Proposition (Newton's approximation). *Let X be a subset of $\mathbf{R}$, let $x_0 \in X$ be a limit point of X, let $f : X \to \mathbf{R}$ be a function, and let L be a real number. Then the following statements are logically equivalent:*

(a) *f is differentiable at x_0 on X with derivative L.*

(b) *For every $\varepsilon > 0$, there exists a $\delta > 0$ such that $f(x)$ is $\varepsilon|x - x_0|$-close to $f(x_0) + L(x - x_0)$ whenever $x \in X$ is δ-close to x_0, i.e., we have*

$$|f(x) - (f(x_0) + L(x - x_0))| \leq \varepsilon|x - x_0|$$

whenever $x \in X$ and $|x - x_0| \leq \delta$.

Theorem There exists a real continuous function on the real line which is nowhere differentiable.

Proof

Define

$$\varphi(x) = |x| \qquad (-1 \le x \le 1)$$

and extend the definition of $\varphi(x)$ to all real x by requiring that

$$\varphi(x+2) = \varphi(x).$$

Then, for all s and t,

$$|\varphi(s) - \varphi(t)| \le |s - t|. \tag{36}$$

In particular, φ is continuous on R^1. Define

$$f(x) = \sum_{n=0}^{\infty} \left(\tfrac{3}{4}\right)^n \varphi(4^n x). \tag{37}$$

Since $0 \le \varphi \le 1$, Theorem 7.10 shows that the series (37) converges uniformly on R^1. By Theorem 7.12, f is continuous on R^1.

Now fix a real number x and a positive integer m. Put

$$\delta_m = \pm \tfrac{1}{2} \cdot 4^{-m}$$

where the sign is so chosen that no integer lies between $4^m x$ and $4^m(x + \delta_m)$. This can be done, since $4^m |\delta_m| = \frac{1}{2}$. Define

$$\gamma_n = \frac{\varphi(4^n(x + \delta_m)) - \varphi(4^n x)}{\delta_m}.$$

When $n > m$, then $4^n \delta_m$ is an even integer, so that $\gamma_n = 0$. When $0 \le n \le m$, (36) implies that $|\gamma_n| \le 4^n$.

Since $|\gamma_m| = 4^m$, we conclude that

$$\left|\frac{f(x+\delta_m) - f(x)}{\delta_m}\right| = \left|\sum_{n=0}^{m} \left(\frac{3}{4}\right)^n \gamma_n\right| \ge 3^m - \sum_{n=0}^{m-1} 3^n = \tfrac{1}{2}(3^m + 1).$$

As $m \to \infty$, $\delta_m \to 0$. It follows that f is not differentiable at x.

Definition (Pointwise convergence). Let $(f^{(n)})_{n=1}^{\infty}$ be a sequence of functions from one metric space (X, d_X) to another (Y, d_Y), and let $f : X \to Y$ be another function. We say that $(f^{(n)})_{n=1}^{\infty}$ *converges pointwise to f on X* if we have

$$\lim_{n\to\infty} f^{(n)}(x) = f(x)$$

for all $x \in X$, i.e.

$$\lim_{n\to\infty} d_Y(f^{(n)}(x), f(x)) = 0.$$

Or in other words, for every x and every $\varepsilon > 0$ there exists $N > 0$ such that $d_Y(f^{(n)}(x), f(x)) < \varepsilon$ for every $n > N$. We call the function f the *pointwise limit* of the functions $f^{(n)}$.

Remark Note that $f^{(n)}(x)$ and $f(x)$ are points in Y, rather than functions, so we are using our prior notion of convergence in metric spaces to determine convergence of functions. Also note that we are not really using the fact that (X, d_X) is a metric space (i.e., we are not using the metric d_X); for this definition it would suffice for X to just be a plain old set with no metric structure. However, later on we shall want to restrict our attention to *continuous* functions from X to Y, and in order to do so we need a metric on X (and on Y), or at least a topological structure. Also when we introduce the concept of *uniform convergence*, then we will definitely need a metric structure on X and Y; there is no comparable notion for topological spaces.

Definition (Uniform convergence). Let $(f^{(n)})_{n=1}^{\infty}$ be a sequence of functions from one metric space (X, d_X) to another (Y, d_Y), and let $f : X \to Y$ be another function. We say that $(f^{(n)})_{n=1}^{\infty}$ *converges uniformly to f on X* if for every $\varepsilon > 0$ there exists $N > 0$ such that $d_Y(f^{(n)}(x), f(x)) < \varepsilon$ for every $n > N$ and $x \in X$. We call the function f the *uniform limit* of the functions $f^{(n)}$.

Theorem (Uniform limits preserve continuity I). *Suppose $(f^{(n)})_{n=1}^{\infty}$ is a sequence of functions from one metric space (X, d_X) to another (Y, d_Y), and suppose that this sequence converges uniformly to another function $f : X \to Y$. Let x_0 be a point in X. If the functions $f^{(n)}$ are continuous at x_0 for each n, then the limiting function f is also continuous at x_0.*

Corollary (Uniform limits preserve continuity II). *Let $(f^{(n)})_{n=1}^{\infty}$ be a sequence of functions from one metric space (X, d_X) to another (Y, d_Y), and suppose that this sequence converges uniformly to another function $f : X \to Y$. If the functions $f^{(n)}$ are continuous on X for each n, then the limiting function f is also continuous on X.*

Example. Let X be a metric space, $x_0 \in X$,

$$f(x) := d(x, x_0) \quad (f : X \to \mathbb{R}).$$

Then f is uniformly continuous on X.

Theorem *Let X be a metric space. A function $f = (f_1, \ldots, f_d) : X \to \mathbb{R}^d$ is (uniformly) continuous precisely if all the f_i $(i = 1, \ldots, d)$ are (uniformly) continuous.*

Proposition *Let* $(f^{(n)})_{n=1}^{\infty}$ *be a sequence of continuous functions from one metric space* (X, d_X) *to another* (Y, d_Y)*, and suppose that this sequence converges uniformly to another function* $f : X \to Y$*. Let* $x^{(n)}$ *be a sequence of points in* X *which converge to some limit* x*. Then* $f^{(n)}(x^{(n)})$ *converges (in* Y*) to* $f(x)$*.*

Definition (Bounded functions). A function $f : X \to Y$ from one metric space (X, d_X) to another (Y, d_Y) is *bounded* if $f(X)$ is a bounded set, i.e., there exists a ball $B_{(Y,d_Y)}(y_0, R)$ in Y such that $f(x) \in B_{(Y,d_Y)}(y_0, R)$ for all $x \in X$.

Proposition (Uniform limits preserve boundedness). *Let* $(f^{(n)})_{n=1}^{\infty}$ *be a sequence of functions from one metric space* (X, d_X) *to another* (Y, d_Y), *and suppose that this sequence converges uniformly to another function* $f : X \to Y$. *If the functions* $f^{(n)}$ *are bounded on* X *for each* n, *then the limiting function* f *is also bounded on* X.

Definition Let V be a vector space over $\mathbb{R}$ (or $\mathbb{C}$). A mapping

$$\|\cdot\| : V \to \mathbb{R}$$

is called a norm if the following conditions are fulfilled:

(i) For $v \in V \setminus \{0\}, \|v\| > 0$ (positive definiteness)

(ii) For $v \in V, \lambda \in \mathbb{R}$ resp. $\mathbb{C}, \|\lambda v\| = |\lambda| \|v\|$

(iii) For $v, w \in V, \|v + w\| \leq \|v\| + \|w\|$ (triangle inequality).

A vector space V equipped with a norm $\|\cdot\|$ is called a normed vector space $(V, \|\cdot\|)$.

If we define the distance between two vectors u, v in (V, $\|\cdot\|$) to be

$$\|u - v\|,$$

then V becomes a metric space.

A sequence $(v_n)_{n\in\mathbb{N}} \subset V$ is said to converge (relative to $\|\cdot\|$) to v if

$$\lim_{n\to\infty} \|v_n - v\| = 0\,.$$

Examples. The absolute value $|\cdot|$ in $\mathbb{R}$ can be generalized to a norm in $\mathbb{R}^d$ by putting for $x = (x^1, \ldots, x^d) \in \mathbb{R}^d$

$$\|x\|_p := \left(\sum_{i=1}^{d} (|x^i|)^p\right)^{\frac{1}{p}}, \quad \text{with } 1 \leq p < \infty,$$

or by

$$\|x\|_\infty := \max_{i=1,\ldots,d} |x^i|.$$

Definition Let K be a set and $f : K \to \mathbb{R}$ (or $\mathbb{C}$) a function.

$$\|f\|_K := \sup\{|f(x)| : x \in K\}.$$

Lemma *$\|\cdot\|_K$ is a norm on the vector space of bounded real (resp. complex) valued functions on K.*

Theorem *$f_n : K \to \mathbb{R}$ (or $\mathbb{C}$) converges uniformly to $f : K \to \mathbb{R}$ (resp. $\mathbb{C}$) if and only if*

$$\lim_{n\to\infty} \|f_n - f\|_K = 0.$$

Proof $\quad f_n \quad$ converges uniformly to $\quad f$

$$\Leftrightarrow \forall\, \varepsilon > 0\ \exists\, N \in \mathbb{N}\ \forall\, n \geq N, x \in K : |f_n(x) - f(x)| < \varepsilon$$

$$\Leftrightarrow \forall\, \varepsilon > 0\ \exists\, N \in \mathbb{N}\ \forall\, n \geq N : \sup\{|f_n(x) - f(x)| : x \in K\} < \varepsilon$$

$$\Leftrightarrow \forall\, \varepsilon > 0\ \exists\, N \in \mathbb{N}\ \forall\, n \geq N : \|f_n - f\|_K < \varepsilon$$

$$\Leftrightarrow \lim_{n\to\infty} \|f_n - f\|_K = 0. \qquad \square$$

Lemma *Let $(X, \|\cdot\|)$ be a normed space and let $(x_n)_{n\in\mathbb{N}} \subset X$ converge to $x \in X$. Then*

$$\lim_{n\to\infty} \|x_n\| = \|x\|$$

i.e. the norm is continuous.

Proof. By the triangle inequality we have

$$|\|x_n\| - \|x\|| \leq \|x_n - x\|,$$

and by assumption, the right hand side tends to 0. $\square$

Equicontinuous families of functions

7.19 Definition Let $\{f_n\}$ be a sequence of functions defined on a set E.

We say that $\{f_n\}$ is *pointwise bounded* on E if the sequence $\{f_n(x)\}$ is bounded for every $x \in E$, that is, if there exists a finite-valued function ϕ defined on E such that

$$|f_n(x)| < \phi(x) \qquad (x \in E, n = 1, 2, 3, \ldots).$$

We say that $\{f_n\}$ is *uniformly bounded* on E if there exists a number M such that

$$|f_n(x)| < M \qquad (x \in E, n = 1, 2, 3, \ldots).$$

Remark We shall prove that if $\{f_n\}$ is pointwise bounded on E and E_1 is a countable subset of E, it is always possible to find a subsequence $\{f_{n_k}\}$ such that $\{f_{n_k}(x)\}$ converges for every $x \in E_1$.

However, even if $\{f_n\}$ is a uniformly bounded sequence of continuous functions on a compact set E, there need not exist a subsequence which converges pointwise on E.

Another question is whether every convergent sequence contains a uniformly convergent subsequence. The next example will show that this need not be so, even if the sequence is uniformly bounded on a compact set

7.21 Example Let

$$f_n(x) = \frac{x^2}{x^2 + (1 - nx)^2} \qquad (0 \le x \le 1, n = 1, 2, 3, \ldots).$$

Then $|f_n(x)| \le 1$, so that $\{f_n\}$ is uniformly bounded on $[0, 1]$. Also

$$\lim_{n\to\infty} f_n(x) = 0 \qquad (0 \le x \le 1),$$

but

$$f_n\left(\frac{1}{n}\right) = 1 \qquad (n = 1, 2, 3, \ldots),$$

so that no subsequence can converge uniformly on $[0, 1]$.

7.22 Definition A family $\mathscr{F}$ of complex functions f defined on a set E in a metric space X is said to be *equicontinuous* on E if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x) - f(y)| < \varepsilon$$

whenever $d(x, y) < \delta$, $x \in E$, $y \in E$, and $f \in \mathscr{F}$. Here d denotes the metric of X.

Remark Every member of an equicontinuous family is uniformly continuous.

7.23 Theorem *If $\{f_n\}$ is a pointwise bounded sequence of complex functions on a countable set E, then $\{f_n\}$ has a subsequence $\{f_{n_k}\}$ such that $\{f_{n_k}(x)\}$ converges for every $x \in E$.*

Proof Let $\{x_i\}$, $i = 1, 2, 3, \ldots$, be the points of E, arranged in a sequence. Since $\{f_n(x_1)\}$ is bounded, there exists a subsequence, which we shall denote by $\{f_{1,k}\}$, such that $\{f_{1,k}(x_1)\}$ converges as $k \to \infty$.

Let us now consider sequences $S_1, S_2, S_3, \ldots$, which we represent by

$$\begin{array}{llllll} S_1: & f_{1,1} & f_{1,2} & f_{1,3} & f_{1,4} & \cdots \\ S_2: & f_{2,1} & f_{2,2} & f_{2,3} & f_{2,4} & \cdots \\ S_3: & f_{3,1} & f_{3,2} & f_{3,3} & f_{3,4} & \cdots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{array}$$

and which have the following properties:

(*a*) S_n is a subsequence of S_{n-1}, for $n = 2, 3, 4, \ldots$.
(*b*) $\{f_{n,k}(x_n)\}$ converges, as $k \to \infty$ (the boundedness of $\{f_n(x_n)\}$ makes it possible to choose S_n in this way);

We now go down the diagonal of the array; i.e., we consider the sequence

$$S: \; f_{1,1} \quad f_{2,2} \quad f_{3,3} \quad f_{4,4} \cdots.$$

Then, $\{f_{n,n}(x_i)\}$ converges, as $n \to \infty$, for every $x_i \in E$.

7.24 Theorem *If K is a compact metric space, if $f_n \in \mathscr{C}(K)$ for $n = 1, 2, 3, \ldots$, and if $\{f_n\}$ converges uniformly on K, then $\{f_n\}$ is equicontinuous on K.*

Proof Let $\varepsilon > 0$ be given. Since $\{f_n\}$ converges uniformly, there is an integer N such that

$$\|f_n - f_N\| < \varepsilon \qquad (n > N). \tag{42}$$

(See Definition 7.14.) Since continuous functions are uniformly continuous on compact sets, there is a $\delta > 0$ such that

$$|f_i(x) - f_i(y)| < \varepsilon \tag{43}$$

if $1 \leq i \leq N$ and $d(x, y) < \delta$.

If $n > N$ and $d(x, y) < \delta$, it follows that

$$|f_n(x) - f_n(y)| \leq |f_n(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f_n(y)| < 3\varepsilon.$$

In conjunction with (43), this proves the theorem.

7.25 Theorem *If K is compact, if $f_n \in \mathscr{C}(K)$ for $n = 1, 2, 3, \ldots$, and if $\{f_n\}$ is pointwise bounded and equicontinuous on K, then*

> *(a) $\{f_n\}$ is uniformly bounded on K,*
> *(b) $\{f_n\}$ contains a uniformly convergent subsequence.*

Proof (*a*) Let $\varepsilon > 0$ be given and choose $\delta > 0$ so that

(44) $$|f_n(x) - f_n(y)| < \varepsilon \qquad \text{for all } n, \text{ provided that } d(x, y) < \delta.$$

Since K is compact, there are finitely many points $p_1, \ldots, p_r$ in K such that to every $x \in K$ corresponds at least one p_i with $d(x, p_i) < \delta$. Since $\{f_n\}$ is pointwise bounded, there exist $M_i < \infty$ such that $|f_n(p_i)| < M_i$ for all n. If $M = \max(M_1, \ldots, M_r)$, then $|f_n(x)| < M + \varepsilon$ for every $x \in K$. This proves (*a*).

(*b*) Let E be a countable dense subset of K. Theorem 7.23 shows that $\{f_n\}$ has a subsequence $\{f_{n_i}\}$ such that $\{f_{n_i}(x)\}$ converges for every $x \in E$.

Put $g_i = f_{n_i}$. We shall prove that $\{g_i\}$ converges uniformly on K.

Let $\varepsilon > 0$, and pick $\delta > 0$ as in the beginning of this proof. Let $V(x, \delta)$ be the set of all $y \in K$ with $d(x, y) < \delta$. Since E is dense in K, and K is compact, there are finitely many points $x_1, \ldots, x_m$ in E such that

(45) $$K \subset V(x_1, \delta) \cup \cdots \cup V(x_m, \delta).$$

Since $\{g_i(x)\}$ converges for every $x \in E$, there is an integer N such that

$$|g_i(x_s) - g_j(x_s)| < \varepsilon \tag{46}$$

whenever $i \geq N, j \geq N, 1 \leq s \leq m$.

If $x \in K$, (45) shows that $x \in V(x_s, \delta)$ for some s, so that

$$|g_i(x) - g_i(x_s)| < \varepsilon$$

for every i. If $i \geq N$ and $j \geq N$, it follows from (46) that

$$\begin{aligned} |g_i(x) - g_j(x)| &\leq |g_i(x) - g_i(x_s)| + |g_i(x_s) - g_j(x_s)| + |g_j(x_s) - g_j(x)| \\ &< 3\varepsilon. \end{aligned}$$

This completes the proof.

THE STONE-WEIERSTRASS THEOREM

7.26 Theorem *If f is a continuous complex function on $[a, b]$, there exists a sequence of polynomials P_n such that*

$$\lim_{n \to \infty} P_n(x) = f(x)$$

uniformly on $[a, b]$. If f is real, the P_n may be taken real.

Proof We may assume, without loss of generality, that $[a, b] = [0, 1]$. We may also assume that $f(0) = f(1) = 0$. For if the theorem is proved for this case, consider

$$g(x) = f(x) - f(0) - x[f(1) - f(0)] \qquad (0 \le x \le 1).$$

Here $g(0) = g(1) = 0$, and if g can be obtained as the limit of a uniformly convergent sequence of polynomials, it is clear that the same is true for f, since $f - g$ is a polynomial.

Furthermore, we define $f(x)$ to be zero for x outside $[0, 1]$. Then f is uniformly continuous on the whole line.

We put

(47) $$Q_n(x) = c_n(1 - x^2)^n \qquad (n = 1, 2, 3, \ldots),$$

where c_n is chosen so that

(48) $$\int_{-1}^{1} Q_n(x)\, dx = 1 \qquad (n = 1, 2, 3, \ldots).$$

Since

$$\int_{-1}^{1}(1-x^2)^n\,dx = 2\int_0^1 (1-x^2)^n\,dx \ge 2\int_0^{1/\sqrt{n}}(1-x^2)^n\,dx$$
$$\ge 2\int_0^{1/\sqrt{n}}(1-nx^2)\,dx = \frac{4}{3\sqrt{n}} > \frac{1}{\sqrt{n}},$$

it follows from (48) that

(49) $$c_n < \sqrt{n}.$$

The inequality $(1-x^2)^n \ge 1-nx^2$ which we used above is easily shown to be true by considering the function

$$(1-x^2)^n - 1 + nx^2$$

which is zero at $x=0$ and whose derivative is positive in $(0, 1)$.

For any $\delta > 0$, (49) implies

(50) $$Q_n(x) \le \sqrt{n}\,(1-\delta^2)^n \qquad (\delta \le |x| \le 1),$$

so that $Q_n \to 0$ uniformly in $\delta \le |x| \le 1$.

Now set

(51) $$P_n(x) = \int_{-1}^{1} f(x+t)Q_n(t)\,dt \qquad (0 \le x \le 1).$$

Our assumptions about f show, by a simple change of variable, that

$$P_n(x) = \int_{-x}^{1-x} f(x+t)Q_n(t)\,dt = \int_0^1 f(t)Q_n(t-x)\,dt,$$

and the last integral is clearly a polynomial in x. Thus $\{P_n\}$ is a sequence of polynomials, which are real if f is real.

Given $\varepsilon > 0$, we choose $\delta > 0$ such that $|y-x| < \delta$ implies

$$|f(y)-f(x)| < \frac{\varepsilon}{2}.$$

Let $M = \sup |f(x)|$. Using (48), (50), and the fact that $Q_n(x) \ge 0$, we see that for $0 \le x \le 1$,

$$\begin{aligned}
|P_n(x)-f(x)| &= \left|\int_{-1}^{1}[f(x+t)-f(x)]Q_n(t)\,dt\right| \\
&\le \int_{-1}^{1}|f(x+t)-f(x)|Q_n(t)\,dt \\
&\le 2M\int_{-1}^{-\delta}Q_n(t)\,dt + \frac{\varepsilon}{2}\int_{-\delta}^{\delta}Q_n(t)\,dt + 2M\int_{\delta}^{1}Q_n(t)\,dt \\
&\le 4M\sqrt{n}\,(1-\delta^2)^n + \frac{\varepsilon}{2} \\
&< \varepsilon
\end{aligned}$$

for all large enough n, which proves the theorem.

9.2 Theorem *Let r be a positive integer. If a vector space X is spanned by a set of r vectors, then* $\dim X \le r$.

Proof If this is false, there is a vector space X which contains an independent set $Q = \{\mathbf{y}_1, \ldots, \mathbf{y}_{r+1}\}$ and which is spanned by a set S_0 consisting of r vectors.

Suppose $0 \le i < r$, and suppose a set S_i has been constructed which spans X and which consists of all $\mathbf{y}_j$ with $1 \le j \le i$ plus a certain collection of $r - i$ members of S_0, say $\mathbf{x}_1, \ldots, \mathbf{x}_{r-i}$. (In other words, S_i is obtained from S_0 by replacing i of its elements by members of Q, without altering the span.) Since S_i spans X, $\mathbf{y}_{i+1}$ is in the span of S_i; hence there are scalars $a_1, \ldots, a_{i+1}, b_1, \ldots, b_{r-i}$, with $a_{i+1} = 1$, such that

$$\sum_{j=1}^{i+1} a_j \mathbf{y}_j + \sum_{k=1}^{r-i} b_k \mathbf{x}_k = \mathbf{0}.$$

If all b_k's were 0, the independence of Q would force all a_j's to be 0, a contradiction. It follows that some $\mathbf{x}_k \in S_i$ is a linear combination of the other members of $T_i = S_i \cup \{\mathbf{y}_{i+1}\}$. Remove this $\mathbf{x}_k$ from T_i and call the remaining set S_{i+1}. Then S_{i+1} spans the same set as T_i, namely X, so that S_{i+1} has the properties postulated for S_i with $i + 1$ in place of i.

Starting with S_0, we thus construct sets $S_1, \ldots, S_r$. The last of these consists of $\mathbf{y}_1, \ldots, \mathbf{y}_r$, and our construction shows that it spans X. But Q is independent; hence $\mathbf{y}_{r+1}$ is not in the span of S_r. This contradiction establishes the theorem.

9.3 Theorem *Suppose X is a vector space, and* $\dim X = n$.

(a) *A set E of n vectors in X spans X if and only if E is independent.*

(b) *X has a basis, and every basis consists of n vectors.*

(c) *If $1 \leq r \leq n$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$ is an independent set in X, then X has a basis containing $\{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$.*

Linear transformations

9.4 Definitions A mapping A of a vector space X into a vector space Y is said to be a *linear transformation* if

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2, \qquad A(c\mathbf{x}) = cA\mathbf{x}$$

for all $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in X$ and all scalars c. Note that one often writes $A\mathbf{x}$ instead of $A(\mathbf{x})$ if A is linear.

Linear transformations of X into X are often called *linear operators* on X. If A is a linear operator on X which (i) is one-to-one and (ii) maps X onto X, we say that A is *invertible*. In this case we can define an operator A^{-1} on X by requiring that $A^{-1}(A\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in X$. It is trivial to verify that we then also have $A(A^{-1}\mathbf{x}) = \mathbf{x}$, for all $\mathbf{x} \in X$, and that A^{-1} is linear.

9.5 Theorem *A linear operator A on a finite-dimensional vector space X is one-to-one if and only if the range of A is all of X.*

9.6 Definitions

(*a*) Let $L(X, Y)$ be the set of all linear transformations of the vector space X into the vector space Y. Instead of $L(X, X)$, we shall simply write $L(X)$. If $A_1, A_2 \in L(X, Y)$ and if c_1, c_2 are scalars, define $c_1A_1 + c_2A_2$ by

$$(c_1A_1 + c_2A_2)\mathbf{x} = c_1A_1\mathbf{x} + c_2A_2\mathbf{x} \qquad (\mathbf{x} \in X).$$

It is then clear that $c_1A_1 + c_2A_2 \in L(X, Y)$.

(*b*) If X, Y, Z are vector spaces, and if $A \in L(X, Y)$ and $B \in L(Y, Z)$, we define their *product* BA to be the composition of A and B:

$$(BA)\mathbf{x} = B(A\mathbf{x}) \qquad (\mathbf{x} \in X).$$

Then $BA \in L(X, Z)$.

(*c*) For $A \in L(R^n, R^m)$, define the *norm* $\|A\|$ of A to be the sup of all numbers $|A\mathbf{x}|$, where $\mathbf{x}$ ranges over all vectors in R^n with $|\mathbf{x}| \leq 1$.

Observe that the inequality

$$|A\mathbf{x}| \leq \|A\| \, |\mathbf{x}|$$

holds for all $\mathbf{x} \in R^n$. Also, if λ is such that $|A\mathbf{x}| \leq \lambda|\mathbf{x}|$ for all $\mathbf{x} \in R^n$, then $\|A\| \leq \lambda$.

9.7 Theorem

(a) *If $A \in L(R^n, R^m)$, then $\|A\| < \infty$ and A is a uniformly continuous mapping of R^n into R^m.*

(b) *If $A, B \in L(R^n, R^m)$ and c is a scalar, then*

$$\|A + B\| \le \|A\| + \|B\|, \qquad \|cA\| = |c|\ \|A\|.$$

With the distance between A and B defined as $\|A - B\|$, $L(R^n, R^m)$ is a metric space.

(c) *If $A \in L(R^n, R^m)$ and $B \in L(R^m, R^k)$, then*

$$\|BA\| \le \|B\|\ \|A\|.$$

Proof

(a) Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis in R^n and suppose $\mathbf{x} = \Sigma c_i \mathbf{e}_i$, $|\mathbf{x}| \le 1$, so that $|c_i| \le 1$ for $i = 1, \ldots, n$. Then

$$|A\mathbf{x}| = \left|\sum c_i A\mathbf{e}_i\right| \le \sum |c_i|\ |A\mathbf{e}_i| \le \sum |A\mathbf{e}_i|$$

so that

$$\|A\| \le \sum_{i=1}^{n} |A\mathbf{e}_i| < \infty.$$

Since $|A\mathbf{x} - A\mathbf{y}| \le \|A\|\ |\mathbf{x} - \mathbf{y}|$ if $\mathbf{x}, \mathbf{y} \in R^n$, we see that A is uniformly continuous.

(*b*) The inequality in (*b*) follows from

$$|(A+B)\mathbf{x}| = |A\mathbf{x}+B\mathbf{x}| \le |A\mathbf{x}| + |B\mathbf{x}| \le (\|A\| + \|B\|)\,|\mathbf{x}|.$$

The second part of (*b*) is proved in the same manner. If

$$A, B, C \in L(R^n, R^m),$$

we have the triangle inequality

$$\|A - C\| = \|(A-B) + (B-C)\| \le \|A - B\| + \|B - C\|,$$

and it is easily verified that $\|A - B\|$ has the other properties of a metric

(*c*) Finally, (*c*) follows from

$$|(BA)\mathbf{x}| = |B(A\mathbf{x})| \le \|B\|\,|A\mathbf{x}| \le \|B\|\,\|A\|\,|\mathbf{x}|.$$

Since we now have metrics in the spaces $L(R^n, R^m)$, the concepts of open set, continuity, etc., make sense for these spaces.

9.8 Theorem *Let Ω be the set of all invertible linear operators on R^n.*

(a) If $A \in \Omega$, $B \in L(R^n)$, and

$$\|B - A\| \cdot \|A^{-1}\| < 1,$$

then $B \in \Omega$.

(b) Ω is an open subset of $L(R^n)$, and the mapping $A \to A^{-1}$ is continuous on Ω.

Proof

(*a*) Put $\|A^{-1}\| = 1/\alpha$, put $\|B - A\| = \beta$. Then $\beta < \alpha$. For every $\mathbf{x} \in R^n$,

$$\begin{aligned} \alpha|\mathbf{x}| &= \alpha|A^{-1}A\mathbf{x}| \le \alpha\|A^{-1}\| \cdot |A\mathbf{x}| \\ &= |A\mathbf{x}| \le |(A - B)\mathbf{x}| + |B\mathbf{x}| \le \beta|\mathbf{x}| + |B\mathbf{x}|, \end{aligned}$$

so that

(1) $$(\alpha - \beta)|\mathbf{x}| \le |B\mathbf{x}| \qquad (\mathbf{x} \in R^n).$$

Since $\alpha - \beta > 0$, (1) shows that $B\mathbf{x} \neq 0$ if $\mathbf{x} \neq 0$. Hence B is $1 - 1$. By Theorem 9.5, $B \in \Omega$. This holds for all B with $\|B - A\| < \alpha$. Thus we have (*a*) and the fact that Ω is open.

(*b*) Next, replace $\mathbf{x}$ by $B^{-1}\mathbf{y}$ in (1). The resulting inequality

(2) $$(\alpha - \beta)|B^{-1}\mathbf{y}| \le |BB^{-1}\mathbf{y}| = |\mathbf{y}| \qquad (\mathbf{y} \in R^n)$$

shows that $\|B^{-1}\| \le (\alpha - \beta)^{-1}$. The identity

$$B^{-1} - A^{-1} = B^{-1}(A - B)A^{-1},$$

combined with Theorem 9.7(*c*), implies therefore that

$$\|B^{-1} - A^{-1}\| \le \|B^{-1}\|\,\|A - B\|\,\|A^{-1}\| \le \frac{\beta}{\alpha(\alpha - \beta)}.$$

This establishes the continuity assertion made in (*b*), since $\beta \to 0$ as $B \to A$.

Differentiation

Introduction

If f is a real function with domain $(a, b) \subset R^1$ and if $x \in (a, b)$, then $f'(x)$ is usually defined to be the real number

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}, \tag{7}$$

provided, of course, that this limit exists. Thus

$$f(x+h) - f(x) = f'(x)h + r(h) \tag{8}$$

where the "remainder" $r(h)$ is small, in the sense that

$$\lim_{h \to 0} \frac{r(h)}{h} = 0. \tag{9}$$

Note that (8) expresses the difference $f(x+h) - f(x)$ as the sum of the *linear function* that takes h to $f'(x)h$, plus a small remainder.

We can therefore regard the derivative of f at x, not as a real number, but as the linear operator on R^1 that takes h to $f'(x)h$.

Let us next consider a function $\mathbf{f}$ that maps $(a, b) \subset R^1$ into R^m. In that case, $\mathbf{f}'(x)$ was defined to be that vector $\mathbf{y} \in R^m$ (if there is one) for which

$$\lim_{h \to 0} \left\{ \frac{\mathbf{f}(x+h) - \mathbf{f}(x)}{h} - \mathbf{y} \right\} = \mathbf{0}. \tag{10}$$

We can again rewrite this in the form

$$\mathbf{f}(x+h) - \mathbf{f}(x) = h\mathbf{y} + \mathbf{r}(h), \tag{11}$$

where $\mathbf{r}(h)/h \to \mathbf{0}$ as $h \to 0$. The main term on the right side of (11) is again a *linear* function of h. Every $\mathbf{y} \in R^m$ induces a linear transformation of R^1 into R^m, by associating to each $h \in R^1$ the vector $h\mathbf{y} \in R^m$. This identification of R^m with $L(R^1, R^m)$ allows us to regard $\mathbf{f}'(x)$ as a member of $L(R^1, R^m)$.

Thus, if $\mathbf{f}$ is a differentiable mapping of $(a, b) \subset R^1$ into R^m, and if $x \in (a, b)$, then $\mathbf{f}'(x)$ is the linear transformation of R^1 into R^m that satisfies

$$\lim_{h \to 0} \frac{\mathbf{f}(x+h) - \mathbf{f}(x) - \mathbf{f}'(x)h}{h} = \mathbf{0}, \tag{12}$$

or, equivalently,

$$\lim_{h \to 0} \frac{|\mathbf{f}(x+h) - \mathbf{f}(x) - \mathbf{f}'(x)h|}{|h|} = 0. \tag{13}$$

9.11 Definition Suppose E is an open set in R^n, $\mathbf{f}$ maps E into R^m, and $\mathbf{x} \in E$. If there exists a linear transformation A of R^n into R^m such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - A\mathbf{h}|}{|\mathbf{h}|} = 0, \tag{14}$$

then we say that $\mathbf{f}$ is *differentiable at* $\mathbf{x}$, and we write

$$\mathbf{f}'(\mathbf{x}) = A. \tag{15}$$

If $\mathbf{f}$ is differentiable at every $\mathbf{x} \in E$, we say that $\mathbf{f}$ is *differentiable in* E.

9.12 Theorem *Suppose E and $\mathbf{f}$ are as in Definition* 9.11, $\mathbf{x} \in E$, *and* (14) *holds with $A = A_1$ and with $A = A_2$. Then $A_1 = A_2$.*

Proof If $B = A_1 - A_2$, the inequality

$$|B\mathbf{h}| \le |\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - A_1\mathbf{h}| + |\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - A_2\mathbf{h}|$$

shows that $|B\mathbf{h}|/|\mathbf{h}| \to 0$ as $\mathbf{h} \to \mathbf{0}$. For fixed $\mathbf{h} \neq \mathbf{0}$, it follows that

$$\frac{|B(t\mathbf{h})|}{|t\mathbf{h}|} \to 0 \qquad \text{as} \quad t \to 0. \tag{16}$$

The linearity of B shows that the left side of (16) is independent of t. Thus $B\mathbf{h} = 0$ for every $\mathbf{h} \in R^n$. Hence $B = 0$.

9.15 Theorem *Suppose E is an open set in R^n, $\mathbf{f}$ maps E into R^m, $\mathbf{f}$ is differentiable at $\mathbf{x}_0 \in E$, $\mathbf{g}$ maps an open set containing $\mathbf{f}(E)$ into R^k, and $\mathbf{g}$ is differentiable at $\mathbf{f}(\mathbf{x}_0)$. Then the mapping $\mathbf{F}$ of E into R^k defined by*

$$\mathbf{F}(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x}))$$

is differentiable at $\mathbf{x}_0$, and

$$\mathbf{F}'(\mathbf{x}_0) = \mathbf{g}'(\mathbf{f}(\mathbf{x}_0))\mathbf{f}'(\mathbf{x}_0). \tag{21}$$

Proof Put $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$, $A = \mathbf{f}'(\mathbf{x}_0)$, $B = \mathbf{g}'(\mathbf{y}_0)$, and define

$$\mathbf{u}(\mathbf{h}) = \mathbf{f}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{f}(\mathbf{x}_0) - A\mathbf{h},$$

$$\mathbf{v}(\mathbf{k}) = \mathbf{g}(\mathbf{y}_0 + \mathbf{k}) - \mathbf{g}(\mathbf{y}_0) - B\mathbf{k},$$

for all $\mathbf{h} \in R^n$ and $\mathbf{k} \in R^m$ for which $\mathbf{f}(\mathbf{x}_0 + \mathbf{h})$ and $\mathbf{g}(\mathbf{y}_0 + \mathbf{k})$ are defined. Then

$$|\mathbf{u}(\mathbf{h})| = \varepsilon(\mathbf{h})|\mathbf{h}|, \qquad |\mathbf{v}(\mathbf{k})| = \eta(\mathbf{k})|\mathbf{k}|, \tag{22}$$

where $\varepsilon(\mathbf{h}) \to 0$ as $\mathbf{h} \to \mathbf{0}$ and $\eta(\mathbf{k}) \to 0$ as $\mathbf{k} \to \mathbf{0}$.

Given $\mathbf{h}$, put $\mathbf{k} = \mathbf{f}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{f}(\mathbf{x}_0)$. Then

$$|\mathbf{k}| = |A\mathbf{h} + \mathbf{u}(\mathbf{h})| \le [\|A\| + \varepsilon(\mathbf{h})]\,|\mathbf{h}|, \tag{23}$$

$$\begin{aligned}\mathbf{F}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{F}(\mathbf{x}_0) - BA\mathbf{h} &= \mathbf{g}(\mathbf{y}_0 + \mathbf{k}) - \mathbf{g}(\mathbf{y}_0) - BA\mathbf{h} \\ &= B(\mathbf{k} - A\mathbf{h}) + \mathbf{v}(\mathbf{k}) \\ &= B\mathbf{u}(\mathbf{h}) + \mathbf{v}(\mathbf{k}).\end{aligned}$$

Hence (22) and (23) imply, for $\mathbf{h} \ne \mathbf{0}$, that

$$\frac{|\mathbf{F}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{F}(\mathbf{x}_0) - BA\mathbf{h}|}{|\mathbf{h}|} \le \|B\|\,\varepsilon(\mathbf{h}) + [\|A\| + \varepsilon(\mathbf{h})]\eta(\mathbf{k}).$$

Let $\mathbf{h} \to \mathbf{0}$. Then $\varepsilon(\mathbf{h}) \to 0$. Also, $\mathbf{k} \to \mathbf{0}$, by (23), so that $\eta(\mathbf{k}) \to 0$. It follows that $\mathbf{F}'(\mathbf{x}_0) = BA$, which is what (21) asserts.