MATH 450 - NOTES

April 10, 2009

Contents

1 METRIC SPACES

1.1 Definition and Examples

Definition 1 *Let X be a set. A function $d : X \times X \to \mathbb{R}$ is called a* **metric** *or distance function if for all $x, y, z \in X$*

1. $d(x, y) \geq 0$, *and* $d(x, y) = 0$ *if and only if* $x = y$.
2. $d(x, y) = d(y, x)$
3. $d(x, z) \leq d(x, y) + d(y, z)$

The pair (X, d) is called a **metric space**.

If it is clear from the context which metric we refer to, we will not explicitly mention the metric and just call X a metric space. For the remaining part of the course we will assume that we have a metric space, even if its not explicitly mentioned.

Some Examples:

1. On $\mathbb{R}$ we can define $d(x, y) = |x - y|$. This is a metric and makes $\mathbb{R}$ into a metric space.

2. Other metrics on $\mathbb{R}$:
$$d_1(x, y) = \frac{|x - y|}{1 + |x - y|}.$$
$$d_2(x, y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases} \tag{1}$$
The second metric can be used to make any set into a metric space.

3. Metrics on $\mathbb{R}^n$, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{x} = (x_1, \ldots, x_n)$
$$\begin{aligned} d_2(\mathbf{x}, \mathbf{y}) &= \left(\sum_{i=1}^{n} |x_i - y_i|^2 \right)^{\frac{1}{2}} && (2) \\ d_1(\mathbf{x}, \mathbf{y}) &= \sum_{i=1}^{n} |x_i - y_i| && (3) \\ d_\infty(\mathbf{x}, \mathbf{y}) &= \max \{|x_i - y_i| : i = 1, \ldots, n\} && (4) \end{aligned}$$

4. Let $C((a, b))$ be the set of all real valued continuous functions on (a, b).
$$\begin{aligned} d_2(f, g) &= \left(\int_a^b |f - g|^2 \, dx \right)^{\frac{1}{2}} && (5) \\ d_1(f, g) &= \int_a^b |f - g| \, dx && (6) \\ d_\infty(f, g) &= \sup \{|f(x) - g(x)| : x \in (a, b)\} && (7) \end{aligned}$$
are all metrics on this space.

1.2 Open Balls and Sets:

Definition 2 *Let (X, d) be a metric space and $r > 0$. The set*
$$B_r(x_0) = \{x \in X : d(x, x_0) < r\}$$
is called the **open ball** *of radius r around x_0*

Example: In $\mathbb{R}^2$ with the metric d_2, the open ball of radius 1 around the origin is just the interior of the circle of radius 1 centered at the origin. With the metric d_1 the ball is the interior of the diamond with vertices at $(1, 0), (0, 1), (-1, 0), (0, -1)$. With the metric d_∞ the ball is the square $(-1, 1) \times (-1, 1)$.

It is important to note that balls can come in many shapes or forms. Often we consider balls with their centers removed. We will use the following notation $\dot{B}_r(x) = B_r(x) \setminus \{x\}$.

Definition 3 *Let (X, d) be a metric space. A set $O \subset X$ is open if for any $x \in O$ there is a $r > 0$ such that $B_r(x) \subset O$.*

The first thing to note is that we used the term open in two different ways, once in the connection with open balls, the next time in connection with open sets . We need to show that open balls are open sets. To to this consider the open ball $B_r(x_0)$. Let $x \in B_r(x_0)$. Let

$$\delta = \frac{r - d(x, x_0)}{2}.$$

We will show that $B_\delta(x) \subset B_r(x_0)$. To do this let $y \in B_\delta(x)$. Then

$$d(y, x_0) \leq d(y, x) + d(x, x_0) < \delta + d(x, x_0) = \frac{r}{2} - \frac{d(x, x_0)}{2} + d(x, x_0) < r,$$

so $y \in B_r(x_0)$.

Proposition 1 *Let (X, d) be a metric space. Then*

1. *X and $\emptyset$ are open.*
2. *If $\{O_\alpha : \alpha \in A\}$ is an arbitrary collection of open sets, then $\bigcup_{\alpha \in A} O_\alpha$ is also open.*
3. *If O_1 and O_2 are open, then so is $O_1 \cap O_2$.*

Proof: To show the first assertion, clearly $B_1(x) \subset X$ for all $x \in X$. For any $x \in \emptyset$ (there are none) $B_1(x) subset \emptyset$. For the second assertion let $x \in \bigcup_{\alpha \in A} O_\alpha$, then there is an α_0 such that $x \in O_{\alpha_0}$. Since this set is open there is a $r > 0$ such that $B_r(x) \subset O_{\alpha_0} \subset \bigcup_{\alpha \in A} O_\alpha$. Thus the union is open. For the intersection observe that if $x \in O_1 \cap O_2$ then there is $r_1, r_2 > 0$ such that $B_{r_1}(x) \subset O_1$ and $B_{r_2}(x) \subset O_2$. Let $r = \min\{r_1, r_2\}$ then $B_r(x) \subset B_{r_1}(x) \cap B_{r-2}(x) \subset O_1 \cap O_2$.

Note that the intersection of arbitrarily many open sets may not be open. If we consider the collection of open balls around a point p with radii $\frac{1}{n}$, then $B_{\frac{1}{n}}(p)$ is open for every $n \in \mathbb{N}$, but

$$\bigcap_{n=1}^{\infty} B_{\frac{1}{n}}(p) = \{p\},$$

which is not open.

Proposition 2 *Let (X, d) be a metric space. Then X has the , i.e. for any $p, q \in X$, $p \neq q$ there exists open sets U_p and U_q such that $p \in U_p$, $q \in U_q$ and $U_p \cap U_q = \emptyset$.*

Proof: Since $p \neq q$ we have $r = d(p, q) > 0$. Then $B_{\frac{r}{3}}(p) \cap B_{\frac{r}{3}}(q) = \emptyset$.

1.3 Closed Sets, Interior and Limit Points:

Definition 4 *A subset F of a metric space is closed , if its complement is open.*

Proposition 3 *Let (X, d) be a metric space. Then X and the empty set are closed. The union of finitely many closed sets is closed, and the intersection of arbitrarily many closed sets is closed.*

Proof: Left as homework.

Definition 5 *Let A be a subset of a metric space X. $p \in X$ is an interior point of A if there exists an open ball $B_r(p)$ such that $B_r(p) \subset A$. $p \in X$ is a limit point of A if for any $r > 0$ $\dot{B}_r(x) \cap A \neq \emptyset$.*

It is clear from this that every element of an open set is an interior point.

Proposition 4 *Let f be a closed set and p be a limit point of F. Then $p \in F$.*

Proof: Assume that $p \notin F$. Then $p \in F^C$, which is open. Therefore there is an $r > 0$ such that $B_r(p) \subset F^C$, i.e. $B_r(p) \cup F = \emptyset$. Thus p cannot be a limit point of F.

1.4 Assignment 1

1. Show that the functions in equations (3), (4), (6), (7) are metrics.
2. Proof Proposition 3.
3. Define l^2 to be the set of all real sequences (a_n) such that $\sum_{k=1}^{\infty} a_n^2 < \infty$. For two sequences $(a_n), (b_n) \in l^2$ define
$$d((a_n), (b_n)) = \left(\sum_{k=1}^{\infty} (a_n - b_n)^2 \right)^{\frac{1}{2}}.$$
Show that this is a metric on l^2.

1.5 Sequences in Metric Spaces

Definition 6 *Let $\{p_n\}$ a sequence of points in a metric space X. $\{p_n\}$ converges to p, if for every $\epsilon > 0$ there is a N such that $p_n \in B_\epsilon(p)$ for all $n \geq N$. $\{p_n\}$ is a Cauchy sequence, if for every $\epsilon > 0$ there is N such that $d(p_n, p_M) < \epsilon$ for all $n, m \geq N$. X is complete if every Cauchy sequence converges.*

Proposition 5 *A sequence in a metric space has at most one limit.*

Proof: Assume that p_n converges to two limits p and q. For any $\epsilon > 0$ there is N_1 such that $d(p_n, p) < \epsilon/2$ for all $n \geq N_1$ and an N_2 such that $d(p_n, q) < epsilon/2$ for all $n \geq N_2$. For any $n \geq \max\{N_1, N_2\}$ we have.

$$d(p, q) \leq d(p, p_n) + d(p_n, q) < \epsilon.$$

Since this is true for any $\epsilon > 0$, it follows that $d(p, q) = 0$ and so $p = q$.

Definition 7 *A set A is bounded if $\sup\{d(p, q) : p, q \in A\}$ is finite. A set C is compact if every open cover of C has a finite subcover.*

Proposition 6 *Let C be a compact subset of a metric space X. Then C is closed and bounded.*

Proof: We first show closed. To do this we show that C^c is open. Let $p \in C^c$. For every $q \in C$ there exists balls $B_{\epsilon_q}(q)$ and $B_{\epsilon_q}(p)$ such that $B_{\epsilon_q}(p) \cap B_{\epsilon_q}(q) = \emptyset$. The collection of balls $B_{\epsilon_q}(q)$ form an open cover of C. Since C is compact there are points $q_1, \ldots, q_n$ such that

$$C \subset B_{\epsilon_{q_1}}(q_1) \cup \cdots \cup B_{\epsilon_{q_n}}(q_n)$$

Let $\epsilon = \min\{\epsilon_{q_1}, \ldots, \epsilon_{q_n}\}$, then $B_\epsilon(p) \cap B_{\epsilon_{q_j}}(q_j) = \emptyset$ and therefore $B_\epsilon(p) \cap C = \emptyset$. Hence, $B_\epsilon(p) \subset C^c$, and p is an iterior point of C^c. Since p was arbitrary it follows that C^c is open and C is closed. For the boundedness, consider the collectio of open balls $\{B_1(q) : q \in C\}$, This is an open cover, and has therefore a finite subcover. $B_1(q_1), \ldots, B_1(q_n)$. Let $M = \max\{d(q_i, q_j) : i, j = 1, \ldots, n\}$. For any pair of points p, q there exist q_i, q_j such that $p \in B_(q_i)$ and $q \in B_1(q_j)$. Thus

$$d(p, q) \leq d(p, q_i) + d(q_i, q_j) + d(q_j, q) < 1 + M + 1 = M + 2.$$

Thus the set is bounded. It is important to note that the converse of this proposition is not true in general.

However, it is true in $\mathbb{R}^n$ as we will see later. In metric space compactness can also be characterized by limit points and sequences. We have the following Theorem.

Theorem 1 *Let X be a metric space and $C \subset X$. The following are equivalent:*

1. *C is compact.*
2. *Every infinite subset of C has a limit point (C is limit point compact)*
3. *Every sequence in C has a convergent subsequence (C is sequentially compact)*

Proof: We first show that (1) implies (2): To do this we prove the contra positive. Let $A \subset C$ with no limit point. Since A contains all its limit points it is closed. For every $p \in A$ we may find an open ball $B(p)$ such that $B(p) \cap A = \{p\}$. The collection of these balls together with A^c form an open cover of C. So there exists

a finite subcover. Since A^C does not contain any points in A, each of the balls must be included in the finite subcover. This implies there were only finitely many balls to start with, and A is finite.
Next we prove that (2) implies (3). Given a sequence $\{p_n\}$ in C, consider the set $A = \{p_n : n \in \mathbb{N}\}$. If this set is finite there is a point p such that $p_n = p$ for infinitely many values of n. This is a constant and therefore convergent subsequence. If A is infinite, then A has a limit point p. Now each $k \in \mathbb{N}$ choose

$$p_{n_k} \in B_{1/k}(p)$$

and $n_k > n_{k-1}$. This subsequence converges to p.
The last step is to prove that (3) implies (1). This is the hard part. We first show the following

Lemma 1 *If C is sequentially compact and $\mathcal{O}$ an open cover of C, then there exists a number $\delta > 0$ such that each subset of C with diameter less than δ is contained in an open set $O \in \mathcal{O}$. δ is called the Lebesgue number.*

Proof of the Lemma: (by contradiction) Given the open cover $\mathcal{O}$ assume that there is no such number δ. In particular for every integer n there is a set of diameter less than $1/n$, which is not contained in any open set in the cover. Denote these sets by C_n. Choose any point $p_n \in C_n$. Since C is sequentially compact there exists a subsequence $\{p_{n_j}\}$ that converges to some point p. $p \in O$ for some open set O in the cover. There exists an $\epsilon > 0$ such that $B_\epsilon(p) \subset O$. For suffitiently large n_j with $1/n_j < \epsilon/2$ we have

$$C_{n_j} \subset B_{\epsilon/2}(p_{n_j})$$

and

$$p_{n_j} \in B_{\epsilon/2}(p).$$

But then

$$C_{n_j} \subset B_\epsilon(p) \subset O.$$

which is a contradiction.

Next we prove:

Lemma 2 *If C is sequentially compact and $\epsilon >$, there exist points $p_1, \ldots, p_n$ such that*

$$C \subset B_\epsilon(p_1) \cup \cdots \cup B_\epsilon(p_n).$$

Proof of the Lemma: Pick

$$p_1 \in C, p_2 \in C \setminus B_\epsilon(p_1), p_k \in C \setminus (B_\epsilon(p_1) \cup \cdots \cup B_\epsilon(p_{k-1}))$$

This sequence must end after finitely many steps, since otherwise we would have an infinite sequence with $d(p_{n+1}, p_j) \geq \epsilon$ for all $j = 1, \ldots, n$, which has no convergent subsequence.

We can now finally prove that (3) implies (1): Given any open cover $\mathcal{O}$ let δ be the number from the first Lemma. (This called the Lebesgue number of the cover). Let $\epsilon = \delta/3$. The second lemma implies that there exist $p_1, \ldots p_n$ such that

$$C \subset B_\epsilon(p_1) \cup \cdots \cup B_\epsilon(p_n).$$

Now observe that the diameter of these balls is $2\delta/3 < \delta$. I.e. for each p_j there is an $O_j \in \mathcal{O}$ such that $B_\epsilon(p_j) \subset O_j$. But then

$$C \subset O_1 \cup \cdots \cup O_n.$$

Hence, C is compact.

1.6 Assignment 2

1. Prove that any finite set is compact.
2. Let A be closed and bounded subset in $\mathbb{R}^2$. Prove that the set $A_1 = \{x \in \mathbb{R} : (x, s) \in A\}$ is also closed and bounded.
3. Is the set A in the previous problem compact?
4. Prove that a closed subset of a compact set is compact.
5. Consider the unit circle in $\mathbb{R}^2$. Any point is uniquely determined by the angle θ. In radians consider the sequence $\theta_0 = 0$, $\theta_{n+1} = \theta_n$. These all lie on the circle. Show that the set of these points is dense in the circle.

1.7 Continuous Functions:

Definition 8 *Let (X, d) and (Y, d') be metric spaces and $A \subset X$. A function $f : A \to Y$ is continuous at $x \in A$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $d'(f(y), f(x)) < \epsilon$ for all $y \in B_\delta(x) \cap A$. A function f is continuous on A, if it is continuous at every $x \in A$.*

Observe that this can be easily rephrased in the following ways:

1. A function $f : A \to Y$ is continuous at $x \in A$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $f(y) \in B_\epsilon(f(x))$ for all $y \in B_\delta(x) \cap A$.
2. A function $f : A \to Y$ is continuous at $x \in A$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $B_\delta(x) \cap A \subset f^{-1}(B_\epsilon(f(x)))$.

One of the most important characterisations of continuous functions is the following:

Proposition 7 *Let (X,d) and (Y,d') be metric spaces. A function $f : X \to Y$ is continuous on X if and only if $f^{-1}(U)$ is open for every open set $U \subset Y$.*

Proof: Let f be continuous on X and U be an open subset of Y. Let $x \in f^{-1}(U)$. We need to show that x is an interior point of this set. Now $f(x) \in U$, which is open, thus there exists an $\epsilon > 0$ such that $B_\epsilon(f(x)) \subset U$. Since f is continuous there exists a $\delta > 0$ such that

$$B_\delta(x) \subset f^{-1}(B_\epsilon(f(x))) \subset f^{-1}(U),$$

and x is an interior point of the latter set. Since this holds for every $x \in f^{-1}(U)$ this set is open.
Conversely, assume that $f^{-1}(U)$ is open for every open subset U of Y. Let $x \in X$ and $\epsilon > 0$. The set $B_\epsilon(f(x))$ is open an thus is $f^{-1}(B_\epsilon(f(x)))$. Hence, $x \in f^{-1}(B_\epsilon(f(x)))$ is an interior point and there exists a $\delta > 0$ such that

$$B_\delta(x) \subset f^{-1}(B_\epsilon(f(x))) \subset f^{-1}(U),$$

and f is continuous at x. Since x was arbitrary, this hols for every $x \in X$.

Remark: The above proposition also holds for functions $f : A \to Y$ for $A \subset X$, if we say $f^{-1}(U)$ is open **relative to** A for every open subset u of Y. Recall that a set O is open relative to A if $O = A \cap V$ for some open subset V of X.

1.8 Properties of Continuous Functions:

Continuous functions have many important properties. In general one can say that continuous functions preserve many topological qualities the two most important ones for now are compactness and connectedness.

Proposition 8 *Let $f : X \to Y$ be continuous and $A \subset X$ connected. Then $f(A)$ is connected.*

Proof: We will prove the contrapositive. Assume that $f(A)$ is not connected, then there exist open non-empty sets U and V such that $f(A) \subset U \cup V$ and $U \cap V = \emptyset$. Now $f^{-1}(U)$ and $f^{-1}(V)$ are open and

$$A \subset f^{-1}(f(A)) \subset f^{-1}(U \cup V) = f^{-1}(U) \cup f^{-1}(V),$$

and

$$f^{-1}(U) \cap f^{-1}(V) = f^{-1}(U \cap V) = \emptyset.$$

Thus A is not connected.

Proposition 9 *Let $f : X \to Y$ be continuous and $A \subset X$ compact. Then $f(A)$ is compact.*

Proof: Let $\{O_\alpha : \alpha \in \Lambda\}$ be an open covering of $f(A)$. Since

$$A \subset f^{-1}(f(A)) \subset f^{-1}(\bigcup_{\alpha \in \Lambda} O_\alpha) = \bigcup_{\alpha \in \Lambda} f^{-1}(O_\alpha)$$

the collection $\{f^{-1}(O_\alpha) : \alpha \in \Lambda\}$ is an open cover of A. Hence there exists a finite subcover $f^{-1}(O_1), \ldots, f^{-1}(O_n)$. But, since

$$A \subset f^{-1}(O_1) \cup \ldots \cup f^{-1}(O_n),$$

we have

$$f(A) \subset O_1 \cup \ldots \cup O_n,$$

and $f(A)$ is compact.

Another important property of continuous functions is that they preserve convergence of sequences.

Proposition 10 *A function $f : X \to Y$ is continuous at $x \in X$ if and only if for every sequence $\{x_n\}$ which converges to x we have $\lim_{n\to\infty} f(x_n) = f(x)$.*

Proof: Assume that f is continuous and that $x_n \to x$. Let $\epsilon > 0$ then there exists a $\delta > 0$ such that $f(y) \in B_\epsilon(f(x))$ for any $y \in B_\delta(x)$. For δ there exists an N such that $x_n \in B_\delta(x)$ for any $n \geq N$. Combining these two statements we get that for $\epsilon > 0$ there exists a N such that $f(x_n) \in B_\epsilon(f(x))$ for all $n \geq N$. Hence $f(x_n) \to f(x)$.
Next assume that f is not continuous at x. I.e there is an $\epsilon_0 > 0$ such that for any $\delta > 0$ there is a $x_\delta \in B_\delta(x)$ such that $f(x_\delta) \notin B_{\epsilon_0}(f(x))$. Since we can choose δ freely, we find x_δ for $\delta = 1/n$ for any $n \in \mathbb{N}$ and label it x_n. Since $x_n \in B_{1/n}(x)$ for any $n \in \mathbb{N}$, $x_n \to x$. However, $f(x_n) \notin B_{\epsilon_0}(f(x)$ for all $n \in \mathbb{N}$ and so $f(x_n)$ cannot converge to $f(x)$. We have thus shown the contrapositive of the statement if $f(x_n)$ converges to $f(x)$ for any sequence $x_n \to x$ then f is continuous at x.

We finish the discussion of continuous functions by looking at uniform continuity.

Definition 9 *$f : X \to Y$ is on $A \subset X$ if for every $\epsilon > 0$ there exist a $\delta > 0$ such that $d'(f(x), f(y)) < \epsilon$* **for all** *$x, y \in A$ with $d(x, y) < \delta$.*

The important result here is

Proposition 11 *Let $C \subset X$ be compact and $f : C \to Y$ be continuous on C. Then f is uniformly continuous on C.*

Proof: Let $\epsilon > 0$ For every $x \in C$ there is a δ_x such that $f(y) \in B_{\epsilon/2}(x)$ for all $y \in B_{\delta_x}(x)$. The collection $\{B_{\delta_x}(x)\}$ forms an open cover of C. Since C is compact, this open cover has a Lebesgue number $\delta > 0$. Now let $d(x,y) < \delta$, then x, y lie in an open set u of diameter smaller than δ and hence there exists a set in the open covering which contains U. I.e. there is a $x_0 \in C$ such that $x, y \in U \subset B_{\delta_{x_0}}(x_0)$, thus

$$d'(f(x), f(y)) \leq d'(f(x), f(x_0)) + d'(f(x_0), f(y)) < \epsilon/2 + \epsilon/2 = \epsilon.$$

I.e. f is uniformly continuous.

1.9 Some Odds and Ends:

Definition 10 *Let X be a space with two metrics d and d'. The two metrics are equivalent if there exist constants C_1 and C_2 such that*

$$d(x,y) \leq C_1 d'(x,y) \qquad \text{and} \qquad d'(x,y) \leq C_2 d(x,y).$$

Proposition 12 *Let X be a space with two metrics d and d' which are equivalent. Then*

1. *$U \subset X$ is open in the metric d if and only if U is open in d'.*
2. *$U \subset X$ is closed in the metric d if and only if U is closed in d'.*
3. *$U \subset X$ is bounded in the metric d if and only if U is bounded in d'.*
4. *$U \subset X$ is compact in the metric d if and only if U is compact in d'.*

Proof: See assignment 3.

This shows that we can often choose the most convenient metric to prove some result. Since continuity only depends on open sets (thanks to a previous proposition) a change from one metric to an equivalent one will not change the continuity of a function. The discussion about equivalent metrics leads immediately to the following

Definition 11 *Let (X,d) and (Y,d') be two metric spaces. A function $\phi : X \to Y$ is called a homeomorphism if it is a continuous bijection with a continuous inverse function. If there is a homeomorphism between X and Y, the two spaces are called homeomorphic.*

It is very easy to confuse this term with the term homomorphism, which is a purely algebraic concept.
Example: The map $(x,y) \mapsto (-y,x)$ is an isometry on $\mathbb{R}^2$ with the standard Euclidean metric. Any isometry is automatically a homeomorphism.
The fact that the three metrics d_1, d_2, and d_∞ are equivalent on $\mathbb{R}^n$ will be very useful later on, as it allows us to use the metric of our choice to prove certain results.

1.10 Assignement 3

1. Prove the last Proposition.
2. Consider the metrics d_1 and d_2 on $\mathbb{R}^n$ from the first day. Prove that
$$d_1(p,q) \geq d_2(p,q),$$
and
$$d_2(p,q) \geq \frac{1}{\sqrt{n}} d_1(p,q).$$
3. Consider the metrics d_∞ and d_2 on $\mathbb{R}^n$ from the first day. Prove that
$$d_2(p,q) \geq d_\infty(p,q),$$
and
$$d_\infty(p,q) \geq \frac{1}{\sqrt{n}} d_2(p,q).$$
4. Consider $\mathbb{R}^n$ with your favorite metric and $\mathbb{R}$ with the standard metric. For a point $(x_1, \ldots, x_n) \in \mathbb{R}^n$ for $k = 1, \ldots, n$ define $\pi_k : \mathbb{R}^n \to \mathbb{R}$ by
$$\pi_k(x_1, \ldots, x_n) = x_k.$$
π_k is called the k-th projection. Prove that π_k is continuous.
5. Consider $\mathbb{R}^2$ and the map $(x, y) \mapsto (-y, x)$. Show that this is an isometry.

2 DIFFERENTIAL CALCULUS ON $\mathbb{R}^n$

In this section we explore derivatives of functions of several variables. We will generally assume that $\mathbb{R}^n$ is endowed with the Euclidean metric d_2, but use the other metrics if it is more convenient to do so. For $\mathbf{x} \in \mathbb{R}^n$ we denote by $|\mathbf{x}| = d_2(\mathbf{x}, 0)$ the norm of the vector $\mathbf{x}$. The components of the vector $\mathbf{x}$ are usually denoted by $(x_1, \ldots x_n)$ or its transpose.

2.1 The Topology of $\mathbb{R}^n$

Before exploring multivariable differential calculus we will discuss briefly the topology of $\mathbb{R}^n$. As seen in the homework and the examples $\mathbb{R}^n$ is a metric space, and we may use any of the three metrics used int the examples. However, $\mathbb{R}^n$ has some special properties which will be discussed shortly. To begin we recall the Heine-Borel Theorem on the real line. It states that a subset of $\mathbb{R}$ is compact if and only if it is closed and bounded. We will use this to prove the Heine-Borel Theorem for $\mathbb{R}^N$:

Proposition 13 **Heine-Borel Theorem** *Let $C \subset \mathbb{R}^n$ then C is compact if and only if it is closed and bounded.*

Proof: Every comp[act subset of a metric space is closed and bounded, so we only have to prove the other direction. Let C be a closed and bounded subset of $\mathbb{R}^n$. For $j = 1, \dots, n$ define $a_j = \inf\{x_j : (x_1, \dots, x_j, \dots, x_n) \in C\}$ and $b_j = \sup\{x_j : (x_1, \dots, x_j, \dots, x_n) \in C\}$. Then $C \subset [a_1, b_1] \times \cdots \times [a_n, b_n]$ and it suffices to show the compactness of the latter set. To do this we show the sequential compactness of this set. In order to make the notations easier we assume that $n = 2$. and then procede with induction. Let $\{(\alpha_k, \beta_k\}$ be a sequence in $[a_1, b_1] \times [a-2, b_2]$, then $\{\alpha_k\}$ is a sequence in $[a_1, b_1]$ and has a converging subsequence $\{alpha_{k_l}\}$ which converges to a limit α, since $[a-1, b_1]$ is compact. Next consider the subsequence $\{\alpha_{k_l}, \beta_{k_l}\}$ of the original sequence. Then $\{\beta_{k_l}\}$ is a sequence in $[a_2, b_2]$ with a converging subsequence $\{\beta_{k_{l_j}}\}$ which converges to $\beta \in [a_2, b_2]$. Hence the sequence $\{(\alpha_{k_{l_j}}, \beta_{k_{l_j}})\}$ converges to $(\alpha, \beta) \in [a_1, b_1] \times [a_2, b_2]$.
Now having shown this result for $n = 2$ it is easy to see that the same proof works to show that $A \times B$ is compact for any two compact sets A, B. This allows us to proceed by induction since

$$[a_1, b_1] \times \cdots \times [a_n, b_n] = ([a_1, b_1] \times \cdots \times [a_{n-1}, b_{n-1}]) \times [a_n, b_n],$$

i.e. the product of two compact sets by induction hypothesis. Finally since C is a closed subset of the compact set $[a_1, b_1] \times \cdots \times [a_n, b_n]$ it is it self compact.

The section about metric spaces covers everything we need to know about continuous functions. However, we must be careful when trying to evaluate a limit as the following example shows.
Example: Let

$$f(x, y) = \begin{cases} 1 & |y| > x^2 \\ 0 & 0 < |y| < x^2 \\ 1 & |y| = 0 \end{cases}$$

Then if we approach the origin along any straight line $y = ax$ we get $\lim_{x \to 0} f(x, ax) = 1$. The same is true when we approach the origin along the y-axis. But If we onsider the curve $y = \frac{x^2}{2}$, we get $\lim_{x \to 0} f(x, \frac{x^2}{2}) = 0$ so the limit cannot possibly exist. Unlike in one dimension where we can compute one sided limits and compare them, in several dimensions it is not even sufficient to compute directional limits (as the limits along all straight lines would be). The only certain way to state that a certain number is the limit is to revert back to the definition using ϵ and δ. However, there is some help, as the sum and product rules for limits still work for real valued functions as does the composition.

2.2 The Derivative of a function of several variables:

Recall the definition of the derivative of a function of one variable, i.e. if $A \subset \mathbb{R}$, $f : A \to \mathbb{R}$ and a an interior point of A then

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h},$$

if this limit exists. Unfortunately, this definition cannot be readily adapted to $\mathbb{R}^n$ for $n > 1$, since it uses a quotient, and vector spaces do not allow for division. So we rewrite this definition as follows:

Let A, f, and a be as above, f is differentiable at a if there exists a real number $f'(a)$ and a function $\xi : (-\epsilon, \epsilon) \to \mathbb{R}$ such that $\lim_{t\to 0} \frac{\xi(t)}{t} = 0$ and

$$|f(a+h) - f(a) - f'(a) \cdot h| \leq |\xi(|h|)| \,,$$

for all $|h| < \epsilon$. To see that this is equivalent assume first that f is differentiable at a in the traditional sense. Then define $\xi(t) = f(a+t) - f(a) - f'(a) \cdot t$ This function has clearly the right properties and thus f satisfies this new condition. Conversely if f satisfies the new condition at a. Then

$$0 \leq \left| \frac{f(a+h) - f(a)}{h} - f'(a) \right| \leq \frac{|\xi(|h|)|}{|h|},$$

and thus

$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h} = f'(a),$$

by the squeeze theorem.

We are now ready to define the derivative of a function in several variables.

Definition 12 *Let $A \subset \mathbb{R}^n$ be open, $\mathbf{f} : A \to \mathbb{R}^m$ a function and $\mathbf{a} \in A$. Then $\mathbf{f}$ is differentiable at $\mathbf{a}$ if there is a linear map $T(\mathbf{a}) : \mathbb{R}^n \to \mathbb{R}^m$, and a function $\xi : (-\epsilon, \epsilon) \to \mathbb{R}$ such that $\lim_{t\to 0} \frac{\xi(t)}{t} = 0$ and*

$$|\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - T(\mathbf{a}) \cdot \mathbf{h}| \leq |\xi(|\mathbf{h}|),$$

for all $|\mathbf{h}| < \epsilon$. The operator $T(\mathbf{a})$ is called the derivative of $\mathbf{f}$ at $\mathbf{a}$, and is denoted by $D_{\mathbf{f}}(\mathbf{a})$.

If we choose bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, $D_{\mathbf{f}}(\mathbf{a})$ is of course represented by an $m \times n$ matrix. In general we will assume that we work in the canonical bases, but it is important to realize that the derivative is a linear operator on vector spaces. In the case that $m = 1$ $D_f(\mathbf{a})$ is of ocurse represented by a vector.

Examples:

1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a constant map. I. e. $f(\mathbf{x}) = c$ for all $\mathbf{x} \in \mathbb{R}^n$. Then f is differentiable and $D_f(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathbb{R}^n$. To see this let $\xi(t) = 0$, Then $|f\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - 0 \cdot \mathbf{h}| = 0 \leq \xi(|\mathbf{h}|)$ for all $\mathbf{h}$.

2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a linear map, i. e. $f(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$ for some $\mathbf{a} \in \mathbb{R}^n$ and all $\mathbf{x} \in \mathbb{R}^n$. Then f is differentiable and $D_f(\mathbf{x}) = \mathbf{a}$ for all $\mathbf{x} \in \mathbb{R}^n$. To see this let $\xi(t) = 0$. Then

$$|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - \mathbf{a} \cdot \mathbf{h}| = 0$$

3. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = x^2 + y^2$. f is differentiable with $D_f(x, y) = 2(x, y)$. Let $\mathbf{h} = (h, k)$, then

$$\begin{aligned} &\left|(x+h)^2 + (y+k)^2 - x^2 - y^2 - 2(x, y) \cdot (h, k)\right| \\ &\quad = \left|x^2 + y^2 + 2xh + 2yk + h^2 + k^2 - x^2 - y^2 - 2xh - 2yk\right| \\ &\quad = \left|h^2 + k^2\right| \end{aligned}$$

So we may use $\xi(t) = t^2$ and show that the function is differentiable. The same argument works for $n > 2$.

The examples presented above were all cases of real valued functions. As we will see it is sufficient to understand this case

Proposition 14 *Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ and let $\mathbf{f} = (f_1, \cdots, f_m)$ in the standard basis of R^m. Then $\mathbf{f}$ is differentiable at $\mathbf{x}$ if and only if f_j is differentiable at $\mathbf{x}$ for $j = 1, \dots, m$.*

Proof: Consider the j-th component of

$$(\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - D_{\mathbf{f}}(\mathbf{x}) \cdot \mathbf{h})_j = f_j(\mathbf{x}+\mathbf{h}) - f_j(\mathbf{x}) - (D_{\mathbf{f}})_j \cdot \mathbf{h}.$$

Using the d_∞ metric on $\mathbb{R}^m$ it is clear that

$$|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - D_{\mathbf{f}}(\mathbf{x}) \cdot \mathbf{h}|_\infty \geq |f_j(\mathbf{x}+\mathbf{h}) - f_j(\mathbf{x}) - (D_{\mathbf{f}})_j \cdot \mathbf{h}|,$$

and

$$|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - D_{\mathbf{f}}(\mathbf{x}) \cdot \mathbf{h}|_\infty \leq \max_{j=1,\dots,m} |f_j(\mathbf{x}+\mathbf{h}) - f_j(\mathbf{x}) - (D_{\mathbf{f}})_j \cdot \mathbf{h}|.$$

So we can use the same function ξ for showing the differentiability of the components and the function as a whole. Moreover, $D_{f_j}(\mathbf{x}) = (D_{\mathbf{f}}(\mathbf{x}))_j$.

2.3 Some Properties of Derivatives

Without proof we can state the linearity properties of the derivative:

Proposition 15 *Let f, g be differentiable at some point $\mathbf{x} \in \mathbb{R}^n$. Then for any $a, b \in \mathbb{R}$, $af + bg$ is differentiable at $\mathbf{x}$ and*

$$D_{af+bg}(\mathbf{x}) = aD_f(\mathbf{x}) + bD_g(\mathbf{x}).$$

The proof of this is left as an exercise. More important is the following result:

Proposition 16 *Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $\mathbf{x} \in \mathbb{R}^n$. Then $\mathbf{f}$ is continuous at $\mathbf{x}$*

Proof: Let $\mathbf{h} = \mathbf{y} - \mathbf{x}$ then

$$|\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})| \leq |\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - D_{\mathbf{f}}(\mathbf{x}) \cdot \mathbf{h}| + |D_{\mathbf{f}}(\mathbf{x}) \cdot \mathbf{h}| \leq \xi(|\mathbf{h}|) + |D_{\mathbf{f}}(\mathbf{x}) \cdot \mathbf{h}|.$$

The result follows now immediately from the definition of differentiability and the fact that for any linear map T on $\mathbb{R}^n$ there is a positive constant C such that

$$|T \cdot \mathbf{h}| \leq C|\mathbf{h}|.$$

Proposition 17 *Let f, g be real valued and differentiable at some $\mathbf{x} \in \mathbb{R}^n$ then fg is differentiable and*

$$D_{fg}(\mathbf{x}) = gD_f(\mathbf{x}) + fD_g(\mathbf{x}).$$

Moreover, if $g(\mathbf{x}) \neq 0$, then $\frac{f}{g}$ is differentiable at $\mathbf{x}$ and

$$D_{f/g} = \frac{1}{g^2}\left(gD_f - fD_g\right).$$

Proof: For the first assertion observe that

$$\begin{aligned}
&|f(\mathbf{x}+\mathbf{h})g(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})g(\mathbf{x}) - (g(\mathbf{x})D_f(\mathbf{x}) + f(\mathbf{x})D_g(\mathbf{x}) \cdot \mathbf{h})| \\
&\quad = |g(\mathbf{x}+\mathbf{h})\left(f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})\right) + f(\mathbf{x})\left(g(\mathbf{x}+\mathbf{h}) - g(\mathbf{x})\right) - (g(\mathbf{x})D_f(\mathbf{x}) \cdot \mathbf{h} + f(\mathbf{x})D_g(\mathbf{x}) \cdot \mathbf{h})| \\
&\quad \leq |g(\mathbf{x}+\mathbf{h})\left(f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - D_f(\mathbf{x}) \cdot \mathbf{h}\right)| \\
&\qquad + |f(\mathbf{x})\left(g(\mathbf{x}+\mathbf{h}) - g(\mathbf{x}) - D_g(\mathbf{x}) \cdot \mathbf{h}\right)) \\
&\qquad\quad + |f(\mathbf{x})\left(g(\mathbf{x}) - g(\mathbf{x}+\mathbf{h})\right)|.
\end{aligned}$$

The first two terms converge to 0 as $|\mathbf{h}| \to 0$, because of the differentiability of f and g, the last term converges to) because of the continuity of g. The second part of the Theorem is left as an exercise.

Proposition 18 **Chain Rule** *Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $\mathbf{x}$ with derivative $D_{\mathbf{g}}(\mathbf{x})$, and $f : \mathbb{R}^m \to \mathbb{R}$ be differentiable at $\mathbf{y} = \mathbf{g}(\mathbf{x})$ with derivative $D_f(\mathbf{y})$. Then $f \circ \mathbf{g}$ is differentiable at $\mathbf{x}$, and*

$$D_{f\circ\mathbf{g}}(\mathbf{x}) = D_f(\mathbf{y}) \cdot D_{\mathbf{g}}(\mathbf{x}).$$

Proof: Let $\mathbf{k} = \mathbf{g}(\mathbf{x}+\mathbf{h}) - \mathbf{g}(\mathbf{x})$, Then $g(\mathbf{x}+\mathbf{h}) = \mathbf{y} + \mathbf{k}$. We have

$$\begin{aligned}
f(\mathbf{g}(\mathbf{x}+\mathbf{h})) - f(\mathbf{g}(\mathbf{x})) &= f(\mathbf{y}+\mathbf{k}) - f(\mathbf{y}) \\
&= D_f(\mathbf{y}) \cdot \mathbf{k} + \xi_1(|\mathbf{k}|) \\
&= D_f(\mathbf{y}) \cdot (\mathbf{g}(\mathbf{x}+\mathbf{h}) - \mathbf{g}(\mathbf{x})) + \xi_1(|\mathbf{k}|) \\
&= D\left(D_{\mathbf{g}} \cdot \mathbf{h} + \xi_2(|\mathbf{h}|)\right) + \xi_1(|\mathbf{k}|).
\end{aligned}$$

Now define

$$\xi(|\mathbf{h}| = D_f(\mathbf{y}) \cdot \xi_2(|\mathbf{h}|) + \xi_1(|\mathbf{k}|),$$

which satisfies the right properties since $|\mathbf{k}| \to 0$ as $|\mathbf{h}| \to 0$. The result follows immediately.

2.4 Assignment 4

1. Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ be both differentiable at $\mathbf{x}$. Show that the skalar product $\mathbf{f} \cdot \mathbf{g}$ is also differentiable at $\mathbf{x}$.
2. Prove the linearity of the derivative Proposition 15.
3. Prove the quotient rule in Proposition 17.
4. Let $f : \mathbb{R}^4 \to \mathbb{R}$ be given by $f(x_1, x_2, x_3, x_4) = x_1x_4 - x_2x_3$. Prove that f is differentiable and that $D_f(x_1, x_2, x_3, x_4) = (x_4, -x_3, -x_2, x_1)$.
5. The space of 2×2 matrices over $\mathbb{R}$ is a four dimensional vector space in a natural way. Show that the determinant is a differentiable function from this space to the real numbers.
6. Show that the set of invertible 2×2 real matrices is an open subset of the space of 2×2 real matrices.

2.5 Directional and Partial Derivatives

Definition 13 *Let* $\mathbf{u}$ *be a unit vector in* $\mathbb{R}^n$*, and* $f : \mathbb{R}^n \to \mathbb{R}$ *be a function. The directional derivative in direction* $\mathbf{u}$ *of* f *at* $\mathbf{x}$ *is defined as*

$$\partial_{\mathbf{u}} f(\mathbf{x}) = \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x})}{t}.$$

If $\mathbf{u}$ *is a unitvector in the canonical basis of* $\mathbb{R}^n$*, i. e.* $\mathbf{u} = \mathbf{e}_j$ *for some* $1 \leq j \leq n$*, then this is called the partial derivative with respect to* x_j *and denoted by*

$$\frac{\partial f(\mathbf{x})}{\partial x_j} \qquad \text{or} \qquad f_{x_j}(\mathbf{x})$$

An easy consequence is:

Lemma 3 *Suppose* $f : \mathbb{R}^n \to \mathbb{R}$ *is differentiable at* $\mathbf{x}$*. Then for any unit vector* $\mathbf{u} \in \mathbb{R}^n$ *the directional derivative in direction* $\mathbf{u}$ *exists and*

$$\partial_{\mathbf{u}} f(\mathbf{x}) = D_f(\mathbf{x}) \cdot \mathbf{u}$$

Proof: Consider the function $\mathbf{g} : \mathbb{R} \to \mathbb{R}^n$ given by $\mathbf{g} : t \mapsto \mathbf{x} + t\mathbf{u}$. The directional derivative at $\mathbf{x}$ is the derivative of the function $F = f \circ \mathbf{g} : \mathbb{R} \to \mathbb{R}$ at 0. since $\mathbf{g}$ is differentiable and $D_{\mathbf{g}}(0) = \mathbf{u}$ the result follows immediately from the chain rule proven above.

In the interest of computing derivatives it would be good to have the converse of this last Lemma. I. e. we would like to say that f is differentiable if all partial derivatives exist. Unfortunately that is not the case as the next example shows.

Example: Let f be defined as follows:

$$f(x_1,x_2)=\begin{cases}0 & |x_2|\le\frac{1}{3}|x_1|\\ 1 & \frac{1}{3}|x_1|<|x_2|<3|x_1|\\ 0 & 3|x_1|\le|x_2|\end{cases}$$

Then f is clearly not differentiable at $(0,0)$ since it isn't even continuous. But

$$\frac{\partial f}{\partial x_1}(0,0)=\frac{\partial f}{\partial x_2}(0,0)=0.$$

Before we can prove a partial converse we will introduce a mean value theorem. The mean value theorems of single variable calculus do not apply since there is no natural ordering on $\mathbb{R}^n$ that is compatible with the metric space topology. We start with the following

Definition 14 *A set $C\subset\mathbb{R}^n$ is called convex if for every $\mathbf{x},\mathbf{y}\in C$ and every $s\in[0,1]$ we have*

$$\mathbf{x}(1-s)+\mathbf{y}s\in C.$$

Losely speaking this means that for any two points $\mathbf{x}$ and $\mathbf{y}$ in C, the line segment connecting these two points is also in C.

Proposition 19 *Let C be a convex subset of $\mathbb{R}^n$ and $f:C\to\mathbb{R}$ be differentiable in C with continuous derivative $D_f(\mathbf{x})$. Letf $\mathbf{x},\mathbf{y}\in C$. Then there exist an $s\in[0,1]$ such that*

$$f(\mathbf{x})-f(\mathbf{y})=D_f(\mathbf{x}s+\mathbf{y}(1-s))\cdot(\mathbf{x}-\mathbf{y}).$$

Proof: Consider the function $F(t)=f(\mathbf{x}t+\mathbf{y}f(1-t))$ on $[0,1]$ and apply chain rule and mean value theorem for 1 variable.

Theorem 2 *Let f be continuous on a ball $B_\epsilon(\mathbf{x})\subset\mathbb{R}^n$, with continuous partial derivatives on $B_\epsilon(\mathbf{x})$. Then f is differentiable at $\mathbf{x}$. Moreover, $D_f(\mathbf{x})=(f_{x_1}(\mathbf{x}),\ldots,f_{x_n}(\mathbf{x}))$.*

Proof: We will do the proof in two variables, the general case workds exactly the same way, but is messier to write down. Let $\mathbf{h}=(h_1,h_2)$ and $|\mathbf{h}|<\epsilon$. Observe that the ball of raidus ϵ is convex. By the mean value theorem (used twice) there exist $s,t\in(0,1)$ such that

$$\begin{aligned}f(\mathbf{x}+\mathbf{h})-f(\mathbf{x}) &= f(x_1+h_1,x_2+h_2)-f(x_1,x_2)\\ &= f(x_1+h_1,x_2+h_2)-f(x_1,x_2+h_2)+f(x_1,x_2+h_2)-f(x_1,x_2)\\ &= f_{x_1}(x_1+sh_1,x_2+h_2)h_1+f_{x_2}(x_1,x_2+th_2)h_2\end{aligned}$$

Thus

$$\begin{aligned}
&f(\mathbf{x}+\mathbf{h})-f(\mathbf{x})-f_{x_1}(x_1,x_2)h_1-f_{x_2}(x_1,x_2)h_2\\
&\qquad=(f_{x_1}(x_1+sh_1,x_2+h_2)-f_{x_1}(x_1,x_2))h_1\\
&\qquad\quad+(f_{x_2}(x_1,x_2+th_2)-f_{x_2}(x_1,x_2))h_2\\
&\qquad=\xi(h_1,h_2).
\end{aligned}$$

Next observe that the right hand side of this equation satisfies

$$|\xi(h_1,h_2)|\leq(|f_{x_1}(x_1+sh_1,x_2+h_2)-f_{x_1}(x_1,x_2)|+|f_{x_2}(x_1,x_2+th_2)-f_{x_2}(x_1,x_2)|)\,|\mathbf{h}|,$$

and thus

$$\lim_{|\mathbf{h}|\to 0}\frac{|\xi(h_1,h_2)|}{|\mathbf{h}|}=0,$$

by the continuity of the partial derivatives.

This theorem now allows us to easily compute derivatives for differentiable functions, by computing the partial derivatives.

Proposition 20 *Suppose that f is differentiable on $B_\epsilon(\mathbf{x}_0)\subset\mathbb{R}^n$. If f has a local extremum at $\mathbf{x}_0$ then $D_f(\mathbf{x}_0)=0$.*

Proof: We assume that f has a local maximum at $\mathbf{x}_0$, and that $f(\mathbf{x}_0)\geq f(\mathbf{x})$ for all $\mathbf{x}\in B_\epsilon(\mathbf{x}_0)$. Let $\mathbf{u}$ be any unit vector in $\mathbb{R}^n$ and $0<h<\epsilon$. Then

$$0\geq f(\mathbf{x}_0+\mathbf{u}h)-f(\mathbf{x}_0)=D_f(\mathbf{x}_0+\lambda\mathbf{u}h)\cdot\mathbf{u}h,$$

for some $\lambda\in(0,1)$. By changing h to $-h$ we have

$$0\geq f(\mathbf{x}_0-\mathbf{u}h)-f(\mathbf{x}_0)=D_f(\mathbf{x}_0+\lambda\mathbf{u}(-h))\cdot\mathbf{u}(-h),$$

Letting $h\to 0$ we get

$$\partial_{\mathbf{u}}f(\mathbf{x}_0)\leq 0$$

from the first inequality, and

$$\partial_{\mathbf{u}}f(\mathbf{x}_0)\geq 0$$

from the second one. Thus

$$D_f(\mathbf{x}_0)\cdot\mathbf{u}=0,$$

for all unit vectors $\mathbf{u}\in\mathbb{R}^n$, and the result follows.

2.6 Assignment 5:

1. Let $\phi : \mathbb{R} \to [0, \infty)$ be a differentiable function. Define $F : \mathbb{R}^3 \to [0, \infty)$ by $F(x_1, x_2, x_3) = \phi(\sqrt{x_1^2 + x_2^2 + x_3^2})$ Show that
$$\phi'(\sqrt{x_1^2 + x_2^2 + x_3^2}) = \left(\left(\frac{\partial F}{\partial x_1}\right)^2 + \left(\frac{\partial F}{\partial x_2}\right)^2 + \left(\frac{\partial F}{\partial x_3}\right)^2\right)^{\frac{1}{2}}$$

2. Let
$$f(x_1, x_2) = \begin{cases} \frac{x_1^2 x_2}{x_1^4 + x_2^2} & x_2 \neq 0 \\ 0 & x_2 = 0 \end{cases}$$
Prove that all directional derivatives of f exist at $(0, 0)$ but that f is neither dfifferentiable nor continuous at this point.

3. Suppose $y = F(x_1, x_2)$ is differentiable at (p_1, p_2) with $F_{x_2}(p_1, p_2) \neq 0$. Let I be an open interval containing p_1. If $f : I \to \mathbb{R}$ is differentiable and $F(x_1, f(x_1)) = 0$ for all $x_1 \in I$ then
$$f'(p_1) = \frac{-F_{x_1}(p_1, p_2)}{F_{x_2}(p_1, p_2)}.$$

4. Show that
$$f(x_1, x_2) = \begin{cases} (x_1 x_2)^\alpha \log(x_1^2 + x_2^2) & (x_1, x_2) \neq (0, 0) \\ 0 & (x_1, x_2) = (0, 0) \end{cases}$$
is differentiable on $\mathbb{R}^2$ for all $\alpha > 1/2$.

5. Prove that
$$f(x_1, x_2) = \begin{cases} \frac{x_1^4 + x_2^4}{(x_1^2 + x_2^2)^\alpha} & (x_1, x_2) \neq (0, 0) \\ 0 & (x_1, x_2) = (0, 0) \end{cases}$$
is differentiable for all $\alpha < 3/2$.

6. Investigate the case when $\alpha = 1/2$ in problem 4 and $\alpha = 3/2$ in problem 5.

2.7 Higher order derivatives:

Let $O \subset \mathbb{R}^n$ be open and $\mathbf{f} : O \to \mathbb{R}^m$ be a function that is differentiable for all $\mathbf{x} \in O$, then $D_\mathbf{f}$ is itself a function defined on O. The values of this function are $m \times n$ matrices, which can be thought of as elements of $\mathbb{R}^{mn}$. We can thus look at the differentiability of this function. We observe, that everytime we take a derivative the dimension of the range increases by a factor n, and things will get confusing quickly. If f is areal valued function on $\mathbb{R}^3$ then its fourth derivative will have vaues in $\mathbb{R}^{81}$! Mindful of this we restrict our discussion to derivatives up to the second order, and real valued functions. For a real-valued function on $\mathbb{R}^n$ the second derivative is an $n \times n$ matrix. We will prove that under certain conditions this matrix is symmetric. Of course we also have higher order directional derivatives an partial derivatives. At this time we make the following definition which really belongs into a linear algebra course.

Definition 15 *Let V be a vector space over $\mathbb{R}$ an n-linear form is a map.*

$$\phi : V^n \to \mathbb{R},$$

which is linear in each of its components. That is for $j = 1, \ldots, n$, and $a, b \in \mathbb{R}$

$$\phi(\mathbf{v}_1, \ldots, a\mathbf{v}_j + b\mathbf{u}_j, \ldots, \mathbf{v}_n) = a\phi(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n) + b\phi(\mathbf{v}_1, \ldots, \mathbf{u}_j, \ldots, \mathbf{v}_n)$$

Now it easy to see that for $f : \mathbb{R}^n \to \mathbb{R}$ the first derivative is a 1-linear form, the second a bilinear form and in general the n-th an n-linear form. Also it is lclear that in any given basis, a 1-linear form is given by the inner product by a vector so

$$\phi(\mathbf{v}) = \mathbf{a} \cdot \mathbf{v}$$

for some fixed vector $\mathbf{a}$ and all $\mathbf{v} \in V$. Similarly a bilinear form is givesn by a matrix A such that:

$$\phi(\mathbf{v}, \mathbf{u}) = \mathbf{v}^T \cdot A\mathbf{u}.$$

For higher order derivatives it is not so easy to see how it is represented. But in order to make the notation easier we write

$$D_f^{(k)}(\mathbf{x})$$

for the k-th derivative of f at $\mathbf{x}$ and keep in mid that this is a k-linear form. If f is k times differentiable then the k-th directional derivative at $\mathbf{x}$ in direction $\mathbf{u}$ is given by

$$D_f^{(k)}(\mathbf{x})(\mathbf{u}, \ldots, \mathbf{u}) = D_f^{(k)}(\mathbf{x})(\mathbf{u}),$$

i.e all k arguments of the k-linear form are evaluated at $\mathbf{u}$.

The crucial result in this section is

Theorem 3 **Schwartz's Theorem**. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 (i.e. continuously differentiable and if*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}$$

exists and is continuous on an open set $U \subset \mathbb{R}^n$, then

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x})$$

exists and

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})$$

for all $\mathbf{x} \in U$.

Proof: Suppose that $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})$ exists at $\mathbf{x} \in U$. Let $\epsilon > 0$ such that $B_\epsilon(\mathbf{x}) \subset U$ in the d_∞ metric. For $|h_i|, |h_j| < \epsilon$ define

$$\begin{aligned}\Delta(h_i, h_j) &= f(x_1, \ldots, x_i + h_i, \ldots, x_j + h_j, \ldots, x_n) - f(x_1, \ldots, x_i + h_i, \ldots, x_j, \ldots, x_n) \\ &\quad - f(x_1, \ldots, x_i, \ldots, x_j + h_j, \ldots, x_n) + f(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n).\end{aligned}$$

Applying the Mean Value Theorem twice yields that there is $t, s \in (0,1)$ such that

$$\begin{aligned}\Delta(h_i, h_j) &= h_j \frac{\partial f}{\partial x_j}(x_1, \ldots, x_i + h_i, \ldots, x_j + th_j, \ldots, x_n) - h_j \frac{\partial f}{\partial x_j}(x_1, \ldots, x_i, \ldots, x_j + th_j, \ldots, x_n) \\ &= h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \ldots, x_i + sh_i, \ldots, x_j + th_j, \ldots, x_n). \frac{\Delta(h_i, h_j)}{h_i h_j} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}).\end{aligned}$$

Since this is mixed partial is continuous we have that

$$\lim_{h_j \to 0} \lim_{h_i \to 0} \frac{\Delta(h_i, h_j)}{h_i h_j} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}).$$

On the other hand we may use the Mean Value Theorem to get

$$\Delta(h_i, h_j) = h_i \left(\frac{\partial f}{\partial x_i}(x_1, \ldots, x_i + uh_i, \ldots, x_j + h_j, \ldots, x_n) - \frac{\partial f}{\partial x_i}(x_1, \ldots, x_i + uh_i, \ldots, x_j, \ldots, x_n) \right),$$

for some $u \in (0,1)$ and

$$\lim_{h_i \to 0} \frac{\Delta(h_i, h_j}{h_i} = \left(\frac{\partial f}{\partial x_i}(x_1, \ldots, x_i, \ldots, x_j + h_j, \ldots, x_n) - \frac{\partial f}{\partial x_i}(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) \right),$$

by the continuity of $\frac{\partial f}{\partial x_i}$ Thus

$$\lim_{h_j \to 0} \lim_{h_i \to 0} \frac{\Delta(h_i, h_j)}{h_i h_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}),$$

and the statement is proven.

2.8 Taylor's Formula:

Theorem 4 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be k-times continuously differentiable on an open convex set S containing the point a. Then for every $\mathbf{x} \in S$ there is a $t \in [0,1]$ such that*

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{j=1}^{k-1} \frac{1}{j!} D_f^{(j)}(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \frac{1}{k!} D_f^{(k)}((1-t)\mathbf{a} + t\mathbf{x}))(\mathbf{x} - \mathbf{a}).$$

Proof: Let

$$\mathbf{u} = \frac{\mathbf{x} - \mathbf{a}}{|\mathbf{x} - \mathbf{a}|},$$

and consider $F(s) = f(\mathbf{a} + \mathbf{u}s)$. This function is k times continuously differentiable on an open interval $(-r, R)$ which contains $[0, |\mathbf{x} - \mathbf{a}|]$. Moreover, $F(|\mathbf{x} - \mathbf{a}|) = f(\mathbf{x})$. The Taylor Formula for functions on the reals implies that

$$F(|\mathbf{x} - \mathbf{a}|) = F(0) + \sum_{j=1}^{k-1} \frac{1}{j!} F^{(j)}(0)(|\mathbf{x} - \mathbf{a}|)^j + \frac{1}{k!} F^{(k)}(\xi)(|\mathbf{x} - \mathbf{a}|)^k,$$

for some $\xi \in [0, |\mathbf{x}-\mathbf{a}|]$. From the form of the directional derivatives explained earlier we get

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{j=1}^{k-1} \frac{1}{j!} D_f^{(j)}(\mathbf{a})(\mathbf{u})(|\mathbf{x}-\mathbf{a}|)^j + \frac{1}{k!} D_f^{(k)}(\mathbf{a}+\mathbf{u}\xi)(\mathbf{u})(|\mathbf{x}-\mathbf{a}|)^k.$$

The result now follows immediately from the multilinearity of the derivatives and the fact that

$$\mathbf{a} + \xi\mathbf{u} = (1-t)\mathbf{a} + t\mathbf{x},$$

for some $t \in [0,1]$.

2.9 Assignment 6:

1. Consider the map $[0,\infty) \times [0, 2\pi) \to \mathbb{R}^2$, defined by $(x(r,\theta), y(r,\theta)) = (r\cos\theta, r\sin\theta)$. Show that this map is differentiable and compute its derivative. Moreover, prove that the map is onto. Is it also one-to-one?
2. Let $f(x,y)$ be a twice differentiable function on $\mathbb{R}^2$ such that $f_{xx} + f_{yy} = 0$ on $\mathbb{R}^2$. Define $\phi(r,\theta) = f(r\cos\theta, r\sin\theta)$.
   (a) Compute
   $$\frac{\partial\phi}{\partial r} \quad \text{and} \quad \frac{\partial\phi}{\partial\theta}$$
   (b) Prove that
   $$\frac{1}{r^2}\frac{\partial^2\phi}{\partial\theta^2} + \frac{1}{r}\frac{\partial\phi}{\partial r} + \frac{\partial^2\phi}{\partial r^2} = 0.$$
3. Consider $f(x_1,x_2) = \sin(x_1^2 + x_2^2)$. Compute the first three terms of the Taylor expansion of this function.

3 THE INVERSE AND IMPLICIT FUNCTION THEOREMS

3.1 The Inverse Function Theorem

If $f : (a,b) \to \mathbb{R}$ is a differentiable function such that $f'(x) \neq 0$ for all $x \in (a,b)$, we know that f is one-to-one and onto $f((a,b))$ with a differentiable inverse function and $(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}$. We would like to generalize this to higher dimensions. First we need to have an idea how to generalize the statemant $f'(x) \neq 0$. Let us start by looking at the following example:

Example:

Let $f(x_1,x_2) = x_1^2 + x_2^2$ then $f'(x_1,x_2) = (2x_1, 2x_2) \neq (0,0)$ except at the origin. But this function is clearly not a one-to-one function. If however, we interpret the derivative as a linear map $\mathbb{R}^n \to \mathbb{R}^m$ then we can interpret the result for $\mathbb{R}$ also as the map $f'(x) : \mathbb{R} \to \mathbb{R}$ defined by $f'(x) : h \to f'(x) \cdot h$, is not singular.

Definition 16 *Let V and W be vector spaces. A linear map $T : V \to W$ is non-singular if* $\ker T = \{0\}$.

An immediate consequence is that if $T : V \to W$ is non-singular and onto then V and W must have the same dimension. Moreover, non singular linear maps are one-to-one. This suggests that the actual property of the derivative which guarantees the injectiveness is the non-singularity of this linear map. A direct consequence is that the result from calculus can atmost be generalize to functions where the domain and the range have the same dimension. We continue with another example.

Example:

Let $\mathbf{f} : (0, \infty) \times (-2\pi, 2\pi)$ be given by $(r, \theta) \mapsto (r\cos\theta, r\sin\theta)$. The the derivative is given by the matrix

$$D_{\mathbf{f}}(r,\theta) \begin{bmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{bmatrix}.$$

Observe that $\det D_{\mathbf{f}}(r,\theta) = r \neq 0$, and therefore the derivative is not singular, but the function is certainly not one-to-one, since, for example, $\mathbf{f}(1, -\pi) = (-1, 0) = \mathbf{f}(1, \pi)$.

This last example shows that we can certainly not have a generalization of the one dimensional case to higher dimensions which gives us global injectivity (like in one dimension) the best we can hope for is a local result.

Theorem 5 Inverse Function Theorem *Let $V \subset \mathbb{R}^n$ be open and $\mathbf{f} : V \to \mathbb{R}^n$ be continuously differentiable on v. Let $\mathbf{x}_0 \in V$ such that $D_{\mathbf{f}}(\mathbf{x}_0)$ is non-sinsgular. The there exist open set $V_0 \subset V$ and $W_0 \subset \mathbf{f}(V)$ such that*

1. *$\mathbf{x}_0 \in V_0$ and $\mathbf{f}(\mathbf{x}_0) \in W_)$.*
2. *$\mathbf{f}$ is a bijection from V_0 onto W_0, and $\mathbf{f}^{-1}$ is a bijection from V_0 onto W_0.*
3. *$\mathbf{f}^{-1}$ is continuously differentiable on W_0, and*
4. *for each $\mathbf{y} = \mathbf{f}(\mathbf{x}) \in W_0$ we have*

   $$D_{\mathbf{f}^{-1}}(\mathbf{y}) - [D_{\mathbf{f}}(\mathbf{x})]^{-1},$$

   where the expression on the right denotes the inverse operator.

Before proving the theorem let me remark that we really do not want much more than this. The Theorem guarantees that the equation

$$\mathbf{f}(\mathbf{x}) = \mathbf{y},$$

has a unique solution in neighborhood of $\mathbf{y}_0$. In practice we may want to solve this equation using Newton's method (see down below). And the theorem guarantees that errors will propagate in a predictable way.

Proof of the Theorem: The proof of this Theorem is rather lengthy and we dicide into several steps. In the first step we reduce the problem to one which has asimpler derivative.

STEP 1: Consider the function $\Phi(\mathbf{x}) = (D_{\mathbf{f}}(\mathbf{x}_0))^{-1} \circ \mathbf{f}(\mathbf{x})$. Then Φ is one-to-one if and only if $\mathbf{f}$ is and $D_\Phi(\mathbf{x}_0) = I$, the identity operator on $\mathbb{R}^n$. So we assume without loss of generality that $D_{\mathbf{f}}(\mathbf{x}_0) = I$.

STEP 2: It follows that if $\mathbf{f}(\mathbf{x}_0+\mathbf{h})=\mathbf{f}(\mathbf{x}_0)$ we have

$$\frac{|\mathbf{f}(\mathbf{x}_0+\mathbf{h})=\mathbf{f}(\mathbf{x}_0)-I\cdot\mathbf{h}|}{|\mathbf{h}|}=\frac{|\mathbf{h}|}{|\mathbf{h}|}=1.$$

But since $\mathbf{f}$ is differentiable at $\mathbf{x}_0$ with derivative I, we must have

$$\lim_{|\mathbf{h}|\to 0}\frac{|\mathbf{f}(\mathbf{x}_0+\mathbf{h})=\mathbf{f}(\mathbf{x}_0)-I\cdot\mathbf{h}|}{|\mathbf{h}|}=0.$$

Thus there is a closed rectangle R such that

$$\mathbf{f}(\mathbf{x})\neq\mathbf{f}(\mathbf{x}_0) \tag{8}$$

for all $\mathbf{x}$ in the interior of R. Since the determinant is a continuous function and $\det D_{\mathbf{f}}(\mathbf{x}_0)=1$ there exists an open set around $\mathbf{x_0}$ such that

$$\det D_{\mathbf{f}}(\mathbf{x})\neq 0, \tag{9}$$

on that open set. Finally, since the partial derivatives of $\mathbf{f}$ are continuous at $\mathbf{x}_0$ there is an open set such that

$$\left|\frac{\partial f_i}{\partial x_j}(\mathbf{x})-\frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)\right|<\frac{1}{2n^2}, \tag{10}$$

for all $i,j=1,\ldots,n$. Without loss of generality we assume that (8-10) all hold on the interior of R.

STEP 3: (10) also implies that

$$\left|\frac{\partial f_i}{\partial x_j}(\mathbf{x})\right|<1+\frac{1}{2}n^2,$$

on R. Now let $\mathbf{g}(\mathbf{x})=\mathbf{f}(\mathbf{x})-\mathbf{x})$, then (10) implies that

$$\left|\frac{\partial g_i}{\partial x_j}(\mathbf{x})\right|<\frac{1}{2n^2}$$

and by the mean value theorem

$$|\mathbf{g}(\mathbf{x}_1)-\mathbf{g}(\mathbf{x}_2)|\leq\frac{1}{2}|\mathbf{x}_1-\mathbf{x}_2|. \tag{11}$$

From the triangle inequality we get:

$$\begin{aligned}|\mathbf{x}_1-\mathbf{x}_2|-|\mathbf{f}(\mathbf{x}_1)-\mathbf{f}(\mathbf{x}_2)| &\leq |\mathbf{f}(\mathbf{x}_1)-\mathbf{x}_1-(\mathbf{f}(\mathbf{x}_2)-\mathbf{x}_2)|\\ &\leq \frac{1}{2}|\mathbf{x}_1-\mathbf{x}_2|.\end{aligned}$$

It immediately follows that for $\mathbf{x}_1,\mathbf{x}_2\in R$ we have

$$|\mathbf{x}_1-\mathbf{x}_2|\leq 2|\mathbf{f}(\mathbf{x}_1)-\mathbf{f}(\mathbf{x}_2)|. \tag{12}$$

Observe that this last inequality will directly imply that the inverse function is Lipschitz continuous, if it exists. We will now show that $\mathbf{f}$ is one-to-one on a subset of R. First we observe that the boundary of R is compact, and therefore $\mathbf{f}(\partial R)$ is compact as well. Moreover, by (8), $\mathbf{f}(\mathbf{x}_0)\notin f(\partial R)$. Thus there is a number $d.0$ such that

$$|\mathbf{f}(\mathbf{x}_0)-\mathbf{f}(\mathbf{x})|\geq d,$$

for all $\mathbf{x} \in \partial R$. Define

$$W = \left\{ \mathbf{y} : |\mathbf{y} - \mathbf{f}(\mathbf{x}_0)| < \frac{d}{2} \right\}.$$

For any $\mathbf{y} \in W$ and any $\mathbf{x} \in \partial R$ we have

$$|\mathbf{y} - \mathbf{f}(\mathbf{x}_0)| < |\mathbf{y} - \mathbf{f}(\mathbf{x})|. \tag{13}$$

STEP 4:

For a given $\mathbf{y} \in W$ define

$$g(\mathbf{x}) = |\mathbf{y} - \mathbf{f}(\mathbf{x})|^2.$$

This is a continuously differentiable function on R and therefore has a minimum on R. By (13) we have

$$g(\mathbf{x}_0) < g(\mathbf{x})$$

for all $\mathbf{x} \in \partial R$, and therefore this minimum msut occur in the interior of R, and at this minimum we have $D_g(x) = 0$. A simple computation gives

$$D_g(x) = (\mathbf{y} - \mathbf{f}(\mathbf{x})) \cdot D_{\mathbf{f}}(\mathbf{x}), \tag{14}$$

and since $D_{\mathbf{f}}(\mathbf{x})$ is non-singular this expression can only vanish if $\mathbf{y} - \mathbf{f}(\mathbf{x}) = 0$, i.e. if there exists an $\mathbf{x} \in R$ such that $\mathbf{y} = \mathbf{f}(\mathbf{x})$. (12) immediately implies the uniqueness of such a solution. Define

$$V = \operatorname{int} R \cap \mathbf{f}^{-1}(W).$$

Then for every $\mathbf{y} \in W$ there is a unique $\mathbf{x} \in V$ such that $\mathbf{y} = \mathbf{f}(\mathbf{x})$, or the function $\mathbf{f}$ has an inverse function on W As mentioned above (12) implies that $\mathbf{f}^{-1}$ is continuous on W. It remains to be shown that this function is also differentiable.

STEP 5: To show that $\mathbf{f}^{-1}$ is differentiable observe that for $\mathbf{h}$ sufficiently small

$$\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}) = D_{\mathbf{f}}(\mathbf{x}) \cdot (\mathbf{x} - 1 - \mathbf{x}) + \xi(\mathbf{x}_1 - \mathbf{x})$$

where

$$\lim_{\mathbf{h} \to 0} \frac{|\xi(\mathbf{h})|}{|\mathbf{h}|} = 0. \tag{15}$$

If we apply $(D_{\mathbf{f}}(\mathbf{x}))^{-1}$ to this equation we get;

$$(D_{\mathbf{f}}(\mathbf{x}))^{-1} \cdot (\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x})) = \mathbf{x}_1 - \mathbf{x} + (D_{\mathbf{f}}(\mathbf{x}))^{-1} \xi(\mathbf{x}_1 - x),$$

which we can rewrite as

$$\mathbf{f}^{-1}(\mathbf{y}_1) - \mathbf{f}^{-1}(\mathbf{y}) = (D_{\mathbf{f}}(\mathbf{x}))^{-1} \cdot (\mathbf{y}_1 - \mathbf{y}) - (D_{\mathbf{f}}(\mathbf{x}))^{-1} \xi(\mathbf{f}^{-1}(\mathbf{y}_1) - \mathbf{f}^{-1}(\mathbf{y})),$$

and we are only left to prove:

$$\lim_{\mathbf{y}_1\to\mathbf{y}} \frac{\left|(D_{\mathbf{f}}(\mathbf{x}))^{-1}\xi(\mathbf{f}^{-1}(\mathbf{y}_1)-\mathbf{f}^{-1}(\mathbf{y}))\right|}{|\mathbf{y}_1-\mathbf{y}|}=0.$$

To do this it is clear that this follows if

$$\lim_{\mathbf{y}_1\to\mathbf{y}} \frac{\left|\xi(\mathbf{f}^{-1}(\mathbf{y}_1)-\mathbf{f}^{-1}(\mathbf{y}))\right|}{|\mathbf{y}_1-\mathbf{y}|}=0,$$

since the derivative is a linear operator. Now observe that

$$\frac{\left|\xi(\mathbf{f}^{-1}(\mathbf{y}_1)-\mathbf{f}^{-1}(\mathbf{y}))\right|}{|\mathbf{y}_1-\mathbf{y}|}=\frac{\left|\xi(\mathbf{f}^{-1}(\mathbf{y}_1)-\mathbf{f}^{-1}(\mathbf{y}))\right|}{|\mathbf{f}^{-1}(\mathbf{y}_1)-\mathbf{f}^{-1}(\mathbf{y})|}\cdot\frac{\left|\mathbf{f}^{-1}(\mathbf{y}_1)-\mathbf{f}^{-1}(\mathbf{y})\right|}{|\mathbf{y}_1-\mathbf{y}|} \tag{16}$$

By the continuity of $\mathbf{f}^{-1}$ and (15 the first fraction will converge to zero as $\mathbf{y}_1\to\mathbf{y}$. Finally (13 implies that the second fraction is bounded by 2, and thus the product converges to zero.

3.2 Assignment 7:

1. Let $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^n$ be such that $|D_{\mathbf{f}}(\mathbf{x})\cdot\mathbf{u}|=|\mathbf{u}|$ for all $\mathbf{x},\mathbf{u}\in\mathbb{R}^n$. Prove that $\mathbf{f}$ is one-to-one on all of $\mathbb{R}^n$ and onto. Moreover, show that $\mathbf{f}(\mathbb{R}^n)$ is both open and closed and therefore $\mathbf{f}(\mathbb{R}^n)=\mathbb{R}^n$. Conclude that $\mathbf{f}$ has a global differentiable inverse which is also an isometry.

2. In this problem we generalize the global result from above. Let $S\subset\mathbb{R}^n$ be open and $\mathbf{f}:\S\to\mathbb{R}^n$ such that there exists $k_1,k_2>0$ such that

   $$k_{\leq}|D_{\mathbf{f}}(\mathbf{x}0\cdot\mathbf{u}|\leq k_2$$

   for all $\mathbf{x}\in S$ and all unit vectors $\mathbf{u}\in\mathbb{R}^n$. Prove that $\mathbf{f}(S)$ is open and that $\mathbf{f}$ is one-to-one and onto $\mathbf{f}(S)$ with a differentiable inverse function.

3. Define $\mathbf{f}:\mathbb{R}^2\to\mathbf{r}^2$ by $\mathbf{f}(x,y)=(e^x\cos y,e^x siny)$ show that $D_{\mathbf{f}}(\mathbf{x})$ is non-singular for all $\mathbf{x}=(x,y)\in\mathbb{R}^2$, but $\mathbf{f}$ is not one-to-one.

3.3 The Implicit Function Theorem

We continue next to the Implicit Function Theorem. To start we all recall problems from calculus of the following form:

Let $y^3-x^2=2$ Find the slope of the tangent line to this curve at the point $(-2,2)$. We learned to solve them by differentiating this equation under the assumption that y is a differentiable function of x to get:

$$3y^2y'(x)-2x=0$$

Next we evaluated x and y at the given point to get $12y'(-2) + 4 = 0$ and get $y'(-2) = -1/3$. The Implicit Function Theorem will give us specific conditions when we can do this and generalize this to higher dimensions. We need to introduce some new notation to simplify this situation.

Let $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be differentable at a point $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^m$. We denote by

$$\partial_{\mathbf{x}} \mathbf{f}(\mathbf{x}, \mathbf{y}) \qquad \text{and} \qquad \partial_{\mathbf{y}} \mathbf{f}(\mathbf{x}, \mathbf{y})$$

the drivatives of $\mathbf{f}$ with respect to the $\mathbf{x}$ and $\mathbf{y}$ variables. If we write $D_{\mathbf{f}}(\mathbf{x}, \mathbf{y})$ as a matrix of partial derivatives, these represent the submatrices consisting of the first n and the last m columns, respectively. And

$$D_{\mathbf{f}}(\mathbf{x}, \mathbf{y}) = (\partial_{\mathbf{x}} \mathbf{f}(\mathbf{x}, \mathbf{y}), \partial_{\mathbf{y}} \mathbf{f}(\mathbf{x}, \mathbf{y})).$$

Using this notation we may formulate the first version of the theorem.

Theorem 6 Implicit Function Theorem. *Let*

$$\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$$

be a continuously differentiable at $(\mathbf{x}_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$ *such that* $\mathbf{f}(\mathbf{x}_0, \mathbf{y}_0) = 0$ *and* $\partial_{\mathbf{y}} \mathbf{f}(\mathbf{x}_0, \mathbf{y}_0)$ *is non-singular. Then there exists an open set* $O \subset \mathbb{R}^n$ *and and open set* $w \subset \mathbb{R}^m$ *such that* $(\mathbf{x}_0, \mathbf{y}_0) \in O \times W$ *and for each* $\mathbf{x} \in O$ *there exists a unique* $\mathbf{g}(\mathbf{x}) \in W$ *such that* $\mathbf{f}(\mathbf{x}, \mathbf{g}(\mathbf{x})) = 0$. *Moreover, the function* $\mathbf{g} : O \to W$ *is differentiable and*

$$D_{\mathbf{g}}(\mathbf{x}) = -(\partial_{\mathbf{y}} \mathbf{f}(\mathbf{x}, \mathbf{y}))^{-1} \cdot \partial_{\mathbf{x}} \mathbf{f}(\mathbf{x}, \mathbf{y}), \qquad \mathbf{y} = \mathbf{g}(\mathbf{x}).$$

Proof: The proof of this theorem is essentially a smart application of the inverse function theorem. To do this we must first construct a function to which the inverse function theorem can be applied. Define $\mathbf{F} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ by

$$\mathbf{F}(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{f}(\mathbf{x}, \mathbf{y})).$$

Observe that this function is differentiable with the derivative given by

$$D_{\mathbf{F}}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} \mathbf{I} & 0 \\ \partial_{\mathbf{x}} \mathbf{f} & \partial_{\mathbf{y}} \mathbf{f} \end{pmatrix}$$

It is easy to see that $D_{\mathbf{F}}(\mathbf{x}_0, \mathbf{y}_0)$ is non-singular and we may therefore apply the inverse function theorem. Thus there exists an open set which we may choose to be of the form $O \times W$ containing the point $(\mathbf{x}_0, \mathbf{y}_0)$ and an open set $U \subset \mathbb{R}^M$ with $0 \in U$, such that $\mathbf{F}$ has a differentiable inverse $\mathbf{F}^{-1} : O \times U \to O \times W$. For $(\mathbf{z}, \mathbf{w}) \in O \times W$ this inverse will be of the form

$$\mathbf{F}^{-1}(\mathbf{z}, \mathbf{w}) = (\mathbf{z}, \mathbf{k}(\mathbf{z}, \mathbf{w})).$$

Finally define

$$\mathbf{g}(\mathbf{z}) = \mathbf{k}(\mathbf{z}, 0).$$

Observe that for all $\mathbf{x} \in O$

$$\mathbf{F}(\mathbf{x}, \mathbf{g}(\mathbf{x})) = \mathbf{F}(\mathbf{x}, \mathbf{k}(\mathbf{x}, 0)) = \mathbf{F}(\mathbf{F}^{-1}(\mathbf{x}, 0)) = (\mathbf{x}, 0),$$

and thus

$$\mathbf{f}(\mathbf{x}, \mathbf{g}(x)) = 0.$$

Since $\mathbf{F}^{-1}$ is differentable, so is $\mathbf{g}$, and by taking the derivative of this last equation using the chain rule we arrive at:

$$0 = \partial_{\mathbf{x}}\mathbf{f} + (\partial_{\mathbf{y}}\mathbf{f}) \cdot D_{\mathbf{g}},$$

which immediately implies the formula for the derivative.

It is clear that the position of the $\mathbf{x}$ and $\mathbf{y}$ variables does not matter. The theorem can therefore be significantly generalized. To do this let $\mathbf{F} : \mathbb{R}^{n+m} \to \mathbb{R}^m$. If $\operatorname{rank} D_{\mathbf{f}}(\mathbf{x}_0, y_0) = m$ we can find m functions $g_j : \mathbb{R}^n \to \mathbb{R}$ such that m variables $x_{i_1}, \ldots, x_{i_m}$ can be expressed $x_{i_j} = g_j(\mathbf{y})$, where $\mathbf{y}$ the n variables which are not in the set $\{x_{i_1}, \ldots, x_{i_m}\}$.

3.4 Assignment 8:

1. Apply the Implicit function theorem to prove the following Theorem:

   Theorem 7 *Let $f : \mathbb{R}^n \to R$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable functions for $i = 1, \ldots, k$, and $U \subset \mathbb{R}^n$ be open. Assume that f attains a local maximum or minumum at $\mathbf{x}^*$ on the set*

   $$\mathcal{D} = U \cap \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) = 0, \qquad i = 1, \ldots, k\}$$

   Let $\mathbf{g} = (g_1, \ldots, g_k)$ and assume that

   $$\operatorname{rank} D_{\mathbf{g}}(\mathbf{x}^*) = k$$

   Then there exists a vector $\Lambda = (\lambda_1, \ldots, \lambda_k) \in \mathbb{R}^k$ such that

   $$D_f(\mathbf{x}^*) + \sum_{i=1}^{k} \lambda_i D_{g_i}(\mathbf{x}^*) = 0.$$

   Hint: Apply the Implicit function theorem to the function $\mathbf{g}$, replace the variables in f using the functions created by the theorem. The resulting function has a local extremum at an interior point of an open subset of $\mathbb{R}^{n-k}$. Supply all the details and give a formula for Λ.

2. Let $\Omega \in \mathbb{R}^n$ be an open set and $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function such that $f(x_1, \ldots, x_n) = 0$ and $\frac{\partial f}{\partial x_J}(x_1, \ldots, x_n) \neq 0$ for all $j = 1, \ldots, n$ and all $(x_1, \ldots, x_n) \in \Omega$. Prove that

   $$\frac{\partial x_1}{\partial x_2}\frac{\partial x_2}{\partial x_3} \cdots \frac{\partial x_{n-1}}{\partial x_n}\frac{\partial x_n}{\partial x_1} = (-1)^n,$$

   for all $(x_1, \ldots, x_n) \in \Omega$.

4 THE BANACH SPACE $C(X)$

Let X be a metric space, then $C(X)$ denotes the set of all **continuous** real valued functions on X. The set of all continuous functions with values in $\mathbb{R}^n$ is denoted by $(C(X))^n$. We will concetrate our efforts on the space of real valued functions as the vector valued case can be treated by treating the components separately.

Proposition 21 *$C(X)$ is a real vector space with a* **norm** *defined by*

$$\|f\|_\infty = \sup_{x \in X} |f(x)|$$

Proof: Clearly if $f, g \in C(X)$ and $a, b \in \mathbb{R}$ then $af + bg$ is real valued continuous function on X and therefore $af + bg \in C(X)$. To show that there is a norm, observe that

$$\|f\|_\infty \geq 0,$$

and if $\|f\|_\infty = 0$ we must have $|f(x)| =)$ for all $x \in X$ and thus $f = 0$. Next, let $a \in \mathbb{R}$ then

$$\|af\|_\infty = \sup_{x \in X} |af(x)| = |a| \sup_{x \in X} |f(x)| = |a| \|f\|_\infty.$$

Finally, observe that for every $x \in X$

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_\infty + \|g\|_\infty.$$

Taking the least upper bound on the left hand side we get

$$\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty.$$

Remarks:

1. This norm is called the uniform norm. Any norm generates a metric in the usual way. The topology generated by this metric is called the uniform topology on $C(X)$.

2. Continuity is not required for this last proposition. However, we restrict ourselves to continuous functions, as $C(X)$ will inherit some nice properties.

3. The set $C(X)$ is actually more than a vectorspace, it is a commutative ring, for those of us who like algebra.

4. In the case of vector valued functions the absolute value is replaced by the Euclidean norm on $\mathbb{R}^n$.

5. If X is a finite set with n elements, then $C(X)$ is isomorphic to $\mathbb{R}^n$, but for infinite sets X the space is infinite dimensional. We can see this easily in the case when X is an interval. The polynomials are elements of $C(X)$, but for any finite set of polynomials one can construct a new polynomial which is

linearly independent. Let $p_1, \ldots, p_n$ denote a set of n linearly independent polynomials. Let m be the largest degree of these n polynomials. Then x^{m+1} cannot be written as a linear combination of the original n polynomials. Thus no finite set of of polynomials can span the vectorspace of polynomials and this is an infinite dimensional subspace of $C(X)$

This Proposition estabished $C(X)$ as a metric space. We can and will now look at all the topological properites of this space. Most importantly we have:

Theorem 8 *$C(X)$ is complete.*

Proof: The proof of this theorem is surprisingly simple. Let $\{f_n\}$ be a Cauchy sequence in $C(X)$. This implies that for any $x \in X$ the sequence $\{f_n(x)\}$ is a Cauchy sequence of real numbers. Since the real numbers are complete this sequence converges to a unique limit. So for every $x \in X$ we can define

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

Next we need to show that this sequence converges uniformly. Let $\epsilon > 0$, then for each $x \in X$ there is a number N_x such that

$$|f(x) - f_n(x)| < \frac{\epsilon}{2}$$

for all $n \geq N_x$. Furthermore, there exists M such that

$$\|f_n - f_m\|_\infty < \frac{\epsilon}{2}$$

for all $n, m \geq M$. Now let $n \geq M$ then for each $x \in X$ there is a $m \geq \max\{M, N_x\}$ such that

$$|f(x) - f_n(x)| \leq |f(x) - f_m(x)| + |f_m(x) - f_n(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2}.$$

Since this inequality holds for all $x \in X$ and the right hand side is independent of x we may take the least upper bound on the left and get

$$\|f - f_n\|_\infty < \epsilon$$

for all $n \geq M$. Finally, we will prove that f is continuous. To do this let $\epsilon > 0$ and $x_0 \in X$ then There is a M such that $\|f - f_n\|_\infty < \frac{\epsilon}{3}$ for all $n \geq M$. Pick a $n \geq M$ then there exists a $\delta > 0$ such that $|f_n(x) - f_n(x_0)| < \frac{\epsilon}{3}$ for all $x \in B_\delta(x_0)$. Finally, observe that

$$|f(x) - f(x_0)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f(x_0) - f_n(x_0)| < \epsilon$$

for all $x \in B_\delta(x_0)$.

Remarks:

1. In the proof of the Theorem we only used the contiuity of f_n to sow that the limit is continuous. WE can thus extend the statement to the following:

   Let $\{f_n\}$ be a sequence of real valued functions on a metric space X which is a Cauchy sequence in the uniform metric. Then the sequence converges to a function on X.

2. Complete, normed vector spaces are usually called **Banach Spaces**, in honor of the Polish mathematician Stefan Banach.

Examples:

1. Consider series

   $$\sum_{j=0}^{\infty} x^j$$

   for $j \in (-p, p)$ where $0 < p < 1$ If we let $f_n(x) = \sum_{j=0}^{n} x^j$ then this is a sequence of continuous functions on $(-p, p)$. We know that this sequece converges for every $x \in (-p, p)$. We will show that it converges uniformly. To do this let $f(x) = \lim_{n\to\infty} f_n(x)$ then

   $$|f(x) - f_n(x)| = \left| \sum_{j=n+1}^{\infty} x^j \right| \leq \sum_{j=n+1}^{\infty} |x|^j < \sum_{j=n+1}^{\infty} p^j = \frac{p^{n+1}}{1-p}$$

   The right most term is independent of x and converges to 0 as $n \to \infty$. So we take the least upper bound on the left to get uniform convergence.

2. Let $f_n(x) = x^n$ then $\{f_n\}$ is not uniformly convergent on $[0, 1]$, but it is uniformly convergent on $[0, p)$ for any $0 < p < 1$. To see this observe that $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in [0, 1)$, but $\lim_{n\to\infty} f_n(1) = 1$. Thus the limit function is not continuous, as it would be if the convergence was uniform. For $x \in [0, p)$ observe that $|f(x) - f_n(x)| < p^n$.

It seems like the first example can be extended to any power series. Moreover, if $\{f_n\}$ is a sequence of real valued functions on the metric space x we can let $g_n(x) = f_{n+1}(x) - f_n(x)$ and $g_0(x) = f_1(x)$ which gives

$$f_n(x) = \sum_{j=0}^{n-1} g_j(x),$$

i.e. we can assume that any given sequence is actually a series. And for series there is an excellent tool to test whether it converges uniformly or not.

Theorem 9 Weierstrass M-Test. *Let $\{g_n\}$ be a sequence of functions on X. If there exists a sequence of non-negative real numbers $\{a_n\}$ suct that the series*

$$\sum_{j=0}^{\infty} a_n$$

converges and $\|g_j\|_\infty \leq a_j$ *for all* $j \in \mathbb{N}$ *then the series*

$$\sum_{j=0}^{\infty} g_j(x)$$

converges uniformly on X*.*

Proof: The proof essentially follows example 1. Let $\epsilon > 0$, since

$$\sum_{j=0}^{\infty} a_j$$

converges the exists a N such that

$$\sum_{j=n+1}^{m} a_j < \epsilon$$

for all $n, m \geq N$. Now

$$\left\| \sum_{j=n+1}^{m} g_j \right\| \leq \sum_{j=n+1}^{m} \|g_j\| \leq \sum_{j=n+1}^{m} a_j < \epsilon$$

for all $n, m \geq N$. Thus the sequence of partial sums is a Cauchy sequence and converges.

4.1 Consequences and Examples

Proposition 22 *Let* $O \subset X$ *and* x_0 *be a limit point of* O *(not necessarily in* O*). Let* $\{f_n\}$ *be a sequence of continuous functions on* O *such that*

$$\lim_{x \to x_0} f_n(x) = a_n.$$

If the sequence of functions converges uniformly to a function f *on* O *and the sequence* a_n *converges to a number* a*, then*

$$\lim_{x \to x_0} f(x) = \lim_{n \to \infty} a_n$$

Proof: Observe that

$$|f(x) - a| \leq \|f - f_n\|_\infty + |f_n(x) - a_n| + |a_n - a|.$$

For any $\epsilon > 0$ there exists N such that

$$\|f - f_n\|_\infty < \frac{\epsilon}{3} \qquad \text{and} \qquad |a_n - a| < \frac{\epsilon}{3},$$

for any $n \geq N$. We can also find $\delta > 0$ such that

$$|f_n(x) - a_n| < \frac{\epsilon}{3}$$

for all $x \in B_\delta(x_0) \cap O$. Combining these yields the desired result.

Derivatives and Integrals of Sequences:

If we have a given sequence of differentiable functions in $C([a,b])$ is the limit function necessarily differentiable. We will investigate this and related questions next. We start by looking at integration.

Example:

Let

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

Then this is a sequence of continuous Riemann integrable functions on $(0,1)$. This sequence converges to $f(x) = 0$, and

$$\int_0^1 f_n\,dx = \frac{1}{2},$$

for all n Thus

$$\lim_{n\to\infty} \int_0^1 f_n\,dx = \frac{1}{2} \neq 0 = \int_0^1 \lim_{n\to\infty} f_n\,dx.$$

Proposition 23 *Let $\{f_n\}$ a sequence of Riemann integrable functions on the interval $[a,b]$, that uniformly coverges to a function f. Then f is Riemann integrable and*

$$\lim_{n\to\infty} \int_a^b f_n\,dx = \int_a^b f\,dx$$

Proof: Let $\epsilon > 0$ since the sequence converges uniformly there exists a $N > 0$ such that

$$\|f_n - f\|_\infty < \frac{\epsilon}{6(b-a)},$$

for all $n \geq N$. In particular this inequality holds for $n = N$. Morevoer, since f_N is integrable there exists a partition P such that

$$\overline{S}(f_N;P) - \underline{S}(f_N;P) < \frac{\epsilon}{3}.$$

For the ith subinterval of a partition let $M_i(f) = \sup\{f(x) : x_{i-1} \leq x \leq x_i\}$, and $m_i(f) = \inf\{f(x) : x_{i-1} \leq x \leq x_i\}$. It follows from the uniform convergence and a problem in the next assignment that

$$|M_i(f) - M_i(f_N)| < \frac{\epsilon}{3(b-a)} \qquad \text{and} \qquad |m_i(f) - m_i(f_N)| < \frac{\epsilon}{3(b-a)}.$$

Thus

$$\begin{aligned} \left|\overline{S}(f;P) - \overline{S}(f_N;P)\right| &\leq \sum_{i=1}^n |M_i(f) - M_i(f_N)|\Delta x_i \\ &< \frac{\epsilon}{3(b-a)} \sum_{i=1}^n \Delta x_i = \frac{\epsilon}{3} \end{aligned}$$

and similarly

$$|\underline{S}(f;P) - \underline{S}(f_N;P)| < \frac{\epsilon}{3}.$$

Finally,

$$\left|\overline{S}(f;P) - \underline{S}(f;P)\right| \leq \left|\overline{S}(f;P) - \overline{S}(f_N;P)\right| + \left|\overline{S}(f_N;P) - \underline{S}(f_N;P)\right| + \left|\underline{S}(f;P) - \underline{S}(f_N;P)\right| < \epsilon,$$

and thus f is Riemann integrable. To show the second part of the proposition observe that for $\epsilon > o$ there is a N such that

$$f_n(x) - \frac{\epsilon}{(b-a)} < f(x) < f_n(x) + \frac{\epsilon}{(b-a)},$$

for all $x \in [a,b]$ and all $n \geq N$. Integration yields

$$\int_a^b f_n\,dx - \epsilon < \int_a^b f\,dx < \int_a^b f_n\,dx + \epsilon$$

which yields the result.

The next example shows that uniform convergence is a sufficien condition, but it is not a necessary condition.

Example:

The sequence of functions $\{f_n(x)\} = \{x^n\}$ converges on the interval $[0,1]$, but since the limit fuction is not continuous this convegerce is not uniform. The limit function f is identically equal to zero on $[0,1)$ and $f(1) = 1$. Thus

$$\int_0^1 f\,dx = 0,$$

and

$$\lim_{n\to\infty}\int_0^1 f_n\,dx = \lim_{n\to\infty}\frac{1}{n+1} = 0.$$

The next proposition has a somewhat lengthy proof, but it is a very powerful statement. We will show that if f_n is a sequence of differentiable functions and whose derivatives converge uniformly, then f_n converges uniformly to a differentiable function, provided it converges at at least one point in its domain.

Proposition 24 *Let $\{f_n\}$ be a sequence of differentiable functions on (a,b) such that the sequence $\{f'_n\}$ converges uniformly to a function g, and there is a point $x_0 \in (a,b)$ such that $\{f_n(x_0)\}$ converges. Then the sequence $\{f_n\}$ converges uniformly to a differentiable function f and $f' = g$.*

Proof: We start with proving that the series converges uniformly. Let $\epsilon > 0$, then there exists a $N > 0$ such that

$$|f_n(x_0) - f_m(x_0)| < \frac{\epsilon}{2}$$

for all $n, m \geq N$ and

$$\|f'_n - f'_m\| < \frac{\epsilon}{2(b-a)},$$

for all $n, m \geq N$. Now we have

$$\begin{aligned} |f_n(x) - f_m(x)| &= |f_n(x) - f_n(x_0) + f_n(x_0) - f_m(x_0) + f_m(x_0) - f_m(x)| \\ &\leq |f_n(x_0) - f_m(x_0)| + |(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))| \\ &\leq \frac{\epsilon}{2} + \|f_n' - \mathbf{f}_m'\| \ |b-a| \\ &< \epsilon, \end{aligned}$$

where we used the Mean Value Theorem for the second term on the right of the second line and for the function $f_n - f_m$. Since this holds for any $x \in [a, b]$ the sequence is a uniform Cauchy sequence and converges. Let $f(x) = \lim_{n\to\infty} f_n(x)$.

Next, fix $x \in (a, b)$ and define

$$h_n(t) = \frac{f_n(t) - f_n(x)}{t - x} \qquad \text{and} \qquad h(t) = \frac{f(t) - f(x)}{t - x} \qquad \text{for all} \qquad t \in [a, b] \setminus \{x\}$$

To continue observe that by using the Mean Value Theorem again we have

$$\begin{aligned} |h_n(t) - h_m(t)| &= \frac{1}{|t-x|} \, |(f_n(t) - f_m(t)) - (f_n(x) - f_m(x))| \\ &\leq \frac{1}{|t-x|} \, \|f_n' - f_m'| \ |t-x| \end{aligned}$$

which imeiately make h_n into a Cauchy sequence which converges uniformly if $t \neq x$. But since $f_n(t)$ converges to $f(t)$ for all $t \neq x$, $h_n(t) \to h(t)$. By a previous proposition we have

$$\lim_{t\to x} \left(\lim_{n\to\infty} h_n(t) \right) = \lim_{n\to\infty} \left(\lim_{t\to x} H_n(t) \right),$$

and thus

$$f'(x) = \lim_{t\to x} h(t) = \lim_{n\to\infty} f_n'(x).$$

4.2 A nowhere differentiable continuous function

Our next step is to give an example of a function which is continuous on $\mathbb{R}$ but nowhere differentiable.

Proposition 25 *There exists a continuous $f : \mathbb{R} \to \mathbb{R}$ which is nowhere differentiable.*

Proof: We begin by defining $\phi : \mathbb{R} \to \mathbb{R}$ as follows. $\phi(x) = |x|$, for $|x| \leq frac12$. and $\phi(x + n) = \phi(x)$ for all integers n. This defines ϕ on all of $\mathbb{R}$ ϕ is continuous and its graph consists of line segments which alternate between slope 1 and -1. Moreover, $0 \leq \phi(x) \leq \frac{1}{2}$ for all $x \in \mathbb{R}$. Next let

$$f_j(x) = \frac{\phi(4^j x)}{4^j}.$$

Then f_j is continuous and $|f_j(x)| \leq \frac{1}{2}4^{-j}$, and $f_j(x+4^j) = f_j(x)$. Thus the sum

$$\sum_{j=0}^{\infty} f_j(x)$$

converges uniformly to a continuous function $f(x)$ To show that this function is nowhere differentiable, fix $x \in \mathbb{R}$ an let $h_j = \pm 4^{-j-1}$, where the sign is chosen such that $4^j x$ and $4^j(x+h_j)$ lie in the same interval $[\frac{k}{2}, \frac{k+1}{2}$. I.e the graph at these two points are the same line segment. Observe that

$$f_k(x) = f_k(x+h_j)$$

for all $k > j$. Furthermore,

$$f_k(x) - f_k(x+h_j) = \pm h_j$$

for all $k \leq j$. This is easy to see for $k = j$. For $k < j$, the points $4^k x$ and $4^k(x+h_j)$ also lie in the same interval $[\frac{m}{2}, \frac{m+1}{2}]$. Therefore, we have

$$\frac{f(x+h_j) - f(x)}{h_j} = \sum_{k=0}^{j} \frac{f_k(x+h_j) - f_k(x)}{j} = \sum_{k=0}^{j} \gamma_k$$

where $\gamma = \pm 1$.Thus if $j+1$ is even we have either an even number of $+1$'s and even number of -1's, or an odd number of $+1$'s and an odd number of -1's. In either case the sum is an even number. By the same reasoning the sum is an odd number if $j+1$ is odd. Thus the limit

$$\lim_{j\to\infty} \frac{f(x+h_j) - f(x)}{h_j}$$

cannot exist, and neither can the derivative.

4.3 Assignment 9:

1. Find an example of asequence of functions that is not continuous at any point in $[0,1]$ but converges uniformly to a continuous function.

2. Define
$$f_n(x) = \begin{cases} \frac{1}{n} & , \quad \frac{1}{2^n} < x \leq \frac{1}{2^{n-1}} \\ 0 & \quad \text{otherwise} \end{cases}$$
Show that the series $\sum_{n=1}^{\infty} f_n$ converges uniformly, but the Weierstrass M-test fails.

3. Prove that if $\sum_{n=1}^{\infty} f_n$ converges uniformly on E, then $\{f_n\}$ converges to zero uniformly on E. Is the converse true?

4. Prove that

(a) If f_n converges uniformly to f and if f_n is bounded, then f is bounded as well.

(b) If f_n converges uniformly to f on $[a,b]$ and if $\|f_N - f\| < \epsilon$, then $|M(f_N) - M(f)| < 2\epsilon$, where $M(f) = \max\{f(x); x \in [a,b]\}$.

4.4 Dense subsets of $C(X)$:

While continuous functions seem very reasonable it is often necessary to approximate continuous functions with another group of functions which are easier to evaluate. In fact the only functions which we can evaluat exactly are the ones which only involve the four basic operations of arithmetic. All others require approximations. For notation we say that a set A is adense subset of a metric space X, if the closure of A is equal to X. In other words if A is a dense subset, every $x \in X$ can be approximated by a sequence of elements of A. We start with a simple result.

Proposition 26 *Let L denote the set of piece wise linear functions on the interval $[a,b]$. Then L is dense in $C([a,b])$.*

Proof: Without loss of generality we assume $[a,b] = [0,1]$. Let $f \in C([0,1])$ be a given continuous function and $\epsilon > 0$. Then there exists a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon/2$ for all $|x - y| < \delta$. Choose N such that $1/N < \delta$, and let $x_0 = 0$ and $x_j = j/N$ for $j = 1, \ldots, N$. Define

$$l(x) = f(x_j)\frac{x - x_{j-1}}{x_j - x_{j-1}} + f(x_{j-1})\frac{x_j - x}{x_j - x_{j-1}},$$

for $x \in [x_{j-1}, x_j]$, $j = 1, \ldots, N$. Then l is continuous and piecewise linear and for every $x \in [x_{j-1}, x_j]$ and every $j = 1, dots, N$ we have:

$$\begin{aligned} |l(x) - f(x)| &= \left| f(x_j)\frac{x - x_{j-1}}{x_j - x_{j-1}} + f(x_{j-1})\frac{x_j - x}{x_j - x_{j-1}} - f(x)\frac{x - x_{j-1}}{x_j - x_{j-1}} + f(x)\frac{x_j - x}{x_j - x_{j-1}} \right| \\ &\leq \left| (f(x_j) - f(x))\frac{x - x_{j-1}}{x_j - x_{j-1}} \right| + \left| (f(x_{j-1}) - f(x))\frac{x_j - x}{x_j - x_{j-1}} \right| < \epsilon. \end{aligned}$$

Hence every continuous function can be uniformly approximated by a piecewise linear function, and the piecewise linear functions are dens in $C([a,b])$.

Instead of proving such results for every possible subset we would like to have a general criterium for density. This will be given by the following two theorems, Stone's Theorem, and the Stone-Weierstrass Theorem. To formulate these theorem we introduce the following notation. Let f and g be two continuous functions on a metric space X, then $(f \vee g)(x) = \max\{f(x), g(x)\}$, and $(f \wedge g)(x) = \min\{f(x), g(x)\}$. It is clear that $f \vee g$ and $f \wedge g$ are both continuous.

Theorem 10 Stone's Theorem *Let X be a compact metric space, and let $\mathcal{L} \subset C(X)$ have the following properties:*

1. *if $f, g \in \mathcal{L}$ then $af + bg \in \mathcal{L}$. i.e. $\mathcal{L}$ is a vector space.*
2. *if $f, g \in \mathcal{L}$, then $f \vee g, f \wedge g \in \mathcal{L}$*
3. *for any $x, y \in X$, with $x \neq y$, there exists $f \in \mathcal{L}$ with $f(x) \neq f(y)$, i.e. $\mathcal{L}$ separates points*
4. *every constant function belongs to $\mathcal{L}$.*

Then $\mathcal{L}$ is dense in $C(X)$.

Proof: First if $x \neq y$ and $a, b \in \mathbb{R}$ there is an $f \in \mathcal{L}$ such that $f(x) = a$, and $f(y) = b$. This is obvious if $a = b$ and if $a \neq b$ there is a $g \in \mathcal{L}$ such that $g(x) = \alpha$ and $g(y) = \beta$. And one of the numbers α, β is not zero. Assume $\alpha \neq 0$, then

$$\frac{a}{\alpha} g + (b - \frac{a}{\alpha}\beta)$$

satisfies the desired property.

Now let $f \in C(X)$ and $\epsilon > 0$ we will constaruct a function $g \in \mathcal{L}$ such that $f(x) - \epsilon, g(x) < f(x) + \epsilon$ for all $x \in X$. Now for every pair $x, y \in X$ there exists a function $g_{xy} \in \mathcal{L}$ such that $g_{xy}(x) = f(x)$ and $g_{xy}(y) = f(y)$. By continuity of g_{xy} there exists a number δ_{xy} such that $g_{xy}(t) < f(t) + \epsilon$ for all $t \in B_{\delta_{xy}}(y)$. The collection of open balls $B_{\delta_{xy}}(y)$ forms an open cover of X, and by compactness there exist points $y_1, \ldots y_N$ such that $X = B_{\delta_{xy_1}}(y_1) \cup \cdots \cup B_{\delta_{xy_N}}(y_N)$. Now we define for every $x \in X$

$$g_x = g_{xy_1} \wedge \cdots \wedge g_{xy_N} \in \mathcal{L}.$$

Then $g_x(t) < f(t) + \epsilon$ for all $t \in x$ and $g_x(x) = f(x)$. Again by continuity there exists numbers δ_x such that $g_x(t) > f(t) - \epsilon$ for all $t \in B_{\delta_x}(x)$. Compactness again applies that there are finitely many points $x_1, \ldots, x_M$ whose balls cover X. Define

$$g = g_{x_1} \vee \cdots \vee g_{x_M}.$$

Then $f(t) - \epsilon < g(t) < f(t) + \epsilon$ for all $t \in X$ and the proof is complete.

This is a powerful theorem which has the proposition on piecewise linear functions as a direct consequence. Since piecewise linear functions on $\mathbb{R}^n$ satisfy all the hypotheses, as do piecewise polynomial functions. However, we often wnat to use functions which are not defined piecewise, like polynomials or trigonometric polynomials etc. And these functions do not satisfy all hypotheses, since $p \vee q$ is not a polynomial if p, q are polynomials for example.

Before continuing to an improved theorem we will explore the condition of maxima further. Observe that $|a| = a \vee (-a)$ and thus

$$|a - b| = a \vee b - a \wedge b \qquad \text{and} \qquad a + b = a \vee b + a \wedge b.$$

So if $f + g \in \mathcal{L}$ and $|g| \in \mathcal{L}$ for every $g \in \mathcal{L}$ the condition (2) of the theorem holds. We have the following

Lemma 4 *There exists a sequence of polynomials $\{p_n\}$ which uniformly converges to $|x|$ on $[-1,1]$.*

Proof: Let $p_0 = 1$ and recursively define

$$p_{j+1}(x) = \frac{1}{2}\left(x^2 + 2p_J - p_j^2\right)$$

If $|x| \le p_j \le 1$ we have

$$p_j - p_{j+1} = \frac{1}{2}\left(p_j^2 - |x|^2\right) \ge 0,$$

and

$$p_{j+1} - |x| = \frac{1}{2}\left((1-|x|)^2 - (1-p_j)^2\right) \ge 0.$$

Since p_0 satisfies $|x| \le p_0 \le 1$ itfollows by induction that $|x| \le p_{j+1} \le p_j \le 1$, and so for every $x \in [-1,1]$ $\{p_j(x)\}$ is a bounded decreasing sequence that converges. Let $\{p\}$ be the limit of this sequence, then p satisfies.

$$p = \frac{1}{2}\left(|x|^2 + 2p - p^2\right),$$

and therefore

$$p^2 = |x|^2.$$

Since p is nongegative we have $p = |x|$. The uniform convergence follows by a theorem that says if a sequence of continuous functions converges monotonically to a continuous limit on a compact set, the convergence is uniform (see assignment 10).

Theorem 11 Stone-Weierstrass Theorem *Let X be a compact metric space, and let $\mathcal{S} \subset C(X)$ have the following properties:*

1. *if $f, g \in \mathcal{S}$ then $af + bg \in \mathcal{S}$. i.e. $\mathcal{S}$ is a vector space.*
2. *if $f, g \in \mathcal{S}$, then $fg \in \mathcal{S}$*
3. *for any $x, y \in X$, with $x \ne y$, there exists $f \in \mathcal{S}$ with $f(x) \ne f(y)$, i.e. $\mathcal{S}$ separates points*
4. *every constant function belongs to $\mathcal{S}$.*

Then $\mathcal{S}$ is dense in $C(X)$.

Proof: It is clear, that if p is a polynomial then $p(f) \in \mathcal{S}$ for every $f \in \mathcal{S}$. Now for every compact interval $I \subset \mathbb{R}$ there is a sequence of polynomials $\{p_n\}$ which converges uniformly to $|x|$ on I. Given $f \in \mathcal{S}$. There is a compact interval I such that $f(X) \subset I$ and a sequence $\{p_n(f)\}$ in $\mathcal{S}$ that converges to $|f|$. Thus $|f|$ is in the closure of $\mathcal{S}$. The closure of $\mathcal{S}$ now satisfies all the hypotheses of Stone's Theorem and is therefore dense in $C(X)$. But since the closure is closed it must equal $C(X)$ and hence $\mathcal{S}$ is dense in $C(X)$.
An immediate consequece is this classical result.

Corollary 1 Weierstrass Approximation Theorem. *Let X be a compact subset of $\mathbb{R}^n$. Then every function $f \in C(X)$ can be uniformly approximated by a sequence of polynomials.*

Of course there are many other consequences. Differentiable functions are dense, and for every $n \in \mathbb{N}$ n-times differentiable functions are dense.

4.5 Compactness in $C(X)$:

Another important question is when does a sequence of functions have a uniformly convergent subsequences. In $\mathbb{R}^n$ we have the Bolzano-Weierstrass Theorem which gives a condition that can be easily checked. The situation in $C(X)$ is somewhat more complicated.

Definition 17 *Let X be a metric space and $\mathcal{F} \subset C(X)$. The set $\mathcal{F}$ is said to be equi-continuous if for every $\epsilon > 0$ and every $x \in X$ there exists a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ for all $y \in B_\delta(x)$ and all $f \in \mathcal{F}$.*

This condition is a condition on a set of functions, not on a single function. We now prove:

Theorem 12 Classical Arzela-Ascoli Theorem. *Let X be a metric space which has a countable dense subset Q. And let $\mathcal{G} \subset C(X)$ have the following properties:*

1. *There exists a $M > 0$ such that $|g| < M$ for all $g \in \mathcal{G}$*
2. *$\mathcal{G}$ is equicontinuous*

Then every sequence of functions in $\mathcal{G}$ has a (pointwise) convergent subsequence.

Proof: Let $Q = \{p1, p_2, \ldots\}$. Then $\{g_n(p_1)\}$ is a bounded sequence of real numbers, and therefore has a convergent subsequence $\{g_n^1(p_1)\}$ whose limit we call $g(p_1)$. We ingore all terms of $\{g_n\}$ which are not part of $\{g_n^1\}$. $\{g_n^1(p_2\}$ is again abounded sequence of real numbers with a convergent subsequence $\{g_n^2(p_2)\}$ observe that

$$g_n^2(p_2) \to g(p_2), \qquad \text{and} \qquad g_n^2(p_1) \to g(p_1)$$

in this manner we keep extracting subsequences $g_n^{j+1}(p_{j+1})$of $g_n^j(p_{j+1}$ with the property that

$$\lim_{n\to\infty} g_n^{j+1}(p_i) = g(p_i) \qquad \text{for all} \qquad i = 1, \ldots, j+1.$$

We next construct a new sequence by diagonalization. I.e, we set

$$h_j = g_j^j,$$

i.e the jth term of this sequence is given tby the jth term of the jth subsequence. In this way $\{h_n\}$ is a subsequence of all the sequences constructed above, and thus

$$\lim_{n\to\infty} h_n(p_j) = g(p_j),$$

for all $p_j \in Q$. We therefore have established a sequence which converges at every point in P. Next let $\epsilon > 0$ and $x \in X$, then there exists a $\delta > 0$ such that $|h_n(x) - h_n(y)|, \epsilon/3$ for all n and all $y \in B_\delta(x)$. Since Q is dense, there is a $q \in Q$ such that $q \in B_\delta(x)$. For this q, there exists a $M > 0$ such that

$$|h_n(q) - h_m(q)| < \frac{\epsilon}{3},$$

for all $m, n \geq M$. Now

$$|h_n(x) - h_m(x)| \leq |h_n(x) - h_n(q)| + |h_n((q) - h_m(q)| + |h_m(q) - h_m(x)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,$$

for all $m, n \geq M$. Therefore $h_n(x)$ is a Cauchy sequence of real numbers and converges to some real number $h(x)$.

A consequence is

Theorem 13 *Let X be a compact metric space, and let $\mathcal{G} \subset C(X)$ have the following properties:*

1. *There exists a $M > 0$ such that $|g| < M$ for all $g \in \mathcal{G}$*
2. *$\mathcal{G}$ is equicontinuous*

Then the closure of $\mathcal{G}$ is a compact subset of $C(X)$.

Proof: In the compact case we observe that we only need to construct finitely many subsequences in order to have subsequence which converges at the centers of finitely many balls of radius δ which cover X. Then we can use a number M which works for all theese centers. And this will give us uniform convergence.

4.6 Assignment 10:

Prove the following Theorem:

Theorem 14 *Let X be a compact metric space, and let $\{f_n\}$ be a monotonically increasing sequence of continuous functions. I.e. $f_{n+1}(x) \geq f_n(x)$ for all $x \in X$.*

1. *If there exists a number M such that $f_n(x) \leq M$ for all $x \in X$, this sequence converges pointwise to a limit function $f(x)$.*
2. *If f (the limit function) is continuous, then the convergence is uniform.*

5 POWER SERIES AND FOURIER SERIES

5.1 Number series with non-negative terms

In this section we will apply some of the principles developed in the previous sections to the most common sequences of functions. Before doing that we need to prove some staments about series of non-negative terms.

Proposition 27 *Let $\{a_n\}$ and $\{b_n\}$ be sequences of non-negative numbers, such that*

$$0 < l \leq \underline{\lim}_{n\to\infty} \frac{a_n}{b_n} \leq \overline{\lim}_{n\to\infty} \frac{a_n}{b_n} = L < \infty$$

Then

$$\sum_{n=1}^{\infty} a_n \qquad \text{and} \qquad \sum_{n=1}^{\infty} b_n$$

converge or diverge together.

Proof: Assume that $\sum_{n=1}^{\infty} b_n$ converges. The statement about the limit implies, that for a given $\epsilon > 0$ there are only finitely many terms of $\frac{a_n}{b_n}$ that are greater than $L + \epsilon$. Thus there exists an $N > 0$ such that'

$$a_n < (L + \epsilon) b_n$$

for all $n \geq N$. Hence, for all $K > N$

$$\sum_{n=N}^{K} a_n < (L+\epsilon) \sum_{n=N}^{K} b_n \leq (L+\epsilon) \sum_{n=1}^{\infty} b_n,$$

and thus

$$\sum_{n=N}^{\infty} a_n$$

converges, and so does

$$\sum_{n=1}^{\infty} a_n.$$

A similar argument using the left inequality proves that the convergence of $\sum_{n=1}^{\infty} a_n$ implies the convergence of $\sum_{n=1}^{\infty} b_n$ and the statement about the divergence.

Unfortunately this result requires that we can find an appropriate series for comparison in order to determine whether or not the series converges. The next two results will do away with this requirement.

Proposition 28 Root Test *Let $\{a_n\}$ be a sequence of non-negative terms, and $L = \overline{\lim} a_n^{\frac{1}{n}}$. Then*

$$\sum_{n=0}^{\infty} a_n$$

converges if $L < 1$ and diverges if $L > 1$.

Proof: Assume first that $L < 1$. Then for any $\epsilon > 0$ there is at most finitely many terms of $a_n^{\frac{1}{n}}$ which are larger than $L + \epsilon$. In particular, there exist aN such that

$$a_n^{\frac{1}{n}} \leq \frac{1+L}{2} < 1,$$

for all $n \geq N$. Therefore,

$$a_n \leq \left(\frac{L+1}{2}\right)^n,$$

for all $n \geq N$ and

$$\sum_{n=N}^{\infty} a_n \leq \sum_{n=N}^{\infty} \left(\frac{L+1}{2}\right)^n < \infty,$$

and the series converges.

Next, assume that $L > 1$. Then the sequence contains a subsequence $\{a_{n_k}\}$ such that there is a $N > 0$ with

$$a_{n_k} > \left(\frac{L+1}{2}\right)^{n_k} \geq \left(\frac{L+1}{2}\right)^k,$$

for all $k \geq N$. Thus

$$\sum_{k=N}^{\infty} a_{n_k} > \sum_{k=N}^{\infty} \left(\frac{L+1}{2}\right)^k,$$

which is known to diverge. This immediately implies the result.

Proposition 29 Ratio Test *Let $\{a_n\}$ be a sequence of non-negative terms, and $L = \overline{\lim} \frac{a_{n+1}}{a_n}$ and $K = \underline{\lim}_{n\to\infty} \frac{a_{n+1}}{a_n}$. Then*

$$\sum_{n=0}^{\infty} a_n$$

converges if $L < 1$ and diverges if $K > 1$.

Proof: Assume that $L < 1$, as in the previous proof there exists a number N such that'

$$\frac{a_{n+1}}{a_n} \leq \frac{L+1}{2}$$

for all $n \geq N$. This implies:

$$a_{N+k} \leq a_{N+k-1}\left(\frac{L+1}{2}\right) \leq a_{N+k-2}\left(\frac{L+1}{2}\right)^2 \leq \cdots \leq a_N \left(\frac{L+1}{2}\right)^k$$

Thus

$$\sum_{k=0}^{\infty} a_{N+k} \leq a_N \sum_{k=0}^{\infty} \left(\frac{L+1}{2}\right)^k,$$

which is known to converge. On the other hand, if $K > 1$ there exists A $N > 0$ such that

$$\frac{a_{n+1}}{a_n} > \frac{K+1}{2}$$

for all $n \geq N$. It follows that

$$a_{N+k} > a_N \left(\frac{K+1}{2}\right)^k.$$

The series diverges by comparison to the divergent geometric series

$$\sum_{k=0}^{\infty} \left(\frac{K+1}{2}\right)^k$$

5.2 Power Series

Definition 18 *Let $a \in \mathbb{R}$ a* **power series** *at a is a function of the form*

$$f(x) = \sum_{k=0}^{\infty} a_k (x-a)^k,$$

where $\{a_k\}$ is a given sequence of real numbers.

We remark that a power series converges at at least one point $x = a$. Applying the root test to this we get that the series converges if

$$\overline{\lim}_{k\to\infty} \left(|a_k|\,|x-a|^k\right)^{\frac{1}{k}} < 1.$$

Simple algebra implies

$$|x-a| < \frac{1}{\overline{\lim}_{k\to\infty} |a_k|^{\frac{1}{k}}} = R.$$

This number R is called the radius of convergence of the series. The Stone Weierstrass Theorem already tells us that every continuous function can be approximated by a sequence of polynomials on a compact interval. A power series is the limit of a sequence of polynomials. We will now study the uniform convergence of power series.

Proposition 30 *Let f be given by the power series*

$$f(x) = \sum_{k=0}^{\infty} a_k (x-a)^k,$$

with apositive radius of convergence R. Then the series converges uniformly on any compact subset of $(a-R, a+R)$.

Proof: Let C be a compact subset of $(a-R, a+R)$. Then there exists a $\rho \in (0, R)$ such that $C \subset [a-\rho, a+\rho] \subset (a-R, a+R)$. Now for any $x \in C$, we have

$$\left|a_k (x-a)^k\right| < |a_k| \rho^k.$$

The result now follows immediatedly from the Weierstrass M-Test.

Power series can formally be differentiated an infinite number of times, however, the derivatives may not converge. A special case is if all derivatives converge.

Definition 19 *f is called an* **analytic function** *on $(a-R, a+R)$ if it can be differentiated infinitely many times and has a convergent power series at A with radius of convergence R.*

In this case we have

$$a_k = \frac{f^{(k)}(a)}{k!},$$

as can easily be checked by direct computation. That is the power series is actually the Taylor series of this function.

The function $f(x) = 0$ is analytic and has the power series $\sum_{k=0}^{\infty} 0x^k = 0$ at 0, and this is clearly the only power series it has at this point. For if there would be another power series

$$\sum_{k=0}^{\infty} a_k x^k$$

we get

$$0 = f^{(k)}(0) = k! a_k,$$

and thus $a_k = 0$. An immediate consequence is

Corollary 2 *Suppose that f is analytic on an open neighborhood of a. Then its power series at a is unique.*

We would like to extend this to higher dimensions. Recall the definition of an k-linear form

$$\Lambda : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}.$$

A power series in $\mathbb{R}^n$ can now easily be defined by

$$\sum_{k=0}^{\infty} \Lambda_k(x-a, \ldots, x-a). \tag{17}$$

where Λ_k is a k-linear form. Next, we define, a norm for k-linear forms

$$\|\Lambda_k\| = \max\{|\Lambda_k(u_1, \ldots, u_k)| : u_j \in \mathbb{R}^n, |u_j| = 1\}$$

It is easily checked thta this is actually a norm. In fact it is the same as the infinity norm for continuous functions. Next, observe that for $x \neq a$

$$|\Lambda_k(x-a, \ldots, x-a)| = |x-a|^k \left|\Lambda_k(\frac{(x-a)}{|x-a|}, \ldots, \frac{(x-a)}{|x-a|})\right| \leq \|\Lambda_k\| \; |x-a|^k.$$

Therefore, we can set

$$\frac{1}{R} = \overline{\lim}_{k\to\infty} \|\Lambda_k\|^{\frac{1}{k}},$$

and see that (17) converges if $|x-a|<R$. In fact, since we use absolute values or norms for the determination of convergence we can easily extend this idea to more complex vectorspaces than $\mathbb{R}^n$ as we will show in the next examples.

Examples:

1. Let A be an $n\times n$ matrix. Define

   $$\|A\| = \max\{|Ax| : x\in\mathbb{R}^n, |x|=1\}$$

   and let

   $$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

   This series converges for any matrix since the underlying series

   $$\sum_{k=0}^{\infty} \frac{x^k}{k!}$$

   converges for any $x\in\mathbb{R}$. The matrix series is clearly a linear operator on $\mathbb{R}^n$. Finally, consider the linear system of first order differential equations:

   $$\begin{aligned} y'(t) &= Ay(t) \\ y(0) &= y_0 \end{aligned}$$

   It is easy to see that the function

   $$y(t) = e^{At}y_0.$$

   solves this initial value problem.

2. Let P be an $n\times n$ matrix with $\|P\|<1$. Then the matrix $(I-P)$ is invertable, since the sum

   $$\sum_{k=0}^{\infty} P^k$$

   converges.

5.3 Assignment 11

1. It is well known that the series $\sum_{n=1}^{\infty}\frac{1}{n}$ does not converge. Show that the sequence

   $$a_n = \sum_{k=1}^{n} - \ln n,$$

   converges. Hint: $\ln n = \int_1^n \frac{1}{x}\,dx$, and draw a picture, and think MVT.

2. Let

$$f(x) = \begin{cases} e^{\frac{1}{|x|}} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Show that this function has derivatives of arbitrarily high orders at 0, but does not have a power series with positive radius of convergence at 0.

3. Let Λ_k be a k-linear form on $\mathbb{R}^2$. Such that

$$\Lambda_k(X_1, \cdots, X_k) = (x_1 + y_1) \times \ldots \times (x_k + y_k),$$

where $X_j = (x_j, y_j)$.

   (a) Show that this is indeed a k-linear form.

   (b) Compute $\|\Lambda_k\|$.

   (c) Compute the radius of confergence of

$$\sum_{k=1}^{\infty} \Lambda_k(X, \ldots, X).$$

5.4 Fourier Series and Hilbert Spaces

For the following discussion we will actually consider complex valued functions which greatly simplify the computational aspects. Formally, a complex valued function is just the linear compination of two real valued functions. Since the major operations and properites of functions are linear operations, they work as well with complex functions. The same is of course true with products and quotients. For the function $f(x) = g(x)+ih(x)$ we define its complex conjugate $\overline{f}(x) = g(x) - ih(x)$ and

$$|f(x)| = \sqrt{f(x)\overline{f}(x)} = \sqrt{g^2(x) + h^2(x)}.$$

g and h are called the real and imaginary parts of f, $g = \Re(f)$, $h = \Im(f)$.

Lemma 5 *Let $f : [a, b] \to \mathbb{C}$ be Riemann integrable. Then*

$$\left| \int_a^b f(t)\, dt \right| \leq \int_a^b |f(t)|\, dt.$$

Proof: Let $\mu \in \mathbb{C}$ such that $|\mu| = 1$ and $\mu \int_a^b f(t)\, dt \geq 0$. Then we have

$$\begin{aligned} \left| \int_a^b f(t)\, dt \right| &= \mu \int_a^b f(t)\, dt = \int_a^b \mu f(t)\, dt \\ &= \Re \int_a^b \mu f(t)\, dt = \int_a^b \Re(\mu f(t))\, dt \\ &\leq \int_a^b |\mu f(t)|\, dt = \int_a^b |f(t)|\, dt \end{aligned}$$

A function $f : \mathbb{R} \to X$, where X is any metric, is called **periodic** with period P, if $f(x+P) = f(x)$ for any $x \in \mathbb{R}$. If f is also continuous consider the set

$$\{P \geq 0 : f(x+P) = f(x) \quad \text{for all} \quad x \in \mathbb{R}\}$$

this set is not empty and thus has a least upper bound P_0. If $P_0 = 0$ the function is obviously constant. We will concentrate our discussion on functions whose smallest period is 1. Examples of these are **trigonometric polynomials** of the form

$$P_N(x) = a_0 + \sum_{n=1}^{N} \left(a_n \cos(2\pi n x) + b_n \sin(2\pi n x)\right).$$

Recall that

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}, \qquad \text{and} \qquad \sin x = \frac{e^{ix} - e^{-ix}}{2i},$$

substitution of these expressions into the trigonometric polynomials yields:

$$P_N(x) = \sum_{n=-N}^{N} c_n e^{2\pi i n x},$$

where

$$c_n = \begin{cases} a_0 & n = 0 \\ \frac{1}{2}(a_n + ib_n) & n > 0 \\ \frac{1}{2}(a_n - ib_n) & n < 0 \end{cases}$$

Let $\mathcal{P}(a,b)$ denote the set of all trigonometric polynomials us a subset of $C([a,b])$. Then clearly this is a vector space, and it is closed under multiplication. It also separates points and contains every constant function. Thus the Stone-Weierstrass Theorem implies that it is dense in $C([a,b])$ or

Proposition 31 *Let $f : [a,b] \to \mathbb{R}$ be a contiuous function, then there exists a sequence of trigonometric polynomials that converges uniformly to f.*

Lemma 6 *Let P be a trigonometric polynomial.*

$$P(x) = \sum_{n=-N}^{N} c_n e^{2\pi i n x}$$

then for any $a \in \mathbb{R}$ and $-N \leq n \leq N$

$$c_n = \int_a^{a+1} P(x) e^{-2\pi i n x}\, dx$$

Proof: First observe that for any periodic function f with period 1 we have

$$\int_a^{a+1} f(x)\, dx = \int_0^1 f(x)\, dx$$

and for any integer k

$$\int_0^1 e^{2\pi k x}\,dx = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

The result follows immediately.

Definition 20 *Define $\mathcal{C}$ be the set of continuous 1-periodic functions.*

Clearly this space is a complex vectorspaces, however, it actually has more structure, as the next Proposition shows.

Proposition 32 *For $f, g \in \mathcal{C}$ define*

$$\langle f, g\rangle = \int_0^1 f(x)\overline{g(x)}\,dx.$$

The this operation has the properties of an inner product

Proof: Clearly $\langle f, g\rangle = \overline{\langle g, f\rangle}$, and the operator has the linearity properties of an inner product. Finally $\langle f, f\rangle \geq 0$ and $\langle f, f\rangle = 0$ if and only if $f = 0$. As usual this inner product will produce a norm on $\mathcal{C}$.

$$\|f\|_2 = \sqrt{\langle f, f\rangle},$$

is a norm on $\mathcal{C}$ as already observed in Assignment 1, thus $\mathcal{C}$ is a metric space, and a normed inner product space. This norm is called the L^2**-norm**.

Lemma 7 *$\mathcal{C}$ is not complete in the metric generated from the inner product.*

Proof: Let

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

Althuogh f_n is not yet a periodic function. Define

$$g_n(x) = \begin{cases} f_n(2x) & 0 \leq x \leq \frac{1}{2} \\ f_n(2(1-x)) & \frac{1}{2} \leq x \leq 1 \end{cases}$$

and extend it periodically. Then $g_n \in \mathcal{C}$. g_n converges pointwise to a step function. We will show that it also converges in the norm defined by the inner product to a step function. Clearly it suffices to show that

$$\int_0^1 |f_n - f|^2\,dx \to 0,$$

where

$$f(x) = \begin{cases} 1 & 0 \leq x \leq \frac{1}{2} \\ -1 & \frac{1}{2} \leq x \leq 1 \end{cases}$$

Observe, that

$$\begin{aligned} \int_0^1 |f_n - f|^2\,dx &= \int_{\frac{1}{2}-\frac{1}{n}}^{\frac{1}{2}+\frac{1}{n}} (f_n - f)^2\,dx \\ &\leq \int_{\frac{1}{2}-\frac{1}{n}}^{\frac{1}{2}+\frac{1}{n}} 1\,dx \\ &= \frac{2}{n}. \end{aligned}$$

And, hence g_n converges to a step function in the L^2-norm.

We next do a step which is similar to the step from the rational numbers to the real numbers.

Theorem 15 *For any Cauchy sequence $\{f_n\}$ in $\mathcal{C}$, let f be the function such that*

$$\|f_n - f\|_2 \to 0.$$

Let X be the set of all such functions and define an equivalence relation $\sim$ on X as follows

$$f \sim g$$

if and only if

$$\|f - g\|_2 = 0.$$

Then the space

$$\mathcal{L}^2 = X/\sim$$

of equivalence classes is a complete inner product space.

Proof: Let $\{F_n\}$ be a Cauchy sequence in $\mathcal{L}^2$ and $\epsilon > 0$. Then there exists a $N > 0$ such that

$$\|F_n - F_m\|_2 < \frac{\epsilon}{3}$$

for all $m, n \geq N$. For each $n \geq N$ there exists a $f_n \in \mathcal{C}$ such that

$$\|f_n - F_n\|_2 < \frac{\epsilon}{3}.$$

Now observe that

$$\|f_n - f_m\|_2 \leq \|f_n - F_n\|_2 + \|F_n - F_m\|_2 + \|F_m - f_m\|_2 < \epsilon,$$

for all $n, m \geq N$. Thus $\{f_n\}$ is a Cauchy sequence in $\mathcal{C}$ and therefore has a limit $f \in \mathcal{L}^2$. But

$$\|F_n - f\|_2 \leq \|F_n - f_n\|_2 + \|f_n - f\|_2$$

and thus $\{F_n\}$ converges to f. Hence, $\mathcal{L}^2$ is complete.

Definition 21 *A* **Hilbert Space** *is a complete inner product space.*

Definition 22 *For each integer let*

$$\phi_(x) = e^{2\pi inx}$$

For each $f \in \mathcal{L}^2$ *define* $\hat{f} : \mathbb{Z} \to \mathbb{C}$ *by*

$$\hat{f}_n = \int_0^1 f\overline{\phi_n}\, dx = \langle f, \phi_n \rangle.$$

$\hat{f}$ *is called the* **Fourier Transform** *of* f*.* $\hat{f}_n$ *is called the n-th Fourier coefficient of* f*. The series*

$$s(x) = \sum_{n=-\infty}^{infty} \hat{f}_n \phi_n,$$

is called the **Fourier Series** *of* f*.*

Proposition 33 *Let* Y *be the vectorspace of complex sequences. The map*

$$\Phi : \mathcal{L}^2 \to Y, \qquad \Phi : f \mapsto \hat{f}$$

is a linear operator between two complex vector spaces.

Proof: See Assignment 12.

Theorem 16 *Let* $f \in \mathcal{C}$*, and* $\hat{f}_n = 0$ *for all* $n \in \mathbb{Z}$*, then* $f = 0$*.*

Proof: Let $f \neq 0$. We may assume that f is real valued and $f(0) > 0$. There exist positive numbers δ, ϵ such that $f(x) \geq \epsilon$ for all $x \in (-\delta, \delta)$. Define

$$g(x) = 1 + \cos 2\pi x - \cos 2\pi\delta.$$

and $g_n = g^n$. Observe that $g_n(x) \to \infty$ for all $x \in (-\delta, \delta)$ and $g_n(x) \to 0$ for all $x \in [-\frac{1}{2}, \frac{1}{2}] \setminus [-\delta, delta]$. Thus $\int_{-\delta}^{\delta} f(x) g_n(x)\, dx \to \infty$, and so does

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} f(x) g_n(x)\, dx.$$

Since every g^n is a trigonometric polynomial there exist $n \in \mathbb{Z}$ such that $\hat{f}_n \neq 0$.

We next turn our attention to the space $\mathcal{L}^2$.

Lemma 8 **Bessel's Inequality** *For* $f \in \mathcal{L}^2$ *define*

$$s_N(x) = \sum_{n=-N}^{N} \hat{f}_n \phi_n(x).$$

We have

$$\langle s_N, s_N \rangle \leq \langle f, f \rangle,$$

I.e.

$$\sum_{n=-N}^{N} |\hat{f}_n|^2 \leq \|f\|_2^2.$$

Proof: Observe that $\langle \phi_n, \phi_m \rangle = 0$ if $m \neq n$, and $\langle \phi_n, \phi_n \rangle = 1$

$$\begin{aligned} 0 &\leq \|f - s_N\|_2^2 = \langle f - s_N, f - s_N \rangle \\ &= \langle f, f \rangle - \langle s_N, f \rangle - \langle f, s_N \rangle + \langle s_N, s_N \rangle \\ &= \langle f, f \rangle - \sum_{n=-N}^{N} \hat{f}_n \langle \phi_n, f \rangle - \sum_{n=-N}^{N} \overline{\hat{f}}_n \langle f, \phi_n \rangle + \sum_{n=-N}^{N} \sum_{k=-N}^{N} \hat{f}_n \overline{\hat{f}}_k \langle \phi_n, \phi_k \rangle \\ &= \langle f, f \rangle - \sum_{n=-N}^{N} |\hat{f}_n|^2. \end{aligned}$$

If f is a real valued function then

$$\hat{f}_{-n} = \overline{\hat{f}_n},$$

and for $n \neq 0$ we have

$$\hat{f}_n = a_n + ib_n$$

then the last lemma implies

$$\sum_{n=1}^{N} a_n^2 + \sum_{n=1}^{N} b_n^2 \leq \|f\|_2^2$$

Another important consequence is

Corollary 3 Riemann-Lebesgue Lemma *If $f \in \mathcal{L}^2$ then*

$$\hat{f}_n \to 0$$

as $|n| \to \infty$.

Definition 23 *We define $l^2(\mathbb{C})$ to be the space of all complex valued sequences $\{c_n\}_{n=-\infty}^{\infty}$ which satisfy*

$$\sum_{n=-\infty}^{\infty} |c_n|^2 < \infty,$$

with the norm

$$\|\{c_n\}\| = \left(\sum_{n=-\infty}^{\infty} |c_n|^2 \right)^{\frac{1}{2}}.$$

Lemma 9 *$l^2(\mathbb{C})$ is a Hilbert space.*

The prove of this is left as an exercise.

We will now prove our first result on the convergence, remember that the Stone-Weierstrass Theorem implies that for any $f \in \mathcal{C}$ there is a sequence of trigonometric polynomials that converges to f uniformly. But it does not imply that this sequence is the sequence of partial sums of the Fourier series of f. For a given $f \in \mathcal{C}$ and $\epsilon > 0$ let

$$p = \sum_{n=-N}^{N} d_n \phi_n$$

be a trigonometric polynomial such that

$$\|f - p\|_\infty < \epsilon$$

Observe that for any $n \in \{-N, N\}$ we have

$$|d_n - \hat{f}_n| = \langle p - f, \phi_{-n} \rangle \leq \|p - f\|_2 \|\phi_n\|_2 \leq \|p - f\|_\infty < \epsilon. \tag{18}$$

Since this holds for any $\epsilon > 0$, we have that the sequence of trigonometric polynomials provide by the Stone Weierstrass Theorem is in fact the Fourier series, and we have shown

Theorem 17 *For every $f \in \mathcal{C}$, the Fourier series converges uniformly to f.*

We next focus our attention to $\mathcal{L}^2$. We cannot possibly expect uniform convergence here.

Theorem 18 *Let $f \in \mathcal{L}^2$. Then there exists a sequence of trigonometric polynomials that converges to f in the L^2-norm.*

Proof: Let $\epsilon > 0$. Since $\mathcal{L}^2$ is the completion of $\mathcal{C}$ in the L^2-metric, we can find a $g \in \mathcal{C}$ such that

$$\|f - g\|_2 < \frac{\epsilon}{2}$$

and for g there exists a trigonometric polynomial p such that

$$\|p - g\|_2 < \frac{\epsilon}{2}.$$

Combining these two inequalities implies the desired result.

Theorem 19 *For every $f \in {}^2$, the Fourier series of f converges to f in the L^2-norm, and*

$$\|f\|_2^2 = \sum_{-\infty}^{\infty} \left|\hat{f}_n\right|^2 \qquad \textbf{Parseval's Relation}$$

Proof: We refer to (18) to get that the coefficients of the trig polynomial that converges to f are in fact the Fourier coefficients of f. Moreover from the proof of Bessel's inequality we get

$$\|f - s_N\|_2^2 = \|f\|_2^2 - \sum_{n=-N}^{N} |\hat{f}_n|^2,$$

and as the left hand side goe to zero as $N \to \infty$ Parseval's Identity follows.

We finish this section by

Corollary 4 *There exists an isometric, isomorphism*

$$\Phi : \mathcal{L}^2 \to l^2(\mathbb{C}).$$

Proof: The map that takes $f \mapsto \hat{f}$ clearly is a linear map. Moreover, its kernel is $\{0\}$, so it is one-to-one. For any sequence $\{c_n\} \in l^2(\mathbb{C})$. The series

$$\sum_{-\infty}^{\infty} c_n \phi_n$$

has finite L^2-norm. Thus the map is onto. Parseval's Identity provides the isometry.

5.5 Assignment 12

1. Derive the integral formulas for a_n and b_n in the trigonometric polynomial

$$P(x) = \frac{a_0}{2} + \sum_{n=1}^{N} (a_n \cos 2\pi n x + b_n \sin 2\pi n x).$$

2. Prove that Φ is a linear map.
3. Define f_a by $f_a(x) = f(x - a)$. Show that $(\hat{f_a})_n = \overline{\phi_n}(a)\hat{f}_n$
4. Prove that $l^2(\mathbb{C})$ is a Hilbert space.
5. Let $f \in \mathcal{C}$. Show that if the Fourier series of f converges uniformly, then it must converge to f.
6. Prove directly (without the Stone Weierstrass Theorem) that if $f \in \mathcal{C}$ and f is differentiable on $(0, 1)$ with continuous extensions of the derivative to $[0, 1]$. the Fourier series of f converges uniformly to f.

Index