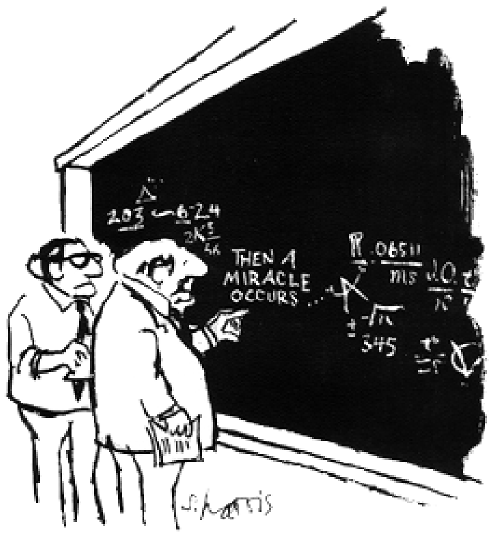
14. Sequences and series of functions.



"I think you should be more explicit here in step two."

The reference for this Chapter is Reed Chapter 5, §1–3, and Example 2 page 197; and Chapter 6, §6.3

14.1. Pointwise and uniform convergence. Study Reed §5.1 carefully; the definitions and all the examples.

14.2. Properties of uniform convergence. Study Reed §5.2 carefully.

The point is that continuity and integration behave well under uniform convergence, but not under pointwise convergence.

Differentiation does not behave well, unless we also require the derivatives to converge uniformly.

Theorem 5.2.4, concerning differentiation under the integral sign, is also important.

14.3. The supremum norm and supremum metric. Study Reed §5.3 and Example 2 page 197 carefully.

One way of measuring the "size" of a bounded function f is by means of the *supremum norm* $\|f\|_\infty$ (Reed page 175). We can then measure the *distance* between two functions f and g (with the same domain E) by using the *sup distance* (or *sup metric*)

$$\rho_\infty(f,g) = \|f-g\|_\infty = \sup_{x\in E} |f(x)-g(x)|.$$

In Reed, $E \subseteq \mathbb{R}$, but exactly the same definition applies for $E \subseteq \mathbb{R}^n$.

In Proposition 5.3.1 it is shown that the sup norm satisfies positivity, the triangle inequality, and is well behaved under scalar multiples.

In Example 2 on page 197 it is shown that as a consequence the sup metric is indeed a metric (i.e. satisfies positivity, symmetry and transitivity — see the first paragraph in Section 10.2).

(See Reed page 177) A sequence of functions (f_n) defined on the same domain E *converges to f in the sup metric*[46] if $\rho_\infty(f_n, f) \to 0$ as $n \to \infty$. (Note that this is completely analogous to Definition 11.1.1 for convergence of a sequence of points in $\mathbb{R}^n$.)

One similarly defines the notion of *Cauchy in the sup metric* (or equivalently *Cauchy in the sup norm*).

One proves (page 178) that if a sequence (f_n) of continuous functions defined on a closed bounded interval $[a, b]$ is Cauchy in the sup metric, then it converges in the sup metric and the limit is also continuous on $[a, b]$. (This is analogous to the fact that if a sequence from $\mathbb{R}^n$ is Cauchy then it converges to a point in $\mathbb{R}^n$.)

The same theorem and proof works if $[a, b]$ is replaced by any set $E \subseteq \mathbb{R}^n$, but then we need to restrict to functions that are continuous and *bounded* (otherwise the definition of $\rho_\infty(f, g)$ may give ∞). (If E is compact we need not assume boundedness, as this follows from continuity by Theorem 13.2.1.)

We say that the set of bounded and continuous functions defined on E is *complete in the sup metric (or norm).*

14.4. Integral norms and metrics. Study Reed pp 179–181, particularly Example 1. We cannot do much on this topic, other than introduce it.

The important "integral" norms for functions f with domain E are defined by

$$\|f\|_p = \left(\int_E |f|^p \right)^{1/p}.$$

for any $1 \le p < \infty$.

The most important are $p = 2$ and $p = 1$ (in that order). If E is not an interval in $\mathbb{R}$ then we need a more general notion of integration; to do it properly requires the Lebesgue integral. But $\int_E g$ is still interpreted as the area between the graph of f and the set $E \subseteq \mathbb{R}$ — being negative where it lies below the axis. In higher dimensional cases we replace "area" by "volume" of the appropriate dimension.)

It is possible to prove for $\|f\|_p$ the three properties of a "norm" in Proposition 5.3.1 of Reed.

Actually, $\|f\|_p$ does not quite give a norm. The problem is that if f is not continuous then $\|f\|_p$ may equal 0 even though f is not the zero function, i.e. not everywhere zero. However f must then be zero "almost everywhere" — in a sense that can be made precise. This problem does not arise if we restrict to continuous functions, since if f is continuous and $\int_a^b |f| = 0$ then $f = 0$ everywhere. But it is usually not desirable to restrict to continuous functions for other reasons.

The metrics corresponding to these norms are

$$\rho_p(f, g) = \|f - p\|_p,$$

again for $1 \le p < \infty$.

Remark 14.4.1. * The set S of continuous functions defined on a closed bounded set is a metric space in any of the metrics ρ_p, but it is not a *complete* metric space (unless $p = \infty$). If we take the "completion" of the set S we are lead to the so-called Lebesgue measurable functions and the theory of Lebesgue integration.

[46] Reed says "converges in the sup norm".

14.5. Series of functions. Study Reed §4.3 carefully.

Suppose the functions f_j all have a common domain (typically $[a, b]$).

The infinite *series* of functions $\sum_{j=1}^{\infty} f_j$ is said to *converge pointwise* to the function f iff the corresponding series of real numbers $\sum_{j=1}^{\infty} f_j(x)$ converges to $f(x)$ for each x in the (common) domain of the f_j.

From the definition of convergence of a series of real numbers, this means that the infinite *sequence* (S_n) of partial sums

$$S_n = f_1 + \cdots + f_n$$

converges in the pointwise sense to f.

We say that the series $\sum_j f$ *converges uniformly* if the sequence of partial sums converges uniformly.

Note the important *Weierstrass M-test*, Theorem 6.3.1 of Reed. This gives a very useful criterion for uniform convergence of a series of functions.

The results about uniform convergence of a sequence of functions in Section14.2 (Reed §5.2) immediately give corresponding results for series (Reed Theorem 6.3.1 — last sentence; Theorem 6.3.2, Theorem 6.3.3). In particular:

- A uniformly convergent series of continuous functions has a continuous limit.
- Integration behaves well in the sense that for a uniformly convergent series of continuous functions on a closed bounded interval, the integral of the sum is the sum of the integrals (i.e. summation and integration can be interchanged).
- If a series of C^1 functions converges uniformly on an interval to a function f, ***and*** if the derivatives also converge uniformly to g say, then f is C^1 and $f' = g$ (i.e. summation and differentiation can be interchanged).

Study example 1 on page 240 of Reed. Example 2 is ***** material (it gives an example of a continuous and nowhere differentiable function).

15. Metric spaces

Study Reed §5.6.

The best way to study convergence and continuity is to do it abstractly by means of metric spaces. This has the advantage of simplicity and generality.

15.1. Definition and examples.

The reference is Reed §5.6 up to the end of the first paragraph on page 201.

The basic idea we need is the notion of a "distance function" or a "metric".

DEFINITION 15.1.1. A metric space is a set $\mathcal{M}$ together with a function $\rho : \mathcal{M} \times \mathcal{M} \to [0, \infty)$ (called a *metric*) which satisfies for all $x, y, z \in \mathcal{M}$

1. $\rho(x, y) \geq 0$ and $\rho(x, y) = 0$ iff $x = y$, (positivity)
2. $\rho(x, y) = \rho(y, x)$, (symmetry)
3. $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$, (triangle inequality)

The idea is that $\rho(x, y)$ measures the "distance (or length of a shortest route) from x to y". The three requirements are natural. In particular, the third might be thought of as saying that one route from x to y is to take a shortest route from x to z and then a shortest route from z to y — but there may be an even shorter route from x to y which does not go via z.

EXAMPLE 15.1.2 (Metrics on $\mathbb{R}^n$). We discussed the Euclidean metric on $\mathbb{R}^n$ in Sections 10.1 and 10.2

Example 3 on page 197 of Reed is important. It can be generalised to give the following metrics on $\mathbb{R}^n$:

$$\rho_p(\mathbf{x}, \mathbf{y}) = ((x_1 - y_1)^p + \dots (x_n - y_n)^p)^{1/p} = \left(\sum_{i=1}^{n} (x_i - y_i)^p \right)^{1/p},$$

$$\rho_{\max}(\mathbf{x}, \mathbf{y}) = \max\{\, |x_i - y_i| : i - 1, \dots, n \,\},$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $1 \leq p < \infty$.

We sometimes write ρ_∞ for $\rho_{\max}$.

Note the diagrams in Reed showing $\{\, \mathbf{x} \in \mathbb{R}^2 | \rho_p(\mathbf{x}, \mathbf{0}) \leq 1 \,\}$ for $p = 1, 2, \infty$.

The proof that ρ_p is a metric has already been given for $p = 2$ (the Euclidean metric). The cases $p = 1, \infty$ are Problem 1 page 202 of Reed in case $n = 2$, but the proofs trivially generalise to $n > 2$. The case of arbitrary p is trickier, and I will set it later as an assignment problem (with hints!).

EXAMPLE 15.1.3 (Metrics on function spaces). These are extremely important. See also Section 14.3.

The basic example is the sup metric ρ_∞ on $\mathcal{C}[a, b]$, the set of continuous functions $f : [a, b] \to \mathbb{R}$. See Reed Example 2 page 197.

More generally, one has the sup metric on

1. $\mathcal{B}(E)$, the set of bounded functions $f : E \ (\subseteq \mathbb{R}^n) \to \mathbb{R}$,
2. $\mathcal{BC}(E)$, the set of bounded continuous functions $f : E \ (\subseteq \mathbb{R}^n) \to \mathbb{R}$.

In the second case, if E is compact, we need only require continuity as this already implies boundedness.

The metrics ρ_p ($1 \leq p < \infty$) on $\mathcal{C}[a, b]$ (and generalisations to other sets than $[a, b]$) have also been discussed, see Section 14.4.

***** If we regard $(a_1, \dots, a_n)$ as a function

$$a : \{1, \dots, n\} \to \mathbb{R},$$

and interpret

$$\int |a|^p \quad \text{as} \quad \sum_{i=1}^{n} |a(i)|^p,$$

(which is indeed the case for an appropriate "measure"), then the metrics ρ_p ($1 \le p < \infty$) on function spaces give the metrics ρ_p on $\mathbb{R}^n$ as a special case.

Example 15.1.4 (Subsets of a metric space). Any subset of a metric space is itself a metric space with the same metric, see the first paragraph of Reed page 199.

See also the first example in Example 4 of Reed, page 199.

Example 15.1.5 (The discrete metric). A simple metric on any set $\mathcal{M}$ is given by

$$d(x,y) = \begin{cases} 1 & x \neq y, \\ 0 & x = y. \end{cases}$$

Check that this is a metric. It is usually called the *discrete metric*. (It is a particular case of what Reed calls a "discrete metric" in Problem 10 page 202.) It is not very useful, other than as a source of counterexamples to possible conjectures.

Example 15.1.6 (Metrics on strings of symbols and DNA sequences). Example 5 in Reed is a discussion about metrics as used to estimate the difference between two DNA molecules. This is important in studying evolutionary trees.

Example 15.1.7 (The natural metric on a normed vector space). Any normed vector space gives rise to a metric space defined by

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

The three properties of a metric follow from the properties of a norm. Examples are the metrics ρ_p for $1 \le p \le \infty$ on both $\mathbb{R}^n$ and the metrics ρ_p for $1 \le p \le \infty$ on spaces of functions.

But metric spaces are much more general. In particular, the discrete metric, and Examples 4 and 5 on pages 199, 200 of Reed are not metrics which arise from norms.

15.2. Convergence in a metric space. Reed page 201 paragraphs 2–4.

Once one has a metric, the following is the natural notion of convergence.

Definition 15.2.1. A sequence (x_n) in a metric space $(\mathcal{M}, \rho)$ converges to $x \in \mathcal{M}$ if $\rho(x_n, x) \to 0$ as $n \to \infty$.

Notice that convergence in a metric space is reduced to the notion of convergence of a sequence of real numbers.

In the same way as for sequences in $\mathbb{R}$, there can be at most one limit.

We discussed convergence in $\mathbb{R}^n$ in Section 11.1. This corresponds to convergence with respect to the Euclidean metric. We will see in the next section that convergence with respect to any of the metrics ρ_p on $\mathbb{R}^n$ is the same.

Convergence of a sequence of functions in the sup norm is the same as convergence in the sup metric. But convergence in the other metrics ρ_p is *not* the same.

For example, let

$$f_n(x) = \begin{cases} 0 & -1 \le x \le 0 \\ nx & 0 \le x \le \frac{1}{n} \\ 1 & \frac{1}{n} \le x \le 1 \end{cases}, \qquad f(x) = \begin{cases} 0 & -1 \le x \le 0 \\ 1 & 0 < x \le 1 \end{cases}.$$

Then $\rho_\infty(f_n, f) = 1$ and $\rho_1(f_n, f) = \frac{1}{2n}$ (*why?*). Hence $f_n \to f$ in $(\mathcal{R}[-1,1], \rho_1)$ (the space of Riemann integrable functions on $[-1,1]$ with the ρ_1 metric). But $f_n \not\to f$ in $(\mathcal{R}[-1,1], \rho_\infty)$[47].

15.3. Uniformly equivalent metrics. The reference here is the Definition and Theorem on Reed page 201, *and* Problem 11 page 203. However, we will only consider *uniformly* equivalent metrics, rather than equivalent metrics.

DEFINITION 15.3.1. Two metrics ρ and σ on the same set $\mathcal{M}$ are *uniformly equivalent*[48] if there are positive numbers c_1 and c_2 such that

$$c_1\rho(x,y) \le \sigma(x,y) \le c_2\rho(x,y)$$

for all $x, y \in \mathcal{M}$.

It follows that

$$c_2^{-1}\sigma(x,y) \le \rho(x,y) \le c_1^{-1}\sigma(x,y).$$

Reed also defines a weaker notion of "equivalent" ("uniformly equivalent" implies "equivalent" but not conversely). We will just need the notion of equivalent. The following theorem has a slightly simpler proof than the analogous one in the text for equivalent metrics.

THEOREM 15.3.2. *Suppose ρ and σ are uniformly equivalent metrics on $\mathcal{M}$. Then $x_n \to x$ with respect to ρ iff $x_n \to x$ with respect to σ.*

PROOF. Since $\sigma(x_n, x) \le c_2\rho(x_n, x)$, it follows that $\rho(x_n, x) \to 0$ implies $\sigma(x_n, x) \to 0$. Similarly for the other implication. □

The following theorem is a generalisation of Problem 12 page 203.

EXAMPLE 15.3.3. *The metrics ρ_p on $\mathbb{R}^n$ are uniformly equivalent to one another.*

PROOF. It is sufficient to show that ρ_p for $1 \le p < \infty$ is equivalent to ρ_∞ (since if two metrics are each uniformly equivalent to a third metric, then from Definition15.3.1 it follows fairly easily that they are uniformly equivalent to one another — *Exercise*).

But

$$\begin{aligned}\rho_p(\mathbf{x},\mathbf{y}) &= \left(|x_1 - y_1|^p + \cdots + |x_n - y_n|^p\right)^{1/p} \\ &\le \left(n \max_{1\le i\le n} |x_i - y_i|^p\right)^{1/p} = n^{1/p}\rho_\infty(\mathbf{x},\mathbf{y}),\end{aligned}$$

and

$$\rho_\infty(\mathbf{x},\mathbf{y}) = \max_{1\le i\le n} |x_i - y_i| \le \rho_p(\mathbf{x},\mathbf{y}).$$

This completes the proof. □

REMARK 15.3.1. *The metrics ρ_p on function spaces are <u>not</u> uniformly equivalent (or even equivalent).*

It is easiest to see this for ρ_1 and ρ_∞ on $\mathcal{R}[a,b]$ (the space of Riemann integrablefunctions on $[a,b]$). It follows from the example in the prevous section, since $f_n \to f$ with respect to ρ_1 but not with respect to ρ_∞.

[47] Remember that "Riemann integrable" implies "bounded" according to our definitions.

[48] What is here called "uniformly equivalent" is usually called "equivalent". But for consistency, I will keep to the convention in Reed.

15.4. Cauchy sequences and completeness. The reference here is Reed §5.7 to the end of page 204.

The definition of a Cauchy sequence is exactly as for Definition 11.5.1 in $\mathbb{R}^p$.

DEFINITION 15.4.1. A sequence $(\mathbf{x}_n)$ is a *Cauchy sequence* if for any $\epsilon > 0$ there is a corresponding N such that

$$m, n \geq N \quad \Rightarrow \quad \rho(\mathbf{x}_m, \mathbf{x}_n) \leq \epsilon.$$

It follows from the Definition in Reed Page 177 that for a sequence of functions "Cauchy in the sup norm" is the same as "Cauchy with respect to the sup metric".

It follows as for convergence, that a sequence is Cauchy with respect to one metric then it is Cauchy with respect to any uniformly equivalent metric.

The definition of a complete metric space is *extremely important* (Reed page 204).

DEFINITION 15.4.2. A metric space $(\mathcal{M}, \rho)$ is complete if every Cauchy sequence from $\mathcal{M}$ converges to an element in $\mathcal{M}$.

Moreover (Problem 10(a) page 209):

THEOREM 15.4.3. *Suppose ρ and σ are uniformly equivalent metrics on $\mathcal{M}$. Then $(\mathcal{M}, \rho)$ is complete iff $(\mathcal{M}, \sigma)$ is complete.*

PROOF. Suppose $(\mathcal{M}, \rho)$ is complete. Let (x_n) be Cauchy with respect to σ. Then it is Cauchy with respect to ρ and hence converges to x (say) with respect to ρ. By Theorem 15.3.2, $x_n \to x$ with respect to σ. It follows that $(\mathcal{M}, \sigma)$ is complete. □

EXAMPLE 15.4.4. The Cauchy Completeness Axiom says that $(\mathbb{R}, \rho_2)$ is complete, and it follows from Proposition 11.5.3 that $(\mathbb{R}^n, \rho_2)$ is also complete.

It then follows from the previous theorem that $(\mathbb{R}^n, \rho_p)$ *is complete for any* $1 \leq p \leq \infty$.

More generally, *if $M \subseteq \mathbb{R}^n$ is closed, then (M, ρ_p) is complete for any $1 \leq p \leq \infty$.* To see this, consider any Cauchy sequence $(\mathbf{x}_i)$ in (M, ρ_2). Since $(\mathbf{x}_i)$ is also a Cauchy sequence in $\mathbb{R}^n$ it converges to $\mathbf{x}$ (say) in $\mathbb{R}^n$. Since M is closed, $\mathbf{x} \in M$. Hence (M, ρ_2) is complete. It follows from the previous theorem that (M, ρ_p) is complete for any $1 \leq p \leq \infty$.

If $M \subseteq \mathbb{R}^n$ is not closed, then (M, ρ_p) is not complete. To see this, choose a sequence $(\mathbf{x}_i)$ in M such that $\mathbf{x}_i \to \mathbf{x} \notin M$. Since $(\mathbf{x}_i)$ converges in $\mathbb{R}^n$ it is Cauchy in $\mathbb{R}^n$ and hence in M, but it does not converge to a member of M. Hence (M, ρ_p) is not complete.

EXAMPLE 15.4.5. $(\mathcal{C}[a, b], \rho_\infty)$ *is complete* from Reed Theorem 5.3.3 page 178. However, $(\mathcal{C}[a, b], \rho_1)$ *is not complete.*

To see this, consider the example in Section 15.2. It is fairly easy to see that the sequence (f_n) is Cauchy in the ρ_1 metric. In fact if $m > n$ then

$$\rho_1(f_n, f_m) = \int_0^{1/n} |f_n - f_m| \leq \int_0^{1/n} |f_n| + |f_m| \leq 2/n.$$

Since $f_n \to f$ in the larger space $(\mathcal{R}[-1, 1], \rho_1)$ of Riemann integrable functions on $[-1, 1]$ with the ρ_1 metric, it follows that (f_n) is Cauchy in this space, and hence is also Cauchy in $(\mathcal{C}[a, b], \rho_1)$. But f_n does not converge to a member of this space[49].

[49] If it did, it would converge to the same function in the larger space, contradicting uniqueness of limits in the larger space

15.5. Contraction Mapping Principle. The reference here is Reed §5.7.

The definition of a contraction in Section 11.6 for a function $F : S \to S$ where $S \subseteq \mathbb{R}^p$ applies with S replaced by $\mathcal{M}$ and d replaced by ρ for any metric space $(\mathcal{M}, \rho)$.

The Contraction Mapping Theorem, Theorem 11.6.1 applies with exactly the same proof, for any *complete* metric space $(\mathcal{M}, \rho)$ instead of the closed set S and the Euclidean metric d. Or see Reed Theorem 5.7.1 page 205.

Completeness is needed in the proof in order to show that the Cauchy sequence of iterates actually converges to a member of $\mathcal{M}$.

In Section 16 we will give some important applications of the Contraction Mapping Principle. In particular, we will apply it to the complete metric space $(\mathcal{C}, [a, b], \rho_\infty)$.

15.6. Topological notions in metric spaces*. In this section I will point out that much of what we proved for $(\mathbb{R}^n, d)$ applies to an arbitrary metric space $(\mathcal{M}, \rho)$ with the same proofs. The extensions are straightforward and I include them for completeness and future reference in the Analysis II course. But we will not actually need the material in this course.

References at about the right level are "Primer of Modern Analysis" by Kennan T. Smith, Chapter 7, and "An Introduction to Analysis" (second edition) by William Wade, Chapter 10.

The definitions of *neighbourhood*, *open ball* $B_r(x)$, and *open set* are analogous to those in Section 10.5. Theorem 10.5.2 remains true with $\mathbb{R}^n$ replaced by $\mathcal{M}$.

The definition of *closed set* is analogous to that in Section 10.6 and Theorem 10.6.2 is true with $\mathbb{R}^n$ replaced by $\mathcal{M}$.

The definitions of *boundary point*, *boundary*, *interior point*, *interior*, *exterior point* and *exterior* are analogous to those in Section 10.8. The three dot points there are true for $S \subseteq \mathcal{M}$.

The *product* of two metric spaces $(\mathcal{M}_1, \rho_1)$ and $(\mathcal{M}_2, \rho_2)$ is the set $\mathcal{M}_1 \times \mathcal{M}_2$ with the metric $\rho_1 \times \rho_2$ defined by

$$(\rho_1 \times \rho_2)\big((x_1, y_1), (x_2, y_2)\big) = \sqrt{\rho_1(x_1, y_1)^2 + \rho_2(x_2, y_2)^2}.$$

If A_1 is open in $(\mathcal{M}_1, \rho_1)$ and A_2 is open in $(\mathcal{M}_2, \rho_2)$ the $A_1 \times A_2$ is open in $(\mathcal{M}_1 \times \mathcal{M}_2, \rho_1 \times \rho_2)$. The proof is the same as Theorem 10.10.1.

A set is closed iff every convergent sequence from the set has its limit in the set; the proof is the same as for Theorem 11.2.1.

The definition of a *bounded set* is the same as in Section 11.3, tha analogue of Proposition 11.3.2 holds (with the origin replaced by any fixed point $a^* \in \mathcal{M}$), convergent sequences are bounded (Proposition 11.3.4).

Bounded sequences do *not* necessarily have a convergent subsequence (c.f. Theorem 11.4.2). The sequence (f_n) in Section 15.2 is bounded in $\mathcal{C}[-1, 1]$ but no subsequence converges. In fact it is not too hard to check that

$$\rho_\infty(f_n, f_m) \geq \frac{1}{2}$$

if $m \geq 2n$. To see this take $x = \frac{1}{m}$. Then

$$|f_m(x) - f_n(x)| = 1 - \frac{n}{m} \geq 1 - \frac{1}{2} = \frac{1}{2}.$$

It follows that no subsequence can be Cauchy, and in particular no subsequence can converge.

Cauchy sequences are defined as in Definition 11.5.1. Convergent sequences are Cauchy, this is proved easily as for sequences of real numbers. But Cauchy sequences will not necesarily converge (to a point in the metric space) unless the metric space is complete.

A metric space $(\mathcal{M}, \rho)$ is *sequentially compact* if every sequence from $\mathcal{M}$ has a convergent subsequence (to a point in $\mathcal{M}$). This corresponds to the set A in Definition 12.1.1 being sequentially compact. The metric space $(\mathcal{M}, \rho)$ is *compact* if every open cover (i.e. a cover of $\mathcal{M}$ by subsets of $\mathcal{M}$ that are open in $\mathcal{M}$) has a finite subcover. This corresponds to the set A in Definition 12.2.2 being compact (although A is covered by sets which are open in $\mathbb{R}^p$, the restriction of these sets to A is a cover of A by sets that are open in A).

A subset of a metric space $(\mathcal{M}, \rho)$ is sequentially compact (compact) iff it is sequentially compact (compact) when regarded as a metric space with the induced metric from $\mathcal{M}$.

A metric space $(\mathcal{M}, \rho)$ is *totally bounded* if for every $\varepsilon > 0$ there is a cover of $\mathcal{M}$ by a *finite* number of open balls (in $\mathcal{M}$) of radius ε. Then it can be proved *a metric space is totally bounded iff every sequence has a Cauchy subsequence*; see Smith Theorem 9.9. (Note that what Smith calls "compact" is what is here, and usually, called "sequentially compact".)

A metric space is complete and totally bounded iff it is sequentially compact iff it is compact. This is the analogue of Theorems 12.1.2 and 12.3.1, where the metric space corresponds to A with the metric induced from $\mathbb{R}^p$.

The proof of the first 'iff" is fairly straightforward. First suppose $(\mathcal{M}, \rho)$ is complete and totally bounded. If (x_n) is a sequence from $\mathcal{M}$ then it has a Cauchy subsequence by the result two paragraphs back, and this subsequence converges to a limit in $\mathcal{M}$ by completeness and so $(\mathcal{M}, \rho)$ is sequentially compact. Next suppose $(\mathcal{M}, \rho)$ is sequentially compact. Then it is totally bounded again by the result two paragraphs back. To show it is complete suppose the sequence (x_n) is Cauchy: first by compactness it has a subsquence which converges to a member of $\mathcal{M}$, and second we use the fact (*Exercise*) that if a Cauchy sequence has a convergent subsequence then the Cauchy sequence itself must converge to the same limit as the subsequence.

The proof of the second "iff" is similar to that on Theorem 12.3.1. The proof that "compact" implies "sequentially compact" is esentially identical. The proof in the other direction uses the existence of a countable dense subset of $\mathcal{M}$. (This replaces the idea of considering those points in $\mathbb{R}^p$ all of whose components are rational.) We say a subset D of $\mathcal{M}$ is *dense* if for each $x \in \mathcal{M}$ and each $\varepsilon > 0$ there is an $x^* \in D$ such that $x \in B_\varepsilon(x^*)$. The existence of a countable dense subset of $\mathcal{M}$ follows from sequential compactness.[50]

If $(\mathcal{M}, \rho)$ and $(\mathcal{N}, \sigma)$ are metric spaces and $f : \mathcal{M} \to \mathcal{N}$ then we say f is *continuous* at $a \in \mathcal{M}$ if

$$(x_n) \subseteq \mathcal{M} \text{ and } x_n \to a \Rightarrow f(x_n) \to f(a).$$

It follows that *f is continuous at a iff for every $\varepsilon > 0$ there is a $\delta > 0$ such that for all $x, y \in \mathcal{M}$:*

$$\rho(x, y) < \delta \Rightarrow \sigma(f(x), f(y)) < \varepsilon.$$

The proof is the same as for Theorem 7.1.2. One defines *uniform continuity* as in Definition 13.1.3.

A continuous function defined on a compact set is bounded above and below and has a maximum and a minimum value. Moreoveor it is uniformly continuous. The

[50]We have already noted that a sequentially compact set is totally bounded. Let A_k be the finite set of points corresponding to $\varepsilon = \frac{1}{k}$ in the definition of total boundedness. Let $A = \bigcup_{k=1}^{\infty} A_k$. Then A is countable. It is also dense, since if $x \in \mathcal{M}$ then there exist points in A as close to x as we wish.

proof is the same as for Theorem 13.2.1, after using the fact that compactness is equivalent to sequential compactness.

Limit points, *isolated points*, the definition of *limit of a function at a point*, and the equivalence of this definition with the ε–δ characterisation, are completely analogous to Definition 13.3.1 and Theorem 13.3.2.

A function $f : \mathcal{M} \to \mathcal{N}$ *is continuous iff the inverse image of every open (closed) set in* $\mathcal{N}$ *is open (closed) in* $\mathcal{M}$. The proof is esentially the same as in Theorem 13.4.3.

The continuous image of a compact set is compact. The proof is the same as in Theorem 13.5.1.

The inverse of a one-to-one continuous function defined on a compact set is continuous. The proof is essentially the same as for Theorem 13.5.2.

16. Some applicatons of the Contraction Mapping Principle

16.1. Markov processes. Recall Theorem 4.0.2:

THEOREM 16.1.1. *If all entries in the probability transition matrix P are greater than 0, and $\mathbf{x}_0$ is a probability vector, then the sequence of vectors*

$$\mathbf{x_0}, P\mathbf{x_0}, P^2\mathbf{x_0}, \dots,$$

converges to a probability vector $\mathbf{x}^*$, and this vector does not depend on $\mathbf{x_0}$.

Moreover, the same results are true even if we just assume P^k has all entries non-zero for some integer $k > 1$.

The vector $\mathbf{x}^*$ is the unique non-zero solution of $(P - I)\mathbf{x}^* = \mathbf{0}$.

We will give a proof that uses the Contraction Mapping Principle. The proof is rather subtle, since P is only a contraction with respect to the ρ_1 metric, and it is also necessary to restrict to the subset of $\mathbb{R}^n$ consisting of those vectors $\mathbf{a}$ such that $\sum a_i = 1$.

LEMMA 16.1.2. *Suppose P is a probability transition matrix all of whose entries are at least ε, for some $\varepsilon > 0$. Then P is a contraction map on*

$$\mathcal{M} = \{\, \mathbf{a} \in \mathbb{R}^n \mid \sum a_i = 1 \,\}.$$

in the ρ_1 metric, with contraction constant $1 - n\varepsilon$.

PROOF. We will prove that

$$\rho_1(P\mathbf{a}, P\mathbf{b}) \le (1 - n\varepsilon)\, \rho_1(\mathbf{a}, \mathbf{b}) \tag{24}$$

for all $\mathbf{a}, \mathbf{b} \in \mathcal{M}$.

Suppose $\mathbf{a}, \mathbf{b} \in \mathcal{M}$. Let $\mathbf{v} = \mathbf{a} - \mathbf{b}$. Then $\mathbf{v} \in H$, where

$$H = \{\, \mathbf{v} \in \mathbb{R}^n \mid \sum v_i = 0 \,\}.$$

Moreover,

$$\begin{aligned} \rho_1(\mathbf{a}, \mathbf{b}) &= \|\mathbf{a} - \mathbf{b}\|_1 = \|\mathbf{v}\|_1, \\ \rho_1(P\mathbf{a}, P\mathbf{b}) &= \|P\mathbf{a} - P\mathbf{b}\|_1 = \|P\mathbf{v}\|_1, \end{aligned}$$

where $\|\mathbf{w}\|_1 = \sum |w_i|$.

Thus in order to prove (24) it is sufficient to show

$$\|P\mathbf{v}\|_1 \le (1 - n\varepsilon)\, \|\mathbf{v}\|_1 \tag{25}$$

for all $\mathbf{v} \in H$.

From the following Lemma, we can write

$$\mathbf{v} = \sum_{ij} \eta_{ij}(\mathbf{e}_i - \mathbf{e}_j), \qquad \text{where } \eta_{ij} > 0,\ \|\mathbf{v}\|_1 = 2\sum \eta_{ij}. \tag{26}$$

Then

$$\begin{aligned} \|P\mathbf{v}\|_1 &= \Big\| \sum_{ij} \eta_{ij} P(\mathbf{e}_i - \mathbf{e}_j) \Big\|_1 \\ &\le \sum_{ij} \eta_{ij}\, \|P(\mathbf{e}_i - \mathbf{e}_j)\|_1 \end{aligned}$$

(by the triangle inequality for $\|\cdot\|_1$ and using $\eta_{ij} > 0$)

$$\begin{aligned} &= \sum_{ij} \eta_{ij} \Big\| \sum_k (P_{ki} - P_{kj})\, \mathbf{e}_k \Big\|_1 \\ &= \sum_{ij} \eta_{ij} \sum_k |P_{ki} - P_{kj}| \end{aligned}$$

$$= \sum_{ij} \eta_{ij} \sum_k \big(P_{ki} + P_{kj} - 2\min\{P_{ki}, P_{kj}\} \big)$$

(since $|a-b| = a+b-2\min\{a,b\}$ for any real numbers a and b, as one sees by checking, without loss of generality, the case $b \le a$)

$$= \sum_{ij} \eta_{ij}(2 - 2\varepsilon)$$

(since the columns in P sum to 1 and each entry is $\ge \varepsilon$)

$$= (1 - n\varepsilon)\|\mathbf{v}\|_1 \quad \text{(from (26))}$$

This completes the proof. □

LEMMA 16.1.3. *Suppose* $\mathbf{v} \in \mathbb{R}^n$ *and* $\sum v_i = 0$. *Then* $\mathbf{v}$ *can be written in the form*

$$\mathbf{v} = \sum_{i,j} \eta_{ij}(\mathbf{e}_i - \mathbf{e}_j), \quad \textit{where } \eta_{ij} > 0,\ \|\mathbf{v}\|_1 = 2\sum \eta_{ij}.$$

PROOF. If $\mathbf{v}$ is in the span of two basis vectors, then after renumbering we have

$$\begin{aligned} \mathbf{v} &= v_1\mathbf{e}_1 + v_2\mathbf{e}_2 \quad (\text{where } \ v_1 + v_2 = 0) \\ &= v_1(\mathbf{e}_1 - \mathbf{e}_2) \\ &= v_2(\mathbf{e}_2 - \mathbf{e}_1). \end{aligned}$$

Since either $v_1 \ge 0$ and then $\|\mathbf{v}\|_1 = 2v_1$, or $v_2 \ge 0$ and then $\|\mathbf{v}\|_1 = 2v_2$, this proves the result in this case.

Suppose the claim is true for any $\mathbf{v} \in H$ which is in the span of k basis vectors. Assume that $\mathbf{v}$ is spanned by $k+1$ basis vectors, and write (after renumbering if necessary)

$$\mathbf{v} = v_1\mathbf{e}_1 + \cdots + v_{k+1}\mathbf{e}_{k+1}.$$

where $v_1 + \cdots + v_{k+1} = 0$.

Choose p so

$$|v_p| = \max\{\, |v_1|, \ldots, |v_{k+1}|\, \}.$$

Choose q so v_q has the opposite sign to v_p (this is possible since $\sum v_i = 0$). Note that

$$|v_p + v_q| = |v_p| - |v_q| \tag{27}$$

since $|v_q| \le |v_p|$ and since v_q has the opposite sign to v_p.

Write (where $\widehat{\ }$ indicates that the relevant term is missing from the sum)

$$\begin{aligned} \mathbf{v} &= \Big(v_1\mathbf{e}_1 + \cdots + \widehat{v_q\mathbf{e}_q} + \cdots + (v_p + v_q)\mathbf{e}_p + \cdots + v_{kn+1}\mathbf{e}_{k+1} \Big) + v_q(\mathbf{e}_q - \mathbf{e}_p) \\ &= \mathbf{v}^* + v_q(\mathbf{e}_q - \mathbf{e}_p). \end{aligned}$$

Since $\mathbf{v}^*$ has no $\mathbf{e}_q$ component, and the sum of the coefficients of $\mathbf{v}^*$ is $\sum v_i = 0$, we can apply the inductive hypothesis to $\mathbf{v}^*$ to write

$$\mathbf{v}^* = \sum \eta_{ij}(\mathbf{e}_i - \mathbf{e}_j), \qquad \text{where } \eta_{ij} > 0, \|\mathbf{v}^*\|_1 = 2\sum \eta_{ij}.$$

In particular,

$$\mathbf{v} = \begin{cases} \sum \eta_{ij}(\mathbf{e}_i - \mathbf{e}_j) + |v_q|(\mathbf{e}_q - \mathbf{e}_p) & v_q \ge 0 \\ \sum \eta_{ij}(\mathbf{e}_i - \mathbf{e}_j) + |v_q|(\mathbf{e}_p - \mathbf{e}_q) & v_q \le 0 \end{cases}$$

All that remains to be proved is

$$\|\mathbf{v}\|_1 = 2\sum \eta_{ij} + 2|v_q|,$$

i.e.

$$\|\mathbf{v}\|_1 = \|\mathbf{v}^*\|_1 + 2|v_q|. \tag{28}$$

But

$$\begin{aligned}
\|\mathbf{v}^*\|_1 &= |v_1| + \cdots + \widehat{|v_q|} + \cdots + |v_p + v_q| + \cdots + |v_{k+1}| \\
&= (|v_1| + \cdots + |v_{k+1}|) - |v_q| - |v_p| + |v_p + v_q| \\
&= (|v_1| + \cdots + |v_{k+1}|) - 2|v_q| \quad \text{by (27)} \\
&= \|\mathbf{v}\|_1 - 2|v_q|.
\end{aligned}$$

This proves (28) and hence the Lemma. □

Proof of Theorem 16.1.1. The first paragraph of the theorem follows from the Contraction Mapping Principle and Lemma 16.1.2.

The last paragraph was proved after the statement of Theorem 4.0.2.

For the second paragraph we consider a sequence $P^i(\mathbf{x}_0)$ with $i \to \infty$. From Lemma 16.1.2, P^k is a contraction map with contraction constant λ (say). For any natural number i we can write $i = mk + j$, where $0 \le j < k$.

Then

$$\begin{aligned}
\rho_1(P^i\mathbf{x}_0, \mathbf{x}^*) &= \rho_1(P^{mk+j}\mathbf{x}_0, \mathbf{x}^*) \\
&\le \rho_1(P^{mk+j}\mathbf{x}_0, P^{mk}\mathbf{x}_0) + \rho_1(P^{mk}\mathbf{x}_0, \mathbf{x}^*) \\
&\le \rho_1\big((P^k)^m P^j\mathbf{x}_0, (P^k)^m\mathbf{x}_0\big) + \rho_1((P^k)^m\mathbf{x}_0, \mathbf{x}^*) \\
&\le \lambda^m \rho_1\big(P^j\mathbf{x}_0, \mathbf{x}_0\big) + \rho_1((P^k)^m\mathbf{x}_0, \mathbf{x}^*) \\
&\le \lambda^m \max_{0\le j<k} \rho_1\big(P^j\mathbf{x}_0, \mathbf{x}_0\big) + \rho_1((P^k)^m\mathbf{x}_0, \mathbf{x}^*).
\end{aligned}$$

(Note that $m \to \infty$ as $i \to \infty$). The first term approaches 0 since $\lambda^m \to 0$ and the second approaches 0 by the Contraction Mapping Principle applied to the contraction map P^k. □

Remark 16.1.1. The fact we only needed to assume some power of P was a contraction map can easily be generalised.

That is, in the Contraction Mapping Principle, Theorem 11.6.1, if we assume $\overset{1}{F} \circ \cdots \circ \overset{k}{F}$ is a contraction map for some k then there is still a unique fixed point. The proof is similar to that above for P.

16.2. Integral Equations. See Reed Section 5.4 for discussion.

Instead of the long proof in Reed of the main theorem, Theorem 5.4.1, we see here that it is a consequence of the Contraction Mapping Theorem.

Theorem 16.2.1. *Let $F(x)$ be a continuous function defined on $[a, b]$. Suppose $K(x, y)$ is continuous on $[a, b] \times [a, b]$ and*

$$M = \max\{|K(x, y)| \mid a \le x \le b,\ a \le y \le b\}.$$

Then there is a unique continuous function $\psi(x)$ defined on $[a, b]$ such that

$$\psi(x) = f(x) + \lambda \int_a^b K(x, y)\, \psi(y)\, dy, \tag{29}$$

provided $|\lambda| < \frac{1}{M(b-a)}$.

Proof. We will use the contraction mapping principle on the complete metric space $(\mathcal{C}[a,b], \rho_\infty)$.

Define

$$T : \mathcal{C}[a,b] \to \mathcal{C}[a,b]$$

by

$$(T\psi)(x) = f(x) + \lambda \int_a^b K(x,y)\,\psi(y)\,dy \tag{30}$$

Note that if $\psi \in \mathcal{C}[a,b]$, then $T\psi$ is certainly a *function* defined on $[a,b]$ — the value of $T\psi$ at $x \in [a,b]$ is obtained by evaluating the right side of (30). For fixed x, the integrand in (30) is a continuous function of y, and so the integral exists.

We next claim that $T\psi$ is *continuous*, and so indeed $T : \mathcal{C}[a,b] \to \mathcal{C}[a,b]$.

The first term on the right side of (30) is continuous, being f.

The second term is also continuous. To see this let $G(x,y) = K(x,y)\,\psi(y)$ and let $g(x) = \int_a^b G(x,y)\,dy$. Then $T\psi = f + \lambda g$. To see that g is continuous on $[a,b]$, first note that

$$|g(x_1) - g(x_2)| = \Big| \int_a^b G(x_1,y) - G(x_2,y)\,dy \Big| \le \int_a^b |G(x_1,y) - G(x_2,y)|\,dy. \tag{31}$$

Now suppose $\epsilon > 0$. Since $G(x,y)$ is continuous on the closed bounded set $[a,b] \times [a,b]$, it is uniformly continuous there (by Theorem 13.2.1). Hence there is a $\delta > 0$ such that

$$|(x_1,y_1) - (x_2,y_2)| < \delta \quad \Rightarrow \quad |G(x_1,y_1) - G(x_2,y_2)| < \epsilon.$$

Using this it follows from (31) that

$$|x_1 - x_2| < \delta \quad \Rightarrow \quad |g(x_1) - g(x_2)| < \epsilon(b-a).$$

Hence g is uniformly continuous on $[a,b]$.

This completes the proof that $T\psi$ is continuous (in fact uniformly) on $[a,b]$. Hence

$$T : \mathcal{C}[a,b] \to \mathcal{C}[a,b].$$

Moreover, *ψ is a fixed point of T iff ψ solves* (29).

We next claim that T is *contraction map* on $\mathcal{C}[a,b]$. To see this we estimate

$$\begin{aligned} |T\psi_1(x) - T\psi_2(x)| &= |\lambda| \, \Big| \int_a^b K(x,y)\,(\psi_1(y) - \psi_2(y))\,dy \Big| \\ &\le |\lambda| \int_a^b |K(x,y)|\,|\psi_1(y) - \psi_2(y)|\,dy \\ &\le |\lambda|\, M(b-a)\,\rho_\infty(\psi_1,\psi_2). \end{aligned}$$

Since this is true for every $x \in [a,b]$, it follows that

$$\rho_\infty(T\psi_1, T\psi_2) \le |\lambda|\, M(b-a)\,\rho_\infty(\psi_1,\psi_2).$$

By the assumption on λ, it follows that T is a contraction map with contraction ratio $|\lambda|\, M(b-a)$, which is < 1.

Because $(\mathcal{C}[a,b], \rho_\infty)$ is a *complete* metric space, it follows that T has a unique fixed point, and so there is a unique continuous function ψ solving (29). □

REMARK 16.2.1. Note that we also have a way of approximating the solution. We can begin with some function such as

$$\psi_0(x) = 0 \quad \text{for } a \le x \le b,$$

and then successively apply T to find better and better approximations to the solution ψ.

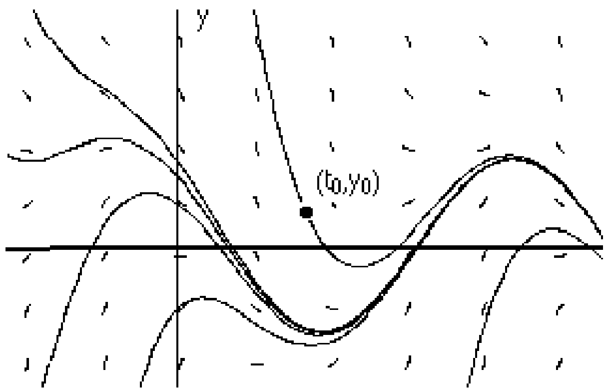
In practice, we apply some type of numerical integration to find approximations to $T\psi_0$, $T^2\psi_0$, $T^3\psi_0, \ldots$.

16.3. Differential Equations. In this section we will prove the *Fundamental Existence and Uniqueness Theorem for Differential Equations*, Theorem 7.1.1 of Reed.

You should first read Chapter 7 of Reed up to the beginning of Theorem 7.1.1.

We will prove Theorem 7.1.1 here, but with a simpler proof using the Contraction Mapping Principle.

Solutions of dy/dt = f(t,y)



The slope of the line segment at (t,y) equals f(t,y)
Every solution of the differential equation is tangent at each point on its graph to the line segment at that point.

THEOREM 16.3.1. *Let f be continuously differentiable on the square*

$$S = [t_0 - \delta, t_0 + \delta] \times [y_0 - \delta, y_0 + \delta].$$

Then there is a $T \le \delta$, and a unique continuously differentiable function $y(t)$ defined on $[t_0 - T, t_0 + T]$, such that

$$\begin{aligned} \frac{dy}{dt} &= f(t,y) \quad \textit{on } [t_0 - T, t_0 + T], \\ y(t_0) &= y_0. \end{aligned} \tag{32}$$

REMARK 16.3.1.

1. The diagram shows the square S centred at (t_0, y_0) and illustrates the situation. We are looking for the unique curve through the point (t_0, y_0) which at every point is tangent to the line segment of slope $f(t,y)$.

   Note that the solution through the point (t_0, y_0) "escapes" through the top of the square S, and this is why we may need to take $T < \delta$.

   From the diagram, the absolute value of the slope of the solution will be $\le M$, where $M = \max\{\, |f(t,y)| \mid (t,y) \in S \,\}$. For this reason, we will require that $MT \le \delta$, i.e. $T \le \delta/M$.

2. The function $y(t)$ must be differentiable for the differential equation to be defined. But it then follows from the differential equation that dy/dt is in fact continuous.

   * In fact more is true. The right side of the differential equation is differentiable (*why*?) and its derivative is even continuous (*why*?). This shows that dy/dt is differentiable with continuous derivative, i.e. $y(t)$ is in fact twice continuously differentiable. If f is infinitely differentiable, it can similarly be shown that $y(t)$ is also infinitely differentiable.

PROOF OF THEOREM.
Step 1: We first reduce the problem to an equivalent integral equation (not the same one as in the last section, however).

It follows by integrating (32) from t_0 to t that any solution $y(t)$ of (32) satisfies

$$y(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\, ds. \tag{33}$$

Conversely, any continuous function $y(t)$ satisfying (33) is differentiable (by the Fundamental Theorem of Calculus) and its derivative satisfies the differential equation in (32). It also satisfies the initial condition $y(t_0) = y_0$ (*why?*).

Step 2: We next show that (33) is the same as finding a certain "'fixed point".

If $y(t)$ is a continuous function defined on $[t_0 - \delta, t_0 + \delta]$, let Fy (often written $F(y)$) be the function defined by

$$(Fy)(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\, ds. \tag{34}$$

For the integral to be defined, we require that $(s, y(s))$ belong to the square S, and for this reason we will need to restrict to functions $y(t)$ defined on some sufficiently small interval $[t_0 - T, t_0 + T]$.

More precisely, since f is continuous and hence bounded on S, it follows that there is a constant M such that

$$(t, y) \in S \quad \Rightarrow \quad |f(t, y)| \leq M.$$

This implies from (34) that

$$|(Fy)(t) - y_0| \leq M\, |t - t_0|.$$

It follows that if the graph of $y(t)$ is in S, then so is the graph of $(Fy)(t)$, provided t satisfies $M\, |t - t_0| \leq \delta$.

For this reason, we impose the restriction on T that $MT \leq \delta$, i.e.

$$T \leq \frac{\delta}{M}. \tag{35}$$

This ensures that $(Fy)(t)$ is defined for $t \in [t_0 - T, t_0 + T]$.

Since $f(s, y(s))$ is continuous, being a composition of continuous functions, it follows from the Fundamental Theorem of Calculus that $(Fy)(t)$ is differentiable, and in particular is continuous. Hence

$$F : \mathcal{C}[t_0 - T, t_0 + T] \to \mathcal{C}[t_0 - T, t_0 + T].$$

Moreover, it follows from the definition (34) that $y(t)$ *is a fixed point of* F *iff* $y(t)$ *is a solution of* (33) *on* $[t_0 - T, t_0 + T]$ *and hence iff* $y(t)$ *is a solution of* (32) *on* $[t_0 - T, t_0 + T]$.

Step 2: We next impose a further restriction on T in order that F be a contraction map. For this we need the fact that since $\frac{\partial f}{\partial t}$ is continuous on S, there is a constant

K such that

$$(t, y) \in S \quad \Rightarrow \quad \left|\frac{\partial f}{\partial y}\right| \leq K.$$

We now compute for any $t \in [t_0 - T, t_0 + T]$ and any $y_1, y_2 \in \mathcal{C}[t_0 - T, t_0 + T]$, that

$$\begin{aligned}
|Fy_1(t) - Fy_2(t)| &= \left|\int_{t_0}^{t} f(s, y_1(s)) - f(s, y_2(s))\ ds\right| \\
&\leq \int_{t_0}^{t} \Big| f(s, y_1(s)) - f(s, y_1(s)) \Big|\ ds \\
&\leq \int_{t_0}^{t} K\,|y_1(s) - y_2(s)|\, ds \qquad \text{by the Mean Value Theorem} \\
&\leq KT\, \rho_\infty(y_1, y_2)
\end{aligned}$$

since $|t - t_0| \leq T$ and since $|y_1(s) - y_2(s)| \leq \rho_\infty(y_1, y_2)$.

Since t is any point in $[t_0 - T, t_0 + T]$, it follows that

$$\rho_\infty(Fy_1, Fy_2) \leq KT\, \rho_\infty(y_1, y_2).$$

We now make the second restriction on T that

$$T \leq \frac{1}{2K}. \tag{36}$$

This guarantees that F is a contraction map on $\mathcal{C}[t_0 - T, t_0 + T]$ with contraction constant $\frac{1}{2}$. It follows that F has a unique fixed point, and that this is then the unique solution of (32). □

Remark 16.3.2. It appears from the diagram that we do not really need the second restriction (36) on T. In other words, we should only need (35). This is indeed the case. One can show by repeatedly applying the previous theorem that the solution can be continued until it either escapes through the top, bottom or sides of S. In fact this works for much more general sets S.

In particular, if the function $f(t, y)$ and the differential equation are defined for all $(t, y) \in \mathbb{R}^2$, then the solution will either approach $+\infty$ or $-\infty$ at some finite time t^*, or will exist for all time. For more discussion see Reed Section 7.2.

Remark 16.3.3. The same proof, with only notational changes, works for general first order systems of differential equations of the form

$$\begin{aligned}
\frac{dy_1}{dt} &= f_1(t, y_1, y_2, \dots, y_n) \\
\frac{dy_2}{dt} &= f_2(t, y_1, y_2, \dots, y_n) \\
&\ \ \vdots \\
\frac{dy_n}{dt} &= f_n(t, y_1, y_2, \dots, y_n).
\end{aligned}$$

Remark 16.3.4. Second and higher order differential equations can be reduced to first order systems by introducing new variables for the lower order derivates.

For example, the second order differential equation

$$y'' = F(t, y, y')$$

is equivalent to the first order system of differential equations

$$y_1' = y_2, \quad y_2' = F(t, y_1, y_2).$$

To see this, suppose $y(t)$ is a solution of the given differential equation and let $y_1 = y,\ y_2 = y'$. Then clearly $y_1(t),\ y_2(t)$ solve the first order system.

Conversely, if $y_1(t),\ y_2(t)$ solve the first order system let $y = y_1$. Then $y' = y_2$ and $y(t)$ solves the second order differential equation.

It follows from the previous remark that a similar Existence and Uniqueness Theorem applies to second order differential equations, provided we specify *both* $y(t_0)$ and $y'(t_0)$.

A similar remark applies to nth order differential equations, except that one must specify the first $n-1$ derivatives at t_0.

Finally, similar remarks also apply to higher order systems of differential equations. There are no new ideas involved, just notation!