THE CONTRACTION MAPPING THEOREM

KEITH CONRAD

1. Introduction

Let $f\colon X \to X$ be a mapping from a set X to itself. We call a point $x \in X$ a *fixed point* of f if $f(x) = x$. For example, if $[a, b]$ is a closed interval then any continuous function $f\colon [a, b] \to [a, b]$ has at least one fixed point. This is a consequence of the intermediate value theorem, as follows. Since $f(a) \geq a$ and $f(b) \leq b$, we have $f(b) - b \leq 0 \leq f(a) - a$. The difference $f(x) - x$ is continuous, so by the intermediate value theorem 0 is a value of $f(x) - x$ for some $x \in [a, b]$, and that x is a fixed point of f. Of course, there could be more than one fixed point.

We will discuss here the most basic fixed-point theorem in analysis. It is due to Banach and appeared in his Ph.D. thesis (1920, published in 1922).

Theorem 1.1. *Let (X, d) be a complete metric space and $f\colon X \to X$ be a map such that*

$$d(f(x), f(x')) \leq c\,d(x, x')$$

for some $0 \leq c < 1$ and all x and x' in X. Then f has a unique fixed point in X. Moreover, for any $x_0 \in X$ the sequence of iterates $x_0, f(x_0), f(f(x_0)), \ldots$ converges to the fixed point of f.

When $d(f(x), f(x')) \leq c\,d(x, x')$ for some $0 \leq c < 1$ and all x and x' in X, f is called a *contraction*. A contraction shrinks distances by a uniform factor c less than 1 for all pairs of points. Theorem 1.1 is called the contraction mapping theorem or Banach's fixed-point theorem.

Example 1.2. A standard procedure to approximate a solution in $\mathbf{R}$ to the numerical equation $g(x) = 0$, where g is differentiable, is to use Newton's method: find an approximate solution x_0 and then compute the recursively defined sequence

$$x_n = x_{n-1} - \frac{g(x_{n-1})}{g'(x_{n-1})}.$$

This recursion amounts to iteration of the function $f(x) = x - g(x)/g'(x)$ starting from $x = x_0$. A solution of $g(x) = 0$ is the same as a solution of $f(x) = x$: a fixed point of f.

To use Newton's method to estimate $\sqrt{3}$, we take $g(x) = x^2 - 3$ and seek a (positive) root of $g(x)$. The Newton recursion is

$$x_n = x_{n-1} - \frac{x_{n-1}^2 - 3}{2x_{n-1}} = \frac{1}{2}\left(x_{n-1} + \frac{3}{x_{n-1}}\right),$$

so $f(x) = (1/2)(x + 3/x)$. The fixed points of f are the square roots of 3. In the following table are iterations of f with three different positive choices of x_0.

n	x_n	x_n	x_n
0	1.5	1.9	10
1	1.75	1.7394736842	5.15
2	1.7321428571	1.7320666454	2.8662621359
3	1.7320508100	1.7320508076	1.9564607317
4	1.7320508075	1.7320508075	1.7449209391
5	1.7320508075	1.7320508075	1.7320982711

All three sequences of iterates in the table appear to be tending to $\sqrt{3} \approx 1.7320508075688$. The last choice of x_0 in the table is quite a bit further away from $\sqrt{3}$ so the iterations take longer to start resembling $\sqrt{3}$.

To justify the application of the contraction mapping theorem to this setting, we need to find a complete metric space on which $f(x) = (1/2)(x+3/x)$ is a contraction. The set $(0, \infty)$ doesn't work, since this is not complete. The closed interval $X_t = [t, \infty)$ is complete for any $t > 0$. For which t is $f(X_t) \subset X_t$ and f a contraction on X_t? Well, the minimum of $f(x)$ on $(0, \infty)$ is at $x = \sqrt{3}$, with $f(\sqrt{3}) = \sqrt{3}$, so for any $t \leq \sqrt{3}$ we have $x \geq t \Longrightarrow f(x) \geq \sqrt{3} \geq t$, hence $f(X_t) \subset X_t$. To find such a t for which f is a contraction on X_t, for any positive x and x' we have

$$f(x) - f(x') = \frac{x - x'}{2}\left(1 - \frac{3}{xx'}\right).$$

If $x \geq t$ and $x' \geq t$ then $1 - 3/t^2 \leq 1 - 3/xx' < 1$. Therefore $|1 - 3/xx'| < 1$ as long as $1 - 3/t^2 > -1$, which is equivalent to $t^2 > 3/2$. Taking $\sqrt{3/2} < t \leq \sqrt{3}$ (*e.g.*, $t = \sqrt{3/2}$ or $t = \sqrt{3}$), we have $f \colon X_t \to X_t$ and $|f(x) - f(x')| \leq (1/2)|x - x'|$ for x and x' in X_t. The contraction mapping theorem says the iterates of f starting at any $x_0 \geq t$ will converge to $\sqrt{3}$. How many iterations are needed to approximate $\sqrt{3}$ to a desired accuracy will be addressed in Section 2 after we prove the contraction mapping theorem.

Although $(0, \infty)$ is not complete, iterations of f starting from any $x > 0$ will converge to $\sqrt{3}$; if we start below $\sqrt{3}$ then applying f will take us above $\sqrt{3}$ (because $f(x) \geq \sqrt{3}$ for all $x > 0$) and the contraction mapping theorem on $[\sqrt{3}, \infty)$ then kicks in to guarantee that further iterations of f will converge to $\sqrt{3}$.

The contraction mapping theorem has many uses in analysis, both pure and applied. After proving the theorem (Section 2) and discussing some generalizations (Section 3), we will give one major application: Picard's theorem (Section 4), which is the basic existence and uniqueness theorem for ordinary differential equations. There are further applications of the contraction mapping theorem to partial differential equations [18], to the Gauss–Seidel method of solving systems of linear equations in numerical analysis [17, p. 269], to a proof of the inverse function theorem [8], [15, pp. 221-223], and to Google's Page Rank algorithm [9], [12], [19]. A basic introductory account of the ideas of iteration and contraction in analysis with a few applications is in [4], and a comprehensive treatment is in [20].

2. Proof of the Contraction Mapping Theorem

Recalling the notation, $f \colon X \to X$ is a contraction with contraction constant c. We want to show f has a unique fixed point, which can be obtained as a limit through iteration of f from any initial value. To show f has at most one fixed point in X, let a and a' be fixed points of f. Then

$$d(a, a') = d(f(a), f(a')) \leq c d(a, a').$$

If $a \neq a'$ then $d(a,a') > 0$ so we can divide by $d(a,a')$ to get $1 \leq c$, which is false. Thus $a = a'$.

Next we want to show, for any $x_0 \in X$, that the recursively defined iterates $x_n = f(x_{n-1})$ for $n \geq 1$ converge to a fixed point of f. How close is x_n to x_{n+1}? For any $n \geq 1$, $d(x_n, x_{n+1}) = d(f(x_{n-1}), f(x_n)) \leq cd(x_{n-1}, x_n)$. Therefore

$$d(x_n, x_{n+1}) \leq cd(x_{n-1}, x_n) \leq c^2 d(x_{n-2}, x_{n-1}) \leq \cdots \leq c^n d(x_0, x_1).$$

Using the expression on the far right as an upper bound on $d(x_n, x_{n+1})$ shows the x_n's are getting consecutively close at a geometric rate. This implies the x_n's are Cauchy: for any $m > n$, using the triangle inequality several times shows

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots + d(x_{m-1}, x_m) \\ &\leq c^n d(x_0, x_1) + c^{n+1} d(x_0, x_1) + \cdots + c^{m-1} d(x_0, x_1) \\ &= (c^n + c^{n+1} + \cdots + c^{m-1}) d(x_0, x_1) \\ &\leq (c^n + c^{n+1} + c^{n+2} + \cdots) d(x_0, x_1) \\ &= \frac{c^n}{1-c} d(x_0, x_1). \qquad (2.1) \end{aligned}$$

To prove from this bound that the x_n's are Cauchy, choose $\varepsilon > 0$ and then pick $N \geq 1$ such that $(c^N/(1-c))d(x_0, x_1) < \varepsilon$. Then for any $m > n \geq N$,

$$d(x_n, x_m) \leq \frac{c^n}{1-c} d(x_0, x_1) \leq \frac{c^N}{1-c} d(x_0, x_1) < \varepsilon.$$

This proves $\{x_n\}$ is a Cauchy sequence. Since X is *complete*, the x_n's converge in X. Set $a = \lim_{n\to\infty} x_n$ in X.

To show $f(a) = a$, we need to know that contractions are continuous. In fact, a contraction is uniformly continuous. This is clear when $c = 0$ since then f is a constant function. If $c > 0$ and we are given $\varepsilon > 0$, setting $\delta = \varepsilon/c$ implies that if $d(x, x') < \delta$ then $d(f(x), f(x')) \leq cd(x, x') < c\delta = \varepsilon$. That proves f is uniformly continuous. Since f is then continuous, from $x_n \to a$ we get $f(x_n) \to f(a)$. Since $f(x_n) = x_{n+1}$, $f(x_n) \to a$ as $n \to \infty$. Then $f(a)$ and a are both limits of $\{x_n\}_{n\geq 0}$. From the uniqueness of limits, $a = f(a)$. This concludes the proof of the contraction mapping theorem.

Remark 2.1. When $f\colon X \to X$ is a contraction with constant c, any iterate f^n is a contraction with constant c^n; the unique fixed point of f will also be the unique fixed point of any f^n.

Remark 2.2. The contraction mapping theorem admits a converse [3], [16, pp. 523–526]. If X is any set (not yet a metric space), $c \in (0,1)$, and $f\colon X \to X$ is a function such that each iterate $f^n\colon X \to X$ has a unique fixed point then there is a metric on X making it a complete metric space such that, for this metric, f is a contraction with contraction constant c. For instance, the function $f\colon \mathbf{R} \to \mathbf{R}$ given by $f(x) = 2x$ has 0 as its unique fixed point, and the same applies to its iterates $f^n(x) = 2^n x$. Therefore there is a metric on $\mathbf{R}$ with respect to which it is *complete* and the function $x \mapsto 2x$ is a contraction, so in particular $2^n \to 0$ in this metric: how strange!

There are examples where f has a unique fixed point but an iterate of f does not, such as $f\colon \mathbf{R} \to \mathbf{R}$ by $f(x) = 1 - x$. Here $f^2(x) = f(f(x)) = x$ for all x in $\mathbf{R}$.

Corollary 2.3. *Let $f\colon X \to X$ be a contraction on a complete metric space and $Y \subset X$ be a closed subset such that $f(Y) \subset Y$. Then the unique fixed point of f is in Y.*

Proof. Since Y is a closed subset of a complete metric space, it is complete. Then we can apply the contraction mapping theorem to $f: Y \to Y$, so f has a fixed point in Y. Since f has only one fixed point in X, it must lie in Y. $\square$

The proof of the contraction mapping theorem yields useful information about the rate of convergence towards the fixed point, as follows.

Corollary 2.4. *Let f be a contraction mapping on a complete metric space X, with contraction constant c and fixed point a. For any $x_0 \in X$, with f-iterates $\{x_n\}$, we have the estimates*

$$d(x_n, a) \leq \frac{c^n}{1-c} d(x_0, f(x_0)), \tag{2.2}$$

$$d(x_n, a) \leq c d(x_{n-1}, a), \tag{2.3}$$

and

$$d(x_n, a) \leq \frac{c}{1-c} d(x_{n-1}, x_n). \tag{2.4}$$

Proof. From (2.1), for $m > n$ we have

$$d(x_n, x_m) \leq \frac{c^n}{1-c} d(x_0, x_1) = \frac{c^n}{1-c} d(x_0, f(x_0)).$$

The right side is independent of m. Let $m \to \infty$ to get $d(x_n, a) \leq (c^n/(1-c)) d(x_0, f(x_0))$.

To derive (2.3), from a being a fixed point we get

$$d(x_n, a) = d(f(x_{n-1}), f(a)) \leq c d(x_{n-1}, a). \tag{2.5}$$

Applying the triangle inequality to $d(x_{n-1}, a)$ on the right side of (2.5) using the three points x_{n-1}, x_n, and a,

$$d(x_n, a) \leq c(d(x_{n-1}, x_n) + d(x_n, a)),$$

and isolating $d(x_n, a)$ yields

$$d(x_n, a) \leq \frac{c}{1-c} d(x_{n-1}, x_n).$$

That's (2.4), so we're done. $\square$

The three inequalities in Corollary 2.4 serve different purposes. The inequality (2.2) tells us, in terms of the distance between x_0 and $f(x_0) = x_1$, how many times we need to iterate f starting from x_0 to be certain that we are within a specified distance from the fixed point. This is an upper bound on how long we need to compute. It is called an *a priori* estimate. Inequality (2.3) shows that once we find a term by iteration within some desired distance of the fixed point, all further iterates will be within that distance. However, (2.3) is not so useful as an error estimate since both sides of (2.3) involve the unknown fixed point. The inequality (2.4) tells us, after each computation, how much closer we are to the fixed point in terms of the previous two iterations. This kind of estimate, called an *a posteriori* estimate, is very important because if two successive iterations are nearly equal, (2.4) guarantees that we are very close to the fixed point. For example, if $c = 2/3$ and $d(x_{n-1}, x_n) < 1/10^{10}$ then (2.4) tells us $d(x_n, a) < 2/10^{10} < 1/10^9$.

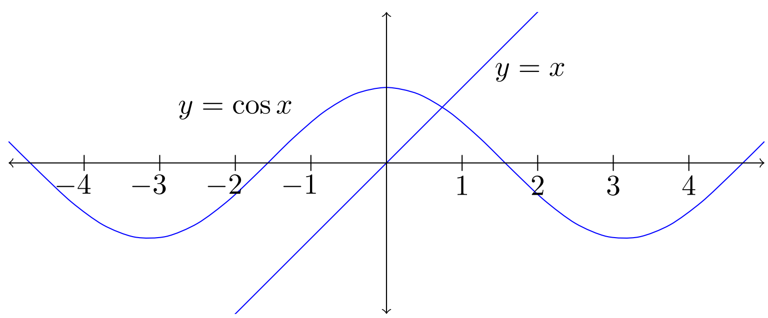
Example 2.5. To compute $\sqrt{3}$ to 4 digits after the decimal point, we use the recursion from Example 1.2: $\sqrt{3}$ is a fixed point of $f(x) = (1/2)(x + 3/x)$. In that example we saw a contraction constant for f on $[3/2, \infty)$ is $c = 1/2$. Taking $x_0 = 3/2$, so $|x_0 - f(x_0)| = 1/4$, (2.2) becomes

$$|x_n - \sqrt{3}| \leq \frac{1}{2^{n+1}}.$$

The right side is less than $1/10^4$ for $n = 13$, which means x_{13} lies within $1/10^4$ of $\sqrt{3}$. However, this iteration gives us that much accuracy a lot sooner: $x_3 \approx 1.732050810$ and $x_4 \approx 1.732050807$, so $|x_3 - x_4| \leq 10^{-8}$. By (2.4) with $c = 1/2$, $|x_4 - \sqrt{3}| \leq |x_4 - x_3|$, so x_4 gives us $\sqrt{3}$ accurately to 4 (and in fact a bit more) places after the decimal point.

The iteration we are using here to approximate $\sqrt{3}$ comes from Newton's method, which has its own error bounds separate from the bounds for general contraction mappings. When Newton's method converges, it does so doubly-exponentially (something like c^{2^n} and not just c^n), so essentially doubling the number of correct digits after each iteration while a typical contraction mapping produces at most one new correct digit after each iteration once the approximations get close enough to the fixed point.

Example 2.6. Graphs of $y = \cos x$ and $y = x$ intersect once (see the figure below), which means the cosine function has a unique fixed point in $\mathbf{R}$. From the graph, this fixed point lies in $[0, 1]$. We will show this point can be approximated through iteration.



The function $\cos x$ is not a contraction mapping on $\mathbf{R}$: there is no $c \in (0, 1)$ such that $|\cos x - \cos y| \leq c|x - y|$ for all x and y. Indeed, if there were such a c then $|(\cos x - \cos y)/(x - y)| \leq c$ whenever $x \neq y$, and letting $y \to x$ implies $|\sin x| \leq c$ for all x, which is false.

However, $\cos x$ *is* a contraction mapping on $[0, 1]$. First let's check $\cos x$ maps $[0, 1]$ to $[0, 1]$. Since cosine is decreasing on $[0, 1]$ and $\cos 1 \approx .54$, $\cos([0, 1]) \subset [0, 1]$. To find $c < 1$ such that $|\cos x - \cos y| \leq c|x - y|$ for all x and y in $[0, 1]$, we can't get by with elementary algebra as we did for the function $(1/2)(x+3/x)$ in Example 1.2. We will use the mean-value theorem: for any differentiable function f, $f(x) - f(y) = f'(t)(x - y)$ for some t between x and y, so bounding $|(\cos t)'| = |\sin t|$ over $[0, 1]$ will give us a contraction constant. Since $\sin t$ is increasing on $[0, 1]$, for all t in $[0, 1]$ we have $|\sin t| = \sin t \leq \sin 1 \approx .84147$. Therefore

$$|\cos x - \cos y| \leq .8415|x - y|,$$

for all x and y in $[0, 1]$. Since $[0, 1]$ is complete, the contraction mapping theorem tells us that there is a unique solution of $\cos x = x$ in $[0, 1]$ and it can be found by iterating the function $\cos x$ starting from any initial point in $[0, 1]$.

Here are some iterations of $\cos x$ starting from $x_0 = 0$:

n	x_n
0	0
1	1
2	.5403023
$\vdots$	$\vdots$
19	.7393038
20	.7389377
21	.7391843
22	.7390182

The estimates (2.2) and (2.4) with $x_0 = 0$ become

$$|x_n - a| \le \frac{.8415^n}{1 - .8415}, \quad |x_n - a| \le \frac{.8415}{1 - .8415}|x_{n-1} - x_n|. \tag{2.6}$$

We can use these bounds to know with certainty a value of n for which $|x_n - a| < .001$. The first bound falls below .001 for the first time when $n = 51$, while the second one falls below .001 much earlier, at $n = 22$.

Taking $n = 22$, the second upper bound on $|x_{22} - a|$ in (2.6) is a little less than .0009, so a lies between

$$x_{22} - .0009 \approx .7381 \text{ and } x_{22} + .0009 \approx .7399.$$

The iterative process achieves the same accuracy in fewer steps if we begin the iteration closer to the fixed point. Let's start at $x_0 = .7$ instead of at $x_0 = 0$. Here are some iterations:

n	x_n
0	.7
1	.7648421
2	.7214916
$\vdots$	$\vdots$
14	.7389324
15	.7391879
16	.7390158

Now (2.2) and (2.4) with $x_0 = .7$ say

$$|x_n - a| \le \frac{.8415^n}{1 - .8415}|.7 - \cos(.7)|, \quad |x_n - a| \le \frac{.8415}{1 - .8415}|x_{n-1} - x_n|.$$

The first upper bound on $|x_n - a|$ falls below .001 for $n = 35$, and the second upper bound falls below .001 when $n = 16$. Using the second bound, a lies between

$$x_{16} - .00092 \approx .7381 \text{ and } x_{16} + .00092 \approx .7399.$$

Example 2.7. We can use the contraction mapping theorem to compute inverses using the computationally "less expensive" operations of addition and multiplication. The linear polynomial $f(x) = mx + b$ has fixed point $b/(1 - m)$, and this function is a contraction mapping on $\mathbf{R}$ when $|m| < 1$. (In fact $x_n = m^n x_0 + (1 + m + m^2 + \cdots + m^{n-1})b = m^n x_0 + \frac{m^n - 1}{m - 1}b$, which implies that the three inequalities in Corollary 2.3 are in this case all equalities, so that corollary is sharp: the convergence of iterates under a contraction mapping can't be any faster *in general* than what Corollary 2.3 says.) For $s \neq 0$, to obtain $1/s$ as a fixed point of $f(x)$ we seek b and m such that $b/(1 - m) = 1/s$ and $|m| < 1$. Choosing b to satisfy $|1 - bs| < 1$, so b is an initial approximation to $1/s$ (their product is not too far from 1), the function $f(x) = (1-bs)x+b$ is a contraction on $\mathbf{R}$ with fixed point $1/s$. For example, if $|s-1| < 1$ then we can use $f(x) = (1-s)x+1$. For this contraction and $x_0 = 0$, we have $x_n = 1+(1-s)+\cdots+(1-s)^n$, which means the convergence of geometric series can be viewed as a special case of the contraction mapping theorem. (Logically this is circular, since we used geometric series in the proof of the contraction mapping theorem; see (2.1).)

We can also make iterations converge to $1/s$ using Newton's method on $g(x) = 1/x - s$, which amount to viewing $1/s$ as the fixed point of $f(x) = x - g(x)/g'(x) = 2x - sx^2$. From the mean-value theorem (or the triangle inequality), check that f is a contraction on $[1/s-\varepsilon, 1/s+\varepsilon]$ when $\varepsilon < 1/(2s)$. So if $|x_0 - 1/s| < 1/(2s)$ then iterates of f starting at x_0 will converge to $1/s$ by the contraction mapping theorem. This method of approximating $1/s$ converges much faster than the iterates of a linear polynomial: check by induction that the nth iterate of $f(x)$ is $1/s-(sx-1)^{2^n}/s$, so if $|x_0-1/s| < 1/(2s)$ then the nth iterate of f at x_0 differs from $1/s$ by less than $(1/2)^{2^n}(1/s)$, which decays to 0 exponentially rather than geometrically.

3. Generalizations

In some situations a function is not a contraction but an iterate of it is. This turns out to suffice to get the conclusion of the contraction mapping theorem for the original function.

Theorem 3.1. *If X is a complete metric space and $f\colon X \to X$ is a mapping such that some iterate $f^N\colon X \to X$ is a contraction, then f has a unique fixed point. Moreover, the fixed point of f can be obtained by iteration of f starting from any $x_0 \in X$.*

Proof. By the contraction mapping theorem, f^N has a unique fixed point. Call it a, so $f^N(a) = a$. To show a is the only possible fixed point of f, observe that a fixed point of f is a fixed point of f^N, and thus must be a. To show a really is a fixed point of f, we note that $f(a) = f(f^N(a)) = f^N(f(a))$, so $f(a)$ is a fixed point of f^N. Therefore $f(a)$ and a are both fixed points of f^N. Since f^N has a unique fixed point, $f(a) = a$.

We now show that for any $x_0 \in X$ the points $f^n(x_0)$ converge to a as $n \to \infty$. Consider the iterates $f^n(x_0)$ as n runs through a congruence class modulo N. That is, pick $0 \le r \le N-1$ and look at the points $f^{Nk+r}(x_0)$ as $k \to \infty$. Since

$$f^{Nk+r}(x_0) = f^{Nk}(f^r(x_0)) = (f^N)^k(f^r(x_0)),$$

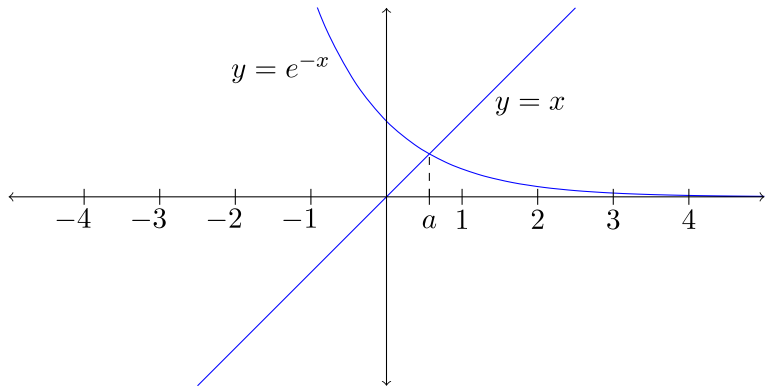
these points can be viewed (for each r) as iterates of f^N starting at the point $y_0 = f^r(x_0)$. Since f^N is a contraction, these iterates of f^N (from any initial point, such as y_0) must tend to a by the contraction mapping theorem. This limit is independent of the value of r in the range $\{1,\ldots,N-1\}$, so all N sequences $\{f^{Nk+r}(x_0)\}_{k\ge 1}$ tend to a as $k \to \infty$. This shows

$$f^n(x_0) \to a \tag{3.1}$$

as $n \to \infty$. □

We will use Theorem 3.1 in Section 4 to solve differential equations.

Example 3.2. The graphs of $y = e^{-x}$ and $y = x$ intersect once, so $e^{-a} = a$ for a unique real number a. See the graph below.



The function $f(x) = e^{-x}$ is not a contraction on $\mathbf{R}$ (for instance, $|f(-2) - f(0)| \approx 6.38 > |-2-0|$), but its second iterate $g(x) = f^2(x) = e^{-e^{-x}}$ is a contraction on $\mathbf{R}$: by the mean-value theorem

$$g(x) - g(y) = g'(t)(x - y)$$

for some t between x and y, where $|g'(t)| = |e^{-e^{-t}}e^{-t}| = e^{-(t+e^{-t})} \leq e^{-1}$ (since $t + e^{-t} \geq 1$ for all real t). Hence f^2 has contraction constant $1/e \approx .367 < 1$. By Theorem 3.1 the solution to $e^{-a} = a$ can be approximated by iteration of f starting with any real number. Iterating f enough times with $x_0 = 0$ suggests $a \approx .567$. To prove this approximation is correct, one can generalize the error estimates in Corollary 2.4 to apply to a function having an iterate as a contraction. Alternatively, check $f([\varepsilon, 1]) \subset [\varepsilon, 1]$ when $0 \leq \varepsilon \leq 1/e$, and when $0 < \varepsilon \leq 1/e$ the function f is a contraction on $[\varepsilon, 1]$. Therefore Corollary 2.4 can be applied directly to f by starting iterates at any $x_0 \in (0, 1]$.

Remark 3.3. When f^N is a contraction, f^N is continuous, but that does not imply f is continuous. For example, let $f\colon [0,1] \to [0,1]$ by $f(x) = 0$ for $0 \leq x \leq 1/2$ and $f(x) = 1/2$ for $1/2 < x \leq 1$. Then $f(f(x)) = 0$ for all x, so f^2 is a contraction but f is discontinuous.

It was important in the proof of the contraction mapping theorem that the contraction constant c be strictly less than 1. That gave us control over the rate of convergence of $f^n(x_0)$ to the fixed point since $c^n \to 0$ as $n \to \infty$. If instead of f being a contraction we suppose $d(f(x), f(x')) < d(x, x')$ whenever $x \neq x'$ in X then we lose that control and indeed a fixed point need not exist.

Example 3.4. Let I be a closed interval in $\mathbf{R}$ and $f\colon I \to I$ be differentiable with $|f'(t)| < 1$ for all t. Then the mean-value theorem implies $|f(x) - f(x')| < |x - x'|$ for $x \neq x'$ in I. The following three functions all fit this condition, where $I = [1, \infty)$ in the first case and $I = \mathbf{R}$ in the second and third cases:

$$f(x) = x + \frac{1}{x}, \quad f(x) = \sqrt{x^2 + 1}, \quad f(x) = \log(1 + e^x).$$

In each case, $f(x) > x$, so none of these functions has a fixed point in I. (One could consider ∞ to be a fixed point, and $f^n(x) \to \infty$ as $n \to \infty$. See [5] for a general theorem in this direction.)

Despite such examples, there is a fixed-point theorem when $d(f(x), f(x')) < d(x, x')$ for all $x \neq x'$ provided the space is *compact*, which is not the case in the previous example. This theorem, which is next, is due to Edelstein [7].

Theorem 3.5. *Let X be a compact metric space. If $f\colon X \to X$ satisfies $d(f(x), f(x')) < d(x, x')$ when $x \neq x'$ in X, then f has a unique fixed point in X and the fixed point can be found as the limit of $f^n(x_0)$ as $n \to \infty$ for any $x_0 \in X$.*

Proof. To show f has at most one fixed point in X, suppose f has two fixed points $a \neq a'$. Then $d(a, a') = d(f(a), f(a')) < d(a, a')$. This is impossible, so $a = a'$.

To prove f actually has a fixed point, we will look at the function $X \to [0, \infty)$ given by $x \mapsto d(x, f(x))$. This measures the distance between each point and its f-value. A fixed point of f is where this function takes the value 0.

Since X is compact, the function $d(x, f(x))$ takes on its minimum value: there is an $a \in X$ such that $d(a, f(a)) \leq d(x, f(x))$ for all $x \in X$. We'll show by contradiction that a is a fixed point for f. If $f(a) \neq a$ then the hypothesis about f in the theorem (taking $x = a$ and $x' = f(a)$) says

$$d(f(a), f(f(a))) < d(a, f(a)),$$

which contradicts the minimality of $d(a, f(a))$ among all numbers $d(x, f(x))$. So $f(a) = a$.

Finally, we show for any $x_0 \in X$ that the sequence $x_n = f^n(x_0)$ converges to a as $n \to \infty$. This can't be done as in the proof of the contraction mapping theorem since we don't have the contraction constant to help us out. Instead we will exploit compactness.

If for some $k \geq 0$ we have $x_k = a$ then $x_{k+1} = f(x_k) = f(a) = a$, and more generally $x_n = a$ for all $n \geq k$, so $x_n \to a$ since the terms of the sequence equal a for all large n. Now we may assume instead that $x_n \neq a$ for all n. Then

$$0 < d(x_{n+1}, a) = d(f(x_n), f(a)) < d(x_n, a),$$

so the sequence of numbers $d(x_n, a)$ is decreasing and positive. Thus it has a limit $\ell = \lim_{n\to\infty} d(x_n, a) \geq 0$. We will show $\ell = 0$ (so $d(x_n, a) \to 0$, which means $x_n \to a$ in X). By compactness of X, the sequence $\{x_n\}$ has a convergent subsequence x_{n_i}, say $x_{n_i} \to y \in X$. The function f is continuous, so $f(x_{n_i}) \to f(y)$, which says $x_{n_i+1} \to f(y)$ as $i \to \infty$. Since $d(x_n, a) \to \ell$ as $n \to \infty$, $d(x_{n_i}, a) \to \ell$ and $d(x_{n_i+1}, a) \to \ell$ as $i \to \infty$. By continuity of the metric, $d(x_{n_i}, a) \to d(y, a)$ and $d(x_{n_i+1}, a) = d(f(x_{n_i}), a) \to d(f(y), a)$. Having already shown these limits are ℓ,

$$d(y, a) = \ell = d(f(y), a) = d(f(y), f(a)). \tag{3.2}$$

If $y \neq a$ then $d(f(y), f(a)) < d(y, a)$, but this contradicts (3.2). So $y = a$, which means $\ell = d(y, a) = 0$. That shows $d(x_n, a) \to 0$ as $n \to \infty$. □

The proof of Edelstein's theorem does not yield any error estimate on the rate of convergence to the fixed point, since we proved convergence to the fixed point without having to make estimates along the way. Edelstein did not assume for his main theorem in [7] that X is compact, as in Theorem 3.5. He assumed instead that there is some $x_0 \in X$ whose sequence of iterates $f^n(x_0)$ has a convergent subsequence and he proved this limit is the unique fixed point of f and that the whole sequence $f^n(x_0)$ tends to that fixed point.

It is natural to wonder if the compactness of X might force f in Theorem 3.5 to be a contraction after all, so the usual contraction mapping theorem would apply. For instance, the ratios $d(f(x), f(x'))/d(x, x')$ for $x \neq x'$ are always less than 1, so they should be less than or equal to some definite constant $c < 1$ from compactness (a continuous real-valued function on a compact set achieves its supremum as a value). But this reasoning is bogus, because $d(f(x), f(x'))/d(x, x')$ is defined not on the compact set $X \times X$ but rather on $X \times X - \{(x, x) : x \in X\}$ where the diagonal is removed, and this is not compact (take a look at the special case $X = [0,1]$, for instance). There is no way to show f in Theorem 3.5 has to be a contraction since there are examples where it isn't.

Example 3.6. Let $f\colon [0,1] \to [0,1]$ by $f(x) = \frac{1}{1+x}$, so $|f(x)-f(y)| = |x-y|/|(1+x)(1+y)|$. When $x \neq y$ we have $|f(x)-f(y)|/|x-y| = 1/(1+x)(1+y) < 1$, so $|f(x)-f(y)| < |x-y|$ and $|f(x)-f(y)|/|x-y|$ gets arbitrarily close to 1 when x and y are sufficiently close to 0. Therefore f is not a contraction on $[0,1]$ with respect to the usual metric.

Theorem 3.5 says f has a unique fixed point in $[0,1]$ and $f^n(x_0)$ tends to this point as $n \to \infty$ for any choice of x_0. Of course, it is easy to find the fixed point: $x = 1/(1+x)$ in $[0,1]$ at $x = (-1+\sqrt{5})/2 \approx .61803$.

This example does not actually require Theorem 3.5. One way around it is to check that if $1/2 \leq x \leq 1$ then $1/2 \leq f(x) \leq 2/3$, so $f\colon [1/2,1] \to [1/2,1]$ with $\max_{x\in[1/2,1]} |f'(x)| = 4/9$. We could apply the contraction mapping theorem to f on $[1/2,1]$ to find the fixed point. A second method is to check that $f^2(x) = (1+x)/(2+x)$ is a contraction on $[0,1]$ (the derivative of $(1+x)/(2+x)$ on $[0,1]$ has absolute value at most $1/4$), so we could apply the contraction mapping theorem to f^2.

Example 3.7. Following [2], let $f\colon [0,1] \to [0,1]$ by $f(x) = x/(1+x)$. Since $|f(x) - f(y)|/|x-y| = 1/(1+x)(1+y)$ for $x \neq y$, the ratio $|f(x)-f(y)|/|x-y|$ for $x \neq y$ can be made arbitrarily close to 1 by taking x and y sufficiently close to 0. What makes this example different from the previous one is that, since 0 is now the fixed point, f does not restrict to a contraction on any neighborhood of its fixed point. Moreover, since $f^n(x) = x/(1+nx)$, for $x \neq y$ the ratio $|f^n(x)-f^n(y)|/|x-y| = 1/(1+nx)(1+ny)$ is arbitrarily close to 1 when x and y are sufficiently close to 0, so no iterate f^n is a contraction on any neighborhood of the fixed point in $[0,1]$.

A similar example on $[0,1]$ with 1 as the fixed point uses $g(x) = 1 - f(1-x) = 1/(2-x)$.

4. Differential Equations

As a significant application of the contraction mapping theorem, we will solve the initial value problem

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t,y), \quad y(t_0) = y_0 \tag{4.1}$$

for "nice" functions f. A solution to (4.1) is a differentiable function $y(t)$ defined on a neighborhood of t_0 such that

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t,y(t)), \quad y(t_0) = y_0.$$

Before stating the basic existence and uniqueness theorem for such initial value problems, we look at some examples to appreciate the scope of the theorem we will prove.

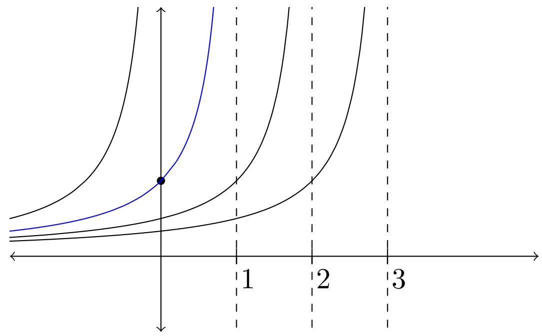
Example 4.1. The initial value problem

$$\frac{\mathrm{d}y}{\mathrm{d}t} = y^2, \quad y(0) = 1$$

has a unique solution passing through the point (0,1) that is found using separation of variables and integration:

$$y(t) = \frac{1}{1-t}.$$

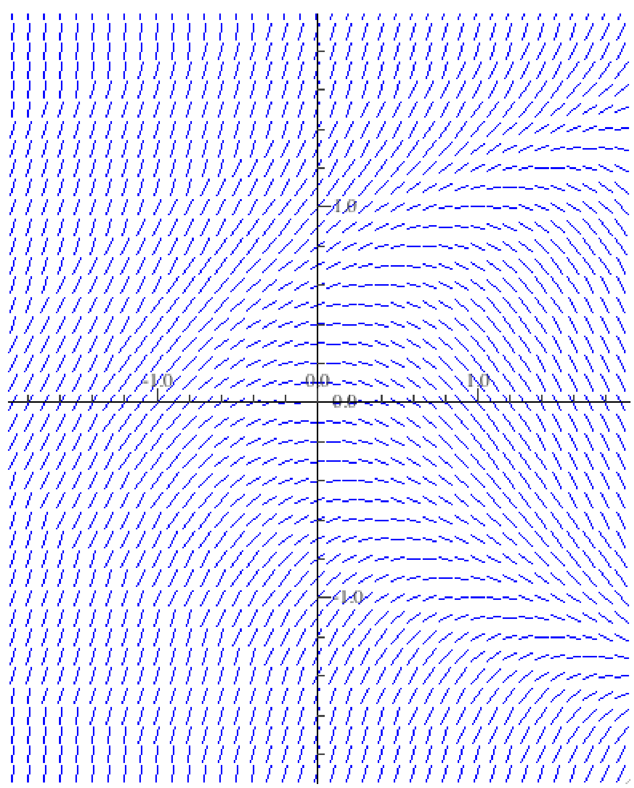
Note that although the right side of the differential equation makes sense everywhere, the solution blows up in finite time (at $t = 1$). See the blue curve in the figure below. Other curves in the figure are other solutions to $dy/dt = y^2$ with $y(0) > 0$ (having the form $y = 1/(c - t)$).



Example 4.2. The initial value problem

$$\frac{\mathrm{d}y}{\mathrm{d}t} = y^2 - t, \quad y(0) = 1$$

has a solution, but the solution can't be expressed in terms of elementary functions or their integrals (theorem of Liouville [14, p. 70 *ff.*]). Like the previous example, the solution blows up in finite time even though the right side makes sense everywhere. Take a look at the slope field digram below to be convinced visually of the finite blow-up time for the solution curve passing through $(0, 1)$; for a proof of this see Appendix A.



Example 4.3. The initial value problem

$$\frac{\mathrm{d}y}{\mathrm{d}t} = y^{2/3}, \quad y(0) = 0,$$

has two solutions: $y(t) = 0$ and $y(t) = t^3/27$.

These examples show that to solve (4.1) in any kind of generality, we will have to be satisfied with a solution that exists only locally (*i.e.*, not everywhere that f makes sense), and some constraint is needed on f to have a unique solution near t_0.

The condition we will introduce on $f(t, y)$ to guarantee a local unique solution is a Lipschitz condition in the second variable. That is, we will assume there is a constant $K > 0$ such that $|f(t, y_1) - f(t, y_2)| \leq K|y_1 - y_2|$ for all y_1 and y_2. The right side doesn't involve t, so this is a Lipschitz condition in the second variable that is uniform in the first variable.

Most functions are not globally Lipschitz. For instance, the single-variable function $h(x) = x^2$ doesn't satisfy $|h(a) - h(b)| \leq K|a - b|$ for some $K > 0$ and *all* a and b. But any function with continuous first derivative is locally Lipschitz:

Lemma 4.4. *Let $U \subset \mathbf{R}^2$ be open and $f\colon U \to \mathbf{R}$ be a C^1-function. Then it is locally Lipschitz in its second variable unifomly in the first: for any $(t_0, y_0) \in U$, there is a constant $K > 0$ such that*

$$|f(t, y_1) - f(t, y_2)| \leq K|y_1 - y_2|$$

for all t near t_0 and all y_1 and y_2 near y_0.

Proof. The open set U contains a rectangle around (t_0, y_0), say $I \times J$. Here I and J are closed intervals (of positive length) with $t_0 \in I$, $y_0 \in J$, and $I \times J \subset U$. Applying the mean-value theorem to f in its second variable, for any $t \in I$ and y_1 and y_2 in J

$$f(t, y_1) - f(t, y_2) = \frac{\partial f}{\partial y}(t, y_3)(y_1 - y_2)$$

for some y_3 between y_1 and y_2, depending possibly on t. Then

$$|f(t, y_1) - f(t, y_2)| = \left|\frac{\partial f}{\partial y}(t, y_3)\right| |y_1 - y_2| \leq \sup_{p \in I \times J} \left|\frac{\partial f}{\partial y}(p)\right| |y_1 - y_2|.$$

Let $K = \sup_{p \in I \times J} |(\partial f/\partial y)(p)|$, which is finite since a continuous function on a compact set is bounded. $\square$

Example 4.5. The function $f(t, y) = \sin(ty)$ is C^1 on all of $\mathbf{R}^2$ and is Lipschitz on any vertical strip $[-R, R] \times \mathbf{R}$: if $|t| \leq R$ then for any y_1 and y_2 in $\mathbf{R}$,

$$\sin(ty_1) - \sin(ty_2) = t\cos(ty_3)(ty_1 - ty_2)$$

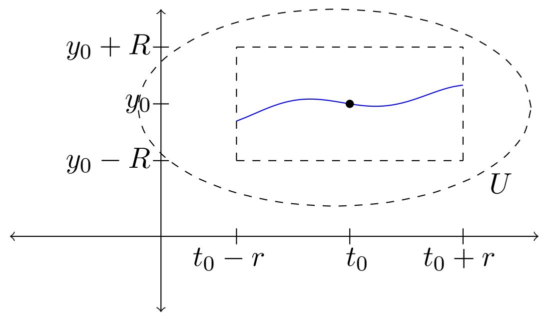
for some y_3 between y_1 and y_2. Therefore

$$|\sin(ty_1) - \sin(ty_2)| \leq t^2|y_1 - y_2| \leq R^2|y_1 - y_2|.$$

Example 4.6. The function $f(t, y) = y^{2/3}$ is not C^1 at $(0, 0)$ and in fact is not even Lipschitz in any neighborhood of $(0, 0)$: if $|y_1^{2/3} - y_2^{2/3}| \leq K|y_1 - y_2|$ for some $K > 0$ and all y_1, y_2 near 0 then taking $y_2 = 0$ shows $|y^{2/3}| \leq K|y|$ for all y near 0, but for $y \neq 0$ that's the same as $|y|^{-1/3} \leq K$, and $|y|^{-1/3}$ is unbounded as $y \to 0$.

Our next lemma controls the growth of a hypothetical solution to (4.1) near t_0.

Lemma 4.7. *Let $U \subset \mathbf{R}^2$ be open and $f\colon U \to \mathbf{R}$ be continuous. For any $(t_0, y_0) \in U$ there are $r, R > 0$ such that the rectangle $[t_0 - r, t_0 + r] \times [y_0 - R, y_0 + R]$ is in U and, for any $\delta \in (0, r]$, a solution $y(t)$ to (4.1) that is defined on $[t_0 - \delta, t_0 + \delta]$ must satisfy $|y(t) - y_0| \le R$ for $|t - t_0| \le \delta$.*



Proof. Suppose $y(t)$ satisfies (4.1) near t_0. For t near t_0, $(t, y(t)) \in U$ because U is open and $y(t)$ is continuous, so $f(t, y(t))$ makes sense. For t near t_0, $\int_{t_0}^{t} f(s, y(s))\,\mathrm{d}s$ has t-derivative $f(t, y(t))$, which is also the t-derivative of $y(t)$, so by (4.1) $y(t)$ and $\int_{t_0}^{t} f(s, y(s))\,\mathrm{d}s$ differ by a constant near t_0. At $t = t_0$ the difference is $y(t_0) - 0 = y_0$, so

$$y(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\,\mathrm{d}s \tag{4.2}$$

for t near t_0. Pick a rectangle

$$[t_0 - r, t_0 + r] \times [y_0 - R, y_0 + R]$$

in U that is centered at (t_0, y_0) and let $B > 0$ be an upper bound of $|f|$ on the rectangle. For t near t_0, (4.2) implies

$$|y(t) - y_0| \le B|t - t_0|. \tag{4.3}$$

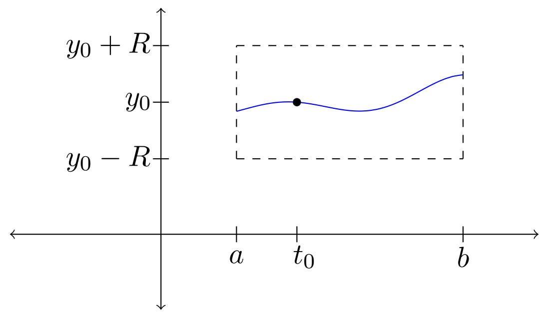
Notice B, r, and R are determined wholly by f and not by any solutions to (4.1).

Shrink r so that $Br \le R$, *i.e.*, replace r with $\min(r, R/B)$. (We could alternatively try to increase R to achieve this inequality, but increasing the size of the rectangle would usually change B and there is also the constraint that the rectangle should be inside of U.) Pick $\delta \in (0, r]$ and suppose there is a solution $y(t)$ to (4.1) for $|t - t_0| \le \delta$. In particular, $(t, y(t)) \in U$ for $|t - t_0| \le \delta$. From (4.3), $|y(t) - y_0| \le Br \le R$, so $(t, y(t))$ is in the rectangle for $|t - t_0| \le r$. □

Here is the key existence and uniqueness theorem for ODEs.

Theorem 4.8 (Picard)**.** *Let $\mathcal{R} = [a, b] \times [y_0 - R, y_0 + R] \subset \mathbf{R}^2$ be a rectangle and $f\colon \mathcal{R} \to \mathbf{R}$ be a continuous function on the rectangle that is Lipschitz in its second variable uniformly in its first variable. Let $B > 0$ be an upper bound on $|f|$ over this rectangle. For $t_0 \in [a, b]$, let $r = \min(|t_0 - a|, |t_0 - b|)$. The initial value problem (4.1) has a unique solution on the interval $[t_0 - r_0, t_0 + r_0]$, where $r_0 = \min(r, R/B)$.*

In particular, if $U \subset \mathbf{R}^2$ is open and $f\colon U \to \mathbf{R}$ is C^1 then (4.1) has a unique local solution passing through any point in its domain.



Proof. Set

$$M = \{h \in C[a,b] : |h(t) - y_0| \leq R \text{ for } t \in [a,b]\}.$$

The constant function $h(t) = y_0$ is in M, so $M \neq \emptyset$. For any $h \in M$, the graph $\{(t, h(t)) : t \in [a,b]\}$ lies in $\mathcal{R}$. Giving $C[a,b]$ the sup-norm $||\cdots||_{\sup}$, M is a closed subspace of $C[a,b]$. Since $C[a,b]$ is complete in the sup-norm, M is also complete in the sup-norm.

Considering the formula (4.2), let $F \colon M \to M$ by

$$(Fh)(t) := y_0 + \int_{t_0}^{t} f(s, h(s))\,\mathrm{d}s. \tag{4.4}$$

The integration makes sense, since $(s, h(s)) \in \mathcal{R}$ for all $s \in [a,b]$, and moreover

$$|(Fh)(t) - y_0| \leq B|t - t_0| \leq Br \leq R,$$

so $F(M) \subset M$. We will show the contraction mapping theorem can be applied to find a fixed point of F in M. A fixed point of F is a solution to (4.1) by the fundamental theorem of calculus.

The solutions of (4.1) and (4.2) are the same. The advantage of (4.2) over (4.1) is that continuous functions have continuous antiderivatives but they might not have a derivative; integration is a "smoother" operation on functions than differentiation.

To illustrate different techniques, we will show how the contraction mapping theorem can be applied to find a fixed point of F in *two* ways:

(1) an iterate F^n is a contraction on M for $n \gg 0$, so Theorem 3.1 applies,
(2) F is a contraction on M using a norm other than the sup-norm, but in which M is still complete.

Proof of (1): For $h_1, h_2 \in M$ and $t \in [a,b]$,

$$(Fh_1)(t) - (Fh_2)(t) = \int_{t_0}^{t} (f(s, h_1(s)) - f(s, h_2(s)))\,\mathrm{d}s,$$

so

$$|(Fh_1)(t) - (Fh_2)(t)| \leq \int_{|[t_0,t]|} |f(s, h_1(s)) - f(s, h_2(s))|\,\mathrm{d}s, \tag{4.5}$$

where $\int_{|[t_0,t]|}$ denotes $\int_{[t_0,t]}$ if $t_0 \leq t$ and $\int_{[t,t_0]}$ if $t < t_0$. We are assuming f is Lipschitz in its second variable uniformly in the first, say $|f(t, y_1) - f(t, y_2)| \leq K|y_1 - y_2|$ for (t, y_1) and

(t, y_2) in $\mathcal{R}$. Then

$$|(Fh_1)(t) - (Fh_2)(t)| \leq \int_{|[t_0,t]|} K|h_1(s) - h_2(s)|\,\mathrm{d}s \tag{4.6}$$

$$\leq K||h_1 - h_2||_{\sup}|t - t_0| \tag{4.7}$$

$$\leq K||h_1 - h_2||_{\sup}(b - a).$$

Thus $||Fh_1 - Fh_2||_{\sup} \leq K(b-a)||h_1 - h_2||_{\sup}$. If $K(b-a) < 1$ then F is a contraction on M. (This already gives us local existence and uniqueness of a solution by shrinking $[a, b]$ small enough around t_0 so that $b - a < 1/K$. But we want to show there is a solution on our original $[a, b]$, so we proceed in a different way.) To cover the case $K(b-a) \geq 1$, we look at how iterates of F shrink distances. Since

$$(F^2h_1)(t) - (F^2h_2)(t) = \int_{t_0}^{t} (f(s, (Fh_1)(s)) - f(s, (Fh_2)(s)))\,\mathrm{d}s,$$

we get

$$\begin{aligned} |(F^2h_1)(t) - (F^2h_2)(t)| &\leq \int_{|[t_0,t]|} K|(Fh_1)(s) - (Fh_2)(s)|\,\mathrm{d}s \quad \text{by the Lipschitz condition} \\ &\leq \int_{|[t_0,t]|} K \cdot K||h_1 - h_2||_{\sup}|s - t_0|\,\mathrm{d}s \quad \text{by (4.7)} \\ &= K^2||h_1 - h_2||_{\sup}\int_{|[t_0,t]|} |s - t_0|\,\mathrm{d}s. \end{aligned}$$

Since $\int_{|[t_0,t]|} |s - t_0|\,\mathrm{d}s = |t - t_0|^2/2$ (check this separately for $t \geq t_0$ and $t < t_0$), we have

$$|(F^2h_1)(t) - (F^2h_2)(t)| \leq K^2||h_1 - h_2||_{\sup}\frac{|t - t_0|^2}{2} = \frac{(K|t - t_0|)^2}{2}||h_1 - h_2||_{\sup}.$$

By induction, using the formula $\int_{|[t_0,t]|} |s - t_0|^n\,\mathrm{d}s = |t - t_0|^{n+1}/(n+1)$ for $n \geq 0$,

$$|(F^nh_1)(t) - (F^nh_2)(t)| \leq \frac{(K|t - t_0|)^n}{n!}||h_1 - h_2||_{\sup}$$

for all $n \geq 0$ and $t \in [a, b]$. Since $|t - t_0| \leq b - a$,

$$||F^nh_1 - F^nh_2||_{\sup} \leq \frac{(K(b-a))^n}{n!}||h_1 - h_2||_{\sup}.$$

When $n \gg 0$, $(K(b-a))^n/n! < 1$, so F^n is a contraction on M in the sup-norm. Thus F has a unique fixed point in M by Theorem 3.1.

Proof of (2): Define a new norm $||\cdot||$ on $C[a, b]$ by

$$||h|| := \sup_{t\in[a,b]} e^{-K|t-t_0|}|h(t)|,$$

where K is the Lipschitz constant for f on $\mathcal{R}$. To check this is a norm on $C[a, b]$, for any $t \in [a, b]$ we have

$$|h_1(t) + h_2(t)| \leq |h_1(t)| + |h_2(t)|,$$

so

$$e^{-K|t-t_0|}|h_1(t) + h_2(t)| \leq e^{-K|t-t_0|}|h_1(t)| + e^{-K|t-t_0|}|h_2(t)| \leq ||h_1|| + ||h_2||.$$

Taking the sup of the left side over all $t \in [a,b]$ shows $||\cdot||$ satisfies the triangle inequality. The other conditions to be a vector space norm are easily checked. How does $||\cdot||$ compare to the sup-norm on $C[a,b]$? For $t \in [a,b]$,

$$e^{-K(b-a)} \leq e^{-K|t-t_0|} \leq 1,$$

so for any $h \in C[a,b]$

$$e^{-K(b-a)}|h(t)| \leq e^{-K|t-t_0|}|h(t)| \leq |h(t)|.$$

Taking the sup over $t \in [a,b]$,

$$e^{-K(b-a)}||h||_{\text{sup}} \leq ||h|| \leq ||h||_{\text{sup}}. \tag{4.8}$$

Thus $||\cdot||$ and the sup-norm bound each other from above and below up to scaling factors, so these two norms have the same open sets and the same convergent sequences in $C[a,b]$. In particular, $C[a,b]$ and its subset M are both complete relative to $||\cdot||$ since they are complete with respect to the sup-norm.

Returning to (4.6), multiply both sides by $e^{-K|t-t_0|}$:

$$\begin{aligned}
e^{-K|t-t_0|}|(Fh_1)(t)-(Fh_2)(t)| &\leq e^{-K|t-t_0|}\int_{|[t_0,t]|} K|h_1(s)-h_2(s)|\,\mathrm{d}s \\
&= e^{-K|t-t_0|}\int_{|[t_0,t]|} Ke^{K|s-t_0|}e^{-K|s-t_0|}|h_1(s)-h_2(s)|\,\mathrm{d}s \\
&\leq e^{-K|t-t_0|}\int_{|[t_0,t]|} Ke^{K|s-t_0|}||h_1-h_2||\,\mathrm{d}s \\
&= ||h_1-h_2||e^{-K|t-t_0|}\int_{|[t_0,t]|} Ke^{K|s-t_0|}\,\mathrm{d}s.
\end{aligned}$$

To compute the integral in this last expression, we take cases depending on whether $t_0 \leq t$ or $t_0 > t$. If $t_0 \leq t$ then $|t-t_0| = t-t_0$ and $|s-t_0| = s-t_0$ for $t_0 \leq s \leq t$, so

$$\int_{|[t_0,t]|} Ke^{K|s-t_0|}\,\mathrm{d}s = \int_{t_0}^{t} Ke^{K(s-t_0)}\,\mathrm{d}s = e^{K(t-t_0)}-1.$$

If $t_0 > t$ then $|t-t_0| = t_0-t$ and $|s-t_0| = t_0-s$ for $t \leq s \leq t_0$, so

$$\int_{|[t_0,t]|} Ke^{K|s-t_0|}\,\mathrm{d}s = \int_{t}^{t_0} Ke^{K(t_0-s)}\,\mathrm{d}s = e^{K(t_0-t)}-1.$$

In either case the value is $e^{K|t-t_0|}-1$, so

$$e^{-K|t_0-t|}\int_{|[t_0,t]|} Ke^{K|s-t_0|}\,\mathrm{d}s = e^{-K|t-t_0|}(e^{K|t-t_0|}-1) = 1-e^{-K|t_0-t|}.$$

Therefore

$$e^{-K|t-t_0|}|(Fh_1)(t)-(Fh_2)(t)| \leq ||h_1-h_2||(1-e^{-K|t-t_0|}) \leq (1-e^{-K(b-a)})||h_1-h_2||.$$

Taking the supremum of the left side over all $t \in [a,b]$,

$$||Fh_1-Fh_2|| \leq (1-e^{-K(b-a)})||h_1-h_2||.$$

Since $K(b-a) > 0$, $1-e^{-K(b-a)} < 1$, so $F\colon M \to M$ is a contraction with respect to $||\cdot||$. Since M is complete with respect to $||\cdot||$, there is a unique fixed point of F in M by the contraction mapping theorem. □

The proof of Picard's theorem says that a solution to (4.1) near t_0 can be found as a limit of the sequence $\{y_n(t)\}$ where $y_0(t) = y_0$ is a constant function and

$$y_{n+1}(t) = y_0 + \int_{t_0}^{t} f(s, y_n(s))\,\mathrm{d}s.$$

These functions $y_n(t)$, defined recursively by integration, are called *Picard iterates*. Picard's proof of Theorem 4.8 appeared in 1890 [13] (30 years before Banach stated the general contraction mapping theorem), and the basic idea was already used in a special case by Liouville in 1837.

Remark 4.9. In the proof of Picard's theorem we introduced an integral operator F in (4.4) and gave two ways of showing it has a fixed point. The same ideas can be used to prove in two ways that for any continuous function $f \in C[a, b]$ the (Volterra) integral equation

$$y(t) = f(t) + \int_{a}^{t} k(t, s)y(s)\,\mathrm{d}s$$

has a unique solution $y \in C[a, b]$ for any given continuous f and k. See [11, pp. 75–76].

Corollary 4.10. *If $f\colon [a, b] \times \mathbf{R} \to \mathbf{R}$ is Lipschitz in its second variable uniformly in its first variable, then the differential equation (4.1) has a unique solution on $[a, b]$ given any initial condition at a point $t_0 \in [a, b]$.*

Proof. In the beginning of the proof of Picard's theorem, the business with R and shrinking the domain on which the differential equation is studied near t_0 was carried out to put ourselves in the position of having a domain where f is Lipschitz in its second variable uniformly in the first variable and such that we define M to contain only continuous functions whose graph remains in this domain for f. Therefore if we assume from the outset that f has the relevant Lipschitz condition and its second variable runs over $\mathbf{R}$, we can drop the preliminary part of the proof of Theorem 4.8 from consideration and just take $M = C[a, b]$. Define $F\colon M \to M$ as in (4.4). □

Example 4.11. The function $f(t, y) = \sin(ty)$ is Lipschitz in y uniformly in t on any set of the form $[-r, r] \times \mathbf{R}$, so the differential equation $\mathrm{d}y/\mathrm{d}t = \sin(ty)$ with $y(0) = 1$ has a unique solution on $[-r, r]$ for all r. Thus the solution y is defined on all of $\mathbf{R}$.

Remark 4.12. If we drop the local Lipschitz condition on f then we lose the uniqueness of the solution to (4.1), as we saw in Example 4.3, but we don't lose existence. It is a theorem of Peano that as long as $f(t, y)$ is continuous, (4.1) will have a solution curve passing through any point (t_0, y_0). In one method of proof [6, pp. 150–154], the basic idea is to express f as the uniform limit of polynomial functions $P_n(t, y)$ using the Stone-Weierstrass theorem. Polynomials are C^1, so by Picard's theorem (essentially the contraction mapping theorem), the initial value problems $y'(t) = P_n(t, y)$, $y(t_0) = y_0$, are each uniquely locally solvable near t_0. Denote the solutions as y_n. It can be shown, using a compactness theorem in spaces of continuous functions, that some subsequence of the y_n's converges to a solution to the original initial value problem $y' = f(t, y)$. For a second proof of Peano's theorem, using a different fixed-point theorem, see [20, Prop. 2.14].

Picard's theorem is applicable to more general first-order differential equations than (4.1), such as

$$t^2 + y^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = 1, \quad y(0) = \frac{1}{2}. \tag{4.9}$$

Setting $t = 0$ yields $1/4 + y'(0)^2 = 1$, so $y'(0) = \pm\sqrt{3}/2$. We have to make a choice of $y'(0)$ before we can pin down a particular solution, since the differential equation does not determine $y'(0)$. Taking $y'(0) = \sqrt{3}/2$, we want to solve

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \sqrt{1 - t^2 - y^2}, \quad y(0) = \frac{1}{2},$$

while taking $y'(0) = -\sqrt{3}/2$ leads to the equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\sqrt{1 - t^2 - y^2}, \quad y(0) = \frac{1}{2}.$$

Each of these has a unique solution near $t = 0$ using Picard's theorem.

More generally, a first-order ordinary differential equation looks like $g(t, y, y') = 0$ with some initial conditions $y(t_0) = y_0$ and $y'(t_0) = z_0$ (so $g(t_0, y_0, z_0) = 0$). Assuming $g(t, y, z)$ is a C^1-function of 3 variables, as long as $(\partial g/\partial z)(t_0, y_0, z_0) \neq 0$ the implicit function theorem tells us there is an open set $U \supset (t_0, y_0)$ in $\mathbf{R}^2$ and a C^1-function $f \colon U \to \mathbf{R}$ such that $z_0 = f(t_0, y_0)$ and for all (t, y) near (t_0, y_0), $g(t, y, z) = 0$ if and only if $z = f(t, y)$. Therefore the differential equation $g(t, y, y') = 0$ locally near (t_0, y_0, z_0) looks like $y' = f(t, y)$, which is exactly the kind of differential equation we solved locally and uniquely by Picard's theorem. So Picard's theorem implies local existence and uniqueness of fairly general first-order ODEs.

Picard's theorem generalizes [1, Chap. 4] to a first-order ordinary differential equation with a vector-valued function:

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = \mathbf{f}(t, \mathbf{y}(t))$$

where $\mathbf{f}$ and $\mathbf{y}$ have values in $\mathbf{R}^n$ and $\mathbf{f}$ is C^1. Essentially the only change needed to extend the proof of Picard's theorem from the 1-dimensional case to higher dimensions is the use of an integral (rather than differential) form of the mean-value theorem to show a C^1-function is locally Lipschitz.

Appendix A. A Differential Inequality

In Example 4.2 it was stated that the initial value problem $y'(t) = y(t)^2 - t$, $y(0) = 1$ has a solution that blows up in finite time. To estimate the blow-up time, let $Y(t) = 1/y(t)$ and see where $Y(t) = 0$. From $Y'(t) = tY(t)^2 - 1$ and $Y(0) = 1$, a computer algebra package has $Y(t) = 0$ at $t \approx 1.125$.

Theorem A.1. *The solution to $y'(t) = y(t)^2 - t$ satisfying $y(0) = 1$ is undefined somewhere before $t = 1.221$.*

This is weaker than what numerics suggest (*i.e.*, the blow-up time is around 1.125), but proving something sharper requires a more careful analysis than we wish to develop. An expert in non-linear ODEs told me that in practice nobody tries to prove very sharp approximations for blow-up times (mere existence of a blow-up time usually suffices).

Proof. We know $y(t)$ is defined for small $t > 0$. Assume $y(t)$ is defined for $0 \leq t < c$. We will show for a suitable c that $y(t) \geq c/(c - t)$ for $0 \leq t < c$, so $y(t) \to \infty$ as $t \to c^+$. Therefore $y(t)$ has to be undefined for some $t \leq c$.

Set $z(t) = c/(c-t)$, with c still to be determined, so

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}(y-z) &= y^2 - t - \frac{\mathrm{d}z}{\mathrm{d}t} \\
&= y^2 - t - \frac{c}{(c-t)^2} \\
&= y^2 - z^2 + \left(1 - \frac{1}{c}\right) z^2 - t \\
&= (y-z)(y+z) + \frac{(c-1)c}{(c-t)^2} - t
\end{aligned}$$

By calculus, $(c-1)c/(c-t)^2 - t \geq 0$ for $0 \leq t < c$ as long as $c - 1 \geq (4/27)c^2$, which happens for c between the two roots of $x - 1 = (4/27)x^2$. The roots are approximately 1.2207 and 5.5292. So taking $c = 1.221$, we have $(y(t) - z(t))' \geq (y(t) - z(t))(y(t) + z(t))$ for $0 \leq t < c$. Using an integrating factor, this differential inequality is the same as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-\int_0^t (y(s)+z(s))\,\mathrm{d}s}(y(t) - z(t))\right) \geq 0.$$

Since $e^{-\int_0^t (y(s)+z(s))\,\mathrm{d}s}(y(t) - z(t))|_{t=0} = 0$, $e^{-\int_0^t (y(s)+z(s))\,\mathrm{d}s}(y(t) - z(t)) \geq 0$ for $t \geq 0$, so $y(t) - z(t) \geq 0$ because the exponential factor is positive. Thus $y(t) \geq z(t) = c/(c-t)$. □

References

[1] V. I. Arnol'd, "Ordinary Differential Equations," Springer-Verlag, New York, 1992.

[2] D. G. Bennett and B. Fisher, *On a Fixed Point Theorem for Compact Metric Spaces*, Math. Magazine **47** (1974), 40–41.

[3] C. Bessaga, *On the Converse of the Banach Fixed-point Principle*, Colloq. Math. **7** (1959), 41–43.

[4] V. Bryant, "Metric Spaces: Iteration and Application," Cambridge Univ. Press, Cambridge, 1985.

[5] A. F. Beardon, *Contractions on the real line*, Amer. Math. Monthly **113** (2006), 557–558

[6] R. C. Buck (ed.), "Studies in Modern Analysis, Vol. 1," Prentice-Hall, Englewood Cliffs, NJ, 1962.

[7] M. Edelstein, *On Fixed and Periodic Points Under Contractive Mappings*, J. London Math. Soc. **37** (1962), 74–79.

[8] M. Freeman, *The Inverse as a Fixed Point in Function Space*, Amer. Math. Monthly **83** (1976), 344–348.

[9] D. J. Higham and A. Taylor, The Sleekest Link Algorithm, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.159.831&rep=rep1&type=pdf.

[10] J. E. Hutchinson, *Fractals and Self-Similarity*, Indiana Univ. Math. J. **30** (1981), 713–747.

[11] A. N. Kolmogorov and S. Fomin, "Introductory Real Analysis," Dover, New York, 1970.

[12] A. Langville and C. Meyer, "Google's PageRank and Beyond: The Science of Search Engine Rankings," Princeton Univ. Press, Princeton, 2006.

[13] E. Picard, *Mémoire sur la théorie des équations aux dérivées partielles et la méthode des approximations successives*, J. Math. Pures et Appl. **6** (1890), 145–210.

[14] J. F. Ritt, "Integration in Finite Terms," Columbia Univ. Press, New York, 1948.

[15] W. Rudin, "Principles of Mathematical Analysis," 3rd ed., McGraw–Hill, New York, 1976.

[16] E. Schechter, "Handbook of Analysis and its Applications," Academic Press, San Diego, 1997.

[17] C. Wagner, *A Generic Approach to Iterative Methods*, Math. Mag. **55** (1982), 259–273.

[18] W. Walter, *An Elementary Proof of the Cauchy–Kowalevsky Theorem*, Amer. Math. Monthly **92** (1985), 115–126.

[19] B. White, How Google Ranks Web Pages, http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=6CDD52D0B67B14A2248967A5164A5981?doi=10.1.1.498.5678.

[20] E. Zeidler, "Nonlinear Functional Analysis and its Applications I: Fixed-Point Theorems," Springer-Verlag New York, 1986.