and the choice of point x. An important consequence of uniform convergence is the next lemma.

Lemma 3.2. *The limit of a uniformly convergent sequence of continuous functions is continuous.*

Proof. Let $u(x)$ denote the limit of $u_n(x)$; we must show that there is a $\delta(\varepsilon, x)$ such that $|u(y) - u(x)| < \varepsilon$ whenever $|y - x| < \delta$. Insert four new terms that sum to zero into this norm:

$$\begin{aligned}|u(y) - u(x)| &= |u(y) - u_n(y) + u_n(y) - u_n(x) + u_n(x) - u(x)| \\ &\leq |u(y) - u_n(y)| + |u_n(y) - u_n(x)| + |u_n(x) - u(x)| .\end{aligned}$$

Since by assumption u_n converges uniformly, then for any $x \in E$ and any $\varepsilon/3$ there is a given N such that $|u_n(x) - u(x)| < \varepsilon/3$ whenever $n > N$. Moreover, since u_n is continuous for any fixed n, there is a $\delta(\varepsilon, x)$ such that $|u_n(x) - u_n(y)| < \varepsilon/3$ for each $y \in \text{int}(B_\delta(x))$. As a consequence,

$$|u(y) - u(x)| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,$$

so u is continuous. □

There is also a uniform version of continuity:

▷ *uniform continuity*: A function f is uniformly continuous on E if for every $x \in E$ and every $\varepsilon > 0$, there is a $\delta(\varepsilon)$, independent of x, such that $|f(y) - f(x)| < \varepsilon$ whenever $|y - x| < \delta$.

It is not too hard to show that when E is a compact set, then every continuous function on E is also uniformly continuous (see Exercise 2).

A generalization of Lemma 3.2 is easily obtained: if each of the elements of a uniformly convergent sequence is uniformly continuous, then the limit is also uniformly continuous.

3.2 Function Space Preliminaries

A function $f : D \to R$ is a map from its domain D to its range R; that is, given any point $x \in D$, there is a unique point $y \in R$, denoted $y = f(x)$. In our applications the domain is often a subset of Euclidean space, $E \subset \mathbb{R}^n$, and the range is $\mathbb{R}^n$; in this case, $f : E \to \mathbb{R}^n$ is given by n components $f_i(x_1, x_2, \ldots, x_n), i = 1, 2, \ldots, n$. The set of functions denoted $C(E)$ or $C^0(E)$ consists of those functions on the domain E whose components are continuous. Colloquially we say "f is C^0" if it is a member of this set. If it is necessary to distinguish different ranges, the set of continuous functions from D to R is denoted $C^0(D, R)$; the second argument is often omitted if it is obvious. When E is an open subset of $\mathbb{R}^n$, a function $f : E \to \mathbb{R}^n$ is differentiable at a point $x \in E$ if there exists an $n \times n$ matrix $Df(x)$ such that

$$\lim_{|h| \to 0} \frac{1}{|h|} |f(x + h) - f(x) - Df(x)h| = 0 \tag{3.2}$$

When it exists, this matrix is unique and is called the *Jacobian matrix*

$$Df(x) \equiv \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix} . \tag{3.3}$$

Conversely, Df exists at x when all its partial derivatives, $\partial f_i / \partial x_j$, exist *and* are continuous in a neighborhood of x. We say f is $C^1(E)$—continuously differentiable—if the elements of $Df(x)$ are continuous on the open set E. Colloquially we will say that f is *smooth* when it is a C^1 function of its arguments.

Spaces of functions, like $C(E)$ and $C^1(E)$, are examples of infinite dimensional linear spaces, or *vector spaces*. Just as for ordinary vectors (recall §2.1), linearity means that whenever f and $g \in C(E)$, then so is $c_1 f + c_2 g$ for any (real) scalars c_1 and c_2. Much of our theoretical analysis will depend upon convergence properties of sequences of functions in some such space. To talk about convergence it is necessary to define a *norm* on the space; such norms will be denoted by $\|f\|$ to distinguish them from the finite dimensional Euclidean norm $|x|$. We already met one such norm, the operator norm, in (2.23). For continuous functions, the *supremum* or *sup-norm*, defined by

$$\|f\| \equiv \sup_{x \in E} |f(x)| \tag{3.4}$$

will often be used. For example, if $E = \mathbb{R}$, and $f = \tanh(x)$, then $\|f\| = 1$. Other norms include the L_p norms,

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

but these will not have much application in this book. This formula becomes the sup-norm in the limit $p \to \infty$, which is why the sup-norm is also called the L_∞ norm and is often denoted $\|f\|_\infty$.

Metric Spaces

A normed space is an example of a metric space. A metric is a distance function $\rho(f, g)$ that takes as arguments two elements of the space and returns a real number, the "distance" between f and g. A metric must satisfy the three properties

1. $\rho(f, g) \geq 0$, and $\rho(f, g) = 0$ only when $f \equiv g$ (positivity),
2. $\rho(f, g) = \rho(g, f)$ (symmetry), and
3. $\rho(f, h) \leq \rho(f, g) + \rho(g, h)$ (triangle inequality).

Associated with any norm $\|f\|$ is a metric defined by $\rho(f,g) = \|f-g\|$. Therefore, a normed vector space is also a metric space; however, metric spaces need not be vector spaces, since in a metric space there is not necessarily a linear structure.

A sequence of functions f_n that are elements of a metric space X is said to *converge* to f^* if $\rho(f_n, f^*) \to 0$ as $n \to \infty$. Since the distance $\rho(f_n, f^*)$ is simply a number, the usual definition of limit can be used for this convergence. Note that the norm (3.4) bounds the Euclidean distance: if we use

$$\rho(f,g) = \|f-g\|_\infty\,, \text{ then } |f(x)-g(x)| \le \rho(f,g).$$

Thus, convergence of a sequence of functions f_n in norm implies that the sequence of points $f_n(x)$ converges *uniformly*.

Another notion often used to discuss convergence is that of

▷ *Cauchy sequence*: Given a metric space X with metric ρ, a sequence $f_n \in X$ is Cauchy if, for every $\varepsilon > 0$ there is an $N(\varepsilon)$ such that $\rho(f_n, f_m) < \varepsilon$ whenever $m, n \ge N(\varepsilon)$.

Informally, a Cauchy sequence satisfies

$$\rho(f_n, f_m) \to 0 \text{ as } m, n \to \infty,$$

where m and n approach infinity independently. One advantage of this idea is that the value of the limit of a sequence need not be known in order to check if it is Cauchy.

It is easy to see that every convergent sequence is a Cauchy sequence. However, it is not necessarily true that every Cauchy sequence converges.

Example: Consider the sequence of functions $f_n(x) = \sin(nx)/n \in C[0,\pi]$, the continuous functions on the interval $[0,\pi]$. This sequence converges to $f^* = 0$ in the sup norm because

$$\|f_n - 0\| = \frac{1}{n} \to 0.$$

The sequence is also Cauchy because

$$\|f_m - f_n\| \le \frac{1}{n} + \frac{1}{m} \le \frac{2}{N} < \frac{3}{N} \quad \forall m, n \ge N.$$

Thus for any ε, we may choose $N(\varepsilon) = 3/\varepsilon$ so that the difference is smaller than ε. ∎

Example: Consider the sequence $f_n = \sum_{j=1}^{n} \frac{x^j}{j}$ of functions in $C(-1,1)$. Assuming that $m > n$, then

$$\|f_m - f_n\| = \left\| \sum_{j=n+1}^{m} \frac{x^j}{j} \right\| = \sum_{j=n+1}^{m} \frac{1}{j} \ge \int_n^m \frac{dy}{y+1} = \ln\left(\frac{m+1}{n+1}\right),$$

since the supremum of $\left|x^j\right|$ on $(-1,1)$ is 1. This does not go to zero for m and n arbitrarily large but otherwise independent. For example, selecting $m = 2N+1$ and $n = N$ gives a difference larger than $\ln 2$. Consequently, the sequence is not Cauchy.

Note that for any fixed $x \in (-1,1)$ this sequence converges to the function $-\ln(1-x)$; however, it does not converge uniformly since the number of terms needed to obtain an accuracy ε depends upon x. Thus in the sense of the L_∞ norm, the sequence does not converge on $C(-1,1)$. ∎

A space X that is nicely behaved with respect to Cauchy sequences is called a

▷ *complete space*: A metric space X is complete if every Cauchy sequence in X converges to an element of X.

For the case of linear spaces a complete space is called a

▷ *Banach space*: A complete normed linear space is a Banach space.

Some spaces, like a closed interval with the Euclidean norm, are complete, and some, like an open interval, are not. The space $C(E)$ with the L_∞ norm is complete when E is compact.[15] However, the continuous functions are not complete in the L_2-norm.

Example: Let $f_n \in C[-1,1]$ be the sequence

$$f_n = \begin{cases} 1, & x \le 0, \\ \dfrac{1}{1+nx}, & x > 0. \end{cases} \tag{3.5}$$

With the L_2-norm, this sequence limits to the function $f = \begin{cases} 1, & x \le 0 \\ 0, & x > 0 \end{cases}$ because

$$\|f_n - f\|_2 = \left(\int_0^1 \frac{dx}{(1+nx)^2} \right)^{1/2} = \frac{1}{\sqrt{1+n}} \underset{n\to\infty}{\to} 0.$$

Note that the limit, however, is not in $C[-1,1]$. In the L_2-norm, the sequence is also a Cauchy sequence:

$$\begin{aligned} \|f_n - f_m\|_2^2 &= \int_0^1 \left(\frac{1}{1+nx} - \frac{1}{1+mx} \right)^2 dx \le \int_0^1 \left[\left(\frac{1}{1+nx} \right)^2 + \left(\frac{1}{1+mx} \right)^2 \right] dx \\ &= \frac{1}{1+n} + \frac{1}{1+m} \le \frac{2}{N}, \end{aligned}$$

for any $n, m \ge N$—of course every convergent sequence is Cauchy. As a consequence, the L_2-norm is not complete on the space $C[-1,1]$. ∎

Example: Now consider the sequence (3.5) with the sup-norm. In this case the sequence does not converge to f, since

$$\|f_n - f\| = \max\left\{ |1-1|, \sup_{x\in(0,1]} \left| \frac{1}{1+nx} \right| \right\} = \max\{0,1\} = 1.$$

[15] The nontrivial proof is given in (Friedman, 1982) and (Guenther and Lee, 1996).

Accordingly, the very definition of convergence can depend upon the choice of norm. Moreover, this sequence is not Cauchy in the sup-norm:

$$\|f_n - f_m\| = \sup_{x\in[0,1]} \left| \frac{1}{1+nx} - \frac{1}{1+mx} \right| = \sup_{x\in[0,1]} \left| \frac{m-n}{(1+nx)(1+mx)} x \right| .$$

Differentiation of this expression shows that its maximum occurs at $x = (mn)^{-1/2}$ and has the value $\|f_n - f_m\| = \left|\frac{\sqrt{m}-\sqrt{n}}{\sqrt{m}+\sqrt{n}}\right|$ that does not approach zero for all $m, n \geq N \to \infty$. For example, $\|f_{4N} - f_N\| = \frac{1}{3}$. This proves that the sequence is not Cauchy. ∎

Since complete spaces are so important, it is worthwhile noting that given one such space we can construct more of them by taking subsets, as in the next lemma.

Lemma 3.3. *A closed subset of a complete metric space is complete.*

Proof. To see this, first note that if $f_j \in Y \subset X$ is a Cauchy sequence on a complete space X, then $f_j \to f^* \in X$. Moreover, since f is a limit point of the sequence f_j, and a closed set Y includes all of its limit points, then $f \in Y$. □

The issues that we have discussed are rather subtle and worthy of a second look—see Exercises 1-3.

Contraction Maps

We have already used the concept of an operator, or map, $T : X \to X$, from a metric space to itself in Chapter 2: an $n \times n$ matrix is a map from $\mathbb{R}^n$ to itself. We will have many more occasions to use maps in our study of dynamical systems, including the proof of the existence and uniqueness theorem in §3.3. This proof will rely heavily on what is perhaps the most important theorem in all of analysis, Stefan Banach's 1922 fixed-point theorem.

Theorem 3.4 (Contraction Mapping). *Let $T : X \to X$ be a map on a complete metric space X. The map T is a contraction if there exists a constant $c < 1$ such that for all $f, g \in X$,*

$$\rho\,(T(f), T(g)) \leq c\rho(f, g). \tag{3.6}$$

In this case T has a unique fixed point, $f^* = T(f^*) \in X$.

Proof. The result will be obtained iteratively. Choose an arbitrary $f_0 \in X$. Define the sequence $f_{n+1} = T(f_n)$. We wish to show that f_n is a Cauchy sequence. Applying (3.6) repeatedly yields

$$\rho(f_{n+1}, f_n) = \rho\,(T(f_n), T(f_{n-1})) \leq c\rho(f_n, f_{n-1}) \leq c^2\rho(f_{n-1}, f_{n-2}) \leq \cdots \leq c^n\rho(f_1, f_0).$$

Therefore, for any integers $m > n$, the triangle inequality implies that

$$\rho(f_m, f_n) \leq \sum_{i=n}^{m-1} \rho(f_{i+1}, f_i) \leq \sum_{i=n}^{m-1} c^i \rho(f_1, f_0) = \frac{1-c^{m-n}}{1-c} c^n \rho(f_1, f_0) \leq Kc^n,$$

where $K = \rho(f_1, f_0)/(1-c)$. Since $c < 1$, then for any $\varepsilon > 0$ there is an N such that for all $m, n \geq N$, $\rho(f_m, f_n) \leq Kc^N < \varepsilon$. This implies that the sequence f_n is Cauchy and, since X is complete, that the sequence converges.

The limit, f^*, is a fixed point of T. Indeed, suppose that N is large enough so that $\rho(f_n, f^*) < \varepsilon$ for all $n > N$, then

$$\begin{aligned}\rho(T(f^*), f^*) &\leq \rho(T(f^*), f_{n+1}) + \rho(f_{n+1}, f^*)\\ &= \rho(T(f^*), T(f_n)) + \rho(f_{n+1}, f^*) < (c+1)\varepsilon.\end{aligned}$$

Because this is true for any ε, the distance is zero and $T(f^*) = f^*$.

Finally, we show that the fixed point is unique. Suppose to the contrary that there are two fixed points $f \neq g$. Then, $\rho(f, g) = \rho(T(f), T(g)) \leq c\rho(f, g)$. Since $c < 1$, this is impossible unless $\rho(f, g) = 0$, but this contradicts the assumption $f \neq g$; thus, the fixed point is unique. □

Example: Consider the space $C^0(\mathbb{S})$ of continuous functions on the circle with circumference one, i.e., continuous functions that are periodic with period one: $f(x+1) = f(x)$. For any $f \in C^0(\mathbb{S})$ define the operator

$$T(f)(x) = \frac{1}{2} f(2x).$$

Note that $T(f) \in C^0(\mathbb{S})$, and, using the sup-norm, that $\|T(f) - T(g)\| = {}^1\!/_2 \|f - g\|$; therefore, T is a contraction map on $C^0(\mathbb{S})$. What is its fixed point? According to the theorem, any initial function will converge to the fixed point under iteration. For example, let $f_0(x) = \sin(2\pi x)$. Then $f_1(x) = {}^1\!/_2 \sin(4\pi x)$, and after n steps, $f_n = \frac{1}{2^n}\sin(2^{n+1}\pi x)$. A previous example showed that this sequence converges to $f^* = 0$ in the sup-norm. In conclusion, $f^* = 0$ is the unique fixed point. ■

Example: As a slightly more interesting example, consider the same function space but let

$$T(f)(x) = \cos(2\pi x) + \frac{1}{2} f(2x). \tag{3.7}$$

Note that T decreases the sup-norm by a factor of ${}^1\!/_2$ as before, so it is still contracting. For example, the sequence starting with the function $f_0(x) = \sin(2\pi x)$ is

$$\begin{aligned}f_1(x) &= \cos(2\pi x) + \frac{1}{2}\sin(4\pi x),\\ f_2(x) &= \cos(2\pi x) + \frac{1}{2}\cos(4\pi x) + \frac{1}{4}\sin(8\pi x),\\ f_j(x) &= \sum_{n=0}^{j-1} \frac{\cos(2^{n+1}\pi x)}{2^n} + \frac{1}{2^j}\sin(2^{j+1}\pi x).\end{aligned}$$

The last term goes to zero in the sup-norm, and by the contraction-mapping theorem, the result is guaranteed to be unique and continuous. The fixed point is not an elementary function but is easy to graph; see Figure 3.1; it is an example of a Weierstrass function (Falconer 1990). ■

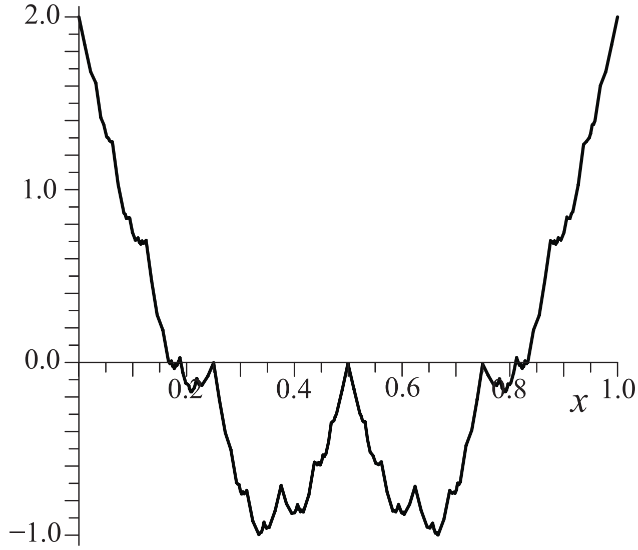

Figure 3.1. *The fixed point of the operator* (3.7).

Lipschitz Functions

Another ingredient that we will need in the existence and uniqueness theorem is a notion that is stronger than continuity but slightly less stringent than differentiability:

> ▷ *Lipschitz*: Suppose (X, ρ_X) and (Y, ρ_Y) are metric spaces with the indicated distance functions. A function $f : X \to Y$ is Lipschitz if for all $x_1, x_2 \in X$, there is a K such that
>
> $$\rho_Y(f(x_1), f(x_2)) \le K\rho_X(x_1, x_2). \tag{3.8}$$
>
> The smallest such K is called the Lipschitz constant for f on X.

For example if $X = Y = \mathbb{R}^2$ and $\rho_X = \rho_Y$ is the Euclidean metric, then when f is Lipschitz, the slope of the chord connecting $(x, f(x))$ and $(y, f(y))$ is at most K in absolute value. More generally the graph of a Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^n$ must be contained in the cone $\{(\xi, \eta) : |\eta - f(x)| \le K|\xi - x|\}$, for each vertex $(x, f(x))$.

The Lipschitz property implies more than continuity, but less than differentiability.

Lemma 3.5. *A Lipschitz function is uniformly continuous.*

Proof. For each ε, set $\delta = \varepsilon/K$. Then whenever $\rho_X(x_1, x_2) \le \delta$, (3.8) implies that $\rho_Y(f(x_1), f(x_2)) \le \varepsilon$. Consequently f is continuous at x_1, say, and moreover, is uniformly so because δ is chosen independently of x_1. □

If the space X is unbounded, then the assumption that f is Lipschitz is very strong. For example, $f = x^2$ is not Lipschitz on $\mathbb{R}$, even though it is Lipschitz on every bounded interval (a, b). A weaker notion is

> ▷ *locally Lipschitz*: $f : X \to Y$ is locally Lipschitz if for every $x \in X$, there is a neighborhood $N(x)$ such that f is Lipschitz on $N(x)$.

Note that for a locally Lipschitz function, the constant K can vary with the point and indeed may become arbitrarily large. Nevertheless, the restriction of any such function to a compact set is (globally) Lipchitz.

Lemma 3.6. *Suppose that $f : X \to Y$ is locally Lipschitz and $A \subset X$ is compact. Then f is Lipschitz on A.*

Proof. By assumption, for each $x_j \in A$, there is a ball $B_{r_j}(x_j)$ on which f is Lipschitz with constant K_j. Since A is compact, there is a finite collection of these balls—even if we decrease their radii by a factor of two—that covers A, i.e., $A \subset \bigcup_{j=1}^n B_{r_j/2}(x_j)$. Since n is finite, there exist positive constants $K = \max_j K_j$ and $\delta = \frac{1}{2} \min_j r_j$. To show f is Lipschitz on A, consider two points $\xi, x \in A$.

First, if $\rho_X(\xi, x) \le \delta$ then there is j such that $\xi, x \in B_{r_j}(x_j)$; indeed there is a j for which $x \in B_{r_j/2}(x_j)$ since these balls cover A, and the triangle inequality then implies

$$\rho_X(\xi, x_j) \le \rho_X(\xi, x) + \rho_X(x, x_j) \le \delta + \frac{r_j}{2} + \le r_j.$$

In this case, we have $\rho_Y(f(\xi), f(x)) \le K\rho_X(\xi, x)$, by the local Lipschitz assumption.

On the other hand, if $\rho_X(\xi, x) > \delta$, then we argue as follows. Since A is compact and f is continuous, there is an M such that $\rho_Y(f(\xi), f(x)) \le M$. Setting $\hat{K} = \max(K, M/\delta)$, we now have $\rho_Y(f(\xi), f(x)) \le \hat{K}\delta \le \hat{K}\rho_X(\xi, x)$. Thus f is Lipschitz on A with constant $\hat{K}$. □

When a function is differentiable on an open set in $\mathbb{R}^n$, it is locally Lipschitz. This is a simple consequence of the following lemma.

Lemma 3.7. *Suppose that $A \subset \mathbb{R}^n$ is compact and convex and $f \in C^1(A, \mathbb{R}^n)$. Then f is Lipschitz with constant $K = \max_{x \in A} \|Df\|$.*

Proof. Since A is convex, the points on a line between any two points $x, y \in A$, are also in A. Accordingly, $\xi(s) = x + s(y - x) \in A$ when $0 \le s \le 1$. Therefore

$$f(y) - f(x) = \int_0^1 \frac{d}{ds}\left(f(\xi(s))\right) ds = \int_0^1 Df\left(\xi(s)\right)(y - x) ds,$$

which amounts to a mean value theorem. Since A is compact and the norm of the Jacobian $\|Df\|$ is continuous, it has a maximum value K, as defined in the lemma. Thus

$$|f(y) - f(x)| \le \int_0^1 \|Df(\xi(s))\| \, |y - x| \, ds \le K |y - x| . \tag{3.9}$$

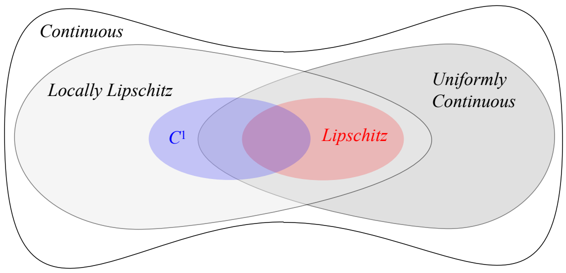

Figure 3.2. *Venn diagram relating continuity, differentiability and the Lipschitz property.*

This is exactly the promised condition. □

Corollary 3.8. *If $E \subset \mathbb{R}^n$ is open and $f \in C^1(E, \mathbb{R}^n)$, then f is locally Lipschitz.*

Proof. For any $x \in E$, there is an r such that $B_r(x) \subset E$. Since $B_r(x)$ is compact and convex, then Lemma 3.7 applies. □

Some of the relationships between continuous, Lipschitz, and smooth functions are summarized in Figure 3.2.

Example: The function $f(x) = |x|$ is continuous and Lipschitz on $\mathbb{R}$. It is obviously C^1 on $\mathbb{R}^+$ and $\mathbb{R}^-$, and if x and y have the same sign, then $|f(x) - f(y)| = |x - y|$. So the only thing to be checked is the Lipschitz condition when the points have the opposite sign. Although this is obvious geometrically, let us be formal: let $x > 0 > y$; then $|f(x) - f(y)| = ||x| - |y|| \leq x + |y| = |x - y|$. So f is Lipschitz with $K = 1$.

However, the function $f(x) = x^{1/2}$ is not Lipschitz on $[0, 1]$ even though it is uniformly continuous. For example, choosing $x = 4\varepsilon$, $y = \varepsilon$, we then have

$$|f(x) - f(y)| = \sqrt{x} - \sqrt{y} = \sqrt{\varepsilon} = \frac{\varepsilon}{\sqrt{\varepsilon}} = \frac{1}{\sqrt{\varepsilon}} \frac{4\varepsilon - \varepsilon}{3} = \frac{1}{3\sqrt{\varepsilon}} |x - y| \,,$$

so that as ε becomes small, the needed value of $K \to \infty$. ■

All these formal definitions have been given to provide us with the tools to prove that solutions to certain ODEs exist and, if the initial values are given, are unique. We are finally ready to begin this analysis.

3.3 Existence and Uniqueness Theorem

Before we can begin to study properties of the solutions of differential equations, we must discover if there *are* solutions in the first place: do solutions exist? The foundation of the theory of differential equations is the theorem proved by the French analyst Charles Emile Picard in 1890 and the Finnish topologist Ernst Leonard Lindelöf in 1894 that guarantees