Elementary Analysis
Math 140C—Spring 1993

Bernard Russo
September 12, 2005

- Course: Mathematics 140C MWF 11:00–11:50 SSTR 100
- Instructor: Bernard Russo PS 270 Office Hours MWF 1:15-2:00 or by appointment
- Discussion section: TuTh 11:00–11:50
- Teaching Assistant: Shandy Hauk

| Grading:               | First midterm | April 30 (Friday of week 4) | 20 percent |
|                       | Second midterm| May 28 (Friday of week 8)   | 20 percent |
| Final Exam            | June 16 (Wednesday) |                | 40 percent |
| Homework              | Due dates: April 15, 27, May 13, 25, June 8 | 20 percent |

- Text: R. C. Buck, Advanced Calculus

- Material Covered

**Schwarz inequality** Theorem 1, page 13

**topology** §1.5 pp 28–33: open, closed, boundary, interior, exterior, closure, neighborhood, cluster point

**compactness** §1.8 pp 64–67: Heine-Borel and Bolzano-Weierstrass properties (Theorems 25, 26, 27, page 65)

**continuity** §§2.2–2.4: Uniform continuity, extreme value theorems (Theorems 1, 2, 6, 10, 11, 13 on pages 73, 74, 84, 90, 91, 93)

**differentiation (of functions)** §3.3: Implies continuity, characterization by approximation (Corollary, page 129 and Theorem 8, page 131)

**integration** §§4.2–4.3: Integrability of continuous functions, fundamental theorem of calculus, mixed partial derivatives (Theorems 1, 4, 7, 11 on pages 169, 176, 182, 189)

**differentiation (of transformations)** §§7.2–7.6: Boundedness of linear transformations, characterization by approximation, chain rule, mean value theorem, inverse function theorem, implicit function theorem (Theorems 5, 8, 10, 11, 12, 16, 17, 18 on pages 335, 338, 344, 346, 350, 358, 363, 364)
1 Week 1

1.1 Monday April 5, 1993

**Theorem 1.1 (Young Inequality)** Let \( \varphi \) be differentiable and strictly increasing on \([0, \infty)\), \( \varphi(0) = 0 \), \( \lim_{u \to \infty} \varphi(u) = \infty \), \( \psi := \varphi^{-1} \), \( \Phi(x) := \int_0^x \varphi(u) \, du \), \( \Psi(x) := \int_0^x \psi(u) \, du \). Then for all \( a, b \in [0, \infty) \),

\[
ab \leq \Phi(a) + \Psi(b).
\]

Moreover, equality holds in (1) if and only if \( b = \varphi(a) \).

**Assignment 1** Give a rigorous proof of Theorem 1.1. More precisely,

Due April 15 First establish, for \( c \in [0, \infty) \), the formula

\[
\int_0^c \varphi(u) \, du + \int_0^{\varphi(c)} \psi(v) \, dv = c\varphi(c).
\]

Due April 27 Use (2) to prove (1).

Due May 13 Prove the “moreover” statement.

**Corollary 1.2** For \( p \in (1, \infty) \), and \( a, b \in [0, \infty) \),

\[
ab \leq \frac{a^p}{p} + \frac{b^q}{q},
\]

where \( q \in (1, \infty) \) is defined by

\[
\frac{1}{p} + \frac{1}{q} = 1.
\]

**Theorem 1.3 (Hölder Inequality)** Let \( x_1, \ldots, x_n \) and \( y_1, \ldots, y_n \) be real numbers and let \( p \in (1, \infty) \). Then with \( q := p/(p-1) \),

\[
\sum_{j=1}^n |x_j y_j| \leq \left( \sum_{j=1}^n |x_j|^p \right)^{1/p} \left( \sum_{j=1}^n |y_j|^q \right)^{1/q}.
\]

**Corollary 1.4 (Schwarz Inequality (Theorem 1, p.13 of [1])** For any real numbers \( x_1, \ldots, x_n \) and \( y_1, \ldots, y_n \),

\[
\sum_{j=1}^n |x_j y_j| \leq \left( \sum_{j=1}^n |x_j|^2 \right)^{1/2} \left( \sum_{j=1}^n |y_j|^2 \right)^{1/2}.
\]

**Assignment 2** (Due April 15)

- [1, §1.2 page 10 #5,10,23]
- [1, §1.3 page 18 #2,3,6]
- [1, §1.4 page 27 #3,15,16]
1.2 Wednesday April 7,1993

1.2.1 Section 1.1 of Buck

In 1, 2, or 3 dimensions you can use geometry, or geometric intuition. For dimensions 4, 5, 6…, ∞ you need algebra and analysis as tools.

1.2.2 Section 1.2 of Buck

The elements of \( \mathbb{R}^n \) := \( \{ p = (x_1, \ldots, x_n) : x_j \in \mathbb{R}, 1 \leq j \leq n \} \) may be considered as vectors (algebraic interpretation) or points (geometric interpretation). \( \mathbb{R} \) is a field which has a nice order structure, in fact, almost all properties of \( \mathbb{R}^n \) depend on those of \( \mathbb{R} \), which in turn depend on the least upper bound property of \( \mathbb{R} \). Unfortunately, no reasonable order can be defined on \( \mathbb{R}^n \) if \( n > 1 \). Although we will not consider the vector space structure of \( \mathbb{R}^n \) until later, we do need the notion of scalar product: for \( p = (x_1, \ldots, x_n), q = (y_1, \ldots, y_n) \in \mathbb{R}^n \),

\[
p \cdot q := \sum_{j=1}^{n} x_j y_j,
\]

and its properties: \( p \cdot (q + q') = p \cdot q + p \cdot q' \), etc.

1.2.3 Section 1.3 of Buck

The length of a vector \( p = (x_1, \ldots, x_n) \in \mathbb{R}^n \) is

\[
|p| = (p \cdot p)^{1/2},
\]

the distance between \( p \) and \( q \) is \( |p - q| \). The famous Schwarz inequality (a true “theorem” recorded as Corollary 1.4 above) can now be phrased compactly as

\[
p \cdot q \leq |p||q|.
\]

Here are two important corollaries (lumped together) to the Schwarz inequality.

**Corollary 1.5** For any two vectors \( p, q \),

**Triangle Inequality** : \(|p + q| \leq |p| + |q|\)

**Backwards Triangle Inequality** : \(|p - q| \geq |p| - |q|\)

Do not waste your time reading about the concepts angle, orthogonal, hyperplane, normal vector, line, convexity, which are discussed in this section. We have no use for them in this class.

A very important type of subset of \( \mathbb{R}^n \) is a ball, which is defined, for a given point \( p \in \mathbb{R}^n \) and \( r > 0 \) by

\[
B(p, r) := \{ q \in \mathbb{R}^n : |p - q| < r \}.
\]
1.3 Friday April 9, 1993

Today we want to prove (the two statements):

\[ \Delta \not\Rightarrow \left\{ \text{open ball is open set} \right\} \Rightarrow \left\{ \text{characterization of interior} \right\} \]

**Definition 1.6** Let \( S \subset \mathbb{R}^n \) and \( q \in \mathbb{R}^n \). The point \( q \) is interior to \( S \) if there exists \( \delta > 0 \) such that \( B(q,\delta) \subset S \). The interior of \( S \) is the set of all points which are interior to \( S \), notation \( \text{int} \, S \), that is

\[ \text{int} \, S = \{ q \in \mathbb{R}^n : \exists \delta > 0 \text{ such that } B(q,\delta) \subset S \}. \]

Finally, \( S \) is an open set if \( S = \text{int} \, S \).

**Proposition 1.7** Let \( p \in \mathbb{R}^n \) and \( r > 0 \). Then the ball \( B(p,r) \) is an open set.

**Proposition 1.8** ((vi) on p.32 of [1]) Let \( S \) be any non-empty subset of \( \mathbb{R}^n \). Then \( \text{int} \, S \) is the largest open subset of \( S \); more precisely

(a) \( \text{int} \, S \) is an open set;

(b) if \( T \) is an arbitrary open subset of \( S \), then \( T \subset \text{int} \, S \).

**Assignment 3** (Due April 15)

- [1, §1.5 page 36 #2,5]
- Fix \( p \in \mathbb{R}^n \). Show that \( \{ q \in \mathbb{R}^n : |q-p| > 2 \} \) is an open set.

2 Week 2

2.1 Monday April 12, 1993

**Definition 2.1** A subset \( S \) of \( \mathbb{R}^n \) is said to be a closed set if its complement \( \mathbb{R}^n \setminus S \) is an open set.

**Remark 2.2** The second part of Assignment 3 shows that the set \( \{ q \in \mathbb{R}^n : |q-p| \leq r \} \) is a closed set for any \( p \in \mathbb{R}^n \) and \( r > 0 \). Needless to say, we call such a set a “closed ball”.

**Definition 2.3** Let \( S \subset \mathbb{R}^n \) and let \( p \in \mathbb{R}^n \). We say that \( p \) is a boundary point of \( S \) if every ball with center \( p \) meets both \( S \) and its complement \( \mathbb{R}^n \setminus S \), that is, for every \( \delta > 0 \), \( B(p,\delta) \cap S \neq \emptyset \) and \( B(p,\delta) \cap (\mathbb{R}^n \setminus S) \neq \emptyset \). The boundary of \( S \), denoted by \( \text{bdy} \, S \), is the set of all boundary points of \( S \). The closure of \( S \), notation \( \overline{S} \) is defined to be \( S \cup \text{bdy} \, S \). Finally, \( p \) is a cluster point of \( S \) if every ball with center \( p \) meets \( S \) in infinitely many points, that is, for every \( \delta > 0 \), the set \( B(p,\delta) \cap S \) contains infinitely many points. We denote the set of cluster points of a set \( S \) by \( \text{cl} \, S \).
Remark 2.4 Although it is hard to believe, the point \( p \in \mathbb{R}^n \) is a cluster point of \( S \subset \mathbb{R}^n \) if and only if every ball with center \( p \) contains at least one point of \( S \) different from \( p \). (Reminder: \( p \) need not be an element of \( S \)).

Here are some examples:

<table>
<thead>
<tr>
<th>( S )</th>
<th>(( a,b ))</th>
<th>(( a,b ))</th>
<th>(( a,b ))</th>
<th>( {5 + 1/n}_{n=1}^\infty )</th>
<th>(( a,b )) \cap \mathbb{Q}</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{bdy } S )</td>
<td>{( a,b )}</td>
<td>{( a,b )}</td>
<td>{( a,b )}</td>
<td>( {5 + 1/n}_{n=1}^\infty )</td>
<td>( [a,b] )</td>
</tr>
<tr>
<td>( \text{cl } S )</td>
<td>{( a,b )}</td>
<td>{( a,b )}</td>
<td>{( a,b )}</td>
<td>( {5 + 1/n}_{n=1}^\infty )</td>
<td>( [a,b] )</td>
</tr>
</tbody>
</table>

Proposition 2.5 [(vii) and (ix) on p.32 of [1]] Let \( S \) be any subset of \( \mathbb{R}^n \).

(a) \( \overline{S} \) is the smallest closed set containing \( S \) (you know what this means)

(b) \( S \) is a closed set if and only if every cluster point of \( S \) belongs to \( S \).

Assignment 4 (Due April 27) Prove the following assertions:

(a) \( \text{int } S = \cup \{G : G \text{ is open }, G \subset S\} \)

(b) \( \overline{S} = \cap \{F : F \text{ is closed }, S \subset F\} \)

Assignment 5 (Due April 27) [1, §1.5 page 36 #6,10,11]

2.2 Wednesday April 14, 1993

Today we devote to the proof of Proposition 2.5.

Step 1: \( \overline{S} \) is a closed set.

Proof: We have to prove that the complement \( \mathbb{R}^n \setminus \overline{S} \) is an open set, so let \( q \in \mathbb{R}^n \setminus \overline{S} \). We must find a ball \( B(q, \delta) \subset \mathbb{R}^n \setminus \overline{S} \). Since \( q \not\in \overline{S} = S \cup \text{bdy } S \), \( q \not\in S \) and \( q \not\in \text{bdy } S \). The latter implies that there is a \( \delta > 0 \) such that either \( B(q, \delta) \cap S = \emptyset \) or \( B(q, \delta) \cap (\mathbb{R}^n \setminus S) = \emptyset \). The point \( q \) belongs to the latter set, so for sure \( B(q, \delta) \cap S = \emptyset \), that is, \( B(q, \delta) \subset \mathbb{R}^n \setminus S \). We complete the proof of Step 1 by showing that in fact \( B(q, \delta) \subset \mathbb{R}^n \setminus \overline{S} \). If this were not true, there would be a point \( q' \in B(q, \delta) \cap \overline{S} \). Since \( B(q, \delta) \subset \mathbb{R}^n \setminus S \), in fact we have \( q' \in B(q, \delta) \cap \text{bdy } S \). Since \( B(q, \delta) \) is an open set, there is \( \epsilon > 0 \) such that \( B(q', \epsilon) \subset B(q, \delta) \). Since \( q' \) is a boundary point of \( S \), \( B(q', \epsilon) \cap S \not= \emptyset \), a contradiction. This proves that \( \overline{S} \) is a closed set.

Step 2: If \( F \) is a closed set and \( S \subset F \), then \( \overline{S} \subset F \).

Proof: Since \( \overline{S} = S \cup \text{bdy } S \), and we are given that \( S \subset F \), we have to show only that \( \text{bdy } S \subset F \). Suppose that \( p \in \text{bdy } S \) and \( p \not\in F \). If we arrive at some contradiction, we will be done. Since \( F \) is closed, \( \mathbb{R}^n \setminus F \) is open, so there exists \( \delta > 0 \) such that \( B(p, \delta) \subset \mathbb{R}^n \setminus F \), that is, \( B(p, \delta) \cap F = \emptyset \). By the definition of boundary point, \( B(p, \delta) \cap S \not= \emptyset \). This is the desired contradiction, since \( B(p, \delta) \cap S \subset B(p, \delta) \cap F \).

Steps 1 and 2 constitute a proof of (vii) of Proposition 2.5.

Step 3: If \( S \) is a closed set, then every cluster point of \( S \) must belong to \( S \).

Proof: Indirect. Suppose \( p \) is a cluster point of the closed set \( S \). If \( p \not\in S \), then since \( \mathbb{R}^n \setminus S \) is open, there exists a ball \( B(p, \delta) \subset \mathbb{R}^n \setminus S \), that is, \( B(p, \delta) \cap S = \emptyset \). But \( B(p, \delta) \cap S \) is an infinite set, contradiction, so step 3 is proved.
Step 4: If a set $S$ contains all of its cluster points, then $S$ is a closed set.

Proof: Let $S$ be a set containing all of its cluster points. We shall show that $\mathbb{R}^n \setminus S$ is open. Let $p \in \mathbb{R}^n \setminus S$, that is, $p \notin S$. It follows from our assumption that $p$ is not a cluster point of $S$. This means that for some $\delta > 0$, the set $B(p, \delta) \cap S$ consists of only finitely many points, say $p_1, \ldots, p_m$. Since these points are in $S$ and $p \notin S$, if we set

$$\delta' = \min\{|p - p_k| : 1 \leq k \leq m\},$$

then $\delta' > 0$. Moreover, $B(p, \delta') \cap S = \emptyset$, that is, $B(p, \delta') \subset \mathbb{R}^n \setminus S$. Thus $\mathbb{R}^n \setminus S$ is open, and $S$ is closed. Step 4 is proved.

Steps 3 and 4 constitute a proof of (ix) of Proposition 2.5.

### 2.3 Friday April 16, 1993

**Definition 2.6** Let $S$ be any subset of $\mathbb{R}^n$.

- **BW** $S$ satisfies the Bolzano-Weierstrass property if every infinite sequence from $S$ has a cluster point in $S$. In other words, if $T = \{p_1, p_2, \ldots\} \subset S$ is infinite, then there exists a point $p \in S$ such that for every $\delta > 0$, $B(p, \delta) \cap T$ is an infinite set.

- **HB** $S$ satisfies the Heine-Borel property if every open cover of $S$ can be reduced to a finite subcover. In other words, if $G_1, G_2, \ldots$ is a sequence of open sets and if $S \subset G_1 \cup G_2 \cup \cdots$, then there is an integer $N$ such that $S \subset G_1 \cup G_2 \cup \cdots \cup G_N$. Another way to write this is: if $S \subset \bigcup_{n=1}^{\infty} G_n$, then for some $N \geq 1$, $S \subset \bigcup_{n=1}^{N} G_n$.

**Examples:**

- $(0, 1)$ does not satisfy BW or HB.
- $[0, \infty)$ does not satisfy BW or HB.
- $[0, 1]$ satisfies BW. This is the Bolzano-Weierstrass theorem, which you learned in Mathematics 140A or 140B. You can also find it in [1, Theorem 21, p. 62].
- $[0, 1]$ satisfies HB. This is [1, Theorem 24, p. 65]. We will discuss this in the next lecture.

We now show that the two properties are equivalent, that is, an arbitrary set $S \subset \mathbb{R}^n$ either satisfies both properties or neither property. This is stated in the next proposition.

**Proposition 2.7** Let $S$ be any subset of $\mathbb{R}^n$. Then $S$ satisfies BW if and only if it satisfies HB.

**Proof:**

Step 1: BW $\Rightarrow$ HB.
Assume that $S$ satisfies BW. Let $S \subseteq G_1 \cup G_2 \cup \cdots$. We must find $N$ such that $S \subseteq G_1 \cup G_2 \cup \cdots \cup G_N$. If this is not true, then for every $n = 1, 2, \ldots$,

$$S \not\subseteq G_1 \cup \cdots \cup G_n.$$  

For each $n$ there is thus a point $p_n \in S$ such that $p_n \not\in \{p_1, \ldots, p_{n-1}\}$ and

$$p_n \not\in G_k \text{ for } 1 \leq k \leq n. \quad (3)$$

Because $S$ satisfies BW, there is a cluster point, say $p$ of the infinite sequence $T = \{p_1, p_2, \ldots, p_n, \ldots\}$ and $p \in S$. Since $p \in S$, there is a $k_0$ such that $p \in G_{k_0}$. Since $G_{k_0}$ is an open set, there is a $\delta > 0$ such that $B(p, \delta) \subseteq G_{k_0}$. Since $p$ is a cluster point of $T$, $B(p, \delta) \cap T$ is infinite, therefore $B(p, \delta) \cap T = \{p_{n_1}, p_{n_2}, \ldots,\}$ is a subsequence, so $n_1 < n_2 < \cdots \to \infty$. We now have a contradiction: take any $n_j > k_0$. Then $p_{n_j} \in G_{k_0}$, which contradicts (3). Step 1 is proved.

3 Week 3

3.1 Monday April 19, 1993

Continuation of the proof of Proposition 2.7:

Step 2: HB $\Rightarrow$ BW.

Let $T = \{p_1, p_2, \ldots,\} \subseteq S$ be an infinite sequence, and suppose that $T$ has no cluster point in $S$. We seek a contradiction, which will then complete the proof of Step 2.

Since no point of $S$ is a cluster point of $T$, there is, for each $p \in S$, a $\delta_p > 0$ such that $B(p, \delta_p) \cap T$ is a finite set. We have

$$T \subseteq S \subseteq \cup_{p \in S} B(p, \delta_p),$$

and by HB, a finite number of the balls $B(p, \delta_p)$ cover $S$, say

$$T \subseteq S \subseteq \cup_{k=1}^m B(p_k, \delta_{p_k}).$$

Then

$$T = T \cap (\cup_{k=1}^m B(p_k, \delta_{p_k})) = \cup_{k=1}^m [T \cap B(p_k, \delta_{p_k})].$$

This is a contradiction, since $T$ is infinite and $\cup_{k=1}^m [T \cap B(p_k, \delta_{p_k})]$ is finite. This proves Step 2 and completes the proof of Proposition 2.7.

**Definition 3.1** Let $S$ be any subset of $\mathbb{R}^n$. We say $S$ is compact if it satisfies BW or HB.

**Assignment 6** (Due April 27) Prove directly the following three assertions. The fourth assertion will be proved in class.

(a) If $S$ satisfies BW, then $S$ is a closed set.
(b) If $S$ satisfies BW, then $S$ is a bounded set.

(c) If $S$ satisfies HB, then $S$ is a bounded set.

(d) (This will be done in class, not part of the homework) If $S$ satisfies HB, then $S$ is a closed set.

These assertions are stated as [1, §1.8 page 69 #1,2]

We now come to a major theorem.  

**Theorem 3.2** Let $S$ be any subset of $\mathbb{R}^n$. If $S$ is closed and bounded, then $S$ is compact.

We shall prove this theorem by showing that a closed and bounded set satisfies BW. In this form, the theorem is known as the Bolzano-Weierstrass theorem (in $\mathbb{R}^n$). Of course you may want to prove this theorem by showing that a closed and bounded set satisfies HB. In that form, the theorem is known as the Heine-Borel theorem (in $\mathbb{R}^n$). You will find the Heine-Borel theorem in [1] as Theorem 24 on page 65 (for $n = 1$) and Theorem 25 on page 65 of [1] for arbitrary $n$.

The following two lemmas, well known facts (by now) about subsequences of sequences of real numbers are the main tools in the proof of Theorem 3.2.

**Lemma 3.3** [Bolzano-Weierstrass theorem in $\mathbb{R}$] Every bounded sequence of real numbers has a convergent subsequence.

**Lemma 3.4** Every subsequence of a convergent sequence of real numbers converges to the same limit as the sequence.

### 3.2 Wednesday April 21, 1993

**Proof of Theorem 3.2:**

Since $S$ is bounded, there is a ball $B(0,M)$ with $S \subseteq B(0,M)$. Obviously

$$B(0,M) \subseteq \cap_{j=1}^{n} \{ p = (a_1, \ldots, a_n) \in \mathbb{R}^n : -M \leq a_j \leq M \}.$$  

Now let $T = \{ p_1, p_2, \ldots \} \subseteq S$ be an infinite sequence. We must find a point $p \in S$ which is a cluster point of $T$.

Choose a subsequence $T_1 = \{ q_1, q_2, \ldots \}$ of $T$ such that the sequence of first coordinates converges (you used Lemma 3.3 here since the first coordinates of $T$ lie in the closed interval $[-M, M]$). Call the limit of the sequence of first coordinates $x_1$.

Now choose a subsequence $T_2 = \{ r_1, r_2, \ldots \}$ of $T_1$ such that the sequence of second coordinates converges (Lemma 3.3 again) and call this limit $x_2$. By Lemma 3.4, the first coordinates of $T_2$ also converge to the previous $x_1$.

Continuing in this way, you obtain subsequences

$$T_n \subseteq T_{n-1} \subseteq \cdots \subseteq T_1 \subseteq T$$

---

1This makes this lecture a very important day in your life
such that the $n$ coordinate sequences of $T_n$ each converge to some number. We have decided to call these numbers $x_1, \ldots, x_n$, and we have thus defined a point $p = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Our proof will be complete as soon as we show that $p$ is a cluster point of $T$. For then, since $T \subset S$, $p$ will be a cluster point of $S$, and since $S$ is closed, $p$ will belong to $S$.

To help us prove that $p$ is a cluster point of $T$, we need some notation. Let $T_n = \{s_1, s_2, \ldots\}$ and let $s_k = (x_1^{(k)}, \ldots, x_n^{(k)})$ for $k = 1, 2, \ldots$, so that

$$
\lim_{k \to \infty} x_j^{(k)} = x_j \quad 1 \leq j \leq n. 
$$

(4)

Let $\delta > 0$. We must show that $B(p, \delta) \cap T$ is infinite. Obviously, it is enough to show that $B(p, \delta) \cap T_n$ is infinite, that is, we must show that $|p - s_k| < \delta$ for infinitely many $k$.

By (4), there exist $N_j$ ($1 \leq j \leq n$) such that $|x_j - x_j^{(k)}| < \delta/\sqrt{n}$ for $k \geq N_j$.

Then for $k \geq N := \max\{N_1, \ldots, N_n\}$ we have $|p - s_k|^2 = \sum_{j=1}^{n} (x_j - x_j^{(k)})^2 \leq n(\delta^2/n) = \delta^2$. Therefore

$$
\{s_N, s_{N+1}, \ldots\} \subset T_n \cap B(p, \delta).
$$

This completes the proof of Theorem 3.2.

### 3.3 Friday April 23, 1993

When you try to prove the false statement “every set is closed”, you find that it helps if you assume that the set is compact.

**Proposition 3.5** Every compact set in $\mathbb{R}^n$ is closed.

**Proof:** Let $S$ be a compact subset of $\mathbb{R}^n$. We show directly that $\mathbb{R}^n \setminus S$ is an open set by using the Heine-Borel property HB. Let $p \in \mathbb{R}^n \setminus S$. For each $q \in S$, let $\delta_p := |p - q|/2$. Since $p \neq q$, $\delta_q > 0$. Now cover $S$:

$$
S \subset \bigcup_{q \in S} B(q, \delta_q).
$$

By HB, there exist finitely many points $q_1, \ldots, q_m \in S$ such that $S \subset \bigcup_{j=1}^{m} B(q_j, \delta_{q_j})$. Then $V := \bigcap_{j=1}^{m} B(p, \delta_{q_j})$ is an open set\(^2\) containing $p$, in fact it is an open ball $B(p, \min\{\delta_{q_j} : 1 \leq j \leq m\})$. Since $B(p, \delta_{q_j})$ is disjoint from $B(q_j, \delta_{q_j})$, it follows that $V$ is disjoint from $\bigcup_{j=1}^{m} B(q_j, \delta_{q_j})$, and hence from $S$, that is, $V \subset \mathbb{R}^n \setminus S$. Thus $S$ is closed. This completes the proof.

\(^2\)because it is a finite intersection!! (this is the beauty of the Heine-Borel property)
Proposition 3.6 (Part of (viii) on page 32 of [1]) For any subset $S$ of $\mathbb{R}^n$, its boundary $\partial S$ is a closed set.

Proof: Just note that for any set $S$, we have the decomposition

$$\mathbb{R}^n = \text{int } S \cup \partial S \cup \text{int } (\mathbb{R}^n \setminus S)$$

of Euclidean space $\mathbb{R}^n$ into three disjoint subsets. It follows that $\partial S = \mathbb{R}^n \setminus (\text{int } S \cup \text{int } (\mathbb{R}^n \setminus S))$ is the complement of an open set. This completes the proof.

MIDTERM ALERT NUMBER 1

The first midterm will take place on Friday April 30 and will cover sections 1.5 (pages 28–33 only) and 1.8 of [1]. Of particular interest are the 10 statements (Propositions) on page 32. You should understand each step in the proofs of these propositions. Some exercises to think about (but not hand in) would be on page 36, #1,3,4,7,8,9,12,14.

For purposes of this midterm, you may ignore Young’s inequality, Hölder’s inequality and the Schwarz inequality (in section 1.3), and the notion of connectedness in section 1.8. We will use the Schwarz inequality in a significant way later in the course but we will not have time to study the important topic of connectedness in this course.

The important results in section 1.8 are the following: you should understand each step in the proofs.

- $\text{BW} \Rightarrow \text{HB}$
- $\text{HB} \Rightarrow \text{BW}$
- $\text{BW} \Rightarrow$ closed
- $\text{HB} \Rightarrow$ closed
- $\text{BW} \Rightarrow$ bounded
- $\text{HB} \Rightarrow$ bounded
- closed and bounded $\Rightarrow \text{BW}$
- closed and bounded $\Rightarrow \text{HB}$

You may ignore Theorems 27,28,29,30 on pages 65–69 of [1]. We shall not discuss them. Note that the proofs of Theorems 24,25,26 on page 65 are contained in the results listed above.

Make sure you understand the homework you turned in on April 15 and April 27.

Reminder: the following homework is due on Tuesday April 27:

\[ \text{Be sure to check this carefully} \]
\[ \text{Do this on your own} \]
\[ \text{we did not discuss this one—this is included in the proof of Theorem 25 in [1, p.67] Of course the result follows from the preceding fact since } \text{BW} \Rightarrow \text{HB} \]
\[ \text{however, you are in a good position to understand them} \]
• Assignment #1 (b) Young’s inequality
• Assignment #4 (a) formula for interior (b) formula for closure
• Assignment #5 page 36 #6,10,11
• Assignment #6 BW ⇒ closed, BW ⇒ bounded, HB ⇒ bounded.

4 Week 4

4.1 Monday April 26, 1993

Proposition 4.1 ((iii) and (iv) on p.32 of [1])

(a) If $A$ and $B$ are closed subset of $\mathbb{R}^n$, then so are $A \cap B$ and $A \cup B$.

(b) If $\{A_k\}_{k=1}^{\infty}$ is a sequence of closed sets, then $\cap_{k=1}^{\infty} A_k$ is closed but $\cup_{k=1}^{\infty} A_k$ need not be closed.

First proof: use De Morgan’s law:

$$\mathbb{R}^n \setminus \cap_{k=1}^{\infty} A_k = \cup_{k=1}^{\infty} (\mathbb{R}^n \setminus A_k).$$

Second proof: Let $S := \cap_{k=1}^{\infty} A_k$ and let $p$ be a cluster point of $S$. We shall show that $p \in S$. Since $S \subset A_k$ for every $k$, for every $\delta > 0$, $B(p, \delta) \cap S \subset B(p, \delta) \cap A_k$. Thus $p$ is a cluster point of $A_k$. Since $A_k$ is closed, $p \in A_k$ for every $k$, that is, $p \in S$.

Proposition 4.2 (Another part of (viii) on p.32 of [1]) For any subset $S$ of $\mathbb{R}^n$,

$$\text{bdy } S = \overline{S} \cap (\mathbb{R}^n \setminus S).$$

Proof:

$$\overline{S} \cap (\mathbb{R}^n \setminus S) = (S \cup \text{bdy } S) \cap ((\mathbb{R}^n \setminus S) \cup \text{bdy } (\mathbb{R}^n \setminus S))$$

$$= (S \cup \text{bdy } S) \cap ((\mathbb{R}^n \setminus S) \cup \text{bdy } S)$$

$$= \text{bdy } S.$$

Here is a preview of our next topic: continuous functions. We need to do only two theorems. The rest is either trivial modification of what you learned in 140AB or consequence of these two theorems.

The main theorems on continuous functions deal with compact sets. They are

• Theorem 13 on page 93 of [1]: The continuous real valued image of a compact subset of $\mathbb{R}^n$ is a compact subset of $\mathbb{R}$.

\footnote{Do not read the proof of Theorem 13 in [1], we will present a better one}
• Theorem 6 on page 84 of [1]: A continuous real valued function on a compact subset of $\mathbb{R}^n$ is uniformly continuous.

Both of these theorems are well known to you in the following form for $n = 1$.

• A continuous function on a closed interval $[a, b]$ is bounded, and assumes a maximum and minimum on $[a, b]$; that is, there exist points $\alpha, \beta \in [a, b]$ (not necessarily unique) such that $f(\alpha) \leq f(x) \leq f(\beta)$ for every $x \in [a, b]$.

(This is stated for functions defined on compact subsets of $\mathbb{R}^n$ as Theorem 10 on page 90 and Theorem 11 on page 91 of [1])

• Theorem 6 on page 84 of [1]: A continuous real valued function on a compact subset of $\mathbb{R}^n$ is uniformly continuous.

4.2 Wednesday April 28, 1993

Definition 4.3 Let $f : D \rightarrow \mathbb{R}$ be a function, where $D$ is any subset of $\mathbb{R}^n$, and let $p_0 \in D$. We say that $f$ is continuous at $p_0$ if

$$\forall \epsilon > 0, \exists \delta > 0$$

such that

$$|f(p) - f(p_0)| < \epsilon$$

for all $p \in D$ with $|p - p_0| < \delta$.

It is important to realize that this lengthy definition can be put in the compact form

$$\forall \epsilon > 0, \exists \delta > 0$$

such that $f[D \cap B(p_0, \delta)] \subset B(f(p_0), \epsilon)$.

Here, we are using the notation

$$f(A) := \{ f(p) : p \in A \}$$

if $A \subset D$.

We refer to $f(A)$ as the image of $A$ under $f$.

Please note that the above definition is a “local” one, that is, concerns a single point $p_0$, together with “neighboring” points. We say $f$ is continuous on $D$ if it is continuous at each point of $D$. This is a “global” definition of continuity.

Here is a description of the first five theorems of Chapter 2 of [1]

Theorems 1,2 page 73-74 These will be discussed in the next lecture (after the first midterm). They concern a characterization of continuity at a point in terms of convergence of sequences, and are extremely useful.

Theorem 3 page 76 This is a global characterization of continuity. It becomes messy if the domain $D$ is not an open set, and for this reason we shall not spend any time on it.

---

8 $\delta$ depends in general on $p_0$ as well as on $\epsilon$

9 no pun intended
Theorem 4 page 77 This concerns the “algebra” of continuous functions, that is sums, products, quotients, and is familiar from elementary calculus. This is important to know but we shall not spend time on it. It is used in [1] to give a proof of the extreme value theorem ([1, Theorem 11, page 91]), but we shall give an independent proof of the extreme value theorem, using only compactness.

Theorem 5 page 78 This involves composite functions and we shall discuss it in connection with our study of the chain rule, later in this course.

Assignment 7 (Due May 13) [1, §2.2 page 80 #1 or 2,3 or 4,7 or 8,12 or 13,14 or 17] You are to hand in 4 problems, one from each of these 5 pairs. You will of course be responsible for all of the problems.

In [1, Section 2.3] we will discuss Definition 2 on page 82 and Theorem 6 on page 84 next week. We will not have time for Definition 3 and Theorem 7, which can be ignored.

Assignment 8 (Due May 13) [1, §2.3 page 88 #1,3 or 4,5 or 6,7]

In [1, Section 2.4] Theorems 10 and 11 follow easily from Theorem 13, as we now show. Before we do that, let us note that Theorems 8,9 and 12 can be skipped (we need Theorem 8 later, but we can wait on that). Theorems 14,15,16 involve connectedness and we have to skip them. As stated earlier, any serious student of analysis needs to study this topic (but wait until after this course is finished).

Theorem 4.4 The continuous image of a compact set is compact. In other words, if \( f : D \to \mathbb{R} \) is a continuous function on \( D \), and \( D \) is a compact subset of \( \mathbb{R}^n \), then \( f(D) \) is a compact subset of \( \mathbb{R} \).

Proof: We choose\(^{10}\) to show that \( f(D) \) satisfies the HB property. We shall use the fact that \( D \) satisfies the HB property.

Let 
\[ f(D) \subset \bigcup_{k=1}^{\infty} G_k \]
be an open cover of \( f(D) \). For each \( p \in D \), \( f(p) \in f(D) \) and so there is a member of the cover, say \( G_{k_p} \), with \( f(p) \in G_{k_p} \). Since the cover is an open cover, \( G_{k_p} \) is an open set so there is \( \epsilon_p > 0 \) such that \( B(f(p), \epsilon_p) \subset G_{k_p} \). Since \( f \) is continuous at every point of \( D \), there exists \( \delta_p > 0 \) such that 
\[ f[B(p, \delta_p) \cap D] \subset B(f(p), \epsilon_p) \]

We can now cover \( D \)\(^{11}\):
\[ D \subset \bigcup_{p \in D} B(p, \delta_p). \]

\(^{10}\)how many choices are there?

\(^{11}\)the redundant cover!
Since $D$ is compact, the HB property tells us there are a finite number of points $p_1, \ldots, p_m$ say, such that
\[
D \subset \bigcup_{j=1}^{m} B(p_j, \delta_{p_j}).
\]
It follows that $D = \bigcup_{j=1}^{m} [B(p_j, \delta_{p_j}) \cap D]$, and therefore that
\[
f(D) = \bigcup_{j=1}^{m} f\left[B(p_j, \delta_{p_j}) \cap D\right] \subset \bigcup_{j=1}^{m} B(f(p_j), \epsilon_{p_j}) \subset \bigcup_{j=1}^{m} G_{p_j}.
\]
We have reduced the cover to a finite subcover, so the proof is complete.

4.3 Friday April 30, 1993
First Midterm—Math 140C—April 30, 1993

Do all problems. However, there is a choice in one of them, number 8
Use only one side of your page and only one problem per page.

**Problem 1 (10 points)** Let $S$ be a closed subset of $\mathbb{R}^n$, that is $\mathbb{R}^n \setminus S$ is an open set. Prove

(A) $\text{bdy } S \subset S$.

(B) $S = \overline{S}$

**Problem 2 (12 points)** Prove rigorously that the set $\mathbb{N} = \{1, 2, \ldots\}$ of natural numbers is a closed subset of $\mathbb{R}^1$. Is it a closed subset of $\mathbb{R}^2$? (Yes or no, no proof required for this part of the question). Is
\[
S := \{(m, k) : m \in \mathbb{N}, k \in \mathbb{N}\}
\]
a closed subset of $\mathbb{R}^2$? (Yes or no, no proof required).

**Problem 3 (12 points)** Find $\text{bdy } S, \text{int } S, \text{ and all cluster points of } S$ if
\[
S = \{(x, y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 < 2\} \cup \{(x, 0) : 0 < x \leq 1/2\}
\]
Just write down your answer, no proof is required.

**Problem 4 (5 points)** Prove or disprove: For every $S \subset \mathbb{R}^n$, $S \setminus \text{int } S = \text{bdy } S$.

**Problem 5 (25 points)** Let $A$ be a bounded subset and $B$ a closed subset of $\mathbb{R}^n$ and suppose that $A \cap B \neq \emptyset$. True or false (3 points for each correct answer, 2 more points for the proof or example)

(A) $A \cap B$ is bounded

(B) $A \cup B$ is bounded

(C) $A \cup B$ is closed
(D) \( A \cap B \) is compact

(E) \( A \cap (\mathbb{R}^n \setminus B) \) is bounded

Problem 6 (10 points) Prove that a compact set is bounded. You may use BW or HB.

Problem 7 (10 points) Let \( A \) be any subset of a compact set \( S \). Assume the Heine-Borel and Bolzano-Weierstrass theorems and their converses.

(A) Prove that if \( A \) is closed, then \( A \) is compact.

(B) Now suppose again that \( A \) is an arbitrary subset of the compact set \( S \). Prove that \( A \), the closure of \( A \), is a compact set.

Problem 8 (16 points) Let \( S \) be an arbitrary subset of \( \mathbb{R}^n \). Do only one of (A) or (B), not both.

(A) Let \( p \in \mathbb{R}^n \) and suppose that for every \( \delta > 0 \), \( B(p, \delta) \cap S \) contains at least one point different from \( p \). Show that \( p \) is a cluster point of \( S \).

(B) Let \( T \) be the set of cluster points of \( S \). Prove that \( T \) is a closed set.

5 Week 5

5.1 Monday May 3, 1993

5.1.1 Sequences of points in \( \mathbb{R}^n \)

Definition 5.1 Let \( \{p_k\}_{k=1}^{\infty} \subset \mathbb{R}^n \) be a subset indexed by the natural numbers, and let \( p \in \mathbb{R}^n \). We say the sequence \( \{p_k\} \) converges to \( p \) if

\[
\lim_{k \to \infty} |p_k - p| = 0,
\]

that is, for every \( \epsilon > 0 \), there exists \( N \) such that

\[
|p_k - p| < \epsilon \text{ for all } k > N.
\]

Notation for this is: \( \lim_{k \to \infty} p_k = p \) or \( \lim_k p_k = p \) or \( \lim p_k = p \) or \( p_k \to p \) as \( k \to \infty \).

Introduce coordinates of the points \( p_k \) and \( p \):

\[
p = (x_1, \ldots, x_n) \text{ and } p_k = (x_1^{(k)}, \ldots, x_n^{(k)}).
\]

Then

\[
|p - p_k|^2 = \sum_{j=1}^{n} (x_j - x_j^{(k)})^2 \geq (x_j - x_j^{(k)})^2 \text{ for all } 1 \leq j \leq n.
\]

This proves the following:
Theorem 5.2 (Theorem 7 on page 42 of [1]) Let \( \{p_k\}_{k=1}^{\infty} \subset \mathbb{R}^n \) be a sequence, and let \( p \in \mathbb{R}^n \). Then
\[
\lim_{k \to \infty} p_k = p,
\]
if and only if
\[
\lim_{k \to \infty} x_j^{(k)} = x_j \text{ for } 1 \leq j \leq n.
\]

Theorem 5.3 (Theorem 3 on page 40 of [1]) A convergent sequence in \( \mathbb{R}^n \) is bounded.

Proof: Let \( p_k \to p \). Choose \( N \) such that \( |p_k - p| < 1 \) if \( k > N \). Then
\[
|p_k| \leq |p_k - p| + |p| < 1 + |p| \text{ for } k > N
\]
and so \( \{p_k\}_{k=1}^{\infty} \subset B(0, M) \) where
\[
M = \max\{1 + |p|, |p_1|, \ldots, |p_N|\},
\]
that is, the sequence is bounded.

Theorem 5.4 (Theorem 1 on page 73 of [1]) Let \( f : D \to \mathbb{R} \), where \( D \subset \mathbb{R}^n \), and suppose that \( f \) is continuous at the point \( p_0 \in D \). Then for every sequence \( p_k \) from \( D \), which converges to \( p_0 \), we have
\[
\lim_{k \to \infty} f(p_k) = f(p_0).
\]

Proof: Let \( \epsilon > 0 \). We have to prove there is an \( N \) such that \( |f(p_k) - f(p_0)| < \epsilon \) for all \( k > N \). Since \( f \) is continuous at \( p_0 \), there exists \( \delta > 0 \) such that
\[
f[D \cap B(p_0, \delta)] \subset B(f(p_0), \epsilon).
\]
Since \( p_k \to p_0 \), and since \( \delta > 0 \), there exists \( N \) such that
\[
p_k \in B(p_0, \delta) \text{ for } k > N.
\]
Putting together (5) and (6) results in \( f(p_k) \in B(f(p_0), \epsilon) \) for \( k > N \).

At this point you are in a position to give a proof of Theorem 4.4 above using the property BW (at both ends). I strongly suggest that you do this as an informal exercise. The following lemma, which we shall use in the extreme value theorem (Theorem 5.7 below) may be helpful in that informal exercise.

Lemma 5.5 For any subset \( S \subset \mathbb{R}^n \), the set of cluster points of \( S \) coincides with the limits of sequences of distinct points from \( S \). In particular, a point is a cluster point of a sequence if and only if it is a limit of a convergent subsequence of the sequence.
**Proof:** Let \( p \) be a cluster point of \( S \). Pick \( p_k \in B(p, \frac{1}{k}) \cap S \). Since this set is infinite, we can certainly assume that \( p_k \not\in \{p_1, \ldots, p_{k-1}\} \). Then \( |p_k - p| < 1/k \to 0 \), so \( p_k \to p \), as required. Conversely if \( p = \lim_{k \to \infty} p_k \) with \( p_k \in S \) all distinct, then for any \( \delta > 0 \), there exists \( N \) such that \( \{p_{N+1}, p_{N+2}, \ldots\} \subset B(p, \delta) \cap S \), so \( B(p, \delta) \cap S \) is an infinite set. DONE

**Theorem 5.6 (Theorem 10 on page 90 of [1])** A continuous function on a compact set is bounded. That is, if \( f : D \to \mathbb{R} \) is continuous on \( D \subset \mathbb{R}^n \) and \( D \) is compact, then \( f \) is a bounded function on \( D \).

**Proof:** This is now trivial, since by Theorem 4.4, \( f(D) \) is compact, hence bounded. (Note that Theorem 4.4 does not depend on Theorem 5.6, so it is OK to use it in the proof).

**Theorem 5.7 (Theorem 11 on page 91 of [1], Extreme values Theorem)** A continuous function \( f \) on a compact set \( D \subset \mathbb{R}^n \) assumes its maximum and its minimum at some points of \( D \).

**Proof:** By Theorem 5.6, \( f \) is bounded, that is \( f(D) \) is a bounded subset of \( \mathbb{R} \). Let

\[
\beta := \sup \{ f(p) : p \in D \},
\]

so that \( \beta \in \mathbb{R} \). By definition of supremum, for each \( k \geq 1 \), there is a point \( p_k \in D \) such that

\[
\beta - \frac{1}{k} \leq f(p_k) \leq \beta. \tag{7}
\]

Since \( D \) is compact, BW implies the existence of a cluster point \( p_0 \) of the sequence \( p_k \), and \( p_0 \in D \). By Lemma 5.5, there is a subsequence \( p_{k_j} \) such that \( \lim_{j \to \infty} p_{k_j} = p_0 \). In particular, from (7), for \( j = 1, 2, \ldots \),

\[
\beta - \frac{1}{k_j} \leq f(p_{k_j}) \leq \beta.
\]

Now let \( j \to \infty \) to get \( \beta \leq f(p_0) \leq \beta \), that is \( f \) assumes its maximum at \( p_0 \in D \).

Similar proof for minimum. DONE

**Assignment 9** (Due May 13) [1, §1.6 page 54 #1 or 2, 3 or 4, 31 or 33, 32 or 35]

### 5.2 Wednesday May 5, 1993

Today we spent most of the time discussing the solutions to the problems on the first midterm. It wasn’t a total loss though, as we were able to discuss the following extremely useful characterization of closed sets.

**Theorem 5.8 (Theorem 5 on page 40 of [1])** Let \( S \) be any subset of \( \mathbb{R}^n \). Then

\[
\overline{S} = \{ \lim_{k \to \infty} p_k : \{p_k\} \subset S, \lim_k p_k \text{ exists} \}. \tag{8}
\]
Proof: Suppose first that \( p = \lim_k p_k \) for some sequence \( p_k \) from \( S \). If \( p \notin \overline{S} = \text{bdy} \, S \cup S \), then \( p \notin S \) and \( p \notin \text{bdy} \, S \). Thus there exists \( \delta > 0 \) such that at least one of \( B(p, \delta) \cap S \) or \( B(p, \delta) \cap (\mathbb{R}^n \setminus S) \) is empty. But the first one is non-empty since it contains some elements of the sequence \( p_k \). Thus the second one is empty, which means \( B(p, \delta) \subset S \). This is a contradiction to \( p \notin S \). We have proved that the right side of (8) is contained in the closure of \( S \).

Now let \( p \in \overline{S} \), and suppose first that \( p \in S \). Then the sequence \( p_k \) defined by \( p_k = p \) for \( k = 1, 2, \ldots \) converges to \( p \). Next suppose that \( p \in \text{bdy} \, S \), so that for every \( k \geq 1 \), \( B(p, \frac{1}{k}) \cap S \neq \emptyset \). Pick a point \( p_k \in B(p, \frac{1}{k}) \cap S \), so that \( p_k \) is a sequence from \( S \) which converges to \( p \) since \( |p - p_k| < 1/k \to 0 \).

Corollary 5.9 (Corollary 2 on page 41 of [1]) A set \( S \) is closed if and only if it contains the limit of each convergent sequence of points from \( S \).

Definition 5.10 (Definition 6 on page 52 of [1]) A sequence \( p_k \) of points in \( \mathbb{R}^n \) is said to be a Cauchy sequence if for every \( \epsilon > 0 \), there exists \( N \) such that \( |p_k - p_j| < \epsilon \) for all \( k \geq N \) and \( j \geq N \).

The following theorem follows easily from the case \( n = 1 \) by considering the sequences of coordinates of all the points involved.

Theorem 5.11 (Corollary on page 63 and exercise 32 on page 56 of [1]) A sequence in \( \mathbb{R}^n \) is convergent if and only if it is a Cauchy sequence.

5.3 Friday May 7, 1993

Definition 5.12 [Definition 2 on page 82 of [1]] A function \( f : E \to \mathbb{R} \), where \( E \subset \mathbb{R}^n \), is uniformly continuous on \( E \) if for every \( \epsilon > 0 \), there exists \( \delta > 0 \) such that \( |f(p) - f(q)| < \epsilon \) whenever \( p, q \in E \) and \( |p - q| < \delta \).

Theorem 5.13 (Theorem 6 on page 84 of [1]) A function which is continuous on a compact set \( D \) is uniformly continuous on \( D \).

Proof: Let \( \epsilon > 0 \). For each \( p \in D \), there exists \( \delta_p > 0 \) such that \( f[B(p, \delta_p) \cap D] \subset B(f(p), \epsilon/2) \). We shall refer to \( B(p, \delta_p) \) as a “continuity ball”. Now cover \( D \) by the corresponding balls with radius halved, that is,

\[ D \subset \bigcup_{p \in D} B(p, \delta_p/2). \]

We can refer to \( B(p, \delta_p/2) \) as a “covering ball”. By compactness, we have \( D \subset \bigcup_{j=1}^{m} B(p_j, \delta_{p_j}/2) \). Now set \( \delta = \min_{1 \leq j \leq m} \{\delta_{p_j}/2\} \). It remains to prove that if \( x, y \in D \) and \( |x - y| < \delta \), then \( |f(x) - f(y)| < \epsilon \).

Since \( x \in D \) there is a \( j \) such that \( x \in B(p_j, \delta_{p_j}/2) \). Since \( |x - y| < \delta \leq \delta_{p_j}/2 \) we have \( |y - p_j| \leq |y - x| + |x - p_j| < \delta + \delta_{p_j}/2 \leq \delta_{p_j} \). In other words, \( x \) and \( y \) both belong to the same continuity ball \( B(p_j, \delta_{p_j}) \). Thus

\[ |f(x) - f(y)| \leq |f(x) - f(p_j)| + |f(p_j) - f(y)| < \epsilon/2 + \epsilon/2 = \epsilon. \]
The proof is complete.

As the next assignment shows, there are non-trivial uniformly continuous functions on non-compact sets.

**Assignment 10 (Due May 13)** *Show that* \( f \) *and* \( g \) *are uniformly continuous on* \( \mathbb{R}^n \), *where*

(A) \( f(p) = |p| \) (Hint: triangle inequality)

(B) \( g(p) = x_1 y_1 + \cdots + x_n y_n \) where \( p = (x_1, \ldots, x_n) \in \mathbb{R}^n \) is a variable point and \( y_1, \ldots, y_n \in \mathbb{R} \) are fixed.

(C) [1, p.88#6], namely, a uniformly continuous function preserves Cauchy sequences.

We shall consider two applications of uniform continuity. Next week we shall study Riemann integration and show that every continuous function on a close rectangle in \( \mathbb{R}^2 \) is integrable (using the fact that it is automatically uniformly continuous, a closed rectangle being a compact set)

Today we consider the another application in the form of a solution to a particular mathematical problem. Let \( S \) be any subset of \( \mathbb{R}^n \) and let \( f : S \rightarrow \mathbb{R} \) be a continuous function. The problem is: can \( f \) be extended to a continuous function, call it \( \tilde{f} \), on the closure \( \overline{S} \) of \( S \)? Stated again, given \( f \) continuous on \( S \), does there exist a continuous function \( \tilde{f} \) on \( \overline{S} \), such that \( \tilde{f}(p) = f(p) \) for \( p \in S \)?

We know already that the answer is no, as the example \( f(x) = 1/x \) on \( S = (0,1) \subset \mathbb{R} \) shows. So to get a positive answer, we must put some restrictions on the function \( f \) and/or on the set \( S \). We will find that if we assume that \( f \) is uniformly continuous on \( S \), then the answer is yes for any set \( S \).

To solve this problem we note first that our hands are tied by Theorems 5.8 and 5.4. That is, we have no choice, we must define the extension \( \tilde{f} \) as follows:

\[
\tilde{f}(p) = \begin{cases} 
  f(p) & \text{if } p \in S; \\
  \lim_{k \to \infty} f(p_k) & \text{if } p \in \overline{S} \setminus S,
\end{cases}
\]

where \( p_k \in S \) is such that \( \lim_{k} p_k = p \).

To make this construction legitimate, we must answer three questions:

- Why does \( \lim_{k} f(p_k) \) exist?
- Why is \( \lim_{k} f(p_k) \) independent of the sequence \( p_k \) chosen in \( S \)?
- Why is \( \tilde{f} \) (which is a function by positive answers to the first two questions) continuous on \( \overline{S} \)?

In order to get affirmative answers to the first and third questions, we have to make an assumption on \( f \), but not on \( S \). The first two questions are easy to answer, so let’s get them out of the way before this lecture ends.
Assume now that $f$ is not merely continuous on $S$, but uniformly continuous on $S$. If $p_k$ is any sequence from $S$ which converges\footnote{Such a sequence exists by Theorem 5.8} to $p \in \overline{S}$, then $p_k$ is a Cauchy sequence, and by uniform continuity of $f$, Assignment 10(C) tells us that $f(p_k)$ is a Cauchy sequence in $\mathbb{R}$. Hence the limit exists and the first question is answered affirmatively.

We now answer the second question. Let $\{p_k\}$ and $\{q_k\}$ be any two sequences from $S$ which converge to $p \in S$. By the answer to the first question, the limits $\alpha := \lim_k f(p_k)$ and $\beta := \lim_k f(q_k)$ exist. We must show that $\alpha = \beta$. To do this, consider a third sequence, obtained by interlacing the two given sequences: $p_1, q_1, p_2, q_2, \ldots$. Obviously, this sequence converges to $p$ also, so the sequence of function values $f(p_1), f(q_1), f(p_2), f(q_2), \ldots$, converges, say to a number $\gamma$. Since every subsequence of this sequence must also converge to $\gamma$, it follows that $\alpha = \gamma$ and $\beta = \gamma$, so $\alpha = \beta$, as required. The second question is answered affirmatively.

6 Week 6

6.1 Monday May 10, 1993

Today’s lecture is devoted to the answer to the third question raised in the last lecture. Let us state this as a theorem.

**Theorem 6.1** Let $f : S \to \mathbb{R}$ be a uniformly continuous function defined on a subset $S$ of $\mathbb{R}^n$. Define a function $\tilde{f} : \overline{S} \to \mathbb{R}$ by

$$
\tilde{f}(p) = \begin{cases} 
  f(p) & \text{if } p \in S; \\
  \lim_{k \to \infty} f(p_k) & \text{if } p \in \overline{S} \setminus S,
\end{cases}
$$

where $p_k \in S$ is such that $\lim_k p_k = p$. Then $\tilde{f}$ is continuous\footnote{The proof will show that actually $\tilde{f}$ is uniformly continuous on $\overline{S}$} on $\overline{S}$.

**Proof:** Let $p \in \overline{S}$ and let $\epsilon > 0$. We shall produce a $\delta > 0$ such that $\tilde{f}[B(p, \delta) \cap \overline{S}] \subset B(\tilde{f}(p), \epsilon)$, that is,

$$
|\tilde{f}(p) - \tilde{f}(q)| < \epsilon \text{ if } q \in \overline{S} \text{ and } |q - p| < \delta.
$$

Discussion (sidebar): here are the basic ideas of the proof. Make sure you understand the reason for each assertion below.

1. The points $p, q (\in \overline{S})$ have “neighbors” $p_k, q_j \in S$: for example $|p - p_k| < 1/k$ and $|q - q_j| < 1/j$.

2. $\tilde{f}(p)$ and $f(p_k)$ are “close”; so are $\tilde{f}(q)$ and $f(q_j)$.

3. if $p_k$ and $q_j$ are close, so are $f(p_k)$ and $f(q_j)$.
4. if \( p \) and \( q \) are close, so are \( p_k \) and \( q_j \).

5. end of sidebar

We now make these statements precise. We begin with the triangle inequality:

\[
|\tilde{f}(p) - \tilde{f}(q)| \leq |\tilde{f}(p) - f(p_k)| + |f(p_k) - f(q_j)| + |f(q_j) - f(q)|.
\]  

(9)

There exists \( N_1 = N_1(\epsilon/3, p) \) such that \( |\tilde{f}(p) - f(p_k)| < \epsilon/3 \) for all \( k > N_1 \) and there exists \( N_2 = N_2(\epsilon/3, q) \) such that \( |\tilde{f}(q) - f(q_j)| < \epsilon/3 \) for all \( j > N_2 \). (This takes care of the first and third terms on the right side of (9)).

There exists \( \delta_1 = \delta_1(f, \epsilon/3, S) \) such that \( |f(x) - f(y)| < \epsilon/3 \) whenever \( x, y \in S \) and \( |x - y| < \delta_1 \). In particular, for the middle term on the right side of (9), \( |f(p_k) - f(q_j)| < \epsilon/3 \) if \( |p_k - q_j| < \delta_1 \).

Now note that (again by the triangle inequality)

\[
|p_k - q_j| \leq |p_k - p| + |p - q| + |q - q_j|.
\]  

(10)

Thus, if we define \( \delta := \delta_1/2 \), then from (10), if \( |p - q| < \delta \), and \( k, j \) are large enough, then \( |p_k - q_j| \) will be less than \( \delta_1 \).

Conclusion: if \( |p - q| < \delta \), where \( \delta = \delta_1(f, \epsilon/3, S) \), then, \( |\tilde{f}(p) - \tilde{f}(q)| < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon \), by (9), where \( k, j \) are chosen so that \( k > N_1, j > N_2 \) and \( 1/k + 1/j < \delta_1 \).

This completes the proof.

**Assignment 11** (Due May 25)

(A) Let \( S \subset \mathbb{R}^n \) be a bounded set and let \( f : S \to \mathbb{R} \) be a continuous function. Prove that \( f \) has a continuous extension to \( \overline{S} \) if and only if \( f \) is uniformly continuous on \( S \).

(B) Let \( f : \mathbb{R} \to \mathbb{R} \) be continuous and suppose that

\[
\lim_{x \to -\infty} f(x) = \lim_{x \to -\infty} f(x) = 0.
\]

Prove that \( f \) is uniformly continuous on \( \mathbb{R} \).

REMINDER: The following is due on Thursday May 13:

- Assignment #1(c): Prove the “moreover” statement in Young’s inequality.
- Assignment #7: [1, §2.2 page 80 #1 or 2,3 or 4,7 or 8,12 or 13,14 or 17] You are to hand in 4 problems, one from each of these 5 pairs. You will of course be responsible for all of the problems.
- Assignment #8 [1, §2.3 page 88 #1,3 or 4,5 or 6,7]
- Assignment #9 [1, §1.6 page 54 #1 or 2,3 or 4,31 or 33,32 or 35]
- Assignment #10 Show that \( f \) and \( g \) are uniformly continuous on \( \mathbb{R}^n \), where
(A) \( f(p) = |p| \) (Hint: triangle inequality)

(B) \( g(p) = x_1 y_1 + \cdots + x_n y_n \) where \( p = (x_1, \ldots, x_n) \in \mathbb{R}^n \) is a variable point and \( y_1, \ldots, y_n \in \mathbb{R} \) are fixed.

(C) [1, p.88#6], namely, a uniformly continuous function preserves Cauchy sequences.

6.2 Wednesday May 12, 1993

We start studying the theory of (Riemann) integration now.

If \( f : [a, b] \rightarrow \mathbb{R} \) is a bounded function on a closed and bounded interval \( I = [a, b] \subset \mathbb{R} \), its Riemann integral, if it exists, can be denoted in several ways, for example:

\[
\int_a^b f(x) \, dx = \int_a^b f = \int_{[a,b]} f = \int_I f.
\]

Similarly, if \( f : [a, b] \times [c, d] \rightarrow \mathbb{R} \) is a bounded function on a closed and bounded rectangle \( R = I \times J = [a, b] \times [c, d] \subset \mathbb{R}^2 \), its Riemann (double) integral, if it exists, can be denoted in several ways, for example:

\[
\int\int_{I \times J} f(x, y) \, dx \, dy = \int\int_R f = \int_R f.
\]

We can even consider triple integrals: if \( f : I \times J \times K \rightarrow \mathbb{R} \) is a bounded function on a closed and bounded box \( B = I \times J \times K \subset \mathbb{R}^3 \), its Riemann (triple) integral, if it exists, can be denoted in several ways, for example:

\[
\int\int\int_{I \times J \times K} f(x, y, z) \, dx \, dy \, dz = \int\int\int_B f = \int_B f.
\]

Being foolish, we decide to consider the Riemann integral of a bounded function defined on an “\( n \)-box” in \( \mathbb{R}^n \), for any \( n \geq 1 \). By an \( n \)-box we mean a product of \( n \) closed intervals: \( B = I_1 \times I_2 \times \cdots \times I_n \subset \mathbb{R}^n \), where \( I_j \) is a closed and bounded interval in \( \mathbb{R} \). Thus, an interval is a 1-box, a rectangle is a 2-box, and a box in \( \mathbb{R}^3 \) is a 3-box.

Let’s get down to business. For simplicity, we start with (you guessed it) \( n = 1 \).

Let \( I = [a, b] \) be a closed and bounded interval in \( \mathbb{R} \). A partition of \( I \) is any finite subset of \( I \) which includes the endpoints \( a \) and \( b \). We write the elements of \( P \) in increasing order: \( P = \{a = x_0 < x_1 < \cdots < x_m = b\} \). Note that \( m + 1 \) is the number of elements of \( P \) and \( m \) is the number of subintervals, that is

\[
I_j = [x_{j-1}, x_j], \quad 1 \leq j \leq m \quad \text{and} \quad I = \cup_{j=1}^m I_j.
\]

Let \( \mathcal{P}(I) \) denote the set of all partitions of \( I \). This is a very large set consisting of all possible partitions with any number \( m = 1, 2, \ldots \) of subintervals.

The length of \( I_j \) is \( \ell(I_j) = x_j - x_{j-1} \). The mesh of the partition \( P \) is denoted by \( d(P) \) and is defined by \( d(P) = \max\{\ell(I_j) : 1 \leq j \leq m\} \). Next we need a choice of points \( C = \{t_1, \ldots, t_m\} \) such that \( t_j \in I_j \) for \( 1 \leq j \leq m \).
Now let $f$ be a function defined on $I$. A Riemann sum of $f$ with respect to a partition $P$ and a choice $C$ is defined by

$$S(f, P, C) = \sum_{j=1}^{m} f(t_j)\ell(I_j).$$

We can now state a fundamental theorem in the theory of Riemann integration.

**Theorem 6.2 (Theorem 1 on page 169 of [1] (n = 1))** If $f$ is a continuous function on a closed bounded interval $I \subset \mathbb{R}$, then there is a unique real number $v$ (depending on $f$ and $I$) with the following property:

For every $\epsilon > 0$, there exists $\delta > 0$ such that for all partitions $P$ with $d(P) < \delta$ and for every choice $C$, we have $|S(f, P, C) - v| < \epsilon$.

We now consider double integrals. Let $R = I \times J$ be a closed rectangle in $\mathbb{R}^2$. A grid of $R$ is a set of the form $N = P \times Q$, where $P$ is a partition of $I$ and $Q$ is a partition of $J$. We can write $P = \{a = x_0 < x_1 < \cdots < x_m = b\}$ and $Q = \{c = y_0 < y_1 < \cdots < y_r = d\}$. Note that there are $mr$ subrectangles $R_{ij} = I_i \times J_j$ of $R$, where

$$I_i = [x_{i-1}, x_i], \quad 1 \leq i \leq m \quad \text{and} \quad J_j = [y_{j-1}, y_j], \quad 1 \leq j \leq r.$$ 

Moreover $R = \bigcup_{j=1}^{r} \bigcup_{i=1}^{m} R_{ij}$. Let $\mathcal{N}(R)$ denote the set of all grids of $R$.

The area of $R_{ij}$ is $A(R_{ij}) = \ell(I_i)\ell(J_j)$. The mesh of the grid $N$ is denoted by $d(N)$ and is defined by

$$d(N) = \max_{1 \leq i \leq m, 1 \leq j \leq r} |(x_i, y_j) - (x_{i-1}, y_{j-1})| = \max_{1 \leq i \leq m, 1 \leq j \leq r} \{(x_i-x_{i-1})^2 + (y_j-y_{j-1})^2\}^{1/2}.$$ 

Next we need a choice of points $C = \{p_{ij} : 1 \leq i \leq m, 1 \leq j \leq r\}$ such that $p_{ij} \in R_{ij}$ for $1 \leq i \leq m, 1 \leq j \leq r$.

Now let $f$ be a function defined on $R$. A Riemann sum of $f$ with respect to a grid $N$ and a choice $C$ is defined by

$$S(f, N, C) = \sum_{j=1}^{r} \sum_{i=1}^{m} f(p_{ij})A(R_{ij}).$$

We can now restate a fundamental theorem in the theory of Riemann integration, this time for $n = 2$

**Theorem 6.3 (Theorem 1 on page 169 of [1] (n = 2))** If $f$ is a continuous function on a closed bounded rectangle $R \subset \mathbb{R}^2$, then there is a unique real number $v$ (depending on $f$ and $R$) with the following property:

For every $\epsilon > 0$, there exists $\delta > 0$ such that for all grids $N$ with $d(N) < \delta$ and for every choice $C$, we have $|S(f, N, C) - v| < \epsilon$.

**Assignment 12** (Due May 25) Prove the uniqueness of $v$ in Theorem 6.2 or in Theorem 6.3.

**Definition 6.4** The number $v$ whose existence is guaranteed by Theorem 6.2 is denoted by $\int_{I} f$. The number $v$ whose existence is guaranteed by Theorem 6.3 is denoted by $\int_{R} f$. 

23
6.3 Friday May 14, 1993

Today we shall begin the proof of Theorem 6.3. We are given the “data” \( f, R \) and we shall start with the statement of three lemmas. In the first two, it is only required that \( f \) be a bounded function. This will be important for later when you need to study integration of discontinuous functions\(^{14}\). Only in the third lemma will the continuity of \( f \) be needed. Of course, this continuity and the compactness of \( R \) implies that \( f \) is bounded, and moreover, perhaps more importantly, that \( f \) is uniformly continuous\(^{15}\).

Let’s get down to business. Suppose \( f \) is a bounded function on the closed and bounded rectangle \( R = I \times J \), and let \( N \) be a grid of \( R \). Thus \( N = P \times Q \), where \( P \) is a partition of \( I \) and \( Q \) is a partition of \( J \), say \( P = \{a = x_0 < x_1 < \cdots < x_m = b\} \) and \( Q = \{c = y_0 < y_1 < \cdots < y_r = d\} \). Recall that there are \( mr \) subrectangles \( R_{ij} = I_i \times J_j \) of \( R \), where

\[
I_i = [x_{i-1}, x_i] \quad 1 \leq i \leq m \quad \text{and} \quad J_j = [y_{j-1}, y_j] \quad 1 \leq j \leq r.
\]

Moreover \( R = \bigcup_{j=1}^r \bigcup_{i=1}^m R_{ij} \).

Since \( f \) is bounded on \( R \), it is also bounded on each subrectangle \( R_{ij} \) and we can define

\[
M_{ij} = \sup_{p \in R_{ij}} f(p) \quad \text{and} \quad m_{ij} = \inf_{p \in R_{ij}} f(p).
\]

Notice that for continuous \( f \), by the extreme values theorem, there will exist points \( x_{ij}, y_{ij} \in R_{ij} \) such that \( f(x_{ij}) = m_{ij} \) and \( f(y_{ij}) = M_{ij} \). We shall use this fact in the third lemma below but for the first two lemmas, only the numbers \( m_{ij}, M_{ij} \) are needed.

We now define the upper and lower Riemann sums corresponding to a grid, namely,

\[
\mathcal{S}(N) := \sum_{j=1}^r \sum_{i=1}^m M_{ij} A(R_{ij}) \quad \text{(upper Riemann sum)}
\]

and

\[
\mathcal{L}(N) := \sum_{j=1}^r \sum_{i=1}^m m_{ij} A(R_{ij}) \quad \text{(lower Riemann sum)}
\]

Since \( m_{ij} \leq f(p) \leq M_{ij} \) for every \( p \in R_{ij} \), and \( A(R_{ij}) > 0 \), for every grid \( N \) and every choice \( C \), we have

\[
\mathcal{L}(N) \leq S(f, N, C) \leq \mathcal{S}(N). \tag{11}
\]

We are now ready to state the three lemmas.

**Lemma 6.5 (Lemma 1 on page 170 of [1])** Let \( f \) be a bounded function on a closed and bounded rectangle \( R \subset \mathbb{R}^2 \). Let \( N \) and \( \tilde{N} \) be grids of \( R \) and suppose \( N \subset \tilde{N} \). Then

\(^{14}\)this will be 99% of the time!, real life is not continuous

\(^{15}\)remember, this is supposed to be an application of uniform continuity
(a) $\underline{S}(N) \leq \underline{S}(\tilde{N})$

(b) $\overline{S}(N) \geq \overline{S}(\tilde{N})$

**Lemma 6.6 (Lemma 2 on page 170 of [1])** Let $f$ be a bounded function on a closed and bounded rectangle $R \subset \mathbb{R}^2$.

(a) The following two subsets\(^{16}\) of $\mathbb{R}$ are bounded sets:

$$\{\underline{S}(N) : N \in \mathcal{N}(R)\} \text{ and } \{\overline{S}(N) : N \in \mathcal{N}(R)\}.$$

(b) Let

$$s := \sup \{\underline{S}(N) : N \in \mathcal{N}(R)\} \text{ and } S = \inf \{\overline{S}(N) : N \in \mathcal{N}(R)\}.$$

Then

- $s \leq S$
- for every grid $N$, $S - s \leq \overline{S}(N) - \underline{S}(N)$

**Lemma 6.7 (Lemma 3 on page 171 of [1])** Let $f$ be a continuous function on a closed and bounded rectangle $R \subset \mathbb{R}^2$. For every $\epsilon > 0$, there exists $\delta > 0$ such that

$$\overline{S}(N) - \underline{S}(N) < \epsilon \text{ for every grid } N \text{ with mesh } d(N) < \delta.$$

We shall prove these three lemmas, one after the other. But first, we use them to give a proof of Theorem 6.3.

**Proof of Theorem 6.3:** From Lemmas 6.6 and 6.7, $0 < S - s < \epsilon$ for every $\epsilon > 0$, that is, $S - s = 0$. Let $v$ denote the common value of $s$ and $S$.

The following sequence of statements will complete the proof. Be sure you can supply the justification for each statement.

- $S - \underline{S}(N) \leq \overline{S}(N) - \underline{S}(N)$ for every grid $N$
- $\overline{S}(N) - s \leq \overline{S}(N) - \underline{S}(N)$ for every grid $N$
- $0 < v - \underline{S}(N) < \epsilon$ if $d(N) < \delta$
- $0 < \overline{S}(N) - v < \epsilon$ if $d(N) < \delta$
- $v - S(f, N, C) \leq v - \underline{S}(N) < \epsilon$ if $d(N) < \delta$ and $C$ is any choice of points
- $S(f, N, C) - v \leq \overline{S}(N) - v < \epsilon$ if $d(N) < \delta$ and $C$ is any choice of points
- $|S(f, N, C) - v| < \epsilon$ if $d(N) < \delta$ and $C$ is any choice of points.

\(^{16}\)recall that $\mathcal{N}(R)$ denotes the set of all grids of $R$
This completes the proof of Theorem 6.3.

We now turn to the proofs of the three lemmas.

**Proof of Lemma 6.5:** There are two parts to this proof. First we prove the lemma under the assumption that it holds in the special case where $\tilde{N}$ is obtained from $N = P \times Q$ by adding a single point to either $P$ or $Q$. Then we prove the special case.

**Step 1:** Assume that the lemma is true in the special case. Write

$$N_0 = \tilde{N} \supset N_1 \supset \cdots \supset N_s \supset N_{s+1} := N,$$

where $N_k$ is obtained from $N_{k+1}$ by adding a single point ($0 \leq k \leq s$).

By assumption, for $0 \leq k \leq s$,

$$S(N_{k+1}) \leq S(N_k) \leq S(N_{k-1}) \leq \cdots \leq S(N_1) \leq S(N_0) = S(\tilde{N}).$$

Therefore,

$$S(N) = S(N_{s+1}) \leq S(N_s) \leq \cdots \leq S(N_1) \leq S(N_0) = S(\tilde{N}) \leq S(N_{s-1}) \leq \cdots \leq S(N_2) \leq S(N_1) \leq S(N_0) = S(N).$$

This completes the proof of step 1.

**7 Week 7**

**7.1 Monday May 17, 1993**

We continue with the proof of Lemma 6.5.

**Step 2:** We start with the following simple observation:

Let $\phi : D \to \mathbb{R}$ be a bounded function on a set $D \subset \mathbb{R}^n$ and suppose that $A \subset D$. Then

$$\sup_{p \in A} \phi(p) \leq \sup_{p \in D} \phi(p) \quad \text{and} \quad \inf_{p \in A} \phi(p) \geq \inf_{p \in D} \phi(p).$$

We now assume that $\tilde{N} = (P \cup \{u\}) \times Q$ and define $i_0$ by $x_{i_0-1} < u < x_{i_0}$. We have

$$R_{i_0j} = R'_{i_0j} \cup R''_{i_0j}, \quad \text{where} \quad R'_{i_0j} = [x_{i_0-1}, u] \times [y_{j-1}, y_j] \quad \text{and} \quad R''_{i_0j} = [u, x_{i_0}] \times [y_{j-1}, y_j].$$

Then

$$S(N) = \sum_{j=1}^{r} \sum_{i=1}^{m} m_{ij} A(R_{ij}) = \sum_{j=1}^{r} m_{i_0j} A(R'_{i_0j}) + \sum_{j=1}^{r} \sum_{i=1, i \neq i_0}^{m} m_{ij} A(R_{ij}),$$

and

$$S(\tilde{N}) = \sum_{j=1}^{r} [m'_{i_0j} A(R'_{i_0j}) + m''_{i_0j} A(R''_{i_0j})] + \sum_{j=1}^{r} \sum_{i=1, i \neq i_0}^{m} m_{ij} A(R_{ij}).$$
Thus, \( S(N) \leq S(\bar{N}) \) if for each \( j \), \( m_{i_0j}A(R_{i_0j}) \leq m'_{i_0j}A(R'_{i_0j}) + m''_{i_0j}A(R''_{i_0j}) \). This last statement is true since \( A(R_{i_0j}) = A(R'_{i_0j}) + A(R''_{i_0j}) \), and by virtue of the observation above, \( m_{i_0j} \leq m'_{i_0j} \), \( m_{i_0j} \leq m''_{i_0j} \).

This completes the proof of (a) in case the new point \( u \) occurs on the “\( x \)-axis”. You need a similar proof in case the new point occurs on the “\( y \)-axis”. Then you need to prove (b) in each of these two cases. These proofs can be omitted since no new ideas are needed for them.

**Definition 7.1** A bounded function on a closed and bounded rectangle \( R \subset \mathbb{R}^2 \) is integrable on \( R \) if it satisfies the condition of Theorem 6.3, that is, there is a unique real number \( v \) (depending on \( f \) and \( R \)) with the following property:

For every \( \epsilon > 0 \), there exists \( \delta > 0 \) such that for all grids \( N \) with \( d(N) < \delta \) and for every choice \( C \), we have \( |S(f, N, C) - v| < \epsilon \).

We can restate Theorem 6.3 as: every continuous function on a compact rectangle is integrable on that rectangle. The question arises: does the converse hold? The answer is no. A discontinuous function can be integrable. There are non-integrable functions, necessarily discontinuous. The next two examples illustrate these two facts.

**Example 1:** Let \( f : [0, 2] \to \mathbb{R} \) be defined by

\[
    f(x) = \begin{cases}
      5 & 0 \leq x < 1 \\
      \alpha & x = 1 \\
      0 & 1 < x \leq 2
    \end{cases}
\]

where \( \alpha \in \mathbb{R} \) is arbitrary. Then \( f \) is integrable on \([0, 2]\) and \( \int_{[0,2]} f = 5 \).

**Proof:** Let \( P = \{0 = x_0 < x_1 < \cdots < x_m = 2\} \) be any partition of \([0, 2]\). The number 1 falls in a unique subinterval \((x_{j-1}, x_j)\), so \( x_{j-1} < 1 \leq x_j \). Let \( C = \{t_1, \ldots, t_m\} \) be any choice of points with \( t_k \in I_k = [x_{k-1}, x_k] \) for \( 1 \leq k \leq m \). Then \( S(f, P, C) = 5\ell(I_1) + 5\ell(I_2) + \cdots + \ell(I_{j-1}) + f(t_j)\ell(I_j) + 0\cdot \ell(I_{j+1}) + \cdots + 0\cdot \ell(I_m) \), and therefore

\[
S(f, P, C) - 5 = 5(x_{j-1} - x_0) + f(t_j)(x_j - x_{j-1}).
\]  

(12)

Now let \( \epsilon > 0 \). Let \( M := \max\{5, |\alpha|\} \) and choose \( \delta = \epsilon / (5 + M) \). By (12), if \( d(P) < \delta \) (and \( C \) is arbitrary),

\[
|S(f, P, C) - 5| \leq 5|x_{j-1} - x_0| + |f(t_j)|(x_j - x_{j-1}) < 5\delta + M\delta < \epsilon.
\]

**Example 2:** Let \( f : [a, b] \to \mathbb{R} \) be defined by

\[
    f(x) = \begin{cases}
      5 & x \text{ rational} \\
      4 & x \text{ irrational}
    \end{cases}
\]

Then \( f \) is not integrable on \([a, b]\).

**Proof:** Given in discussion, May 18.

**Assignment 13** (Due May 25, 1993) [1, §4.2 page 178 #3,6,8]. In problem 6, take \( D = R \), a compact rectangle.
7.2 Wednesday May 19, 1993

A lemma a day keeps the doctor away. Today we prove Lemma 6.6.

**Proof of Lemma 6.6**: Every grid

\[ N = \{ a = x_0 < \cdots < x_m = b \} \times \{ c = y_0 < \cdots < y_r = d \} \]

contains the trivial grid \( N_0 = \{ a, b \} \times \{ c, d \} \). Therefore, with \( m := \inf \{ f(p) : p \in R \} \) and \( M := \sup \{ f(p) : p \in R \} \), by Lemma 6.5,

\[ mA(R) = \underline{S}(N_0) \leq \underline{S}(N) \leq \overline{S}(N) \leq \overline{S}(N_0) = MA(R). \]

Thus the two sets \( \{ \underline{S}(N) : N \in \mathcal{N}(R) \} \) and \( \{ \overline{S}(N) : N \in \mathcal{N}(R) \} \) are bounded, and the numbers \( s \) and \( \overline{S} \) exist.

For every \( N \), since \( \underline{S}(N) \leq s \) we have \( -\overline{S}(N) \geq -s \). Add this inequality to the inequality \( \overline{S}(N) \geq \overline{S} \) and you get \( \overline{S}(N) - \underline{S}(N) \geq \overline{S} - s \), which proves the second statement of the lemma.

To prove the first statement of the lemma, we shall make use of the following:

**Claim**: for any two grids \( N_1, N_2 \), \( \underline{S}(N_1) \leq \overline{S}(N_2) \).

Let’s assume this claim for the moment. Thinking of \( N_2 \) as fixed and \( N_1 \) as varying, and taking the supremum over \( N_1 \), you get, for every \( N_2 \),

\[ \sup_{N_1 \in \mathcal{N}(R)} \underline{S}(N_1) \leq \overline{S}(N_2). \]

Thus \( s \leq \overline{S}(N) \) for every grid \( N \) so taking the infimum over all grids \( N \), you get

\[ s \leq \inf_{N \in \mathcal{N}(R)} \overline{S}(N), \]

which proves the first statement.

To prove the claim, we use Lemma 6.5 again. Given any two grids \( N_1, N_2 \), let \( N = N_1 \cup N_2 \). Then \( N_1 \subset N \) and \( N_2 \subset N \), so that by Lemma 6.5,

\[ \underline{S}(N_1) \leq \underline{S}(N) \leq \overline{S}(N) \leq \overline{S}(N_2). \]

This completes the proof of Lemma 6.6.

We still need to prove Lemma 6.7. We shall do that in the next lecture.

The following theorem differs from Theorem 4 on page 176 of [1] in the following respects. In [1, Theorem 4, page 176], \( f \) and \( g \) are assumed continuous, and the integration is over arbitrary sets, not necessarily compact rectangles. Moreover, there is a fifth statement, which I shall state and prove separately, but only (for convenience) for \( n = 1 \) (see Theorem 7.3 below).

**Theorem 7.2 (Theorem 4 on page 176 of [1])** Let \( f \) and \( g \) be integrable functions on the compact rectangle \( R \subset \mathbb{R}^2 \). Let \( c \) be any real number. Then:
1. $f + g$ is integrable on $\mathbb{R}$ and $\int_{\mathbb{R}} (f + g) = \int_{\mathbb{R}} f + \int_{\mathbb{R}} g$

2. $cf$ is integrable on $\mathbb{R}$ and $\int_{\mathbb{R}} cf = c \int_{\mathbb{R}} f$

3. If $f(x) \geq 0$ for all $x \in \mathbb{R}$, then $\int_{\mathbb{R}} f \geq 0$

4. $|f|$ is integrable on $\mathbb{R}$ and $|\int_{\mathbb{R}} f| \leq \int_{\mathbb{R}} |f|$.

**Proof:** It is trivial to verify that for any grid $N$ and choice $C$,

$$S(f + g, N, C) = S(f, N, C) + S(g, N, C).$$

For $\epsilon > 0$, choose $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$|S(f, N, C) - \int_{\mathbb{R}} f| < \frac{\epsilon}{2}$$

for all $C$ and all $N$ with $d(N) < \delta_1$,

and

$$|S(g, N, C) - \int_{\mathbb{R}} g| < \frac{\epsilon}{2}$$

for all $C$ and all $N$ with $d(N) < \delta_2$.

Then, with $\delta = \min\{\delta_1, \delta_2\}$, we have, for all choices $C$ and all grids $N$ with $d(N) < \delta$,

$$|S(f + g, N, C) - \int_{\mathbb{R}} f - \int_{\mathbb{R}} g| = |S(f, N, C) + S(g, N, C) - \int_{\mathbb{R}} f - \int_{\mathbb{R}} g|$$

$$\leq |S(f, N, C) - \int_{\mathbb{R}} f| + |S(g, N, C) - \int_{\mathbb{R}} g|$$

$$< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This proves the first statement and the proof of the second statement is similar. We refer to [1, page 177] for the proofs of the third and fourth statements.

The following theorem contains two important properties of the integral. The first one will be part of a homework assignment (see Assignment 14 below). The second one will be proved in the next lecture.

**Theorem 7.3** Let $[a, b]$ be a compact interval in $\mathbb{R}$ and let $f : [a, b] \rightarrow \mathbb{R}$ be a bounded function.

(a) If $f$ is integrable on $[a, b]$ then $f$ is integrable on any compact subinterval $[c, d] \subset [a, b]$.

(b) If $a < c < b$ and if $f$ is integrable on $[a, c]$ and on $[c, b]$, then $f$ is integrable on $[a, b]$ and

$$\int_{a}^{b} f = \int_{a}^{c} f + \int_{c}^{b} f.$$

**Assignment 14** (Due May 25, 1993) Prove the following statements, which will result in a proof of (a) of Theorem 7.3. (These will be stated for $n = 1$ but both the statements and proofs are valid for $n = 2$ and in fact for any $n$)
(A) Prove that \( f \) is integrable on \([a, b]\) if for every \( \epsilon > 0 \) there is a \( \delta > 0 \) such that \( S(P) - S(P) < \epsilon \) for all partitions \( P \) with \( d(P) < \delta \). (Hint: The proof is contained in the proof of Theorem 6.3 given above.)

(B) (converse of (A)) Suppose that \( f \) is integrable on \([a, b]\). Prove that for every \( \epsilon > 0 \) there is a \( \delta > 0 \) such that \( S(P) - S(P) < \epsilon \) for all partitions \( P \) with \( d(P) < \delta \).

Hint: I shall give an incorrect proof of this statement, which however, will give the correct idea. For any partition \( P \), with \([a, b] = \bigcup_{i=1}^{m} I_i \), pick \( t_i' \) and \( t_i'' \) in \( I_i \) such that \( m_i := \inf\{f(x) : x \in I_i\} = f(t_i') \) and \( M_i := \sup\{f(x) : x \in I_i\} = f(t_i'') \). Then \( S(P) = S(f, P, C) \) where \( C = \{t_1', \ldots, t_m'\} \) and \( S(P) = S(f, P, C') \) where \( C' = \{t_1'', \ldots, t_m''\} \).

We now have, for \( d(P) < \delta \),

\[
|S(P) - S(P)| = |S(f, P, C') - S(f, P, C'')| \\
= |S(f, P, C') - \int_I f| + |\int_I f - S(f, P, C'')| \\
\leq |S(f, P, C') - \int_I f + \int_I f - S(f, P, C'')| \\
< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\]

The problem with this proof is that \( f \) is not necessarily continuous, so there is no guarantee that the points \( t_i', t_i'' \) exist. This can be corrected by just using the definition of the numbers \( m_i, M_i \)—that is what you have to do.

(C) Let \([c, d] \subset [a, b]\) and let \( P \) be a partition of \([a, b]\) which includes the two points \( c, d \). Let \( P_2 := P \cap [c, d] \) (which is a partition of \([c, d]\)). Prove that

\[
S(P_2) - S(P_2) \leq S(P) - S(P).
\]

Note that, for example, \( S(P_2) \) is an upper Riemann sum for \( f \) on the interval \([c, d]\), and \( S(P) \) is an upper Riemann sum for \( f \) on the interval \([a, b]\). You can use the notation \( S(P_2) = S(P_2, [c, d]) \) and \( S(P) = S(P, [a, b]) \) to remind yourself of these facts. I will use a similar notation in the proof of Theorem 7.3(b) below.

(D) Use (A) and (C) to prove (a) of Theorem 7.3.

Assignment 15 (Due May 25, 1993) Let \( f \) be an integrable function on \([a, b]\).

(A) Prove that \( f \) has at least one point where it is continuous.

(Do not worry if you have difficulty with this problem. Trying it will give you a great appreciation for and understanding of Riemann’s definition! Don’t waste too much time on it; just use it to prove part (B))
(B) Prove that if \( f(x) > 0 \) for every \( x \in [a, b] \), then \( \int_a^b f > 0 \). (Note that by Theorem 7.2, \( \int_a^b f \geq 0 \); the point is to prove that \( \int_a^b f \neq 0 \).)

### 7.3 Friday May 21, 1993

A lemma a day keeps the doctor away. Today we prove Lemma 6.7.

**Proof of Lemma 6.7:** Since \( f \) is continuous on the compact rectangle \( R \), it is uniformly continuous on \( R \). For any \( \epsilon > 0 \), let \( \delta = \delta(\epsilon/A(R), f, R) \), that is,

\[
|f(p) - f(q)| < \epsilon/A(R) \quad \text{for all} \quad p, q \in R \quad \text{with} \quad |p - q| < \delta.
\]

Let \( N \) be any grid with \( d(N) < \delta \). Since \( f \) is continuous on the each compact subrectangle \( R_{ij} \) of \( R \), by the extreme values theorem, there exist points \( p_{ij}, q_{ij} \in R_{ij} \) such that \( M_{ij} = f(p_{ij}) \) and \( m_{ij} = f(q_{ij}) \). Since \( p_{ij}, q_{ij} \in R_{ij} \), \( |p_{ij} - q_{ij}| < \delta \), and so \( M_{ij} - m_{ij} = f(p_{ij}) - f(q_{ij}) < \epsilon/A(R) \). We now have

\[
0 \leq S(N) - \underline{S}(N) = \sum_{i,j} (M_{ij} - m_{ij})A(R_{ij}) \leq \epsilon/A(R) \sum_{i,j} A(R_{ij}) = [\epsilon/A(R)] \cdot A(R) = \epsilon.
\]

This proves Lemma 6.7.

Having proved Lemmas 6.5, 6.6, 6.7, the proof of Theorem 6.3 is now complete.\(^{17}\)

We now turn to the proof of (b) of Theorem 7.3.

**Proof of (b) of Theorem 7.3:** Let \( v_1 := \int_a^c f \) and \( v_2 := \int_c^b f \), which are assumed to exist. This means that for every \( \epsilon > 0 \) there exist \( \delta_1 > 0, \delta_2 > 0 \) such that

\[
|S(f, P_1, C_1, [a, c]) - v_1| < \epsilon \quad \text{if} \quad d(P_1) < \delta_1, \forall C_1.
\]

and

\[
|S(f, P_2, C_2, [c, b]) - v_2| < \epsilon \quad \text{if} \quad d(P_2) < \delta_2, \forall C_2.
\]

Here, \( P_1 \) is a partition of \([a, c]\) and \( C_1 \) is a choice of points corresponding to the subintervals of \( P_1 \). Similarly for \( P_2, C_2 \) on \([c, b]\).

We have to prove that for every \( \epsilon > 0 \), there exists \( \delta > 0 \) such that

\[
|S(f, P, C, [a, b]) - v_1 - v_2| < \epsilon \quad \text{if} \quad d(P) < \delta, \forall C.
\]

Here, \( P \) is a partition of \([a, b]\) and \( C \) is a choice of points corresponding to the subintervals of \( P \).

Let’s get down to business. Let \( P \) be any partition of \([a, b]\). Write \( P = \{a = x_0 < x_1 < \cdots < x_m = b\} \) and let \( C = \{t_1, \ldots, t_m\} \) be any choice of points corresponding to \( P \). There is a unique \( k \) such that \( x_{k-1} \leq c < x_k \). Define partitions \( P_1 \) of \([a, c]\) and \( P_2 \) of \([c, b]\) by

\[P_1 = \{a = x_0 < x_1 < \cdots < x_{k-1}\} \cup \{c\} \quad \text{and} \quad P_2 = \{c = x_k < x_{k+1} < \cdots < x_m = b\}.
\]

\(^{17}\)Hallelujah!
Define choices $C_1$ and $C_2$ corresponding to $P_1$ and $P_2$ respectively as follows:

$$C_1 = \{t_1, \ldots, t_{k-1}, t'_k\} \quad \text{and} \quad C_2 = \{t''_k, t_{k+1}, \ldots, t_m\},$$

where $t'_k$ and $t''_k$ are chosen so that $\{t'_k, t''_k\} = \{c, t_k\}$, that is, if $t_k \leq c$ define $t'_k = t_k$ and $t''_k = c$, whereas, if $c < t_k$, define $t'_k = c$ and $t''_k = t_k$.

Now let’s calculate:

- $S(f, P, C, [a, b]) = \sum_{j=1}^{k-1} f(t_j)\ell(I_j) + f(t_k)\ell(I_k) + \sum_{j=k+1}^{m} f(t_j)\ell(I_j)$
- $S(f, P_1, C_1, [a, c]) = \sum_{j=1}^{k-1} f(t_j)\ell(I_j) + f(t'_k)\ell([x_{k-1}, c])$
- $S(f, P_2, C_2, [c, b]) = f(t''_k)\ell([c, x_k]) + \sum_{j=k+1}^{m} f(t_j)\ell(I_j)$

We have some cancellation here:

$$S(f, P, C, [a, b]) - S(f, P_1, C_1, [a, c]) - S(f, P_2, C_2, [c, b]) = f(t_k)\ell(I_k) - f(t'_k)(c - x_{k-1}) - f(t''_k)(x_k - c). \quad (16)$$

We are now almost done: let $M := \sup\{|f(x)| : x \in [a, b]\}$ and set $\delta = \min\{\delta_1, \delta_2, \epsilon\}$. Then for any $P$ with $d(P) < \delta$ and for any choice $C$, we have (by (13), (14), and (16)),

$$|S(f, P, C, [a, b]) - v_1 - v_2| \leq |S(f, P, C, [a, b]) - S(f, P_1, C_1, [a, c]) - S(f, P_2, C_2, [c, b])| + |S(f, P_1, C_1, [a, c]) - v_1| + |S(f, P_2, C_2, [c, b]) - v_2| < |f(t_k)\ell(I_k) - f(t'_k)(c - x_{k-1}) - f(t''_k)(x_k - c)| + \epsilon + \epsilon < 2M\ell(I_k) + 2\epsilon < 2(M + 1)\epsilon.$$

It is now clear that by a better choice of “$\epsilon$”, we will have (15), and the proof of (b) of Theorem 7.3 is complete.

**Assignment 16** (Due May 25) Let $f(x) = x$ for rational $x$ and $f(x) = 0$ for irrational $x$. If $f$ integrable on $[0, 1]$?

**Assignment 17** (Due May 25) Let $f$ be integrable on $[a, b]$ and suppose that $g$ is a function on $[a, b]$ such that $g(x) = f(x)$ except for finitely many $x$ in $[a, b]$. Show that $g$ is integrable on $[a, b]$ and that $\int_a^b f = \int_a^b g$.

**Assignment 18** (Due May 25) Prove that a bounded function $f$ on $R \subset \mathbb{R}^2$ is integrable on $R$ if and only if for every $\epsilon > 0$, there exists a grid $N$ such that $\mathcal{S}(N) - \mathcal{S}(N) < \epsilon$.

**Assignment 19** (Due June 8)

(a) Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of continuous functions on $[a, b]$, and suppose that $f_n \to f$ uniformly on $[a, b]$. Why is $f$ integrable on $[a, b]$? Prove that

$$\lim_{n \to \infty} \int_a^b f_n = \int_a^b f.$$
(b) Let \( \{f_n\}_{n=1}^\infty \) be a sequence of integrable functions on \([a, b]\), and suppose that \( f_n \to f \) uniformly on \([a, b]\). Prove that \( f \) is integrable on \([a, b]\) and that

\[
\lim_{n \to \infty} \int_a^b f_n = \int_a^b f.
\]

Assignment 20 (Due June 8) Show that \( f \) is integrable on \([-1, 1]\) if

(a) \( f(x) = \sin(1/x) \) for \( x \neq 0 \) and \( f(0) = 0 \)

(b) \( f(x) = x \text{sgn}(\sin(1/x)) \) for \( x \neq 0 \) and \( f(0) = 0 \).

Assignment 21 (Due June 8) For each rational number \( x \), write \( x = p/q \) where \( p, q \) are integers with no common factors and \( q > 0 \). Define \( f(x) = 1/q \) for \( x \) rational and \( f(x) = 0 \) if \( x \) is irrational. Show that \( f \) is integrable on every compact interval \([a, b]\) and that \( \int_a^b f = 0 \).

8 Week 8

8.1 Monday May 24, 1993

8.1.1 Differentiability implies continuity I

There will be another version of this later—see the coordinate-free definition of derivative below.

Let’s begin by recalling the mean value theorem in one variable. We shall use Theorem 8.1 (a theorem in one dimension) in the proof of Theorem 8.4 below (a theorem in \( n \geq 1 \) dimensions).

Theorem 8.1 (Mean Value Theorem in one variable) If \( f : (a, b) \to \mathbb{R} \) is differentiable on \((a, b)\), then for every \( x_1, x_2 \in (a, b) \) with \( x_1 < x_2 \), there exists \( c \in (x_1, x_2) \) such that

\[
\frac{f(x_1) - f(x_2)}{x_1 - x_2} = f'(c).
\]

Rhetorical question: is \( f' \) a continuous function? NO!, in general. However, only the existence of a derivative, not the continuity of the derivative, is required in Theorems 8.1 and 8.2. This is one difference between these two one-dimensional theorems, and the \( n \)-dimensional theorem Theorem 8.4.

Now let’s recall the proof in one variable that differentiability implies continuity.

Theorem 8.2 (Differentiability implies continuity—one variable) If \( f : (a, b) \to \mathbb{R} \) is differentiable on \((a, b)\), then \( f \) is continuous on \((a, b)\).

Proof: If \( f : (a, b) \to \mathbb{R} \) is differentiable on \((a, b)\), then for any fixed \( c \in (a, b) \), and any \( x \neq c \),

\[
f(x) - f(c) = \frac{f(x) - f(c)}{x - c} \cdot (x - c).
\]
Thus, \( f(x) = f(c) + \frac{f(x) - f(c)}{x-c} \cdot (x - c) \) so that
\[
\lim_{x \to c} f(x) = f(c) + f'(c) \cdot 0 = f(c).
\]

We now consider a notion of differentiability for functions \( f : D \to \mathbb{R} \) defined on open subsets \( D \) of \( \mathbb{R}^n \). For such a function and a point \( p_0 = (x_1^0, \ldots, x_n^0) \in D \), the partial derivatives at \( p_0 \) are defined by

\[
D_1 f(p_0) = \lim_{x_1 \to x_1^0} \frac{f(x_1, x_2^0, \ldots, x_n^0) - f(x_1^0, x_2^0, \ldots, x_n^0)}{x_1 - x_1^0} = \frac{d}{dx_1} \bigg|_{x_1=x_1^0} f(x_1, x_2, x_3^0, \ldots, x_n^0),
\]

\[
D_2 f(p_0) = \lim_{x_2 \to x_2^0} \frac{f(x_1^0, x_2, x_3^0, \ldots, x_n^0) - f(x_1^0, x_2^0, \ldots, x_n^0)}{x_2 - x_2^0} = \frac{d}{dx_2} \bigg|_{x_2=x_2^0} f(x_1, x_2, x_3, \ldots, x_n^0),
\]

and so forth, until

\[
D_n f(p_0) = \lim_{x_n \to x_n^0} \frac{f(x_1^0, \ldots, x_{n-1}^0, x_n) - f(x_1^0, x_2^0, \ldots, x_n^0)}{x_n - x_n^0} = \frac{d}{dx_n} \bigg|_{x_n=x_n^0} f(x_1^0, \ldots, x_{n-1}^0, x_n).
\]

Some common notations for this are
\[
D_j f(p_0) = f_j(p_0) = \frac{\partial f}{\partial x_j}(p_0).
\]

You can also write (if you prefer)
\[
\frac{\partial f}{\partial x_j}(p_0) = \lim_{t \to 0} \frac{f(x_1^0, \ldots, x_{j-1}^0, x_j + t, x_{j+1}^0, \ldots, x_n^0) - f(x_1^0, x_2^0, \ldots, x_n^0)}{t}.
\]

Other common notations can be found in [1, page 127].

When you differentiate a function the result is another function, which you can then proceed to (try to) differentiate again. This gives rise to higher derivatives in one variable, \( f, f', f'', f''', \ldots \). We can do the same thing in several variables, where we have a lot more variety. That is, given a function \( f \) on an open set \( D \) in \( \mathbb{R}^n \), its “first” derivatives (when they exist!) are the functions \( D_1 f, D_2 f, \ldots, D_n f \), which are themselves functions on \( D \). Each one of these new functions has \( n \) partial derivatives, so the list of “second” derivatives of \( f \) is very large, and the number of “third” or even higher order derivatives grows very quickly (Question: what is that number?)

Higher order partial derivatives are denoted as follows: for example, for order 2,

\[
D_1(D_j f) = (f_j)_i = f_{ji} = \frac{\partial}{\partial x_i} \left( \frac{\partial f}{\partial x_j} \right) = \frac{\partial^2 f}{\partial x_i \partial x_j},
\]

and if \( i = j \),

\[
D^2 f = D_j(D_j f) = (f_j)_j = f_{jj} = \frac{\partial}{\partial x_j} \left( \frac{\partial f}{\partial x_j} \right) = \frac{\partial^2 f}{\partial x_j^2}.
\]

We want to prove an analog of Theorem 8.2 for functions of \( n \) variables. We will see that it differs both in statement and difficulty of proof from the case \( n = 1 \).
Definition 8.3 Let \( k \) be any positive integer, \( k = 1, 2, \ldots \). A function \( f \) defined on an open set \( D \) in \( \mathbb{R}^n \) is said to be of class \( C^k \) on \( D \), notation \( f \in C^k(D) \), if all of its partial derivatives up to and including order \( k \) exist and are continuous functions on \( D \). A continuous function on \( D \) is said to be of class \( C^0 \).

To be explicit, a function \( f \) is of class \( C^1 \) on \( D \) if the following \( n \) functions are all continuous on \( D \):

\[
D_1 f, \ldots, D_n f.
\]

The function \( f \) is of class \( C^2 \) if the following \( n^2 + n \) functions are all continuous on \( D \):

\[
D_j f(1 \leq j \leq n), \quad D_m(D_i f) (1 \leq i, n, 1 \leq m \leq n).
\]

We have

\[
C^1(D) \supset C^2(D) \supset \cdots \supset C^k(D) \supset C^{k+1}(D) \supset \cdots \tag{17}
\]

In particular, if \( n = 1 \), and \( D \) is an open interval \( I \) in \( \mathbb{R} \), then

\[
C^0(I) \supset C^1(I) \supset C^2(I) \supset \cdots \supset C^k(I) \supset C^{k+1}(I) \supset \cdots \tag{18}
\]

Notice that (18) has an extra inclusion at the beginning, namely \( C^0(I) \supset C^1(I) \), due to Theorem 8.2. We now show that (17) has an extra inclusion too, namely \( C^0(D) \supset C^1(D) \). (Question: how do these two extra inclusion relations differ from each other?)

Theorem 8.4 (Corollary on page 129 of [1]) Let \( f : D \to \mathbb{R} \) be defined on an open subset \( D \) of \( \mathbb{R}^n \), and suppose that \( f \in C^1(D) \). Then \( f \) is continuous on \( D \).

Restated, if \( D_1 f, \ldots, D_n f \) exist and are continuous at all points of \( D \), then \( f \) is continuous on \( D \).

Proof: Fix \( p_0 \in D \) and let \( p \in B(p_0, r) \subset D \) for some \( r > 0 \).

Sidebar: We shall travel from \( p_0 = (x_0^0, \ldots, x_n^0) \) to \( p = (x_1, \ldots, x_n) \) by going parallel to the coordinate axes, one axis at a time, using only the existence of each partial derivative \( f_j \) and the mean value theorem in one variable to obtain an expression of the form

\[
f(p) - f(p_0) = f_1(q_1)(x_1 - x_1^0) + f_2(q_2)(x_2 - x_2^0) + \cdots + f_n(q_n)(x_n - x_n^0) \tag{19}
\]

for certain vectors \( q_1, \ldots, q_n \in B(p_0, r) \).

Next we shall use the continuity of the partial derivatives to get \( |f(p) - f(p_0)| < \epsilon \) for \( |p - p_0| < \delta \).

Let’s get down to business. For simplicity, we do the proof in the case \( n = 3 \) (otherwise we will get lost in the notation, but the proof we shall give works in any dimension). Accordingly, we shall use the notation \( p_0 = (x_0, y_0, z_0) \) and \( p = (x, y, z) \).

\(^{18}\)In [1, Definition 1, page 128], the definition of \( C^k \) requires that \( f \) be continuous. By Theorem 8.4, Buck’s definition of \( C^k \) and our Definition 8.3 are equivalent.
Step 1 Let \( p_1 = (x, y_0, z_0) \). Then by the mean value theorem in one variable
\[
f(p_1) - f(p_0) = \frac{\partial f}{\partial x}(c, y_0, z_0)(x - x_0)
\]
for some \( c \) between \( x \) and \( x_0 \).
(Question: what does \( c \) depend on?)

Step 2 Let \( p_2 = (x, y, z_0) \). Then by the mean value theorem in one variable
\[
f(p_2) - f(p_1) = \frac{\partial f}{\partial y}(x, d, z_0)(y - y_0)
\]
for some \( d \) between \( y \) and \( y_0 \).
(Question: what does \( d \) depend on?)

Step 3 Let \( p_3 = (x, y, z) (= p) \). Then by the mean value theorem in one variable
\[
f(p) - f(p_2) = \frac{\partial f}{\partial z}(x, y, e)(z - z_0)
\]
for some \( e \) between \( z \) and \( z_0 \).
(Question: what does \( e \) depend on?)

Step 4 Letting \( q_1 = (c, y_0, z_0), q_2 = (x, d, z_0), q_3 = (x, y, e) \), we have
\[
f(p) - f(p_0) = \left[ f(p_1) - f(p_0) \right] + \left[ f(p_2) - f(p_1) \right] + \left[ f(p) - f(p_2) \right] \\
= f_1(q_1)(x - x_0) + f_2(q_2)(y - y_0) + f_3(q_3)(z - z_0).
\]
This proves (19).

By construction, \(|q_k - p_0| \leq |p - p_0| \) for \( k = 1, 2, 3 \) and of course \(|x - x_0| \leq |p - p_0|, |y - y_0| \leq |p - p_0|, |z - z_0| \leq |p - p_0| \). The continuity of the partial derivatives, together with (19) now shows that for any \( \epsilon > 0 \) there exists \( \delta > 0 \) such that \(|f(p) - f(p_0)| < \epsilon \) for \(|p - p_0| < \delta \) and \( p \in D \). DONE

We repeat that if \( n = 1 \), you do not have to assume that the derivative is continuous, only the existence is required. For \( n > 1 \), existence and continuity of the derivatives is required\(^{19} \).

Assignment 22 (Due June 8) [1, §3.3 page 134 #4,5,11]

8.1.2 Differential as a Linear approximation (the case of functions)

Let’s examine the equation (19). If we write it in vector notation we get some new insight which leads us to the notion of gradient (or differential) of a function and to the notion of approximating a function by a linear function (namely, the differential of the function). The equation (19) can be rewritten as a dot product of vectors:
\[
f(p) - f(p_0) = (f_1(q_1), f_2(q_2), \ldots, f_n(q_n)) \cdot (x_1 - x_1^0, x_2 - x_2^0, \ldots, x_n - x_n^0),
\]
(20)
or, \( f(p) - f(p_0) = V \cdot (p - p_0) \), where \( V \) is the vector \( V = (f_1(q_1), f_2(q_2), \ldots, f_n(q_n)) \).
Recall that the assumption is that \( f \in C^1(D) \), \( D \) is an open set, \( p_0 \in D \) and the conclusion is that the points \( q_1, \ldots, q_n \) can be chosen in any ball with center \( p_0 \) containing \( p \).

Two questions can be asked in connection with (20).

\(^{19}\)this is a little white lie, see Problem 5 in the next assignment
1. Can we pick the $q_1, \ldots, q_n$ all to be the same point (call it $p^*$) lying on the line segment from $p_0$ to $p$? The answer is: YES! This is the Mean Value Theorem in several variables, see [1, Theorem 16, page 151] and a theorem below in the section on Mean Value Theorems. As in the case of one variable, a mean value theorem may not be so interesting in its own right, but it is an important tool which will be very useful in our lifetime.

2. Carrying the previous question one step further, we can be greedy and ask whether the point $p^*$ can be equal to $p_0$. The answer here is NO! (See Assignment 23)

**Assignment 23** (Due June 8) Give an example for $n = 1$ where $p^*$ cannot be chosen to be $p_0$. (Hint: almost any example works). What about $n = 2$?

The following is a fundamental definition. It has occurred implicitly in the above two questions.

**Definition 8.5** If $f : D \to \mathbb{R}$ is defined on an open set $D \subset \mathbb{R}^n$, the gradient of $f$ at $p \in D$ is the vector $\nabla f(p) = (D_1 f(p), D_2 f(p), \ldots, D_n f(p))$. Of course $\nabla f$ is defined only at those points of $D$ where all first order partial derivatives of $f$ exist.

Notice that $\nabla f$ is an example of a transformation, that is, a function $T : D \to \mathbb{R}^m$, where $D \subset \mathbb{R}^n$. In this case, $m = n$. The main purpose of the rest of this course, (and much of classical and modern mathematics) is to study properties of transformations $T : D \to \mathbb{R}^m$, such as continuity and differentiability (suitably defined).

Even though the answer to the second question above is negative, *something is*, nevertheless true. To see what it is that interests me, let us just write down the fact, in a different way, that a function (of one variable) is differentiable. This will enable us to formulate an analogous property for functions of several variables.

If $f$ is differentiable at the point $c \in (a, b) \subset \mathbb{R}$ with derivative $f'(c)$, then

$$\lim_{x \to c} \frac{f(x) - f(c) - f'(c)(x - c)}{x - c} = 0. \tag{21}$$

This is the same as

$$\lim_{h \to 0} \frac{|f(c + h) - f(c) - f'(c)h|}{|h|} = 0. \tag{22}$$

The following is the analog, for functions of several variables, of (21) and (22). It says that a $C^1$-function can be approximated, in some sense, by a linear function. An interpretation of the following theorem is that the functions $f$ and $g$, where $g(p) := f(p_0) + \nabla f(p_0) \cdot (p - p_0)$, are “asymptotic” at $p_0$. Note that (23) implies the much weaker statement that $|f(p) - f(p_0) - \nabla f(p_0) \cdot (p - p_0)| \to 0$ as $p \to p_0$.

**Theorem 8.6** (Theorem 8 on page 131 of [1]) Let $f$ be of class $C^1$ on an open set $D \subset \mathbb{R}^n$. For any $p_0 \in D$,

$$\lim_{p \to p_0} \frac{|f(p) - f(p_0) - \nabla f(p_0) \cdot (p - p_0)|}{|p - p_0|} = 0.$$
Since we have not used the notation \( \lim_{p \to p_0} \), we should explain that it simply means the following: for every \( \epsilon > 0 \), there exists \( \delta > 0 \) such that

\[
\frac{|f(p) - f(p_0) - \nabla f(p_0) \cdot (p - p_0)|}{|p - p_0|} < \epsilon \quad \text{whenever} \quad p \in B(p_0, \delta) \cap D. \tag{23}
\]

**Proof:** Let \( R := f(p) - f(p_0) - \nabla f(p_0) \cdot (p - p_0) \). By (20) (which is the main point in the proof of Theorem 8.4), \( f(p) - f(p_0) = V \cdot (p - p_0) \), where \( V \) is the vector

\[
V = (f_1(q_1), f_2(q_2), \ldots, f_n(q_n)).
\]

Therefore

\[
R = V \cdot (p - p_0) - \nabla f(p) \cdot (p - p_0) = [V - \nabla f(p_0)] \cdot (p - p_0).
\]

Now use the Schwarz inequality:

\[
|R| = |[V - \nabla f(p_0)] : [p - p_0]| \leq |V - \nabla f(p_0)||p - p_0|,
\]

that is

\[
\frac{|R|}{|p - p_0|} \leq |V - \nabla f(p_0)|, \tag{24}
\]

and if you write out the coordinates of \( V - \nabla f(p_0) \) you will see that \( |V - \nabla f(p_0)| \), and hence by (24) \( |R|/|p - p_0| \), approaches zero as \( p \) approaches \( p_0 \).

Here are the details:

\[
V - \nabla f(p_0) = [f_1(q_1), f_2(q_2), \ldots, f_n(q_n)] - [f_1(p_0), f_2(p_0), \ldots, f_n(p_0)]
\]

\[
= [f_1(q_1) - f_1(p_0), f_2(q_2) - f_2(p_0), \ldots, f_n(q_n) - f_n(p_0)],
\]

so that

\[
|V - \nabla f(p_0)|^2 = (f_1(q_1) - f_1(p_0))^2 + (f_2(q_2) - f_2(p_0))^2 + \cdots + (f_n(q_n) - f_n(p_0))^2. \tag{25}
\]

Since each \( f_j \) is continuous and since \( |q_j - p_0| \leq |p - p_0| \) for each \( j \), we see from (24) and (25) that (23) holds.

### 8.2 Wednesday, May 26, 1993

Today we interrupt our study of differentiation to discuss Assignments 15 and 18.

To deal with Assignment 15\(^{21}\), you can use the following lemma:

**Lemma 8.7** Let \( f \) be integrable on \([a, b]\) and let \( \epsilon > 0 \). Then there exists a closed subinterval \( J \subset [a, b] \) such that

\[
M(f, J) - m(f, J) < \epsilon.
\]

Here, we are using the notation \( M(f, S) = \sup \{ f(x) : x \in S \} \) and \( m(f, S) = \inf \{ f(x) : x \in S \} \).

\(^{20}\)This proof was not discussed in class on May 24

\(^{21}\)I would like to thank Dr. S.-Y. Li for interrupting his lunch to show me this proof
**Proof:** (Sketch) For our given $\epsilon$, choose a partition $P$ such that $\overline{S}(P) - \underline{S}(P) < \epsilon$. Choose $J$ to be the subinterval $I_i$ determined by this partition for which the value $M_i - m_i$ is the smallest, that is,

$$M_i - m_i \leq M_j - m_j \text{ for all } 1 \leq j \leq m.$$ 

Since for all $j$, $(M_i - m_i)(x_j - x_{j-1}) \leq (M_j - m_j)(x_j - x_{j-1})$, we get

$$(b-a)(M_i - m_i) \leq \overline{S}(P) - \underline{S}(P) < \epsilon,$$

so $M_i - m_i < \epsilon/(b-a)$, which is just as good as $M_i - m_i < \epsilon$.

Here is the solution to Assignment 15(A): Apply the lemma successively with $\epsilon = 1/n, n = 1, 2, \ldots$. You get a nested sequence of closed intervals

$$[a, b] \supset J_1 \supset J_2 \supset \cdots \supset J_n \supset \cdots$$

such that $M(f, J_n) - m(f, J_n) < 1/n$ for $n = 1, 2, \ldots$. By a well known property of the real number system, the intersection $\cap_{k=1}^{\infty} J_k$ is not empty. You now need to show that $f$ is continuous at any point $x_0$ of this intersection.

There are several remarks to be made in connection with the foregoing. First of all, if you examine the last part of the solution, you find that you may only get one-sided continuity of $f$ at $x_0$. Not to worry; this is enough to solve part (B) of Assignment 15. On the other hand, the above proof can be modified to show that $f$ is indeed continuous at some point $x_0$. You should verify for yourself these last two statements.

The next remark is that the above proof actually shows that $f$ has infinitely many points of continuity. Do you see why?

I even have a third remark. It would be interesting to prove Assignment 15(B) without using (A)?

We now move to a discussion of Assignment 18. Personal opinion of Bernard Russo: the result contained in Assignment 18 is very useful and must be done. However, it is the only part of integration theory that I find unpleasant, in the sense that the proof is painful and not very illuminating. It is not the kind of proof that you could come up with in a short period of time, as I believe is the case with most of the other proofs in the subject. (End of opinion)

Assignment 18 is hereby striken from the homework assignments, and therefore upgraded to a theorem.

**Theorem 8.8** (Formerly Assignment 18)

(A) A bounded function $f$ on a compact interval $I \subset \mathbb{R}$ is integrable on $I$ if and only if

for every $\epsilon > 0$, there exists a partition $P$ such that $\overline{S}(P) - \underline{S}(P) < \epsilon$. (26)
(B) A bounded function $f$ on a compact rectangle $R \subset \mathbb{R}^2$ is integrable on $R$ if and only if

for every $\epsilon > 0$, there exists a grid $N$ such that $\mathcal{S}(N) - \mathcal{S}(N) < \epsilon$.

**Proof:** We shall write out the proof of the first statement. The second one involves exactly the same ideas, but the notation is more cumbersome. Maybe you should write out the proof of the second statement for practice.

We shall use the criterion established in Assignment 14(A),(B), namely, that $f$ is integrable if and only if for every $\epsilon > 0$ there is a $\delta > 0$ such that $S(P) - \mathcal{S}(P) < \epsilon$ for all partitions $P$ with $d(P) < \delta$.

Step 1: If $f$ is integrable, and $\epsilon > 0$, then choose $\delta$ as in the previous paragraph. Choose any partition $P_0$ with $d(P_0) < \delta$. Then (26) is satisfied. This is the easy part of the proof.

Step 2: Assume that (26) holds. We shall prove that $f$ is integrable by again using the criterion established in Assignment 14(A),(B). So, let $\epsilon > 0$, so that we can get on with finding an appropriate $\delta$. By assumption, there is a partition $P_0$ such that

$$S(P_0) - \mathcal{S}(P_0) < \epsilon/2. \quad (27)$$

Let $m$ be the number of subintervals determined by the partition $P_0$ and let $B$ be a bound for $f$: $|f(x)| \leq B$ for every $x \in I$. Now set $\delta := \epsilon/8mB$. Miraculously, this $\delta$ does the job. To show this we take any partition $P$ with $d(P) < \delta$ and proceed to show that $\mathcal{S}(P) - \mathcal{S}(P) < \epsilon$.

We start with the following

**CLAIM:** With $Q$ defined by $Q = P_0 \cup P$, we have

$$S(Q) - \mathcal{S}(P) \leq 2mB \cdot d(P) \quad (28)$$

and

$$\mathcal{S}(P) - \mathcal{S}(Q) \leq 2mB \cdot d(P).$$

Assume for a moment that this claim has been proved. Then since $d(P) < \delta$, and $\delta = \epsilon/8mB$, (28) implies $S(Q) - \mathcal{S}(P) \leq 2mB \cdot d(P) < \epsilon/4$, and $S(P_0) \leq \mathcal{S}(Q)$ implies $S(P_0) - \mathcal{S}(P) \leq \mathcal{S}(Q) - \mathcal{S}(P) < \epsilon/4$. Similarly $\mathcal{S}(P) - \mathcal{S}(Q) < \epsilon/4$ and $\mathcal{S}(P_0) \geq \mathcal{S}(Q)$ implies $\mathcal{S}(P) - \mathcal{S}(P_0) \leq \mathcal{S}(P) - \mathcal{S}(Q) < \epsilon/4$. So

$$\mathcal{S}(P) - \mathcal{S}(P) < [\epsilon/4 + \mathcal{S}(P_0)] + [\epsilon/4 - \mathcal{S}(P_0)] = \mathcal{S}(P_0) - \mathcal{S}(P_0) + \epsilon/2$$

and by (27), $\mathcal{S}(P) - \mathcal{S}(P) < \epsilon/2 + \epsilon/2 = \epsilon$.

It remains to prove the above claim(s). Well, let’s first establish some notation: with

$$P_0 = \{a = s_0 < s_1 < \cdots < s_{m-1} < s_m = b\}$$

and

$$P = \{a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b\},$$

40
we have

$$P \subset Q_1 := P \cup \{s_1\} \subset Q_2 := P \cup \{s_1, s_2\} \subset \cdots \subset Q_{m-1} = P \cup \{s_1, \ldots, s_{m-1}\} = Q.$$  

Now $S(Q_1) - S(P)$ is of the form (let $u$ denote $s_1$)

$$m(f, [t_{k-1}, u])(u - t_{k-1}) + m(f, [u, t_k])(t_k - u) - m(f, [t_{k-1}, t_k])(t_k - t_{k-1})$$

for some $k$ and so

$$|S(Q_1) - S(P)| \leq B(u - t_{k-1}) + B(t_k - u) + B(t_k - t_{k-1}) \leq 2B \cdot d(P). \quad (29)$$

Similarly,

$$|S(Q_2) - S(Q_1)| \leq 2Bd(Q_1) \leq 2B \cdot d(P) \quad (30)$$

$$\cdots$$

$$|S(Q) - S(Q_{m-2})| = |S(Q_{m-1}) - S(Q_{m-2})| \leq 2Bd(Q_{m-2}) \leq 2B \cdot d(P). \quad (31)$$

Adding up the $m - 1$ inequalities (29)-(31), you get $S(Q) - S(P) \leq 2(m - 1)B \cdot d(P)$. This proves the first statement in the claim. Let’s believe the companion statement so we can stop.

**Assignment 24** (Due June 8, 1993) Show that the following 4 statements (write them out) are each equivalent to the integrability of a bounded function $f$ on a compact interval $I \subset \mathbb{R}$.

1. The definition of integrability of $f$
2. The result of Assignment 14(A) and (B).
3. The result of Assignment 18
4. $s = S$

### 8.3 Friday, May 28, 1993

**Second Midterm**

Do all problems. Use only one side of your page and only one problem per page.

**Question 1** (12 points) Let $f : \mathbb{R}^n \to \mathbb{R}$. Prove or disprove: if $f$ is uniformly continuous, then $f$ is convergence preserving, that is, if $p, p_k \in \mathbb{R}^n$ and $\lim_{k \to \infty} p_k = p$, then $\lim_{k \to \infty} f(p_k) = f(p)$.

**Question 2** (24 points) Let $f : [a, b] \to \mathbb{R}$ be any function and let $G \subset \mathbb{R}^2$ be defined by

$$G := \{(x, f(x)) : x \in [a, b]\}.$$

Prove that $f$ is continuous on $[a, b]$ if and only if $G$ is a compact subset of $\mathbb{R}^2$.  

41
Question 3 (12 points) Let \( f : D \to \mathbb{R} \) be defined and continuous on a compact subset \( D \subset \mathbb{R}^n \). Prove that for every closed set \( C \subset D \), \( f(C) \) is a closed set in \( \mathbb{R} \).

Question 4 (12 points) Prove rigorously that a constant function is integrable over any compact interval and give a formula for the value of the integral.

Question 5 (16 points) Let \( f(x) = \begin{cases} 2, & 0 \leq x \leq 1 \\ -1, & 1 < x \leq 3 \end{cases} \).

Explain carefully why \( f \) is integrable on \([0, 3]\) and why \( \int_0^3 f = 0 \). (Hint: Consider \( f \) on the closed subintervals \([0, 1]\), and \([1, 3]\))

Question 6 (24 points) True or false. No proof required, just true or false. Intelligent guesses are allowed. Let \( f \) be defined on a compact rectangle \( R \subset \mathbb{R}^2 \).

(A) If \( f \) is integrable on \( R \), then so is \(|f|\).

(B) If \(|f| \) is integrable on \( R \), then so is \( f \).

(C) If \(|f| \) is integrable on \( R \), then \( \int_R |f| \geq 0 \).

(D) If \( f \) is integrable on \( R \) and \( \int_R |f| = 0 \), then \( f = 0 \) on \( R \).

(E) If \( \int_R f \neq 0 \), then \( \int_R |f| \neq 0 \).

(F) If \( f \) is integrable and \( f(p) \neq 0 \) for every \( p \in R \), then \( \int_R |f| \neq 0 \).

9 Week 9

9.1 Monday May 31, 1993 (Memorial Day Holiday)

9.2 Tuesday June 1, 1993 (Extra Lecture)

We now begin the study of transformations. First a formal definition.

Definition 9.1 A transformation is any function \( T : D \to \mathbb{R}^m \), where \( D \subset \mathbb{R}^n \).

Here, \( m \geq 1 \) and \( n \geq 1 \), so this includes the special case of a function \( f \) considered up to now (that is, \( m = 1, n \) arbitrary). Every transformation gives rise to coordinate functions as follows: if \( p = (x_1, \ldots, x_n) \in D \), and \( T(p) = (y_1, \ldots, y_m) \in \mathbb{R}^m \), then each \( y_j \) is a function of \( p = (x_1, \ldots, x_n) \), which we can denote by \( f_j \) or \( f_j^{22} \). Thus

\[
T(p) = (f^1(p), \ldots, f^m(p)),
\]

\(^{22}\)the latter notation is preferable in order to avoid confusion with the notation \( f_j \) for a partial derivative of some function \( f \)
where each $f^j : D \to \mathbb{R}$ is a function of $n$ variables $x_1, \ldots, x_n$.

Transformations are the subject of [1, Chapter 7] and their geometric properties are discussed in [1, Section 7.2]. Although these geometric properties are important to know for a better understanding of transformations, we will have to take the moral high ground and concentrate on analytic properties of transformations, that is, continuity, and most importantly, differentiability.

Fortunately, the study of continuity of transformations is no more difficult than the study of continuity of functions of several variables. This will be established in the following assignments, namely Assignments 25 to 32.23

The following is the analog of Definition 4.3

**Definition 9.2** Let $T : D \to \mathbb{R}^m$ be a transformation, where $D$ is any subset of $\mathbb{R}^n$, and let $p_0 \in D$. We say that $f$ is **continuous at** $p_0$ if

$$\forall \epsilon > 0, \exists \delta > 0$$

such that

$$|T(p) - T(p_0)| < \epsilon \text{ for all } p \in D \text{ with } |p - p_0| < \delta.$$  

This definition can be put in the compact form

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } T(D \cap B(p_0, \delta)) \subset B(f(p_0), \epsilon).$$

**Assignment 25** (Due June 8) Let $T(p) = (f^1(p), \ldots, f^m(p))$ be a transformation with coordinate functions $f^1, \ldots, f^m$. Prove that $T$ is continuous at $p_0$ if and only if each coordinate function $f^j$, $1 \leq j \leq m$, is continuous at $p_0$.

The following is the analog of Theorem 4.4.

**Theorem 9.3 (Theorem 4 on page 333 of [1])** The continuous image of a compact set is compact. In other words, if $T : D \to \mathbb{R}^m$ is a continuous transformation on $D$, and $D$ is a compact subset of $\mathbb{R}^n$, then $T(D)$ is a compact subset of $\mathbb{R}^m$.

**Assignment 26** Prove Theorem 9.3.

The following is the analog of Theorem 5.4.

**Theorem 9.4** Let $T : D \to \mathbb{R}^m$, where $D \subset \mathbb{R}^n$, and suppose that $T$ is continuous at the point $p_0 \in D$. Then for every sequence $p_k$ from $D$, which converges to $p_0$, we have

$$\lim_{k \to \infty} T(p_k) = T(p_0).$$

**Assignment 27** Prove Theorem 9.4.

---

23Don’t worry, not all of these assignments will be handed in
Assignment 28 State and prove an analog of the Extreme values theorem, Theorem 5.4. (Hint: Since $\mathbb{R}^m$ has no order structure, you have to express the theorem in terms of $|T(p)|$.)

The following is the analog of Definition 5.12

**Definition 9.5** A transformation $T : E \rightarrow \mathbb{R}^m$, where $E \subset \mathbb{R}^n$, is uniformly continuous on $E$ if for every $\epsilon > 0$, there exists $\delta > 0$ such that $|T(p) - T(q)| < \epsilon$ whenever $p, q \in E$ and $|p - q| < \delta$.

The following is the analog of Theorem 5.13.

**Theorem 9.6** A transformation which is continuous on a compact set $D$ is uniformly continuous on $D$.

Assignment 29 Prove Theorem 9.6.

Assignment 30 Show that a linear transformation (see [1, Section 7.3]) is uniformly continuous. (Hint: Use [1, Theorem 8, page 338])

The following is the analog of Theorem 6.1

**Theorem 9.7** Let $T : D \rightarrow \mathbb{R}^m$ be a uniformly continuous transformation defined on a subset $D$ of $\mathbb{R}^n$. Define a transformation $\tilde{T} : \overline{D} \rightarrow \mathbb{R}^m$ by

$$ \tilde{T}(p) = \begin{cases} T(p) & \text{if } p \in D; \\ \lim_{k \to \infty} T(p_k) & \text{if } p \in \mathbb{R}^m \setminus D, \end{cases} $$

where $p_k \in D$ is such that $\lim_k p_k = p$. Then $\tilde{T}$ exists, is well defined, and is continuous on $\overline{D}$.

Assignment 31 Prove Theorem 9.7.

Assignment 32 Let $D \subset \mathbb{R}^n$ be a bounded set and let $T : D \rightarrow \mathbb{R}^m$ be a continuous transformation. Prove that $T$ has a continuous extension to $\overline{D}$ if and only if $T$ is uniformly continuous on $D$.

9.3 Wednesday June 2, 1993

9.3.1 Approximation by the differential—the case of transformations

Our next main result is the analog for transformations of (23) in Theorem 8.6. First we need to define the replacement for the gradient.
Definition 9.8 If $T : D \to \mathbb{R}^m$ is defined on an open set $D \subset \mathbb{R}^n$, with coordinate functions $f^1, \ldots, f^m$, the Jacobian matrix of $T$ at $p \in D$ is the $m$ by $n$ matrix

$$J_T(p) = \begin{bmatrix}
\frac{\partial f^1}{\partial x_1} & \cdots & \frac{\partial f^1}{\partial x_n} \\
\frac{\partial f^2}{\partial x_1} & \cdots & \frac{\partial f^2}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial f^m}{\partial x_1} & \cdots & \frac{\partial f^m}{\partial x_n}
\end{bmatrix}.$$  

Of course $J_T(p)$ is defined only at those points of $D$ where all first order partial derivatives of each coordinate function $f^i$ exist.

We can also write this in the form

$$J_T(p) = \left[ \frac{\partial f^i}{\partial x_j} \right]_{1 \leq i \leq m, 1 \leq j \leq n} = \left[ D_{j} f^i(p) \right]_{1 \leq i \leq m, 1 \leq j \leq n}$$

We shall use $\times$ to denote matrix multiplication. Thus, for example, if $q$ is any (column) vector in $\mathbb{R}^n$, $J_T(p) \times q$ is a (column) vector in $\mathbb{R}^m$. In particular, for the dot product of two (row) vectors $p, q$, if $q^t$ is the transpose of $q$, then $p \cdot q = p \times q^t$.

Later on, for the inverse function theorem, we will have $m = n$, and it will be very important to consider the Jacobian determinant of $T$, which is defined to be $\det J_T(p)$.

Assignment 33 Prove that a transformation of class $C^1$ is continuous.

Theorem 9.9 (Theorem 10 on page 344 of [1]) Let $T : D \to \mathbb{R}^m$ be a transformation of class $C^1$ on an open set $D \subset \mathbb{R}^n$. For any $p_0 \in D$,

$$\lim_{p \to p_0} \frac{|T(p) - T(p_0) - J_T(p_0) \times (p - p_0)^t|}{|p - p_0|} = 0.$$  

The meaning here is: for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$\frac{|T(p) - T(p_0) - J_T(p_0) \times (p - p_0)^t|}{|p - p_0|} < \epsilon \text{ whenever } p \in B(p_0, \delta) \cap D.$$  

(32)

Proof: Let $T = (f^1, \ldots, f^m)$. By Theorem 8.6, for each $1 \leq i \leq m$

$$\frac{|f^i(p) - f^i(p_0) - \nabla f^i(p_0) \cdot (p - p_0)|}{|p - p_0|} \to 0 \text{ as } p \to p_0.$$  

(33)
Using the notation \( R_i(p) = f^i(p) - f^i(p_0) - \nabla f^i(p_0) \cdot (p - p_0) \), we have

\[
T(p) - T(p_0) = (f^1(p) - f^1(p_0), \ldots, f^m(p) - f^m(p_0))^t \\
= (\nabla f^1(p_0) \cdot (p - p_0), \ldots, \nabla f^m(p_0) \cdot (p - p_0))^t + (R_1(p), \ldots, R_m(p))^t \\
= (\sum_{j=1}^n D_j f^1(p_0)(x_j - x_j^0), \ldots, \sum_{j=1}^n D_j f^m(p_0)(x_j - x_j^0))^t + (R_1(p), \ldots, R_m(p))^t.
\]

On the other hand,

\[
J_T(p_0) \cdot (p - p_0)^t = \begin{bmatrix}
\frac{\partial f^1}{\partial x_1} & \ldots & \frac{\partial f^1}{\partial x_n} \\
\frac{\partial f^2}{\partial x_1} & \ldots & \frac{\partial f^2}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial f^m}{\partial x_1} & \ldots & \frac{\partial f^m}{\partial x_n}
\end{bmatrix} \times \begin{bmatrix}
x_1 - x_1^0 \\
x_2 - x_2^0 \\
\vdots \\
x_n - x_n^0
\end{bmatrix}
= (\sum_{j=1}^n D_j f^1(p_0)(x_j - x_j^0), \ldots, \sum_{j=1}^n D_j f^m(p_0)(x_j - x_j^0))^t.
\]

Now let us subtract the last two equations. We get

\[
T(p) - T(p_0) - J_T(p_0) \cdot (p - p_0)^t = (R_1(p), \ldots, R_m(p))^t.
\]

Now use (33) to obtain

\[
\left| \frac{T(p) - T(p_0)}{||p - p_0||} - J_T(p_0) \cdot (p - p_0)^t \right| = \left( \sum_{i=1}^m \frac{R_i(p)^2}{||p - p_0||^2} \right)^{1/2} \to 0
\]
as \( p \to p_0 \). DONE

9.4 Friday June 4, 1993

9.4.1 The Chain Rule

We begin by recalling the statement and proof of the one-dimensional chain rule that we encounter as freshmen (or as seniors in high school) and use every day (sometimes without realizing it). Here, we are very lucky, since we shall write the proof in one-dimensional in such a way that the proof in arbitrary dimensions of the chain rule for transformations will require only notational changes. The key idea underlying this scheme is to write every formula “horizontally”, or on a line. In other words, you can divide by numbers, but not by vectors.

We denote the composition of functions by \( f \circ g \), that is,

\[
f \circ g(x) = f(g(x)).
\]

In order for this to make sense, the range of \( g \) must be a subset of the domain of \( f \).

**Theorem 9.10 (One-dimensional chain rule)** Let \( g \) be a real valued function defined on an open interval containing \( a \in \mathbb{R} \) and suppose that \( g \) is differentiable at \( a \).
with derivative $g'(a)$. Let $f$ be a real valued function defined on an open interval containing $g(a)$ and suppose that $f$ is differentiable at $g(a)$ with derivative $f'(g(a))$. Then $f \circ g$ is differentiable at $a$ with derivative

$$(f \circ g)'(a) = f'(g(a))g'(a).$$

**Proof:** Since $g$ is differentiable at $a$, $\forall \epsilon' > 0, \exists \delta' > 0$ such that

$$|g(x) - g(a) - g'(a)(x - a)| < \epsilon'|x - a| \quad \text{if } |x - a| < \delta'. \quad (34)$$

Since $f$ is differentiable at $g(a)$, $\forall \epsilon'' > 0, \exists \delta'' > 0$ such that

$$|f(y) - f(g(a)) - f'(g(a))(y - g(a))| < \epsilon''|y - g(a)| \quad \text{if } |y - g(a)| < \delta''. \quad (35)$$

We need to prove: $\forall \epsilon > 0, \exists \delta > 0$ such that

$$|f(g(x)) - f(g(a)) - f'(g(a))g'(a)(x - a)| < \epsilon|x - a| \quad \text{if } |x - a| < \delta. \quad (36)$$

Since $g$ is continuous at $a$, $\exists \delta_c > 0$ such that

$$|g(x) - g(a)| < \delta'' \quad \text{if } |x - a| < \delta_c. \quad (37)$$

Using (37), we may replace $y$ in (35) by $g(x)$ to obtain

$$|f(g(x)) - f(g(a)) - f'(g(a))(g(x) - g(a))| < \epsilon''|g(x) - g(a)| \quad \text{if } |x - a| < \delta_c. \quad (38)$$

Now set $\delta := \min\{\delta_c, \delta'\}$ and $\eta(x) := g(x) - g(a) - g'(a)(x - a)$ so that

$$g(x) - g(a) = g'(a)(x - a) + \eta(x) \quad (39)$$

and by (34),

$$|\eta(x)| < \epsilon'|x - a| \quad \text{if } |x - a| < \delta. \quad (40)$$

Now substitute (39) into (38) (in two places!) and set

$$A := f(g(x)) - f(g(a)) - f'(g(a))[g'(a)(x - a) + \eta(x)] \quad (41)$$

to obtain from (38)

$$|A| < \epsilon''|g'(a)(x - a) + \eta(x)| \quad \text{if } |x - a| < \delta. \quad (42)$$

Finally, if $|x - a| < \delta$, we have,

$$|f(g(x)) - f(g(a)) - f'(g(a))g'(a)(x - a)|$$

$$= |A + f'(g(a))\eta(x)| \quad \text{(by (41))}$$

$$\leq |A| + |f'(g(a))\eta(x)|$$

$$\leq \epsilon''|g'(a)||x - a| + \epsilon''|\eta(x)| + |f'(g(a))||\eta(x)| \quad \text{(by (42))}$$

$$\leq |\epsilon''g'(a)| + \epsilon'' \epsilon' + |f'(g(a))|\epsilon'|x - a| \quad \text{(by (40))}$$

$$< \epsilon|x - a|,$$

the last step provided we simply choose $\epsilon'$ and $\epsilon''$ so that $|\epsilon''|g'(a)| + \epsilon'' \epsilon' + |f'(g(a))|\epsilon'| < \epsilon$. This proves (36). DONE

We now consider composition of transformations and the chain rule in arbitrary dimensions.
Definition 9.11 Let $T$ be a transformation defined on a subset $A$ of $\mathbb{R}^n$ with $T(A) \subset \mathbb{R}^m$. Suppose that $S$ is a transformation defined on a subset $C$ of $\mathbb{R}^m$ with $S(C) \subset \mathbb{R}^k$. We suppose that $C \subset T(A)$. Under these circumstances, the composition of $S$ and $T$ is the transformation $S \circ T$ (also denoted simply by $ST$) defined by

$$S \circ T(p) = S(T(p)) \quad (p \in A).$$

EXAMPLE: If $T(x, y) = (xy, 2x, -y)$ and $S(x, y, z) = (x - y, yz)$, then $ST(x, y) = S(T(x, y)) = S(xy, 2x, -y) = (xy - 2x, -2xy)$. In this case, $TS$ is defined and $TS(x, y, z) = T(S(x, y, z)) = T((x - y)yz, 2(x - y), -yz)$. Note that in this case, $ST \neq TS$.

Theorem 9.12 (Theorem 3, page 333 of [1]) If $S : A \rightarrow \mathbb{R}^m$ is a transformation which is continuous at a point $p_0 \in A \subset \mathbb{R}^n$, and $T : B \rightarrow \mathbb{R}^k$ is a transformation which is continuous at the point $S(p_0) \in B \subset \mathbb{R}^m$, then the composition $T \circ S : A \rightarrow \mathbb{R}^k$ is continuous at the point $p_0$.

Assignment 34 Prove Theorem 9.12.

Before stating the general chain rule we must give a “coordinate-free” definition of derivative and discuss some of its properties.

Definition 9.13 (Coordinate-free definition of derivative) Let $T$ be a transformation defined on a subset $A$ of $\mathbb{R}^n$ with $T(A) \subset \mathbb{R}^m$. We say that $T$ is differentiable at $p_0 \in A$ if there exists a linear transformation $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$, such that

$$\lim_{p \to p_0} \frac{|T(p) - T(p_0) - L(p - p_0)|}{|p - p_0|} = 0. \quad (43)$$

We denote $L$ by $T'(p_0)$ (this is justified by Assignment 35) and call it the derivative of $T$ at $p_0$. (Other names for this are total derivative, differential, Frechét derivative, \ldots; other notations are $dT|_{p_0}$, $DT(p_0)$, \ldots)

Assignment 35 Prove that, for a fixed $p_0$, at most one linear transformation $L$ can satisfy (43). (This is the same as Exercise #10, page 352 in [1])

Since at most one linear transformation can satisfy (43), the notation $T'(p_0)$ is justified, that is, $T'$ is a function (single valued, or well-defined) with domain $\{p \in A : T$ is differentiable at $p\}$, which has its values in the set of all linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$.

The next three remarks can be thought of as examples or as informal exercises. Each one is a special case of its successor.

\footnote{There is some logic to this notation: $fg$ (in place of $f \circ g$) can be confused with the ordinary product of the two functions $f$ and $g$, whereas $ST$ cannot, because you cannot multiply vectors}
Remark 9.14 If \( m = 1 \) and \( n = 1 \), then a transformation \( T \) is just a function \( f : A \rightarrow \mathbb{R} \), where \( A \subset \mathbb{R} \). In this case, if \( f \) is differentiable at \( x_0 \), that is, \( f'(x_0) \) exists, then the transformation \( T \) is differentiable at \( x_0 \), with derivative \( T'(x_0) \) which is the linear transformation \( L : \mathbb{R} \rightarrow \mathbb{R} \) given by \( L(x) = f'(x_0)x \). (What is the justification for this?)

Remark 9.15 If \( m = 1 \) and \( n \geq 1 \), then a transformation \( T \) is just a function \( f : A \rightarrow \mathbb{R} \), where this time \( A \subset \mathbb{R}^n \). In this case, if \( f \) is of class \( C^1 \) on an open set containing \( p_0 \), then the transformation \( T \) is differentiable at \( p_0 \), with derivative \( T'(p_0) \) which is the linear transformation \( L : \mathbb{R}^n \rightarrow \mathbb{R} \) given by \( L(p) = \nabla f(p_0) \cdot p \). (What is the justification for this?)

Remark 9.16 If \( m \geq 1 \) and \( n \geq 1 \), then a transformation \( T \) is just a function \( T : A \rightarrow \mathbb{R}^m \), where \( A \subset \mathbb{R}^n \). In this case, if \( T \) is of class \( C^1 \) on an open set containing \( p_0 \), then the transformation \( T \) is differentiable at \( p_0 \), with derivative \( T'(p_0) \) which is the linear transformation \( L : \mathbb{R}^n \rightarrow \mathbb{R}^m \) given by \( L(p) = J_T(p_0) \times p \). (What is the justification for this?)

We now see one reason for introducing the coordinate-free definition of the derivative of a transformation \( T \). In the first place, it is more general than the “coordinate” definition given by the Jacobian matrix \( J_T \). For, according to Remark 9.16, if \( T \) is of class \( C^1 \), that is, all the first order partial derivatives exist and are continuous, then \( T \) is differentiable with derivative \( T'(p_0) = J_T(p_0) \). On the other hand, for a differentiable transformation, the first order partial derivatives of its coordinate functions all exist (see the next Assignment), but they are not necessarily continuous.

Assignment 36 If \( T = (f^1, \ldots, f^m) \) is a differentiable transformation at \( p_0 \), then the partial derivatives \( \frac{\partial f^i}{\partial x^j}(p_0) \) exist for all \( 1 \leq j \leq n, 1 \leq i \leq m \). In other words, the Jacobian matrix \( J_T(p_0) \) exists. (Hint: In the definition of partial derivative, let \( p = p_0 + te_j \) where \( t \in \mathbb{R} \) and \( e_1 = (1,0,\ldots,0), e_2 = (0,1,0,\ldots,0), \ldots, e_n = (0,\ldots,0,1) \), as in (16').

We are now ready to prove the chain rule for composition of transformations. We only have to assume that the transformations are differentiable (not necessarily of class \( C^1 \)). There is very little work to do, in fact, this proof is a word processor’s dream—just make the notational changes to the proof, already printed above, of Theorem 9.10.

To make life simpler, we shall isolate two lemmas, which are themselves of independent interest. We first met Lemma 9.17 in Assignment 30.

Lemma 9.17 (Theorem 8, page 338 of [1]) A linear transformation \( L \) from \( \mathbb{R}^n \) to \( \mathbb{R}^m \) is continuous. In fact, \( L \) is uniformly continuous and there is a constant \( C \)

\[ p^t \]Recall that \( p^t \) is the transpose of the row vector \( p \)
such that $|L(p)| \leq C|p|$ for every $p \in \mathbb{R}^n$. More precisely, if $L$ is given by a matrix $[a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$ as follows:

$$L(\sum_{j=1}^{n} x_j e_j) = \sum_{j=1}^{n} x_j L(e_j)$$

where $Le_j = \sum_{i=1}^{m} a_{ij} e_i$ and $e_1 = (1,0,\ldots,0)$, $e_2 = (0,1,0,\ldots,0)$, $\ldots$, $e_n = (0,\ldots,0,1)$, then

$$|L(p)| \leq (\sum_{i} \sum_{j} a_{ij}^2)^{1/2}|p|.$$ 

**Proof:** With $p = \sum_{j=1}^{n} x_j e_j$,

$$L(p) = \sum_{j} x_j \sum_{i} a_{ij} e_i = \sum_{i} \left(\sum_{j} x_j a_{ij}\right) e_i,$$

so, by the Schwarz inequality,

$$|L(p)|^2 = \sum_{i} |\sum_{j} x_j a_{ij}|^2 \leq \sum_{i} \left(\sum_{j} x_j^2\right) \left(\sum_{j} a_{ij}^2\right) = \left(\sum_{i} \sum_{j} a_{ij}^2\right) |p|^2.$$ 

DONE

**Lemma 9.18 (Differentiability implies continuity II)** A transformation which is differentiable at a point $p_0$ is continuous at that point.

**Proof:** We know that

$$\lim_{p \to p_0} \frac{|T(p) - T(p_0) - T'(p_0)(p - p_0)|}{|p - p_0|} = 0.$$ 

Let $\epsilon = 365$. Then there exists a $\delta > 0$ such that

$$\frac{|T(p) - T(p_0) - T'(p_0)(p - p_0)|}{|p - p_0|} < 365 \text{ for } |p - p_0| < \delta.$$ 

Writing this “horizontally”, you get

$$|T(p) - T(p_0)| < 365|p - p_0| \text{ for } |p - p_0| < \delta.$$ 

Now write $T(p) - T(p_0) = T(p) - T(p_0) - T'(p_0)(p - p_0) + T'(p_0)(p - p_0)$ to arrive at

$$|T(p) - T(p_0)| \leq |T(p) - T(p_0) - T'(p_0)(p - p_0)| + |T'(p_0)(p - p_0)|$$

$$\leq 365|p - p_0| + C|p - p_0| = (365 + C)|p - p_0|.$$ 

Thus $T$ is continuous at $p_0$.

**Question:** What is the difference between Lemma 9.18 and Assignment 33.

In the proof of the following theorem, the names of the characters were changed to protect the innocent. Any similarity with any characters, living or dead, is purely intentional.
Theorem 9.19 (Chain Rule, Theorem 11, page 346 of [1]) Let $T : D \to \mathbb{R}^m$ be a transformation which is differentiable on an open set $D \subset \mathbb{R}^n$, and let $S : E \to \mathbb{R}^k$ be a differentiable transformation on an open subset $E$ of $\mathbb{R}^m$ containing $T(D)$. Then $S \circ T$ is differentiable on $D$, and if $p \in D$, then

$$(S \circ T)'(p) = S'(T(p)) \circ T'(p).$$

**Proof:** Let $p_0 \in D$. Since $T$ is differentiable at $p_0$, $\forall \epsilon > 0$, $\exists \delta' > 0$ such that

$$|T(p) - T(p_0) - T'(p_0)(p - p_0)| < \epsilon |p - p_0| \text{ if } |p - p_0| < \delta'.$$

(44)

Since $S$ is differentiable at $T(p_0)$, $\forall \epsilon'' > 0$, $\exists \delta'' > 0$ such that

$$|S(q) - S(T(p_0)) - S'(T(p_0))(q - T(p_0))| < \epsilon'' |q - T(p_0)| \text{ if } |q - T(p_0)| < \delta''.$$

(45)

We need to prove: $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$$|S \circ T(p) - S \circ T(p_0) - S'(T(p_0)) \circ T'(p_0)(p - p_0)| < \epsilon |p - p_0| \text{ if } |p - p_0| < \delta.$$

(46)

By Lemma 9.18, $T$ is continuous at $p_0$, so $\exists \delta_c > 0$ such that

$$|T(p) - T(p_0)| < \delta'' \text{ if } |p - p_0| < \delta_c.$$

(47)

Using (47), we may replace $q$ in (45) by $T(p)$ to obtain

$$|S(T(p)) - S(T(p_0)) - S'(T(p_0))(T(p) - T(p_0))| < \epsilon'' |T(p) - T(p_0)| \text{ if } |p - p_0| < \delta_c.$$

(48)

Now set $\delta := \min\{\delta_c, \delta'\}$ and $\eta(p) := T(p) - T(p_0) - T'(p_0)(p - p_0)$ so that

$$T(p) - T(p_0) = T'(p_0)(p - p_0) + \eta(p)$$

(49)

and by (44),

$$|\eta(p)| < \epsilon' |p - p_0| \text{ if } |p - p_0| < \delta.$$

(50)

Now substitute (49) into (48) (in two places!) and set

$$A(p) := S(T(p)) - S(T(p_0)) - S'(T(p_0))[T'(p_0)(p - p_0) + \eta(p)]$$

(51)

to obtain from (48)

$$|A(p)| < \epsilon''|T'(p_0)(p - p_0) = \eta(p)| \text{ if } |p - p_0| < \delta.$$

(52)

Finally, if $|p - p_0| < \delta$, we have,

$$|S(T(p)) - S(T(p_0)) - S'(T(p_0)) \circ T'(p_0)(p - p_0)|$$

$$= |A(p) + S'(T(p_0))\eta(p)| \quad \text{(by (51))}$$

$$\leq |A(p)| + |S'(T(p_0))\eta(p)|$$

$$\leq \epsilon''|T'(p_0)(p - p_0)| + \epsilon''|\eta(p)| + |S'(T(p_0))\eta(p)| \quad \text{(by (52))}$$

$$\leq \epsilon''C_1 |p - p_0| + \epsilon'' \epsilon' |p - p_0| + C_2 \epsilon' |p - p_0| \quad \text{(by (50) and Lemma 9.18)}$$

$$< \epsilon |p - p_0|,$$

51
the last step provided we simply choose $\epsilon'$ and $\epsilon''$ so that $|\epsilon''C_1 + \epsilon''\epsilon' + C_2\epsilon'| < \epsilon$. This proves (46). DONE

The power of Theorem 9.19 is that by setting $m = n = k = 1$ you get the one-dimensional chain rule (Theorem 9.10), and by setting $m = k = 1$ and leaving $n \geq 1$ you subsume the discussion of the chain rule in [1, section 3.4]. To make this last statement really accurate we need to discuss the difference between a transformation being differentiable and being of class $C^1$. This was already broached in an earlier assignment.

First, let’s have some fun with coordinates in the setting of the chain rule. Let $T = (f^1, \ldots, f^m)$, $S = (g^1, \ldots, g^k)$, and $S \circ T = (h^1, \ldots, h^k)$ where, for $1 \leq i \leq m, 1 \leq j \leq k, 1 \leq r \leq k$,

$$f^i : D \to \mathbb{R}, \quad g^j : E \to \mathbb{R}, \text{ and } h^r : D \to \mathbb{R}.$$

Since

$$S \circ T(p) = S(T(p)) = S(f^1(p), \ldots, f^m(p)) = (g^1(f^1(p), \ldots, f^m(p)), \ldots, g^k(f^1(p), \ldots, f^m(p))),$$

we see that $h^r(p) = g^r(f^1(p), \ldots, f^m(p))$ for $1 \leq r \leq k$. Using this you should have no problem with the next assignment.

**Assignment 37** Let $T$ be a transformation which is of class $C^1$ on an open set $D$, and let $S$ be a transformation of class $C^1$ on an open set containing $T(D)$. Then $S \circ T$ is of class $C^1$ on $D$.

The following is the “coordinatized” version of the chain rule. Notice that it requires the stronger assumption of the transformations being of class $C^1$, not just differentiable. Notice also that there is nothing to prove, given Theorem 9.19 and Assignment 37.

**Corollary 9.20** Let $T$ be a transformation which is of class $C^1$ on an open set $D$, and let $S$ be a transformation of class $C^1$ on an open set containing $T(D)$. Then $S \circ T$ is of class $C^1$ on $D$, and if $p \in D$, then

$$J_{S \circ T}(p) = J_S(T(p)) \times J_T(p).$$

As an illustration of the power of Corollary 9.20, we prove the following theorem from [1, section 3.4].

**Theorem 9.21 (Baby chain rule, Theorem 14, page 136 of [1])** Let $F(t) = f(x, y)$, where $x = g(t), y = h(t)$, the functions $g, h$ are assumed to be of class $C^1$ on an open interval containing $t_0 \in \mathbb{R}$, and the function $f$ is assumed to be of class $C^1$ in an open ball with center $p_0 = (x_0, y_0) = (g(t_0), h(t_0))$. Then $F$ is of class $C^1$ on an open interval containing $t_0 \in \mathbb{R}$, and for $t$ in that interval,

$$F'(t) = \frac{\partial f}{\partial x}(g(t), h(t))g'(t) + \frac{\partial f}{\partial y}(g(t), h(t))h'(t).$$
Proof: Set \( T(t) = (g(t), h(t)) \) and \( S(x, y) = f(x, y) \). Then \( F(t) = S \circ T(t) \), and by Corollary 9.20,

\[
F'(t) = J_F(t) = J_{S \circ T}(t) = J_S(T(t)) \times J_T(t)
\]

\[
= \begin{bmatrix}
\frac{\partial f}{\partial x}(g(t), h(t)) & \frac{\partial f}{\partial y}(g(t), h(t))
\end{bmatrix}
\begin{bmatrix}
g'(t) \\
h'(t)
\end{bmatrix}
\]

\[
= \frac{\partial f}{\partial x}(g(t), h(t))g'(t) + \frac{\partial f}{\partial y}(g(t), h(t))h'(t). \text{ DONE}
\]

Assignment 38  
Let \( F(x, y) = f(g(x, y), h(x, y)) \), where \( g : \mathbb{R}^2 \to \mathbb{R} \) and \( h : \mathbb{R}^2 \to \mathbb{R} \). Use Corollary 9.20 to prove that

\[
F_1(x, y) = f_1(g(x, y), h(x, y))g_1(x, y) + f_2(g(x, y), h(x, y))h_1(x, y)
\]

and

\[
F_2(x, y) = f_1(g(x, y), h(x, y))g_2(x, y) + f_2(g(x, y), h(x, y))h_2(x, y).
\]

(Compare using Corollary 9.20 with the method on page 137 of [1].)

Assignment 39  
[1, §7.4 page 351, #2,5,(7 or 8)] (three problems)

HOMEWORK ALERT! (Reminder, Adjustment,...)

The following are the assignments that are due on June 8 (total of eleven assignments—
that’s a lot, but what the heck, some of them are short. Think how much you will
learn by just thinking about them!)

Assignment 19; Assignment 20 or 21—not both; Assignment 22; Assignment 23 or
24—not both; Assignment 25; One of Assignments 26,27,29,31—(just one, any one);
One of Assignments 28,30,32—(just one, any one); Assignment 33 or 34—not both;
Assignment 35 or 36—not both; Assignment 37 or 38—not both, Assignment 39.

10  Week 10

We shall prove four theorems this week. The last one is a famous one, called the
Inverse function theorem. The inverse function theorem is the key tool in the implicit
function theorem, which was the main goal of this course, and is a very useful result in
almost any branch of analysis. Since we did not have enough time to cover the implicit
function theorem, it will be included in an Appendix which you can study after this
course is over. Also included in an appendix will be a discussion of connectedness,
another topic which we did not have time for.

Even in one variable, the inverse function theorem is not so easy. We will recall
the statement (but not the proof) of the one-variable result below for motivation (see
Theorem 10.8).

Here is a preview of the four theorems to be discussed this week. We shall present
them in a slightly different order from that of [1, §7.6]. We shall give each of these
theorems a “nickname".
10.1 Monday June 7, 1993

10.1.1 Mean Value Theorems

Up to now we have used the mean value theorem in one variable (Theorem 8.1). But we mentioned the mean value theorem in several variables above, so we might as well talk about it. There are two several-variable versions, one for functions and one for transformations. We shall state and prove both of them in what follows, and use the one about transformations to give an alternate proof to Theorem 9.9. This is just one application, and there are many others. For example, we shall use it to prove the local invertibility of a $C^1$ transformation (Theorem 10.3).

We note that the version for functions, nicknamed the “Little Mean Value Theorem” will be used in the proof of the version for transformations, nicknamed the “Big Mean Value Theorem”. Also, the “Baby Chain Rule” (Theorem 9.21) is needed in the proof of the “Little Mean Value Theorem”\(^{27}\).

**Theorem 10.1 (“Little” Mean Value Theorem, Theorem 16, page 151 of [1])** Let $f : B(p_0, r) \to \mathbb{R}$ be of class $C^1$ on a ball $B(p_0, r) \subset \mathbb{R}^n$. Then for any two points $p_1, p_2 \in B(p_0, r)$, there is another point $p^*$ on the line segment $L := \{tp_2 + (1-t)p_1 : 0 \leq t \leq 1\}$ connecting $p_1$ and $p_2$ such that

$$f(p_2) - f(p_1) = \nabla f(p^*) \cdot (p_2 - p_1).$$

\(^{27}\)We have a little and big mean value theorem. Question: what is the “tiny mean value theorem”?

\(^{28}\)Note that this line segment is a subset of $B(p_0, r)$
**Proof:** Define a function $F : [0, 1] \to \mathbb{R}$ by

$$F(t) = f(tp_2 + (1 - t)p_1).$$

We note that $F = f \circ \phi$ where $\phi : [0, 1] \to \mathbb{R}^n$ is the function $\phi(t) = tp_2 + (1 - t)p_1$ and that $J_\phi(\lambda) = (p_2 - p_1)^t$.

By the one-variable mean value theorem, since $f(p_2) - f(p_1) = F(1) - F(0)$,

$$f(p_2) - f(p_1) = F'(\lambda)$$

for some $\lambda \in (0, 1)$.

Letting $p^* = \phi(\lambda)$ we get by the chain rule,

$$F'(\lambda) = \nabla f(\phi(\lambda)) \times J_\phi(\lambda) = \nabla f(\phi(\lambda)) \times (p_2 - p_1)^t = \nabla f(p^*) \cdot (p_2 - p_1).$$

(54)

Compare (53) and (54). **DONE**

**Theorem 10.2 ("Big" Mean Value Theorem, Theorem 12, page 350 of [1])**

Let $T : D \to \mathbb{R}^m$ be a transformation of class $C^1$ on an open set $D \subset \mathbb{R}^n$. Let $p', p'' \in D$ and suppose that the line segment $L := \{tp' + (1 - t)p' : 0 \leq t \leq 1\}$ is a subset of $D$. Then there exist points $p_1^*, \ldots, p_m^* \in L$ such that

$$T(p'') - T(p') = M \times (p'' - p')^t,$$

where $M$ is the matrix $(D_j f^i(p_i^*))_{1 \leq i \leq m, 1 \leq j \leq n}$, that is,

$$M = \begin{bmatrix}
\frac{\partial f^1}{\partial x_1}(p_1^*) & \cdots & \frac{\partial f^1}{\partial x_n}(p_1^*) \\
\frac{\partial f^2}{\partial x_1}(p_2^*) & \cdots & \frac{\partial f^2}{\partial x_n}(p_2^*) \\
\vdots & \ddots & \vdots \\
\frac{\partial f^m}{\partial x_1}(p_m^*) & \cdots & \frac{\partial f^m}{\partial x_n}(p_m^*)
\end{bmatrix}.$$

**Proof:** Apply the Little mean value theorem (Theorem 10.1) to each $f^i : D \to \mathbb{R}$: you will get points $p_i^* \in L$ such that

$$f^i(p'') - f^i(p') = \nabla f^i(p_i^*) \cdot (p'' - p') \quad (1 \leq i \leq m).$$

(55)

Now write down the coordinates of the vector $T(p'') - T(p')$ and use (55):

$$T(p'') - T(p') = (f^1(p''), \ldots, f^m(p''))^t - (f^1(p'), \ldots, f^m(p'))^t$$
$$= (f^1(p'') - f^1(p'), \ldots, f^m(p'') - f^m(p'))^t$$
$$= (\nabla f^1(p_1^*) \cdot (p'' - p'), \ldots, \nabla f^m(p_m^*) \cdot (p'' - p'))^t.$$

---

29 Note that in the following equation, vectors of the form $T(p)$ are column vectors

30 How does $M$ differ from the Jacobian matrix of $T$?

31 Note that $M = (\nabla f^1(p_1^*), \ldots, \nabla f^m(p_m^*))^t$
On the other hand,

\[ M \times (p'' - p')^t = \begin{bmatrix} \nabla f^1(p^*_1) \\ \vdots \\ \nabla f^m(P^*_m) \end{bmatrix} \times (p'' - p')^t = \begin{bmatrix} \nabla f^1(p^*_1) \cdot (p'' - p') \\ \vdots \\ \nabla f^m(p^*_m) \cdot (p'' - p') \end{bmatrix}. \]

Now compare the last two displayed equations. DONE

For no particularly good reason, we now give an alternate proof to the approximation property of the Jacobian matrix (Theorem 9.9).

**Second Proof of Theorem 9.9:** By the Big mean value theorem (Theorem 10.2), \( T(p) - T(p_0) = L^* \times (p - p_0)^t \) where \( L^* := (D_jf^i(p^*_i)) \). Look at the matrix entries of \( L^* - J_T(p_0) = (a_{ij}); \) they are \( a_{ij} = D_jf^i(p^*_i) - D_jf^i(p_0) \). By Lemma 9.17, for all column vectors \( q \in \mathbb{R}^n \),

\[ |(L^* - J_T(p_0)) \times q| \leq \left( \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2} |q|. \]

Since \( T(p) - T(p_0) - J_T(p_0) \times (p - p_0)^t = (L^* - J_T(p_0)) \times (p - p_0)^t \), we have,

\[ \frac{|T(p) - T(p_0) - J_T(p_0) \times (p - p_0)^t|}{|p - p_0|} \leq \frac{|(L^* - J_T(p_0)) \times (p - p_0)^t|}{|p - p_0|} \]

\[ \leq \frac{\left( \sum_{i,j} (a_{ij})^2 \right)^{1/2} |p - p_0|}{|p - p_0|} \]

\[ = \left( \sum_{i,j} (D_jf^i(p^*_i) - D_jf^i(p_0))^2 \right)^{1/2} \]

\[ \rightarrow 0 \text{ as } p \rightarrow p_0, \]

because, as \( p \rightarrow p_0 \), each \( p^*_i \rightarrow p_0 \) and \( T \) is of class \( C^1 \). DONE

### 10.1.2 The local invertibility theorem

The following simple one-dimensional illustration gives the flavor of the statement and proof of the local invertibility theorem, Theorem 10.3. Let \( f : D \rightarrow \mathbb{R} \) be differentiable on an open set \( D \subset \mathbb{R} \) and suppose that \( f'(x) \neq 0 \) for every \( x \in D \). Then \( f \) is **locally one-to-one** on \( D \), that is, for every \( x_0 \in D \) there exists \( \delta > 0 \) such that \( B(x_0, \delta) \subset D \) and \( f \) is one-to-one on \( B(x_0, \delta) \). **Proof:** Since \( D \) is open, given \( x_0 \in D \), just choose any interval \( I = B(x_0, \delta) \subset D \) and apply the one-variable mean value theorem: if \( x', x'' \in I \), then for some \( \tilde{x} \) between \( x' \) and \( x'' \),

\[ f(x'') - f(x') = f'(\tilde{x})(x'' - x'). \]  \hspace{1cm} (56)

If \( f(x'') = f(x') \), then since \( f'(\tilde{x}) \neq 0 \), (56) implies \( x'' = x' \).
Theorem 10.3 (Local invertibility, Theorem 14, page 355 of [1]) Let $T : D \to \mathbb{R}^n$ be a transformation of class $C^1$ defined on an open set $D \subset \mathbb{R}^n$ and suppose that $\det J_T(p) \neq 0$ for all $p \in D$.

Then $T$ is locally one-to-one in $D$, in the sense that for every $p_0 \in D$, there is a $\delta > 0$ such that $B(p_0, \delta) \subset D$ and the restriction of $T$ to $B(p_0, \delta)$ is one-to-one on $B(p_0, \delta)$.

Proof: Consider the open set $\Omega := D \times \cdots \times D \subset \mathbb{R}^n \times \cdots \times \mathbb{R}^n$. The set $\Omega$ is a subset of $\mathbb{R}^{n^2}$. Here is the trick: define a function $F : \Omega \to \mathbb{R}$ by

$$F(p_1, \ldots, p_n) = \det[D_j f^i(p_i)]$$
for $p_1, \ldots, p_n \in D$.

We note first that $F$ is a continuous function on $\Omega$ since, each $T$ being of class $C^1$, all of the functions $D_j f^i$ are continuous, and $F$, being a determinant, is a sum of products of these functions.

We note next that the value of $F$ at a special point of $\Omega$ of the form $(p, \ldots, p)$ is given by $F(p, \ldots, p) = \det[D_j f^i(p_j)] = \det J_T(p)$ and so for every $p \in D$, $F(p, \ldots, p) \neq 0$.

It follows from the last two paragraphs that, given a point, let’s call it $p_0$ now, there is a $\delta > 0$ such that $B(p_0, \delta) \subset D$ and

$$F(p_1, \ldots, p_n) \neq 0$$
for every $(p_1, \ldots, p_n) \in B(p_0, \delta) \times \cdots \times B(p_0, \delta)$. (57)

CLAIM: $T$ is one-to-one on $B(p_0, \delta)$

To prove this claim, we use the Mean value theorem for transformations, Theorem 10.2. Let $p', p'' \in B(p_0, \delta)$ and suppose that $T(p') = T(p'')$. We shall prove that $p' = p''$. Now the line segment $L$ connecting $p'$ and $p''$ lies in $B(p_0, \delta)$ and the Mean value theorem tells us that there are points $p_1^*, \ldots, p_n^* \in L$ such that, with $M = [D_j f^i(p_j^*)]$,

$$T(p'') - T(p') = M \times (p'' - p').$$

Now $\det M = F(p_1^*, \ldots, p_n^*) \neq 0$ by (57), so $M$ is non-singular. Since we are assuming $T(p'') = T(p')$, (58) shows $p'' - p' = 0$. DONE

10.2 Wednesday June 9, 1993

10.2.1 The open mapping theorem

In the next theorem, we shall use the following elementary “critical point” result.

Lemma 10.4 (Theorem 11, page 133 of [1]) Let $f : D \to \mathbb{R}$ be of class $C^1$ on an open set $D \subset \mathbb{R}^n$ and suppose that $f$ has a local minimum at a point $p_0 \in D$. Then all the first order partial derivatives of $f$ vanish at $p_0$: $D_j f(p_0) = 0$ for $1 \leq j \leq n$. Stated another way, $\nabla f(p_0) = 0$.

\(^{32}\) Note that $J_T(p)$ is an $n$ by $n$ matrix, so its determinant makes sense.
**Proof:** The meaning of “local minimum” is that there exists a ball \( B(p_0, r) \subset D \) such that \( f(p) \geq f(p_0) \) for all \( p \in B(p_0, r) \). By definition,

\[
D_j f(p_0) = \lim_{t \to 0} \frac{f(p_0 + te_j) - f(p_0)}{t}.
\]  

(59)

In (59), the numerator is non-negative whenever \( p_0 + te_j \in B(p_0, r) \). Thus if we let \( t \) approach zero through negative values, we get \( D_j f(p_0) \geq 0 \), whereas if we let \( t \) approach zero through positive values, we get \( D_j f(p_0) \leq 0 \). Thus \( D_j f(p_0) = 0 \). DONE

We shall also use the following fact about compact sets.

**Assignment 40** Prove that if \( K \) is a compact set in \( \mathbb{R}^n \) and \( q \notin K \), then

\[
\inf \{|p - q| : p \in K\} > 0.
\]

**Theorem 10.5 (Open mapping, Theorem 15, page 356 of [1])** Let \( T : D \to \mathbb{R}^n \)
be a transformation of class \( C^1 \) defined on an open set \( D \subset \mathbb{R}^n \) and suppose that

\[
\det J_T(p) \neq 0 \text{ for all } p \in D.
\]

Then \( T(D) \) is an open subset of \( \mathbb{R}^n \).

**Proof:** Let \( q_0 \in T(D) \). Choose a point \( p_0 \in D \) such that \( q_0 = T(p_0) \). By Theorem 10.3, there is a \( \delta > 0 \) such that \( T \) is one-to-one on \( B(p_0, 2\delta) \subset D \). Thus \( T \) is one-to-one on the closed ball \( N := \{ p \in D : |p - p_0| \leq \delta \} \subset D \). The boundary \( C = \{ p \in D : |p - p_0| = \delta \} \) of \( N \) is a compact set and therefore so is its image \( T(C) \), and clearly \( q_0 \notin T(C) \). Thus by Assignment 40, \( d := \inf\{|q_0 - q| : q \in T(C)\} > 0 \).

**CLAIM 1:** \( B(q_0, d/3) \subset T(D) \).

This claim shows that \( T(D) \) is an open set. Thus we are done if we prove this claim. We shall show that each point \( q_1 \in B(q_0, d/3) \) belongs to \( T(D) \). So fix a point \( q_1 \in B(q_0, d/3) \). Define a function \( \phi : N \to [0, \infty) \) by the rule: \( \phi(p) = |T(p) - q_1|^2 \).

The function \( \phi \) is continuous on the compact set \( N \), so by the extreme values theorem, it attains its minimum at some point, call it \( p^* \in N \). Thus \( \phi(p) \geq \phi(p^*) \) for all \( p \in N \), which can be expressed as:

\[
\forall p \in N, \quad |T(p) - q_1|^2 \geq |T(p^*) - q_1|.
\]  

(60)

**CLAIM 2:** \( p^* \in \text{int } N \), that is, \( p^* \notin C \).

To prove claim 2, note first that, by the definition of \( d \), for all \( p \in C, |T(p) - q_0| \geq d \), and thus by the backwards Schwarz inequality, for \( p \in C, \)

\[
|T(p) - q_1| \geq |T(p) - q_0| - |q_0 - q_1| \geq d - d/3 = 2d/3.
\]  

(61)

Note that \( p_0 \in N, T(p_0) = q_0, \) and \( |q_0 - q_1| < d/3 \). Suppose now that \( p^* \in C \). Then we would have on the one hand, by (61), \( |T(p^*) - q_1| \geq 2d/3 \), and on the other hand, by (60), \( |T(p^*) - q_1| \leq |T(p_0) - q_1| < d/3 \), a contradiction, proving claim 2.
By claim 2, \( p^* \) is an interior point of \( N \) so that by Lemma 10.4, \( D_j \phi(p^*) = 0 \) for \( 1 \leq j \leq n \).

We now need to write down some explicit formulas for the function \( \phi \). At this point, for convenience, we assume that \( n = 2 \). We can write \( T(x, y) = (f(x, y), g(x, y)) \), where \( f \) and \( g \) are the coordinate functions of \( T \), and if we set \( q_1 = (a, b) \) and \( p = (x, y) \), we have

\[
\phi(x, y) = (f(x, y) - a)^2 + (g(x, y) - b)^2
\]

\[
\frac{\partial \phi}{\partial x}(x, y) = 2(f(x, y) - a)\frac{\partial f}{\partial x}(x, y) + 2(g(x, y) - b)\frac{\partial g}{\partial x}(x, y)
\]

\[
\frac{\partial \phi}{\partial y}(x, y) = 2(f(x, y) - a)\frac{\partial f}{\partial y}(x, y) + 2(g(x, y) - b)\frac{\partial g}{\partial y}(x, y)
\]

and so (plugging in \( p^* \))

\[
0 = 2(f(p^*) - a)\frac{\partial f}{\partial x}(p^*) + 2(g(p^*) - b)\frac{\partial g}{\partial x}(p^*)
\]

\[
0 = 2(f(p^*) - a)\frac{\partial f}{\partial y}(p^*) + 2(g(p^*) - b)\frac{\partial g}{\partial y}(p^*)
\]

The matrix of coefficients of this two by two system of linear equations is \( J_T(p^*) \), which has a non-zero determinant by assumption. Thus \( f(p^*) - a = 0 \) and \( g(p^*) - b = 0 \), that is

\[ T(p^*) = (f(p^*), g(p^*)) = (a, b) = q_1, \]

and thus \( q_1 \in T(D) \), as required. DONE

10.3 Friday June 11, 1993

10.3.1 Automatic continuity of the inverse

The special case of Theorem 10.7 in which \( m = n = 1 \) and \( D \) is a compact interval is proved in [2, 18.4, 18.6]. So even in the one-dimensional case, Theorem 10.7 is stronger than [2, 18.4, 18.6].

Before stating and proving the next theorem, let’s state a very simple and very useful lemma.

Lemma 10.6 A sequence of points in \( \mathbb{R}^n \) converges to a point \( p \in \mathbb{R}^n \) if and only if every subsequence of the given sequence has a subsequence which converges to \( p \).

Assignment 41 Prove Lemma 10.6.

Assignment 42 If a transformation preserves convergent sequences, then it is continuous. (Same proof as [1, Theorem 2, page 74].)

Theorem 10.7 (Automatic continuity of inverse, Theorem 13, page 353 of [1])

Let \( T : D \to \mathbb{R}^n \) be a continuous one-to-one transformation defined on a compact set \( D \subset \mathbb{R}^n \). Then the inverse transformation \( T^{-1} \) (which exists since \( T \) is one-to-one) is continuous.
**Proof:** Let \( p_k \) be a sequence from \( D \), let \( p \in D \) and suppose that \( \lim_{k \to \infty} T(p_k) = T(p) \). According to Assignment 42 all we need to do is prove \( \lim_{k \to \infty} p_k = p \). For this we shall use Lemma 10.6. So let \( p_{kj} \) be a subsequence of \( p_k \). By the BW property there is a further subsequence \( p_{kj} \) and a point \( q \in D \) such that

\[
\lim_{l \to \infty} p_{kj_l} = q.
\]

Since \( T \) is continuous, \( \lim_{l \to \infty} T(p_{kj_l}) = T(q) \). But \( T(p_{kj_l}) \) is a subsequence of \( T(p_k) \) so \( T(p_{kj_l}) \to T(p) \). Thus \( T(p) = T(q) \) and since \( T \) is one-to-one, \( p = q \). By Lemma 10.6, \( \lim_k p_k = p \). **DONE**

### 10.3.2 The inverse function theorem

The inverse function theorem (Theorem 10.9 below) is the \( n \)-dimensional analog of the following result in one-variable which we state here for comparison purposes.

**Theorem 10.8 (Theorem 29.9, page 165 of [2])** Let \( f \) be a one-to-one continuous function on an open interval \( I \subset \mathbb{R} \) and let \( J = f(I) \). If \( f \) is differentiable at \( x_0 \in I \), and if \( f'(x_0) \neq 0 \), then \( f^{-1} \) is differentiable at \( f(x_0) \) and

\[
(f^{-1})'(f(x_0)) = \frac{1}{f'(x_0)}.
\]

**Theorem 10.9 (Inverse Function Theorem, Theorem 16, page 358 of [1])** Let \( T : D \to \mathbb{R}^n \) be a transformation of class \( C^1 \) defined on an open set \( D \subset \mathbb{R}^n \) and suppose that

\[
det J_T(p) \neq 0 \text{ for all } p \in D.
\]

Suppose also that \( T \) is one-to-one on \( D \). Then the inverse \( T^{-1} \) (which exists and is defined on the open subset \( T(D) \subset \mathbb{R}^n \)) is of class \( C^1 \) on \( T(D) \) and

\[
J_{T^{-1}}(T(p)) = [J_T(p)]^{-1} \text{ for all } p \in D.
\]

**Proof:** Since \( T \) is of class \( C^1 \), by Theorem 9.9,

\[
T(p) - T(p_0) = J_T(p_0) \times (p - p_0)^t + R(p)
\]

where

\[
\lim_{p \to p_0} \frac{|R(p)|}{|p - p_0|} = 0.
\]

By assumption \( \det J_T(p_0) \neq 0 \) so \( J_T(p_0) \) is non-singular. Multiplying (63) (on the left) by \( [J_T(p_0)]^{-1} \), you get

\[
[J_T(p_0)]^{-1}(T(p) - T(p_0)) = (p - p_0)^t + [J_T(p_0)]^{-1}(R(p)).
\]

Let us now denote by \( q \) and \( q_0 \), the column vectors which are the images of \( p^t \) and \( p_0^t \) under \( T \); that is \( q = T(p^t) \) and \( q_0 = T(p_0^t) \), so that \( p^t = T^{-1}(q) \), \( p_0^t = T^{-1}(q_0) \). Then by (65),

\[
T^{-1}(q) - T^{-1}(q_0) = (p - p_0)^t = [J_T(p_0)]^{-1}(T(p) - T(p_0)) - [J_T(p_0)]^{-1}(R(p))\]
that is, (eliminating the middle person \((p - p_0)^t\)),

\[
T^{-1}(q) - T^{-1}(q_0) - [J_T(p_0)]^{-1}(T(p) - T(p_0)) = -[J_T(p_0)]^{-1}(R(p)).
\] (66)

If we can show that the right hand side of (66) satisfies

\[
\lim_{q \to q_0} \frac{|[J_T(p_0)]^{-1}(R(p))|}{|q - q_0|} = 0,
\] (67)

then (66) will say that (62) is true. So we need to prove (67).

First recall that by Lemma 9.17 there is a constant \(M\) such that \(|[J_T(p_0)]^{-1}(u)| \leq M|u|\) for all \(u \in \mathbb{R}^n\). Therefore,

\[
\frac{|[J_T(p_0)]^{-1}(R(p))|}{|q - q_0|} \leq M|R(p)|.
\] (68)

By (65), \((p - p_0)^t = [J_T(p_0)]^{-1}(T(p) - T(p_0)) - [J_T(p_0)]^{-1}(R(p))\) so

\[
|p - p_0| \leq M|q - q_0| + M|R(p)|,
\] (69)

and by (64),

\[
|R(p)| \leq \epsilon|p - p_0| \text{ for } |p - p_0| < \delta \quad (\delta \text{ depending on } \epsilon).
\] (70)

Therefore, (69) becomes

\[
|p - p_0| \leq M|q - q_0| + M\epsilon|p - p_0|,
\]

or,

\[
(1 - \epsilon M)|p - p_0| \leq M|q - q_0|,
\]

that is,

\[
|p - p_0| \leq \frac{M}{1 - \epsilon M}|q - q_0| \text{ for } |p - p_0| < \delta.
\] (71)

Taking reciprocals in (71) you get

\[
\frac{1}{|q - q_0|} \leq \frac{M}{1 - \epsilon M} \frac{1}{|p - p_0|} \text{ for } |p - p_0| < \delta.
\] (72)

Now by (68),(72), and (70), we have, for \(|p - p_0| < \delta\),

\[
\frac{|[J_T(p_0)]^{-1}(R(p))|}{|q - q_0|} \leq M|R(p)|\frac{M}{|p - p_0|(1 - \epsilon M)} \leq \frac{\epsilon M^2}{1 - \epsilon M}.
\]

The quantity

\[
\frac{\epsilon M^2}{1 - \epsilon M}
\]

is “just as good” as \(\epsilon\) (since it goes to zero as \(\epsilon\) does). Therefore (67) holds. Note that we have used the fact that \(T^{-1}\) is continuous (Theorem 10.7). That is, if \(q \to q_0\), then \(p^t = T^{-1}q \to T^{-1}q_0 = p_0^t\), so \(|R(p)|/|p - p_0| < \epsilon\) if \(|p - p_0| < \delta\).
We still need to prove that $T^{-1}$ is of class $C^1$. To see this, just notice that the matrix entries of $J_T(p)$ are continuous functions by assumption and therefore the entries of the inverse matrix $J_T(p)^{-1}$ are continuous functions (Why?). By (62) then, the entries of $J_{T^{-1}}(T(p))$ are continuous functions of $q = T(p)$. DONE

Assignment 43 [1, §7.6, page 361, #11, 14]

END OF LECTURES

COURSE SUMMARY (from Buck)

1.3 Schwarz inequality—Theorem 1

1.5 topology—open, closed, interior, boundary, closure, cluster point

1.6 sequences—characterization of closure: Theorem 5

1.8 compactness—Bolzano Weierstrass, Heine Borel, Theorem 24, 25, 26, 27.

2.2 continuity—sequential criteria, Theorem 1, 2; composition Theorem 5

2.3 uniform continuity—on compact sets, Theorem 6

2.4 extreme values—Theorem 10, 11, 13

2.6 extension—Theorem 24

3.3 gradient—$D \Rightarrow C$: Corollary (page 129), approximation: Theorem 8

3.4 baby chain rule—Theorem 14

3.5 little mean value theorem—Theorem 16

4.2 integration—of continuous, Theorem 1, properties Theorem 4

7.2 transformations—continuity, compactness Theorem 3, 4

7.3 linear transformation—uniform continuity of them, Theorem 8

7.4 coordinate free derivative—approximation Theorem 10, chain rule Theorem 11

7.5 inverse functions—automatic continuity of inverse Theorem 13, local invertibility
   Theorem 14, open mapping Theorem 15, inverse function Theorem 16

FINAL EXAMINATION: WEDNESDAY JUNE 16, 1993
   10:30 A.M.-12:30 P. M.

SPECIAL OFFICE HOURS OF B. RUSSO:
   MON JUNE 14, AND TUES JUNE 15 1:00-2:00

---

33 After the final examination, you should do [1, §7.6, page 361, #2, 3, 5, 7] for your own good
11 Week 11

11.1 Wednesday June 16, 1993—Final Examination

Use only one side of your page and only one problem per page. Do all five problems in Part I, and do 1 of the 2 problems in each of Part II and Part III. Total of 7 problems (100 points).

Part I—Do all problems in this part—total 52 points

Prob 1 (12 points) Determine all the cluster points of the following sets and decide whether the sets are open, closed, or neither. Use the box to record your answers.

(A) In R: $S_1 =$ all integers; $S_2 =$ the interval $(a, b]$; $S_3 =$ all numbers of the form $2^{-n} + 5^{-m}$ ($m, n = 1, 2, \ldots$); $S_4 =$ all numbers of the form $(-1)^n + 1/m$ ($m, n = 1, 2, \ldots$).

(B) In $R^2$: $S_5 =$ all $(x, y)$ with $x^2 + y^2 > 1$; $S_6 =$ all $(1/n, 1/m)$ ($m, n = 1, 2, \ldots$); $S_7 =$ all $(x, y)$ with $x > 0$.

<table>
<thead>
<tr>
<th></th>
<th>cluster points</th>
<th>open, closed, neither?</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_4$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_6$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_7$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prob 2 (10 points) Let $T : R^n \to R^m$ be a continuous transformation and suppose that $K$ is a compact subset of $R^n$. Prove or disprove: If $C$ is a closed subset of $K$, then $T(C)$ is a closed subset of $R^m$.

Prob 3 (10 points) Define a contraction to be a transformation $T$ with the property that there is a constant $c < 1$ such that for all $p, p'$ in the domain of $T$,

$$|T(p) - T(p')| < c|p - p'|.$$  

Prove or disprove: The transformation $T(x, y) = (3 + \frac{1}{2}x, \frac{3}{2} + \frac{1}{2}y)$ is a contraction.

Prob 4 (10 points) Let $D$ be a compact subset of $R^n$ and let $T : D \to R^m$ be a continuous transformation. Prove that there is a vector $p_0 \in D$ such that $|T(p_0)| = \sup\{|T(p)| : p \in D\}$.

Prob 5 (10 points) Let $T(p) = (f^1(p), \ldots, f^m(p))$ be a transformation from a subset $D \subset R^n$ with coordinate functions $f^1, \ldots, f^m$. Prove that $T$ is uniformly continuous on $D$ if and only if each coordinate function $f^i, 1 \leq i \leq m$ is uniformly continuous on $D$. 
Part II—Do problem 6 or 7 in this part—not both.

Prob 6 (24 points) Let \( \{f_n\}_{n=1}^{\infty} \) be a sequence of strictly positive continuous functions on \([0, 1]\), and suppose that the sequence converges uniformly on \([0, 1]\) to a function \( f \). Suppose that \( \int_0^1 f_n = 1/n \) for \( n = 1, 2, \ldots \)

(A) (8 points) Why is \( f \) integrable on \([0, 1]\).

(B) (8 points) What is the value of \( \int_0^1 f \)? Justify your answer.

(C) (8 points) Prove or disprove: \( f(x) = 0 \) for all \( x \in [0, 1] \).

Prob 7 (24 points) Let \( f \) be defined on a compact rectangle \( R \subset \mathbb{R}^2 \).

(A) (6 points) Explain why \( |f(p)| - |f(q)| \leq |f(p) - f(q)| \) for every \( p, q \in R \)

(B) (6 points) For a subset \( S \subset R \), explain why, for \( p, q \in S \)

\[
|f(p) - f(q)| \leq M(f, S) - m(f, S)
\]

Hint: consider the case \( f(p) \geq f(q) \) first. Recall that \( M(g, S) = \sup_{p \in S} g(p) \) and \( m(g, S) = \inf_{p \in S} g(p) \).

(C) (6 points) Prove that \( M(|f|, S) - m(|f|, S) \leq M(f, S) - m(f, S) \). Hint: For \( \epsilon > 0 \) pick points \( x_0, y_0 \) with \( |f(x_0)| > M(|f|, S) - \epsilon \) and \( |f(y_0)| < m(|f|, S) + \epsilon \)

(D) (6 points) Prove that if \( f \) is integrable on \( R \), then so is \( |f| \).

Part III—Do problem 8 or 9 in this part—not both.

Prob 8 (24 points) Let \( T: \mathbb{R}^2 \to \mathbb{R}^2 \) and \( S: \mathbb{R}^2 \to \mathbb{R}^2 \) be defined by \( T(x, y) = (x^2 + 3, y^2 + x - 1) \) and \( S(x, y) = (y - x^2, x^3) \). Find

(A) (6 points) \( T(1, 2) \).

(B) (6 points) \( S \circ T(1, 2) \).

(C) (6 points) \( J_T(1, 2) \).

(D) (6 points) \( J_{S \circ T}(1, 2) \).

Prob 9 (24 points) Given \( T(x, y) = (\cos x, e^y x^2) \), so \( T: \mathbb{R}^2 \to \mathbb{R}^2 \) is a transformation.

(A) (6 points) Calculate \( J_T(x, y) \).

(B) (6 points) For which points \( (x, y) \) does \( T'(x, y) \) exist?

(C) (6 points) Explain why \( T \) is one-to-one when restricted to some ball centered at \( (\frac{\pi}{2}, 0) \), and why \( T^{-1} \) is differentiable at \( T(\frac{\pi}{2}, 0) \).

(D) (6 points) Find \( (T^{-1})'(T(\frac{\pi}{2}, 0)) \). Hint: \( \begin{pmatrix} a & 0 \\ b & c \end{pmatrix}^{-1} = \begin{pmatrix} a^{-1} & 0 \\ -b(ac)^{-1} & c^{-1} \end{pmatrix} \).
11.2 Appendices

11.2.1 The implicit function theorem

In much of analysis, the linear functions are the easiest to work with\(^{34}\). Let \( F : \mathbb{R}^n \to \mathbb{R} \) be a linear function, that is, there are real numbers \( a_1, \ldots, a_n \) such that

\[
F(x_1, \ldots, x_n) = \sum_{j=1}^n a_j x_j.
\]

Note that for such a function, \( \frac{\partial F}{\partial x_k}(x_1, \ldots, x_n) = a_k \), and moreover, if \( a_k \neq 0 \), we can solve the equation \( F(x_1, \ldots, x_n) = 0 \) for \( x_k \) in terms of the other \( n-1 \) variables. Explicitly,

\[
x_k = -\sum_{j=1, j\neq k}^n \frac{a_j}{a_k} x_j.
\]

Thus we have seen that we can easily solve for one of the variables in terms of the others if the partial derivative with respect to that variable does not vanish. This is the idea behind the implicit function theorem for non-linear functions.

For a second example let \( F(x, y) = x^2 + y^2 - 1 \) for \( (x, y) \in \mathbb{R}^2 \) so that \( F : D \to \mathbb{R} \) where \( D = \mathbb{R}^2 \). Note that \( \frac{\partial F}{\partial y}(x, y) = 2y \).

Suppose that \( (x_0, y_0) \in \mathbb{R}^2 \) is such that \( F(x_0, y_0) = 0 \), that is, \( (x_0, y_0) \) is a point on the unit circle. We wish to find a function \( \phi \), defined in an interval \( (x_0 - r, x_0 + r) \), such that \( y = \phi(x) \) is a solution of the equation \( F(x, y) = 0 \) for every \( x \in (x_0 - r, x_0 + r) \), that is, \( x^2 + (\phi(x))^2 - 1 = 0 \) for every \( x \in (x_0 - r, x_0 + r) \), and \( \phi(x_0) = y_0 \). Moreover we want the function \( \phi \) to have a continuous derivative at every point of \( (x_0 - r, x_0 + r) \).

In this example, it is easy to know when such a function exists and it is also easy to find it. Obviously (draw a circle), we can take \( r = 1 - |x_0| \), and set \( \phi(x) = +\sqrt{1-x^2} \) for \( x \in (x_0 - r, x_0 + r) \). The only problem arises when \( |x_0| = 1 \), that is \( y_0 = 0 \), which is precisely where \( \frac{\partial F}{\partial y} \) vanishes. Another solution is obtained by taking \( \phi(x) = -\sqrt{1-x^2} \). Before we leave this example, let’s note that we can interchange the roles of the variables \( x \) and \( y \) and obtain a function \( x = \psi(y) \) satisfying, among other things \( (\psi(y))^2 + y^2 - 1 = 0 \).

Let’s now consider a third example, which is not so easy (correction: impossible) to solve with our bare hands. Let \( F(x, y) = x + 2y + x^2 y^5 - 8 \), for \( (x, y) \in \mathbb{R}^2 \). Note that \( F(2, 1) = 0 \). We wish to find a solution \( y = \phi(x) \) of the equation \( F(x, y) = 0 \) for all \( x \) in an interval of the form \( (2 - r, 2 + r) \), in such a way that \( \phi(2) = 1 \), and \( \phi \) has a continuous derivative on \( (2 - r, 2 + r) \). For this example, it is not clear that there will be a solution \( y \) of the equation \( x + 2y + x^2 y^5 - 8 = 0 \) for any \( x \) (this is a fifth degree equation in \( y \) for each fixed \( x \)). But we are greedy and want even more. We want a function \( \phi \) which systematically produces a solution \( \phi(x) \) to the equation for a given \( x \), and moreover, we want this function to be continuous, even differentiable, and furthermore, we want the derivative to be continuous.

Let’s return to our second example, that is, \( F(x, y) = x^2 + y^2 - 1 \) for \( (x, y) \in \mathbb{R}^2 \) so that \( F : D \to \mathbb{R} \) where \( D = \mathbb{R}^2 \). Of course \( F \) is a function. Let’s construct a

\(^{34}\)This is not necessarily the case for algebra
related transformation \( T_F : D \to \mathbb{R}^2 \) as follows: \( T_F(x, y) = (x, F(x, y)) \). Note that if we set \( G(x, y) = x \) then \( G \) and \( F \) are the coordinate functions of the transformation \( T_F \), that is \( T_F = (G, F) \). Hereafter, we’ll just write \( T \) instead of \( T_F \).

**Assignment 44** Show that, for \( F = x^2 + y^2 - 1 \), \( T = T_F \) is not one-to-one on \( D = \mathbb{R}^2 \) and \( T(\mathbb{R}^2) \) is not an open subset of \( \mathbb{R}^2 \).

Suppose again that \((x_0, y_0) \in \mathbb{R}^2\) is such that \( F(x_0, y_0) = x_0^2 + y_0^2 - 1 = 0 \), that is, \((x_0, y_0)\) is a point on the unit circle. Note that \( T(x_0, y_0) = (x_0, 0) \). Finally we construct the Jacobian matrix of \( T \):

\[
J_T(x, y) = \begin{pmatrix}
\frac{\partial G}{\partial x}(x, y) & \frac{\partial G}{\partial y}(x, y) \\
\frac{\partial F}{\partial x}(x, y) & \frac{\partial F}{\partial y}(x, y)
\end{pmatrix} = \begin{pmatrix}
1 & 0 \\
\frac{\partial F}{\partial x}(x, y) & \frac{\partial F}{\partial y}(x, y)
\end{pmatrix}.
\]

It follows that the Jacobian determinant is

\[
\det J_T(x, y) = \frac{\partial F}{\partial y}(x, y).
\]

Since we have just introduced most of the ideas in its proof, it seems appropriate now to state the implicit function theorem.

**Theorem 11.1 (Theorem 17, page 363 of [1], "downgraded" to two variables)** Let \( F : D \to \mathbb{R} \) be of class \( C^1 \) on an open set \( D \subset \mathbb{R}^2 \), let \((x_0, y_0) \in D\), and suppose that \( F(x_0, y_0) = 0 \) and \( \frac{\partial F}{\partial y}(x_0, y_0) \neq 0 \). Then there exists a \( r > 0 \) and a function \( \phi : (x_0 - r, x_0 + r) \to \mathbb{R} \) of class \( C^1 \) on \((x_0 - r, x_0 + r)\), such that \( \phi(x_0) = y_0 \) and \( F(x, \phi(x)) = 0 \) for all \( x \in (x_0 - r, x_0 + r) \).

Before going into the proof of Theorem 11.1, let’s reiterate exactly all that it says.

- There is (theoretically!) a function \( \phi \), such that for each \( x \) close enough to \( x_0 \), \( y = \phi(x) \) is a solution\(^{35}\) of the equation \( F(x, y) = 0 \)

- As a function of \( x \), \( \phi \) is continuous

- Actually, \( \phi \) is differentiable

- Actually, the derivative of \( \phi \) is a continuous function\(^{36}\)

- Question: Can we calculate \( \phi'(x) \) by implicit differentiation and the chain rule?\(^{37}\)

**Proof of Theorem 11.1:** Define a transformation \( T = (G, F) \) by setting \( G(x, y) = x \). Let \( p_0 \) denote \((x_0, y_0)\). Since \( J_T(p_0) \neq 0 \), by the "local invertibility theorem", \( T \) is locally one-to-one at \( p_0 \). That is, there is a ball \( B \) with center \( p_0 \) such that the restriction of \( T \) to this ball is one-to-one, so has an inverse transformation \( T^{-1} \):

\(^{35}\)This already says a lot! If you stop here you got a bargain.

\(^{36}\)This statement implies the previous two statements

\(^{37}\)Yes, but it is not entirely satisfactory because the answer is in terms of \( \phi(x) \)
$T(B) \to B$. Since $T$ is of class $C^1$, by making the radius of $B$ even smaller, we may assume that $J_T$ is not zero anywhere in this smaller ball. Thus, if we call this new ball $B'$, then $T$ is one-to-one on $B'$ with inverse $T^{-1}$ on $T(B')$, and by the “open mapping theorem”, $T(B')$ is an open set. Since $(x_0,0) = T(x_0,y_0) \in T(B')$, there is an open ball $B((x_0,0),r) \subset T(B')$. Let us write the inverse transformation $T^{-1}$ in terms of its coordinate functions, call them $g$ and $h$: $T^{-1} = (g,h)$. We have the relation

$$(x,y) = T^{-1} \circ T(x,y) = T^{-1}(T(x,y)) = T^{-1}(x,F(x,y)) = (g(x,F(x,y)), h(x,F(x,y)))$$

for all $(x,y) \in B'$. Therefore, comparing coordinates, for $(x,y) \in B'$,

$$x = g(x,F(x,y)) \text{ and } y = h(x,F(x,y)).$$

But we also have the relation

$$(u,v) = T \circ T^{-1}(u,v) = T(T^{-1}(u,v)) = T(g(u,v), h(u,v)) = (g(u,v), F(g(u,v), h(u,v)))$$

for all $(u,v) \in B((x_0,0),r)$. In particular, $u = g(u,v)$ and

$$v = F(g(u,v), h(u,v)) = F(u, h(u,v)).$$

Substitute for $(u,v)$, any point of the form $(x,0) \in B((x_0,0), r)$. From (73), we have

$$0 = F(x, h(x,0)) \text{ for all } |x - x_0| < r.$$  \(\text{Thus, if we define } \phi(x) = h(x,0) \text{ for } |x - x_0| < r, \text{ we have the desired function } \phi. \)

Note that by the chain rule, $\phi'(x) = \frac{\partial h}{\partial x}(x,0)$ so that $\phi$ is of class $C^1$ on $(x_0-r, x_0+r)$. This completes the proof.

We now state a version of the implicit function theorem in 3 variables. We refer to [1] for the proof, which is not significantly different from the above proof.

Draw a diagram (=graph) for the next theorem. If that seems difficult, draw a diagram for the previous theorem first.

**Theorem 11.2 (Theorem 17, page 363 of [1]—three variables)** Let $F : D \to \mathbb{R}$ be of class $C^1$ on an open set $D \subset \mathbb{R}^3$, let $(x_0,y_0,z_0) \in D$, and suppose that $F(x_0,y_0,z_0) = 0$ and $\frac{\partial F}{\partial x}(x_0,y_0,z_0) \neq 0$. Then there exists a $r > 0$ and a function $\phi : B((x_0,y_0),r) \to \mathbb{R}$ of class $C^1$ on $B((x_0,y_0),r)$, such that $\phi(x_0,y_0) = z_0$ and $F(x,y,\phi(x,y)) = 0$ for all $(x,y) \in B((x_0,y_0),r)$.

It is now easy to state (and prove) a general theorem of implicit function type in any number of variables. There are no new ideas needed to prove this theorem so we do not write the proof here.

---

38 What is the reason for this?
Theorem 11.3 Let $F : D \to \mathbb{R}$ be of class $C^1$ on an open set $D \subset \mathbb{R}^n$, let $(x_0^0, \ldots, x_n^0)$ be a point of $D$, and suppose that

$$F(x_1^0, x_2^0, \ldots, x_n^0) = 0$$

and for some $k$, $\frac{\partial F}{\partial x_k}(x_1^0, x_2^0, \ldots, x_n^0) \neq 0$.

Then there exists $r > 0$ and a function

$$\phi : B((x_1^0, \ldots, x_{k-1}^0, x_{k+1}^0, \ldots, x_n^0), r) \to \mathbb{R}$$

of class $C^1$ on $B((x_1^0, \ldots, x_{k-1}^0, x_{k+1}^0, \ldots, x_n^0), r) \subset \mathbb{R}^{n-1}$, such that

$$\phi(x_1^0, \ldots, x_{k-1}^0, x_{k+1}^0, \ldots, x_n^0) = x_k^0$$

and

$$F(x_1^0, \ldots, x_{k-1}^0, x_{k+1}^0, \ldots, x_n^0, \phi(x_1^0, \ldots, x_{k-1}^0, x_{k+1}^0, \ldots, x_n^0)) = 0$$

for all $(x_1^0, \ldots, x_{k-1}^0, x_{k+1}^0, \ldots, x_n^0) \in B((x_1^0, \ldots, x_{k-1}^0, x_{k+1}^0, \ldots, x_n^0), r)$.

If we introduce a little notation we can make the last theorem easier to read.

Let $F : D \to \mathbb{R}$ be of class $C^1$ on an open set $D \subset \mathbb{R}^n$, let $p_0 = (x_0^0, \ldots, x_n^0)$ be a point of $D$, and suppose that $F(p_0) = 0$ and $\frac{\partial F}{\partial x_k}(p_0) \neq 0$ for some $k$. Let $p_0^{(k)} = (x_0^0, \ldots, x_{k-1}^0, x_{k+1}^0, \ldots, x_n^0)$. Then there exists $r > 0$ and a function $\phi : B(p_0^{(k)}, r) \to \mathbb{R}$ of class $C^1$ on $B(p_0^{(k)}, r) \subset \mathbb{R}^{n-1}$, such that, with $p = (x_1^0, \ldots, x_n^0)$ and $p^{(k)} = (x_1^0, \ldots, x_{k-1}^0, x_{k+1}^0, \ldots, x_n^0)$, we have $\phi(p_0^{(k)}) = x_k^0$ and $F(x_1^0, \ldots, x_{k+1}^0, \phi(p^{(k)}), x_{k+1}^0, \ldots, x_n^0) = 0$ for all $p^{(k)} \in B(p_0^{(k)}, r)$.

There are versions of the implicit function theorem in which more than one of the independent variables $x_1, \ldots, x_n$ can be solved in terms of the remaining variables. The situation is described in [1, Theorem 18, page 364], and the discussion on page 366 of [1]. READ IT!

We now present some examples in the form of exercises.

Assignment 45 Let $F(x, y, z) = x^2 + y^2 + z^2 - 1$ and take a point $(x_0, y_0, z_0)$ on the unit sphere in $\mathbb{R}^3$: $x_0^2 + y_0^2 + z_0^2 = 1$, that is, $F(x_0, y_0, z_0) = 0$. “Prove” that\footnote{Don’t laugh, you need to assume that $x^2 + y^2 < 1$} $z = \phi(x, y) := \sqrt{1 - x^2 - y^2}$ satisfies $F(x, y, \phi(x, y)) = 0$. According to the implicit function theorem, we need $\frac{\partial F}{\partial z}(x_0, y_0, z_0) \neq 0$, that is, $2z_0 \neq 0$, so take, for example $p_0 = (1/\sqrt{2}, 0, 1/\sqrt{2})$. Now find $r > 0$ such that

$$(x - \frac{1}{\sqrt{2}})^2 + (y - 0)^2 < r \Rightarrow x^2 + y^2 < 1.$$
Assignment 46  Let $F(x,y,z) = x^2 + yz^5 - 3xyz + z$, take the point $(1,0,-1)$, and note that $F(1,0,-1) = 0$ and $\frac{\partial F}{\partial z}(1,0,-1) = 1 \neq 0$. Conclude that there exists $r > 0$ and a function $\phi(x,y)$ of class $C^1$ in the ball $|(x,y) - (1,0)| < r$

such that $F(x,y,\phi(x,y)) = 0$ for all $(x,y)$ with $(x-1)^2 + y^2 < r^2$, that is $x^2 + y[\phi(x,y)]^5 - (3xy - 1)\phi(x,y) = 0$.

Assignment 47  Let $F(x,y,z) = \sin xy + e^z - e$, take the point $(x_0,0,1)$, and note that $F(x_0,0,1) = 0$. Also

$$\frac{\partial F}{\partial x}(x_0,0,1) = 0, \quad \frac{\partial F}{\partial y}(x_0,0,1) = x_0, \quad \frac{\partial F}{\partial z}(x_0,0,1) = e.$$ 

What does the implicit function theorem say in this case? Can you solve for any of the three variables without the help of the implicit function theorem?

Assignment 48  Let $F(x,y,z) = (\sin x)e^y + (\cos y)e^{xz} + \sin z$, take the point $(0,\pi/2,\pi)$, and note that $F(0,\pi/2,\pi) = 0$. Also

$$\frac{\partial F}{\partial x}(0,\pi/2,\pi) = e^{\pi/2}, \quad \frac{\partial F}{\partial y}(0,\pi/2,\pi) = -1, \quad \frac{\partial F}{\partial z}(0,\pi/2,\pi) = -1.$$ 

By the implicit function theorem, you have $z = \phi(x,y)$ for $(x,y)$ close to $(0,\pi/2)$, as well as $x = \psi(y,z)$ for $(y,z)$ close to $(\pi/2,\pi)$, etc. Now let $S(x,y) = (x,y,\phi(x,y))$ and apply the chain rule to $F \circ S$ to derive

$$\frac{\partial \phi}{\partial x}(x,y) = -\frac{\frac{\partial F}{\partial x}(x,y,\phi(x,y))}{\frac{\partial F}{\partial z}(x,y,\phi(x,y))},$$

and

$$\frac{\partial \phi}{\partial y}(x,y) = -\frac{\frac{\partial F}{\partial y}(x,y,\phi(x,y))}{\frac{\partial F}{\partial z}(x,y,\phi(x,y))}.$$ 

Assignment 49  [1, §7.6,page 366,#1,2,5]$^{40}$

The following table summarizes a lot of information$^{41}$

---

$^{40}$After the final examination, you should do [1, §7.6,page 366,#3,6,8,9] for your own good

$^{41}$some of the references are to these notes!
11.2.2 Connectedness

To study the concept of connectedness, which is of paramount importance in most of analysis, for example, in complex function theory (Math 220ABC), I shall present an approach based on relatively open sets, and relatively closed sets. The definition in [1, Definition 2, page 34] is difficult to remember and not really intuitive (in other words, I don’t like it).

**Definition 11.4** \[42\] Let \( D \) be any subset of \( \mathbb{R}^n \). A subset \( A \) of \( D \) is said to be *relatively open with respect to \( D \)* if there is an open subset \( G \) of \( \mathbb{R}^n \) such that \( A = G \cap D \). Similarly, a subset \( B \) of \( D \) is said to be *relatively closed with respect to \( D \)* if there is a closed subset \( F \) of \( \mathbb{R}^n \) such that \( B = F \cap D \).

A subset \( D \) of \( \mathbb{R}^n \) is *connected* if it is not possible to write \( D \) as a disjoint union of two non-empty sets, each of which is relatively open with respect to \( D \). A set which is not connected is said to be *disconnected*.

Thus a set \( D \subset \mathbb{R}^n \) is connected if whenever \( D = A \cup B \), with \( A \cap B = \emptyset \) and \( A \) and \( B \) both relatively open with respect to \( D \), then either \( A = \emptyset \) or \( B = \emptyset \).

Note that if \( D = \mathbb{R}^n \), then being relatively open with respect to \( D \) is the same as being open. Similarly for closed.

More generally, we have

\[42\] The definition may seem a little bit abstract, and it is, but it is pretty easy to work with (I hope), see [1, page 77]
Assignment 50 Let $D$ be an open subset of $\mathbb{R}^n$. A subset $A$ of $D$ is relatively open with respect to $D$ if and only if $A$ is open.

Assignment 51 Let $D$ be any subset of $\mathbb{R}^n$. A subset $A$ of $D$ is relatively open with respect to $D$ if and only if $D \setminus A$ is relatively closed with respect to $D$.

Assignment 52 A subset $D$ of $\mathbb{R}^n$ is connected if and only if it is not possible to write $D$ as a disjoint union of two non-empty sets, each of which is relatively closed with respect to $D$.

Assignment 53 A subset $D$ of $\mathbb{R}^n$ is connected if and only if the only subsets of $D$ which are at the same time relatively open with respect to $D$ and relatively closed with respect to $D$ are $D$ and $\emptyset$.

Assignment 54 Show that the above definition of connected set is equivalent to the one given in [1, Definition 2, page 34].

We can consider other kinds of “connectedness-like” properties. We introduce one of them now. The other (pathwise connectedness) will appear shortly.

Definition 11.5 A set $S \subset \mathbb{R}^n$ is polygon connected if, given any two points $p, q \in S$, there exists a finite chain of line segments in $S$ which abut (that is, one starts where another ends) and form a path starting at $p$ and ending at $q$.

Theorem 11.6 (Theorem 2, page 35 of [1]) An open connected set $S \subset \mathbb{R}^n$ is polygon connected.

Proof: Let $q_1, q_2 \in S$ and define a set

$$A = \{p \in S : p \text{ can be joined to } q_1 \text{ by a finite polygon path lying in } S\}.$$ 

Let $B := S \setminus A$. Now show that $A$ and $B$ are open subsets of $\mathbb{R}^n$ and that $A$ is not empty. Therefore $B$ is empty and $S = A$. DONE

Here are some details:
To show that $A$ is open, let $p_0 \in A$ and choose a ball $B(p_0, \delta) \subset S$. Since $p_0 \in A$, $p_0$ can be joined to $q_1$ by a finite polygon path in $S$. But any other point in $B(p_0, \delta)$ can be joined to $p_0$ by a straight line. Thus any point of $B(p_0, \delta)$ can be joined to $q_1$ by a finite polygon path. This proves that $A$ is open.

Finally, to show that $B$ is open, let $p_0 \in B$ and choose a ball $B(p_0, \delta) \subset S$. We wish to prove that $B(p_0, \delta) \subset B$. Suppose not. Then there exists $q' \in B(p_0, \delta) \cap A$. Thus $q_1$ can be joined to $q'$ and $q'$ can be joined to $p_0$, implying that $q_1$ can be joined to $p_0$, contradiction.

To show that $A$ is not empty, note that some ball $B(q_1, r) \subset S$ and obviously $B(q_1, r) \subset A$.

Assignment 55 [1, §1.5, page 36, easy: #1, 2, 3, 4; harder: #15, 16, 17; gluttony: #18, 19]
We can express the continuity of a function in terms of relatively open sets. As with much of what follows the same holds for transformations, but we will stick to functions for simplicity.  

**Assignment 56** A function \( f : D \to \mathbb{R} \) is continuous on a set \( D \subset \mathbb{R}^n \) if and only if the inverse image of every open set in \( \mathbb{R} \) is a relatively open subset of \( D \), that is, if \( G \subset \mathbb{R} \) is an open set, then \( f^{-1}(G) := \{ p \in D : f(p) \in G \} \) is relatively open with respect to \( D \).

**Assignment 57** [1, §2.2, page 80, #5,10]

The following theorem is an immediate consequence of Assignment 56.

**Theorem 11.7 (Theorem 9, page 90 of [1])** Let \( f \) be continuous on a set \( D \subset \mathbb{R}^n \), and let \( c \in \mathbb{R} \). Then the set \( \mathbb{R} \) of all points \( p \in D \) for which \( f(p) > c \) is an open set, relative to \( D \), and the set \( G \) of points \( p \) where \( f(p) = c \) is a closed set, relative to \( D \).

The following theorem is important. It is called the intermediate value theorem. You are familiar with it from one variable calculus \((n = 1)\) when \( S \) is an interval.

**Theorem 11.8 (Theorem 14, page 93 of [1])** Let \( S \subset \mathbb{R}^n \) be a connected set and let \( f : S \to \mathbb{R} \) be a continuous function. If \( a, b \in f(S) \), then for each number \( c \) between \( a \) and \( b \), there is a point \( p \in S \) such that \( f(p) = c \).

**Proof:** Suppose the theorem is not true. Set \( U = \{ p \in S : f(p) > c \} \) and \( V = \{ p \in S : f(p) < c \} \). We have \( S = U \cup V \) and \( U \cap V = \emptyset \). Since \( f \) is continuous, both \( U \) and \( V \) are relatively open with respect to \( S \). One of these sets must contain a point \( p_1 \) with \( f(p_1) = a \) and the other must contain a point \( p_2 \) with \( f(p_2) = b \), that is, neither is empty, a contradiction. DONE

The following theorem is even more general (and more to my taste).

**Theorem 11.9 (Theorem 15, page 94 of [1])** Let \( S \subset \mathbb{R}^n \) be a connected set and let \( f : S \to \mathbb{R} \) be a continuous function. Then \( f(S) \) is a connected subset of \( \mathbb{R} \).

**Proof:** Suppose that \( f(D) = A \cup B \) where \( A \) and \( B \) are disjoint and each relatively closed with respect to \( f(D) \). We will show that one of \( A \) or \( B \) is empty. First check that \( f^{-1}(A) \) and \( f^{-1}(B) \) are disjoint and each relatively closed with respect to \( D \) and that \( D = f^{-1}(A) \cup f^{-1}(B) \). Since \( D \) is connected, either \( f^{-1}(A) \) or \( f^{-1}(B) \) is empty. This implies that one of \( A \) or \( B \) is empty

---

43. Rephrase all further results in this section for transformations; you will be glad you did
44. See also [1, Theorem 4, page 333]
45. What are all the connected subsets of \( \mathbb{R} \)? (See [2, §22])
46. The proof in [1, Theorem 15, page 94] is badly written!
47. You probably need the following exercise for this
Assignment 58 A subset $A$ of $D \subset \mathbb{R}^n$ is relatively closed with respect to $D$ if and only if, for every convergent sequence $p_k$ in $A$, with limit $p_0 = \lim_{k \to \infty} p_k \in D$, it follows that actually $p_0 \in A$. (Read this assignment very carefully!)

Definition 11.10 A set $S \subset \mathbb{R}^n$ is pathwise connected if, given any two points $p, q \in S$, there exists a continuous path $\gamma$ lying entirely in $S$ which joins $p$ and $q$ (that is, the path $\gamma$ starts at $p$ and ends at $q$). More precisely, there exists a continuous function $\gamma : [0, 1] \to \mathbb{R}^n$ such that $\gamma(0) = p$, $\gamma(1) = q$, and $\gamma([0, 1]) \subset S$.

Theorem 11.11 (Theorem 16, page 95 of [1]) A pathwise connected set $S \subset \mathbb{R}^n$ is connected.

Proof: Let $S = A \cup B$ where $A$ and $B$ are disjoint and each relatively closed with respect to $S$ and non-empty. We shall obtain a contradiction.

Take $p \in A$ and $q \in B$ and join $p$ to $q$ by a continuous path $\gamma$. Recall that $\gamma : I \to S$ where $I = [0, 1]$. Define $A_0 = A \cap \gamma(I)$ and $B_0 = B \cap \gamma(I)$. Of course, $\gamma(I) = A_0 \cup B_0$ and $\gamma(I)$ is connected since $\gamma$ is continuous and $I$ is connected. Note that it is a trivial verification that $A_0$ and $B_0$ are relatively closed with respect to $\gamma(I)$. Since $A_0 \cap B_0 = \emptyset$, we must have either $A_0 = \emptyset$ or $B_0 = \emptyset$. But $p \in A_0$ and $q \in B_0$, so we have arrived at the desired contradiction. DONE

The following is a striking application of connectedness. We refer to [1, Theorem 17, page 95] for the proof.

Theorem 11.12 (Theorem 17, page 95 of [1]) No continuous function can map the open unit square $S$ onto the interval $[0, 1]$ in a one-to-one fashion, although this is possible for discontinuous functions.

Assignment 59 [1, §2.4, page 96, easy: #1,2; harder: #8,9]

Here is another application of connectedness. It is the analog of a well known theorem in one variable (which is an application of the mean value theorem). Similarly, its proof uses a multi-variable mean value theorem.

Theorem 11.13 (Theorem 12, page 133 of [1]) If $f$ is of class $C^1$ on a connected open set $S \subset \mathbb{R}^n$, and if $\nabla f(p) = (0, \ldots, 0)$ for every $p \in S$, then $f$ is a constant.

Assignment 60 Write out a complete proof of Theorem 11.13

11.2.3 Integration on $\mathbb{R}^n$

In this section I want to use multi-index notation to give the definition and prove some fundamental results in the theory of Riemann integration over compact boxes in $\mathbb{R}^n$, where $n$ is arbitrary. We previously did this only for $n = 1$ and $n = 2$.

---

48Word processor’s dream
We shall consider the Riemann integral of a bounded function defined on a compact "n-box" in $\mathbb{R}^n$, for any $n \geq 1$. By an compact $n$-box we mean a product of $n$ intervals: $B = I_1 \times I_2 \times \cdots \times I_n \subset \mathbb{R}^n$, where $I_j$ is a closed and bounded interval in $\mathbb{R}$. Thus, an interval is a 1-box, a rectangle is a 2-box, and a box in $\mathbb{R}^3$ is a 3-box. We shall use the word "box" for an $n$-box and denote it by $R$. The "measure", or $n$-dimensional volume of a box $R = I_1 \times I_2 \times \cdots \times I_n$ is defined by

$$\mu(R) = \ell(I_1) \cdot \ell(I_2) \cdots \ell(I_n).$$

Let $R = I_1 \times I_2 \times \cdots \times I_n$ be a compact box in $\mathbb{R}^n$ and let $I_j = [a_j, b_j]$. A grid of $R$ is a set of the form $N = P_1 \times \cdots \times P_n$, where $P_j$ is a partition of $I_j$ for $1 \leq j \leq n$. We can write $P_j = \{a_j = x_{j0} < x_{j1} < \cdots < x_{jm_j} = b_j\}$ for $1 \leq j \leq n$, where $m_j$ is the number of subintervals of $I_j$ determined by $P_j$. Note that there are $m_1m_2 \cdots m_n$ subboxes $R_{i_1,i_2,\ldots,i_n} = I_{i_1} \times I_{i_2} \times \cdots \times I_{i_n}$ of $R$, where $I_{ij} = [x_{ij}, x_{ij+1}]$ for $1 \leq i_j \leq m_j$ and $1 \leq j \leq n$.

Moreover

$$R = \bigcup_{i_1=1}^{m_1} \cdots \bigcup_{i_n=1}^{m_n} R_{i_1,i_2,\ldots,i_n}. \quad (74)$$

Let $\mathcal{N}(R)$ denote the set of all grids of $R$.

Note that the measure of $R_{i_1,i_2,\ldots,i_n}$ is $\mu(R_{i_1,i_2,\ldots,i_n}) = \ell(I_{i_1}) \cdots \ell(I_{i_n})$. The mesh of the grid $N$ is denoted by $d(N)$ and is defined by

$$d(N) = \max_{1 \leq i_1 \leq m_1, \ldots, 1 \leq i_n \leq m_n} [(x_{i_11} - x_{i_11-1})^2 + \cdots + (x_{i_n1} - x_{i_n1-1})^2]^{1/2}. \quad (75)$$

Next we need a choice of points $C = \{p_{i_1,\ldots,i_n} : 1 \leq i_1 \leq m_1, \ldots, 1 \leq i_n \leq m_n\}$ such that $p_{i_1,\ldots,i_n} \in R_{i_1,i_2,\ldots,i_n}$ for $1 \leq i_1 \leq m_1, \ldots, 1 \leq i_n \leq m_n$\(^{49}\).

Now let $f$ be a function defined on $R$. A Riemann sum of $f$ with respect to a grid $N$ and a choice $C$ is defined by

$$S(f,N,C) = \sum_{i_1=1}^{m_1} \cdots \sum_{i_n=1}^{m_n} f(p_{i_1,\ldots,i_n})\mu(R_{i_1,\ldots,i_n}). \quad (76)$$

We can now state a fundamental theorem in the theory of Riemann integration.

**Theorem 11.14 (Theorem 1 on page 169 of [1])** **Arbitrary n** If $f$ is a continuous function on a closed bounded box $R \subset \mathbb{R}^n$, then there is a unique real number $v$ (depending on $f$ and $R$) with the following property:

For every $\epsilon > 0$, there exists $\delta > 0$ such that for all grids $N$ with $d(N) < \delta$ and for every choice $C$, we have $|S(f,N,C) - v| < \epsilon$.

**Assignment 61** Prove the uniqueness of $v$ in Theorem 11.14\(^{49}\)

\(^{49}\)If you are getting tired of the notation, not to worry, it will be improved soon; stay tuned
Definition 11.15 The number $v$ whose existence is guaranteed by Theorem 11.14 is denoted by $\int_R f$.

Before we begin the proof of Theorem 11.14 we are going to introduce some notation which will make our task more pleasant. This is “multi-index notation”, especially useful in the theories of partial differential equations and combinatorics, to name just two theories.

We let $\alpha = (i_1, \ldots, i_n)$ be an $n$-tuple of positive integers satisfying $1 \leq i_j \leq m_j$ for $1 \leq j \leq n$ and some fixed $n$-tuple of positive integers $(m_1, \ldots, m_n)$. Let

$$\Lambda = \{\alpha = (i_1, \ldots, i_n) : 1 \leq i_j \leq m_j \text{ for } 1 \leq j \leq n\}.$$ 

Then for instance we can write, in place of (74),

$$R = \bigcup_{\alpha \in \Lambda} R_{\alpha},$$

where $R_{\alpha} = R_{i_1, \ldots, i_n}$. Note that $\mu(R) = \sum_{\alpha \in \Lambda} \mu(R_{\alpha})$. Also, if we denote $(i_1 - 1, \ldots, i_n - 1)$ by $\alpha - 1$, then we can write (75) in the (much improved!) form

$$d(N) = \max_{\alpha \in \Lambda} |x_{\alpha} - x_{\alpha-1}|.$$ 

Finally, letting $p_{\alpha} = p_{i_1, \ldots, i_n}$, we can rewrite (76) in the form

$$S(f, N, C) = \sum_{\alpha \in \Lambda} f(p_{\alpha})\mu(R_{\alpha}).$$

Now we are ready to begin the proof of Theorem 11.14. We are given the “data” $f, R$ and we shall start with the statement of three lemmas. In the first two, it is only required that $f$ be a bounded function. This will be important for later when you need to study integration of discontinuous functions. Only in the third lemma will the continuity of $f$ be needed. Of course, the continuity of $f$ and the compactness of $R$ implies that $f$ is bounded, and moreover, perhaps more importantly, that $f$ is uniformly continuous.

Let’s get down to business. Suppose $f$ is a bounded function on the closed and bounded box $R$ and let $N$ be a grid of $R$. Keep in mind the construction given above and the new (multi-index) notation for it. Since $f$ is bounded on $R$, it is also bounded on each subbox $R_{\alpha}$ and we can define

$$M_{\alpha} = \sup_{p \in R_{\alpha}} f(p) \text{ and } m_{\alpha} = \inf_{p \in R_{\alpha}} f(p).$$

Notice that for continuous $f$, by the extreme values theorem, there will exist points $x_{\alpha}, y_{\alpha} \in R_{\alpha}$ such that $f(x_{\alpha}) = m_{\alpha}$ and $f(y_{\alpha}) = M_{\alpha}$. We shall use this fact in the third lemma below but for the first two lemmas, only the numbers $m_{\alpha}, M_{\alpha}$ are needed.

We now define the upper and lower Riemann sums corresponding to a grid, namely,

$$\overline{S}(N) := \sum_{\alpha} M_{\alpha}\mu(R_{\alpha}) \quad \text{(upper Riemann sum)}.$$
and
\[ S(N) := \sum_{\alpha} m_\alpha \mu(R_\alpha) \] (lower Riemann sum)

Since \( m_\alpha \leq f(p) \leq M_\alpha \) for every \( p \in R_\alpha \), and \( \mu(R_\alpha) > 0 \), for every grid \( N \) and every choice \( C \), we have
\[ S(N) \leq S(f, N, C) \leq \overline{S}(N) \] (77)

We are now ready to state the three lemmas.

**Lemma 11.16** Let \( f \) be a bounded function on a closed and bounded box \( R \subset \mathbb{R}^n \). Let \( N \) and \( \tilde{N} \) be grids of \( R \) and suppose \( N \subset \tilde{N} \). Then

(a) \[ S(N) \leq \tilde{S}(\tilde{N}) \]

(b) \[ \overline{S}(N) \geq \overline{S}(\tilde{N}) \]

**Lemma 11.17** Let \( f \) be a bounded function on a closed and bounded box \( R \subset \mathbb{R}^n \).

(a) The following two subsets\(^{50}\) of \( R \) are bounded sets:
\[ \{S(N) : N \in \mathcal{N}(R)\} \text{ and } \{\overline{S}(N) : N \in \mathcal{N}(R)\} \]

(b) Let
\[ s := \sup\{S(N) : N \in \mathcal{N}(R)\} \text{ and } S = \inf\{\overline{S}(N) : N \in \mathcal{N}(R)\} \]

Then
- \( s \leq S \)
- for every grid \( N \), \( S - s \leq \overline{S}(N) - S(N) \)

**Lemma 11.18** Let \( f \) be a continuous function on a closed and bounded box \( R \subset \mathbb{R}^n \). For every \( \epsilon > 0 \), there exists \( \delta > 0 \) such that
\[ \overline{S}(N) - S(N) < \epsilon \text{ for every grid } N \text{ with mesh } d(N) < \delta. \]

We shall prove these three lemmas, one after the other. But first, we use them to give a proof of Theorem 11.14.

**Proof of Theorem 11.14:** From Lemmas 11.17 and 11.18, \( 0 < S - s < \epsilon \) for every \( \epsilon > 0 \), that is, \( S - s = 0 \). Let \( v \) denote the common value of \( s \) and \( S \).

The following sequence of statements will complete the proof. Be sure you can supply the justification for each statement.

- \( S - \underline{S}(N) \leq \overline{S}(N) - \underline{S}(N) \) for every grid \( N \)
- \( \overline{S}(N) - s \leq \overline{S}(N) - S(N) \) for every grid \( N \)
- \( 0 < v - \underline{S}(N) < \epsilon \) if \( d(N) < \delta \)

\(^{50}\)recall that \( \mathcal{N}(R) \) denotes the set of all grids of \( R \)
• \(0 < \overline{S}(N) - v < \epsilon\) if \(d(N) < \delta\)
• \(v - S(f, N, C) \leq v - \overline{S}(N) < \epsilon\) if \(d(N) < \delta\) and \(C\) is any choice of points
• \(S(f, N, C) - v \leq \overline{S}(N) - v < \epsilon\) if \(d(N) < \delta\) and \(C\) is any choice of points
• \(|S(f, N, C) - v| < \epsilon\) if \(d(N) < \delta\) and \(C\) is any choice of points.

This completes the proof of Theorem 11.14.

**Definition 11.19** A bounded function on a closed and bounded box \(R \subset \mathbb{R}^n\) is **integrable on** \(R\) if it satisfies the condition of Theorem 11.14, that is, there is a unique real number \(v\) (depending on \(f\) and \(R\)) with the following property:

For every \(\epsilon > 0\), there exists \(\delta > 0\) such that for all grids \(N\) with \(d(N) < \delta\) and for every choice \(C\), we have \(|S(f, N, C) - v| < \epsilon\).

We can restate Theorem 11.14 as: every continuous function on a compact box is integrable on that box. The question arises: does the converse hold? The answer is no. A discontinuous function can be integrable. There are non-integrable functions, necessarily discontinuous. We proved this in the case \(n = 1\).

We now turn to the proofs of the three lemmas.

**Proof of Lemma 11.16:** There are two parts to this proof. First we prove the lemma under the assumption that it holds in the special case where \(\tilde{N}\) is obtained from \(N = P_1 \times \cdots \times P_n\) by adding a single point to one of the \(P_j\). Then we prove the special case.

**Step 1:** Assume that the lemma is true in the special case. Write

\[ N_0 = \tilde{N} \supset N_1 \supset N_2 \supset \cdots \supset N_s \supset N_{s+1} := N, \]

where \(N_k\) is obtained from \(N_{k+1}\) by adding a single point (\(0 \leq k \leq s\)).

By assumption, for \(0 \leq k \leq s\),

\[ \underline{S}(N_{k+1}) \leq \underline{S}(N_k) \leq \overline{S}(N_k) \leq \overline{S}(N_{k+1}). \]

Therefore,

\[ \underline{S}(N) = \underline{S}(N_{s+1}) \leq \underline{S}(N_s) \leq \underline{S}(N_{s-1}) \leq \cdots \leq \underline{S}(N_1) \leq \underline{S}(N_0) = \underline{S}(\tilde{N}) \]

\[ \leq \overline{S}(N_0) \leq \overline{S}(N_1) \leq \overline{S}(N_2) \leq \cdots \leq \overline{S}(N_{s-1}) \leq \overline{S}(N_s) \leq \overline{S}(N_{s+1}) = \overline{S}(N). \]

This completes the proof of step 1.

**Step 2:** We start with the following simple observation:

Let \(\phi : D \to \mathbb{R}\) be a bounded function on a set \(D \subset \mathbb{R}^n\) and suppose that \(A \subset D\). Then

\[ \sup_{p \in A} \phi(p) \leq \sup_{p \in D} \phi(p) \quad \text{and} \quad \inf_{p \in A} \phi(p) \geq \inf_{p \in D} \phi(p). \]
We now assume that $\tilde{N} = (P_1 \cup \{ u \}) \times \cdots \times P_n$ and define $i^*_1$ by $x_{i^*_1 - 1} < u < x_{i^*_1}$.

If we let $\beta = (i_2, \ldots, i_n)$ denote an appropriate $(n - 1)$-tuple of indices, we have $R_{i_1, \beta} = R'_{i_1, \beta} \cup R''_{i_1, \beta}$, where $R'_{i_1, \beta} = [x_{i^*_1 - 1}, u] \times R_\beta$ and $R''_{i_1, \beta} = [u, x_{i^*_1}] \times R_\beta$. Then

$$S(N) = \sum_{\alpha} m_\alpha \mu(R_\alpha) = \sum_\beta m_{i_1, \beta} \mu(R'_{i_1, \beta}) + \sum_{\alpha, i_1 \neq i^*_1} m_\alpha \mu(R_\alpha),$$

and

$$S(\tilde{N}) = \sum_\beta [m'_{i_1, \beta} \mu(R'_{i_1, \beta}) + m''_{i_1, \beta} \mu(R''_{i_1, \beta})] + \sum_{\alpha, i_1 \neq i^*_1} m_\alpha \mu(R_\alpha).$$

Thus we see that $S(N) \leq S(\tilde{N})$ if for each $\beta$, $m_{i_1, \beta} \leq m'_{i_1, \beta} \mu(R'_{i_1, \beta}) + m''_{i_1, \beta} \mu(R''_{i_1, \beta})$. This last statement is true since $\mu(R'_{i_1, \beta}) = \mu(R''_{i_1, \beta}) + \mu(R''_{i_1, \beta})$, and by virtue of the observation above, $m_{i_1, \beta} \leq m_{i_1, \beta}$, and $m_{i_1, \beta} \leq m_{i_1, \beta}$.

This completes the proof of (a) in case the new point $u$ occurs on the “$x_1$-axis”. You need a similar proof in case the new point occurs on the “$x_j$-axis” for some $j = 2, 3, \ldots, n$. Then you need to prove (b) in each of these $n$ cases. These proofs can be omitted since no new ideas are needed for them.

**Proof of Lemma 11.17:** Every grid $N$ contains the trivial grid $N_0 = \{a_1, b_1\} \times \cdots \times \{a_n, b_n\}$. Therefore, with $m := \inf \{f(p) : p \in R\}$ and $M := \sup \{f(p) : p \in R\}$, by Lemma 11.16

$$m \mu(R) = S(N_0) \leq S(N) \leq \overline{S}(N) = M \mu(R).$$

Thus the two sets $\{S(N) : N \in \mathcal{N}(R)\}$ and $\{\overline{S}(N) : N \in \mathcal{N}(R)\}$ are bounded, and the numbers $s$ and $\overline{s}$ exist.

For every $N$, since $\underline{S}(N) \leq s$ we have $-\underline{S}(N) \geq -s$. Add this inequality to the inequality with $S(N) \geq \overline{s}$ and you get $S(N) - \underline{S}(N) \geq \overline{s} - s$, which proves the second statement of the lemma.

To prove the first statement of the lemma, we shall make use of the following:

**Claim:** for any two grids $N_1, N_2$, $S(N_1) \leq \overline{S}(N_2)$.

Let’s assume this claim for the moment. Thinking of $N_2$ as fixed and $N_1$ as varying, and taking the supremum over $N_1$, you get, for every $N_2$,

$$s = \sup_{N_1 \in \mathcal{G}(R)} S(N_1) \leq \overline{S}(N_2).$$

Thus $s \leq \overline{S}(N)$ for every grid $N$ so taking the infimum over all grids $N$, you get

$$s \leq \inf_{N \in \mathcal{G}(R)} \overline{S}(N) = S,$$

which proves the first statement.

---

$^{51}m'_{i_1, \beta} = \inf \{f(p) : p \in R'_{i_1, \beta}\}, \text{etc}$
To prove the claim, we use Lemma 11.16 again. Given any two grids $N_1, N_2$, let $N = N_1 \cup N_2$. Then $N_1 \subset N$ and $N_2 \subset N$, so that by Lemma 11.16,

$$S(N_1) \leq S(N) \leq \overline{S}(N) \leq \overline{S}(N_2).$$

This completes the proof of Lemma 11.17.

**Proof of Lemma 11.18:** Since $f$ is continuous on the compact box $R$, it is uniformly continuous on $R$. For any $\epsilon > 0$, let $\delta = \delta(\epsilon/\mu(R), f, R)$, that is,

$$|f(p) - f(q)| < \epsilon/\mu(R) \text{ for all } p, q \in R \text{ with } |p - q| < \delta.$$

Let $N$ be any grid with $d(N) < \delta$. Since $f$ is continuous on each of the compact subboxes $R_\alpha$ of $R$, by the extreme values theorem, there exist points $p_\alpha, q_\alpha \in R_\alpha$ such that $M_\alpha = f(p_\alpha)$ and $m_\alpha = f(q_\alpha)$. Since $p_\alpha, q_\alpha \in R_\alpha$, $|p_\alpha - q_\alpha| < \delta$, and so $M_\alpha - m_\alpha = f(p_\alpha) - f(q_\alpha) < \epsilon/\mu(R)$. We now have

$$0 \leq \overline{S}(N) - S(N) = \sum_\alpha (M_\alpha - m_\alpha)\mu(R_\alpha) \leq \frac{\epsilon}{\mu(R)} \sum_\alpha \mu(R_\alpha) = [\epsilon/\mu(R)] \cdot \mu(R) = \epsilon.$$

This proves Lemma 11.18.

Having proved Lemmas 11.16, 11.17, 11.18, the proof of Theorem 11.14 is now complete.

The following theorem differs from Theorem 4 on page 176 of [1] in the following respects. In [1, Theorem 4, page 176], $f$ and $g$ are assumed continuous, and the integration is over (more or less) arbitrary sets in $\mathbb{R}^2$, not necessarily compact boxes in $\mathbb{R}^n$. Moreover, there is a fifth statement, which we stated and proved separately, but only (for convenience) for $n = 1$ (see Theorem 7.3 above).

**Theorem 11.20 (Theorem 4 on page 176 of [1], arbitrary $n$):** Let $f$ and $g$ be integrable functions on the compact box $R \subset \mathbb{R}^n$. Let $c$ be any real number. Then:

1. $f + g$ is integrable on $R$ and $\int_R(f + g) = \int_R f + \int_R g$
2. $cf$ is integrable on $R$ and $\int_R cf = c \int_R f$
3. If $f(p) \geq 0$ for all $p \in R$, then $\int_R f \geq 0$
4. $|f|$ is integrable on $R$ and $|\int_R f| \leq \int_R |f|$.

**Proof:** It is trivial to verify that for any grid $N$ and choice $C$,

$$S(f + g, N, C) = S(f, N, C) + S(g, N, C).$$

For $\epsilon > 0$, choose $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$|S(f, N, C) - \int_R f| < \frac{\epsilon}{2} \text{ for all } C \text{ and all } N \text{ with } d(N) < \delta_1,$$
and

\[ |S(g, N, C) - \int_R g| < \frac{\epsilon}{2} \text{ for all } C \text{ and all } N \text{ with } d(N) < \delta_2. \]

Then, with \( \delta = \min\{\delta_1, \delta_2\} \), we have, for all choices \( C \) and all \( N \) with \( d(N) < \delta \),

\[
|S(f + g, N, C) - \int_R f - \int_R g| = |S(f, N, C) + S(g, N, C) - \int_R f - \int_R g| \\
\leq |S(f, N, C) - \int_R f| + |S(g, N, C) - \int_R g| \\
< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\]

This proves the first statement and the proof of the second statement is similar. We refer to [1, page 177] for the proofs of the third and fourth statements\(^5\).

The following theorem contains two important properties of the integral. It is the analog, for \( n > 1 \) of Theorem 7.3. I will state the theorem carefully and leave the proof as an assignment.

**Theorem 11.21** Let \( R \) be a compact box in \( \mathbb{R}^n \) and let \( f : R \to \mathbb{R} \) be a bounded function.

(a) If \( f \) is integrable on \( R \) then \( f \) is integrable on any compact subbox \( R' \subset R \).

(b) If \( N \) is a grid of \( R \) with subboxes \( \{R_\alpha : \alpha \in \Lambda\} \) and if \( f \) is integrable on each \( R_\alpha \), then \( f \) is integrable on \( R \) and

\[
\int_R f = \sum_\alpha \int_{R_\alpha} f.
\]

**Assignment 62** Prove the following statements, which result in a proof of (a) of Theorem 11.21.

(A) Prove that \( f \) is integrable on \( R \) if for every \( \epsilon > 0 \) there is a \( \delta > 0 \) such that \( S(N) - \underline{S}(N) < \epsilon \) for all grids \( N \) with \( d(N) < \delta \). (Hint: The proof is contained in the proof of Theorem 11.14 given above.)

(B) (converse of (A)) Suppose that \( f \) is integrable on \( R \). Prove that for every \( \epsilon > 0 \) there is a \( \delta > 0 \) such that \( \overline{S}(N) - S(N) < \epsilon \) for all grids \( N \) with \( d(N) < \delta \). (See the hint for Assignment 14)

(C) Let \( R' = [a'_1, b'_1] \times \cdots \times [a'_n, b'_n] \subset R \) and let \( N \) be any grid of \( R \) which includes all of the points \( (c_1, \ldots, c_n) \), where \( c_j \in \{a'_j, b'_j\} \) for \( 1 \leq j \leq n \). Let \( N_2 := N \cap R' \) (which is a grid of \( R' \)). Prove that \( \overline{S}(N_2) - S(N_2) \leq \overline{S}(N) - S(N) \).

\(^5\)The integrability of \( |f| \) is not so easy because we are not assuming that \( f \) is continuous, just integrable—see a question on the final examination above.
(Note that, for example, \( \mathcal{S}(N_2) \) is an upper Riemann sum for \( f \) on the box \( R' \), and \( \mathcal{S}(N) \) is an upper Riemann sum for \( f \) on the box \( R \). You can use the notation \( \mathcal{S}(N_2) = \mathcal{S}(N_2, R') \) and \( \mathcal{S}(N) = \mathcal{S}(N, R) \) to remind yourself of these facts. I will use a similar notation in the proof of Theorem 11.21(b) below.)

(D) Use (A) and (C) to prove (a) of Theorem 11.21.

**Assignment 63** Let \( f \) be an integrable function on a compact box \( R \subset \mathbb{R}^n \).

(A) Prove that \( f \) has at least one point where it is continuous\(^{53}\).

(B) Prove that if \( f(p) > 0 \) for every \( p \in R \), then \( \int_R f > 0 \). (Note that by Theorem 11.20, \( \int_R f \geq 0 \); the point is to prove that \( \int_R f \neq 0 \).)

We now turn to the proof of (b) of Theorem 11.21.

**Proof of (b) of Theorem 11.21:** Let \( v_\alpha = \int_{R_\alpha} f \), which are assumed to exist. This means that for every \( \alpha \in \Lambda \) and every \( \epsilon > 0 \) there exist \( \delta_\alpha > 0 \) such that
\[
|S(f, N_\alpha, C_\alpha, R_\alpha) - v_\alpha| < \epsilon \text{ if } d(N_\alpha) < \delta_\alpha, \forall C_\alpha,
\]
(78)
Here, \( N_\alpha \) is a grid of \( R_\alpha \) and \( C_\alpha \) is a choice of points corresponding to the subboxes of \( N_\alpha \).

We have to prove that for every \( \epsilon > 0 \), there exists \( \delta > 0 \) such that
\[
|S(f, N, C, R) - \sum_{\alpha \in \Lambda} v_\alpha| < \epsilon \text{ if } d(N) < \delta, \forall C.
\]
(79)
Here, \( N \) is a grid of \( R \) and \( C \) is a choice of points corresponding to the subboxes of \( N \).

**Assignment 64** Complete the proof of (b) of Theorem 11.21.

In the next three assignments \( R \) denotes a compact box in \( \mathbb{R}^n \).

**Assignment 65** Let \( f \) be integrable on \( R \) and suppose that \( g \) is a function on \( R \) such that \( g(p) = f(p) \) except for finitely many \( p \) in \( R \). Show that \( g \) is integrable on \( R \) and that \( \int_R f = \int_R g \).

**Assignment 66** Prove that a bounded function \( f \) on \( R \subset \mathbb{R}^n \) is integrable on \( R \) if and only if for every \( \epsilon > 0 \), there exists a grid \( N \) such that \( \mathcal{S}(N) - \mathcal{S}(N) < \epsilon \).\(^{54}\)

**Assignment 67** (a) Let \( \{f_n\}_{n=1}^\infty \) be a sequence of continuous functions on \( R \), and suppose that \( f_n \to f \) uniformly on \( R \). Why is \( f \) integrable on \( R \)? Prove that
\[
\lim_{n \to \infty} \int_R f_n = \int_R f.
\]
(b) Let \( \{f_n\}_{n=1}^\infty \) be a sequence of integrable functions on \( R \), and suppose that \( f_n \to f \) uniformly on \( R \). Prove that \( f \) is integrable on \( R \) and that
\[
\lim_{n \to \infty} \int_R f_n = \int_R f.
\]

\(^{53}\)See the solution to Assignment 15

\(^{54}\)See Theorem 8.8 (formerly Assignment 18)
References
