_1) \subset \mathrm{B}(a;r)$, we see that $x \in \mathrm{B}(a;r)$, and the theorem is proved. □

As an immediate consequence we have:

Corollary 2.8.4 *A complete metric space is not a countable union of nowhere dense sets.*

---

[1]Most books refer to meager sets as "sets of first category" while comeager sets are called "residual sets". Sets that are not of first category, are said to be of "second category". Although this is the original terminology of René-Louis Baire (1874-1932) who introduced the concepts, it is in my opinion so nondescriptive that it should be abandoned in favor of more evocative terms.

Baire's Category Theorem is a surprisingly strong tool for proving theorems about sets and families of functions. We shall take a look at some examples.

Definition 2.8.5 *Let (X, d) be a metric space. A family $\mathcal{F}$ of functions $f : X \to \mathbb{R}$ is called* pointwise bounded *if for each $x \in X$, there is a constant $M_x \in \mathbb{R}$ such that $|f(x)| \leq M_x$ for all $f \in \mathcal{F}$.*

Note that the constant M_x may vary from point to point, and that there need not be a constant M such that $|f(x)| \leq M$ for all f and all x (a simple example is $\mathcal{F} = \{f : \mathbb{R} \to \mathbb{R} \,|\, f(x) = kx$ for $k \in [-1, 1\}$, where $M_x = |x|$). The next result shows that although we cannot guarantee boundedness on all X, we can under reasonable assumptions guarantee boundedness on a part of X.

Proposition 2.8.6 *Let (X, d) be a complete metric space, and assume that $\mathcal{F}$ is a pointwise bounded family of continuous functions $f : X \to \mathbb{R}$. Then there exists an open, nonempty set G and a constant $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for all $f \in \mathcal{F}$ and all $x \in G$.*

Proof: For each $n \in \mathbb{N}$ and $f \in \mathcal{F}$, the set $f^{-1}([-n, n])$ is closed as it is the inverse image of a closed set under a continuous function (recall Proposition 1.3.10). As intersections of closed sets are closed (see Exercise 1.3.12)

$$A_n = \bigcap_{f \in \mathcal{F}} f^{-1}([-n, n])$$

is also closed. Since $\mathcal{F}$ is pointwise bounded, $\mathbb{R} = \bigcup_{n \in \mathbb{N}} A_n$, and the corollary above tells us that not all A_n can be nowhere dense. If A_{n_0} is not nowhere dense, there must be an open set G such that all balls inside G contains elements from A_{n_0}. Since A_{n_0} is closed, this means that $G \subset A_{n_0}$ (check!). By definition of A_{n_0}, we see that $|f(x)| \leq n_0$ for all $f \in \mathcal{F}$ and all $x \in G$. □

You may doubt the usefulness of this theorem as we only know that the result holds for *some* open set G, but the point is that if we have extra information on the the family $\mathcal{F}$, the sole existence of such a set may be exactly what we need to pull through a more complex argument. In functional analysis, there is a famous (and most useful) example of this called the *Banach-Steinhaus Theorem.*

For our next application, we first observe that although $\mathbb{R}^n$ is not compact, it can be written as a countable union of compact sets:

$$\mathbb{R}^n = \bigcup_{k \in \mathbb{N}} [-k, k]^n$$

We shall show that this is *not* the case for $C([0,1],\mathbb{R})$ — this space can not be written as a countable union of compact sets. We need a couple of lemmas.

Lemma 2.8.7 *A closed set F is nowhere dense if and only if it does not contain any open balls.*

Proof: If F contains an open ball, it obviously isn't nowhere dense. We therefore assume that F does *not* contain an open ball, and prove that it is nowhere dense. Given a nonempty, open set G, we know that F cannot be contained in G as G contains open balls and F does not. Pick an element x in G that is not in F. Since F is closed, there is a ball $\mathrm{B}(x;r_1)$ around x that does not intersect F. Since G is open, there is a ball $\mathrm{B}(x;r_2)$ around x that is contained in G. If we choose $r=\min\{r_1,r_2\}$, the ball $\mathrm{B}(x;r)$ is contained in G and does not intersect F, and hence F is nowhere dense. □

The next lemma contains the key to the argument.

Lemma 2.8.8 *A compact subset K of $C([0,1],\mathbb{R})$ is nowhere dense.*

Proof: Since compact sets are closed, it suffices (by the previous lemma) to show that each ball $\mathrm{B}(f;\epsilon)$ contains elements that are not in K. By Arzelà-Ascoli's Theorem, we know that compact sets are equicontinuous, and hence we need only prove that $\mathrm{B}(f;\epsilon)$ contains a family of functions that is not equicontinuous. We shall produce such a family by perturbing f by functions that are very steep on small intervals.

For each $n\in\mathbb{N}$, let g_n be the function

$$g_n(x)=\begin{cases} nx & \text{for } x\le\frac{\epsilon}{2n} \\ \\ \frac{\epsilon}{2} & \text{for } x\ge\frac{\epsilon}{2n}\end{cases}$$

Then $f+g_n$ is in $B(f,\epsilon)$, but since $\{f+g_n\}$ is not equicontinuous (see Exercise 9 for help to prove this), all these functions can not be in K, and hence $\mathrm{B}(f;\epsilon)$ contains elements that are not in K. □

Proposition 2.8.9 $C([0,1],\mathbb{R})$ *is not a countable union of compact sets.*

Proof: Since $C([0,1],\mathbb{R})$ is complete, it is not the countable union of nowhere dense sets by Corollary 2.8.4. Since the lemma tells us that all compact sets are nowhere dense, the theorem follows. □

Remark: The basic idea in the proof above is that the compact sets are nowhere dense since we can obtain arbitrarily steep functions by perturbing a given function just a little. The same basic idea can be used to prove more sophisticated results, e.g. that the set of nowhere differentiable functions is comeager in $C([0,1],\mathbb{R})$. The key idea is that starting with any continuous function, we can perturb it into functions with arbitrarily large derivatives by using small, but rapidly oscillating functions. With a little bit of technical work, this implies that the set of functions that are differentiable at at least one point, is meager.

Exercises for Section 2.8

1. Show that $\mathbb{N}$ is a nowhere dense subset of $\mathbb{R}$.
2. Show that the set $A = \{g \in C([0,1],\mathbb{R}) \mid g(0) = 0\}$ is nowhere dense in $C([0,1],\mathbb{R})$.
3. Show that a subset of a nowhere dense set is nowhere dense and that a subset of a meager set is meager.
4. Show that a subset S of a metric space X is nowhere dense if and only if for each open ball $\mathrm{B}(a_0;r_0) \subset X$, there is a ball $\mathrm{B}(x;r) \subset \mathrm{B}(a_0;r_0)$ that does not intersect S.
5. Recall that the closure $\overline{N}$ of a set N consist of N plus all its boundary points.
   a) Show that if N is nowhere dense, so is $\overline{N}$.
   b) Find an example of a meager set M such that $\overline{M}$ is not meager.
   c) Show that a set is nowhere dense if and only if $\overline{N}$ does not contain any open balls.
6. Show that a countable union of meager sets is meager.
7. Show that if $N_1, N_2, \ldots, N_k$ are nowhere dense, so is $N_1 \cup N_2 \cup \ldots N_k$.
8. Prove that $G \subset A_{n_0}$ in the proof of Proposition 2.8.6.
9. In this problem we shall prove that the set $\{f + g_n\}$ in the proof of Lemma 2.8.8 is not equicontinuous.
   a) Show that the set $\{g_n\}$ is not equicontinuous.
   b) Show that if $\{h_n\}$ is an equicontinous family of functions $h_n : [0,1] \to \mathbb{R}$ and $k : [0,1] \to \mathbb{R}$ is continuous, then $\{h_n + k\}$ is equicontinuous.
   c) Prove that the set $\{f + g_n\}$ in the lemma is not equicontinuous. (*Hint:* Assume that the sequence is equicontinuous, and use part b) with $h_n = f + g_n$ and $k = -f$ to get a contradiction with a)).
10. Let $\mathbb{N}$ have the discrete metric. Show that $\mathbb{N}$ is complete and that $\mathbb{N} = \bigcup_{n\in\mathbb{N}}\{n\}$. Why doesn't this contradict Baire's Category Theorem?
11. Let (X,d) be a metric space.
    a) Show that if $G \subset X$ is open and dense, then G^c is nowhere dense.
    b) Assume that (X,d) is complete. Show that if $\{G_n\}$ is a countable collection of open, dense subsets of X, then $\bigcap_{n\in\mathbb{N}} G_n$ is dense in X

12. Assume that a sequence $\{f_n\}$ of continuous functions $f_n : [0,1] \to \mathbb{R}$ converges pointwise to f. Show that f must be bounded on a subinterval of $[0,1]$. Find an example which shows that f need not be bounded on all of $[0,1]$.

13. In this problem we shall study sequences $\{f_n\}$ of functions converging pointwise to 0.

    a) Show that if the functions f_n are continuous, then there exists a nonempty subinterval (a,b) of $[0,1]$ and an $N \in \mathbb{N}$ such that for $n \geq N$, $|f_n(x)| \leq 1$ for all $x \in (a,b)$.

    b) Find a sequence of functions $\{f_n\}$ converging to 0 on $[0,1]$ such that for each nonempty subinterval (a,b) there is for each $N \in \mathbb{N}$ an $x \in (a,b)$ such that $f_N(x) > 1$.

14. Let (X,d) be a metric space. A point $x \in X$ is called *isolated* if there is an $\epsilon > 0$ such that $\mathrm{B}(x;\epsilon) = \{x\}$.

    a) Show that if $x \in X$, the singleton $\{x\}$ is nowhere dense if and only if x is not an isolated point.

    b) Show that if X is a complete metric space without isolated points, then X is uncountable.

    We shall now prove:

    Theorem: The unit interval $[0,1]$ can *not* be written as a countable, disjoint union of closed, proper subintervals $I_n = [a_n, b_n]$.

    c) Assume for contradictions that $[0,1]$ can be written as such a union. Show that the set of all endpoints, $F = \{a_n, b_n \mid n \in \mathbb{N}\}$ is a closed subset of $[0,1]$, and that so is $F_0 = F \setminus \{0,1\}$. Explain that since F_0 is countable and complete in the subspace metric, F_0 must have an isolated point, and use this to force a contradiction.