Analysis in Several Variables
Math 140C—Fall 2005

Bernard Russo

November 26, 2005

Contents

1 Friday September 23—Inequalities of Young, Hölder, and Schwarz

1.1 Course Information

- Course: Mathematics 140C MWF 1:00–1:50 ET 204
- Instructor: Bernard Russo MSTB 263 Office Hours MW 2:30-3:30 or by appointment (a good time for short questions is right after class just outside the classroom)
- Discussion section: TuTh 1:00–1:50 MSTB 118
- Teaching Assistant: Mitchell Khong
- Homework: There will be approximately 35 assignments with at least one week notice before the due date.
- Grading:

First midterm	October 21 (Friday of week 4)	20 percent
Second midterm	November 18 (Friday of week 8)	20 percent
Final Exam	December 7 (Wednesday)	40 percent
Homework	approximately 35 assignments	20 percent

- Holidays: November 11, 24, and 25
- Text: R. C. Buck, Advanced Calculus
- Material to be Covered

Schwarz inequality Theorem 1, page 13 (1 lecture)

topology §1.5 pp 28–33: open, closed, boundary, interior, exterior, closure, neighborhood, cluster point (about 4 lectures)

compactness §1.8 pp 64–67: Heine-Borel and Bolzano-Weierstrass properties (Theorems 25,26,27, page 65) (about 4 lectures)

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) (about 4 lectures)

differentiation (of functions) §3.3: Implies continuity, characterization by approximation (Corollary, page 129 and Theorem 8, page 131) (about 4 lectures)

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) (about 4 lectures)

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) (about 5 lectures)

1.2 Young's, Hölder's, and Schwarz's inequalities

Theorem 1.1 (Young Inequality) *Let φ 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 $a, b \in [0,\infty)$,*

$$ab \le \Phi(a) + \Psi(b). \tag{1}$$

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

Assignment 1 (Due September 30) *Give a rigorous proof of Theorem 1.1. More precisely,*

Step 1 *First establish, for $c \in [0,\infty)$, the formula*

$$\int_0^c \varphi(u)\,du + \int_0^{\varphi(c)} \psi(v)\,dv = c\varphi(c). \tag{2}$$

Step 2 *Use (2) to prove (1).*

Step 3 *Prove the "moreover" statement.*

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

$$ab \le \frac{a^p}{p} + \frac{b^q}{q},$$

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

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Proof: Take $\varphi(u) = u^{p-1}$ in the theorem. □

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 $q := p/(p-1)$,*

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

Proof: Take $a = |x_j|/\|x\|_p$ and $b = |y_j|/\|y\|_p$ in the corollary, where $\|x\|_p$ denotes $\left(\sum_{j=1}^n |x_j|^p\right)^{1/p}$. □

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

$$\sum_{j=1}^n |x_j y_j| \le \left(\sum_{j=1}^n |x_j|^2\right)^{1/2} \left(\sum_{j=1}^n |y_j|^2\right)^{1/2}.$$

Proof: Take $p = 2$ in the theorem. $\square$

Assignment 2 (Due September 30)

1. Read sections 1.2,1.3,1.4 in Buck (The lectures will continue with section 1.5). Do not waste your time reading about the concepts *angle, orthogonal, hyperplane, normal vector, line, convexity*, which are discussed in section 1.3 of Buck. We have no immediate use for them. Thus, you may skip pages 15-18 and 21-27 for now.

2.
- Buck [§1.2 page 10 #5,10,23]
- Buck [§1.3 page 18 #2,3,6]
- Buck [§1.4 page 27 #3,15,16]

2 Monday, September 26—The triangle inequality and open sets

2.1 The triangle inequality

Section 1.1 of Buck In 1,2, or 3 dimensions you can use geometry, or geometric intuition. For dimensions $4, 5, 6 \ldots, \infty$ you need algebra and analysis as tools.

Section 1.2 of Buck The elements of $\mathbf{R}^n := \{p = (x_1, \ldots, x_n) : x_j \in \mathbf{R}, 1 \le j \le n\}$ may be considered as vectors (algebraic interpretation) or points (geometric interpretation). $\mathbf{R}$ is a field which has a nice order structure, in fact, almost all properties of $\mathbf{R}^n$ depend on those of $\mathbf{R}$, which in turn depend on the *least upper bound property* of $\mathbf{R}$. Unfortunately, no reasonable order can be defined on $\mathbf{R}^n$ if $n > 1$. Although we will not consider the vector space structure of $\mathbf{R}^n$ until later, we do need the notion of scalar product: for $p = (x_1, \ldots, x_n), q = (y_1, \ldots, y_n) \in \mathbf{R}^n$,

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

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

Section 1.3 of Buck The *length* of a vector $p = (x_1, \ldots, x_n) \in \mathbf{R}^n$ 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

$$p \cdot q \le |p||q|.$$

Here are two important consequences of the Schwarz inequality.

Corollary 2.1 (Triangle Inequality) *For any two vectors* p, q, $|p+q| \le |p| + |q|$

Proof: $|p+q|^2 = (p+q)\cdot(p+q) = p\cdot p + p\cdot q + q\cdot p + q\cdot q \le |p|^2 + 2|p||q| + |q|^2 = (|p|+|q|)^2$. $\square$

Corollary 2.2 (Backwards Triangle Inequality) $|p-q| \ge ||p| - |q||$

Proof: $|p| = |(p-q)+q| \le |p-q| + |q|$, which proves $|p-q| \ge |p| - |q|$. Now interchange p and q. $\square$

2.2 Open sets

A very important type of subset of $\mathbf{R}^n$ is a *ball*. An *open ball* is defined, for a given point $p \in \mathbf{R}^n$ and $r > 0$ by

$$B(p, r) := \{q \in \mathbf{R}^n : |p - q| < r\}.$$

The *center* of $B(p,r)$ is p and the *radius* is r. Today we want to prove (the two statements):

$$\text{Triangle inequality} \Rightarrow \left\{ \begin{array}{c} \text{open ball} \\ \text{is open set} \end{array} \right\} \Rightarrow \left\{ \begin{array}{c} \text{characterization} \\ \text{of interior} \end{array} \right\}$$

Definition 2.3 Let $S \subset \mathbf{R}^n$ and $q \in \mathbf{R}^n$. 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 $\operatorname{int} S$, that is

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

Finally, S is an *open* set if $S = \operatorname{int} S$.

Proposition 2.4 *Let* $p \in \mathbf{R}^n$ *and* $r > 0$. *Then the ball* $B(p,r)$ *is an open set.*

Proof: Let $x \in B(p,r)$ so that $|x-p| < r$. Choose $\delta := r - |x-p|$. Then the triangle inequality implies that $B(x, \delta) \subset B(p,r)$, showing that every point of $B(p,r)$ is an interior point of $B(p,r)$.

MIDTERM ALERT: It is very important that the 10 propositions (i)-(x) on page 32 of Buck be mastered before the first miderm. Here is one of them.

Proposition 2.5 ((vi) on p.32 of Buck) *Let* S *be any non-empty subset of* $\mathbf{R}^n$. *Then* $\operatorname{int} S$ *is the largest open subset of* S*; more precisely*

(a) $\operatorname{int} S$ *is an open set;*

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

Proof: The assertion of (a) is that $\operatorname{int} S = \operatorname{int}(\operatorname{int} S)$ and it suffices to show only that $\operatorname{int} S \subset \operatorname{int}(\operatorname{int} S)$. If $p \in \operatorname{int} S$, then there exists $\delta > 0$ with $B(p,\delta) \subset S$. Since the ball $B(p, \delta)$ is open, for each point $x \in B(p,\delta)$ there exists $\delta' > 0$ with $B(x, \delta') \subset B(p, \delta)$. However, since $B(p,\delta) \subset S$, we have $B(x,\delta') \subset S$ so that $x \in \operatorname{int} S$, and thus $B(p, \delta) \subset \operatorname{int} S$. By definition then, $p \in \operatorname{int}(\operatorname{int} S)$. This proves (a).

Let $T \subset S$ and let T be an open set. If $x \in T$, then there exists $\delta > 0$ with $B(x, \delta) \subset T$. Therefore $B(x, \delta) \subset S$ and so $T \subset \operatorname{int} S$, proving (b). $\square$

3 Wednesday September 28—More on open sets

Assignment 3 (Due October 7)

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

Remark 3.1 *Every open set in $\mathbf{R}^n$ is the union of (not necessarily disjoint) open balls.*

3.1 The structure of open sets in $\mathbf{R}^1$

Definition 3.2 A set of points in $\mathbf{R}^1$ is said to be *bounded* if it is a subset of a finite interval.

Definition 3.3 Let S be an open set in $\mathbf{R}^1$ and let (a, b) be an open interval which is contained in S, but whose endpoints are not in S. Then (a, b) is called a *component interval* of S.

Lemma 3.4 *Let S be a bounded open set in $\mathbf{R}^1$. Then*

(i) *Each point of S belongs to a component interval of S.*

(ii) *The component intervals of S form a countable (possibly finite) collection of disjoint sets whose union is S*

Proof: Assume $x \in S$. Let

$$a = \inf\{\text{left endpoints of all open intervals } I \text{ such that } x \in I \subset S\},$$

and let b be the sup of the right endpoints of these intervals. Then (a, b) contains all intervals I with $x \in I \subset S$, and in particular $x \in (a, b)$. From the way (a, b) was constructed, it follows that $(a, b) \subset S$ (See Remark 3.5 below) and $a \notin S,\ b \notin S$.

We have associated with each $x \in S$, at least one component interval I_x containing x. If two of these intervals I_x and I_y have a non-empty intersection, they must coincide since their endpoints do not belong to S. This proves (i).

It is now clear that S is the disjoint union of its component intervals. To prove (ii) it remains to show that they form a countable set. For this purpose, let $\{x_1, \ldots, x_n, \ldots\}$ be an enumeration of the rational numbers. Define a function F by means of the equation $F(I_x) = n$, if x_n is the rational number in I_x with the smallest index n. This function is one-to-one since $F(I_x) = F(I_y) = n$ would mean that $x_n \in I_x \cap I_y$, and therefore $I_x = I_y$. Therefore F establishes a one-to-one correspondence between the set of component intervals of S and a subset of $\mathbf{N}$. This proves (ii). □

Remark 3.5 Here is the proof that $(a, b) \subset S$. Let $y \in (a, b)$. There are two cases to consider: either $a < y < x < b$ or $a < x < y < b$. In the first case, by definition of inf, there is an interval $(a', b') \subset S$ such that $a < a' < y < x < b' \leq b$. In the second case, by definition of *sup*, there is an interval $(a'', b'') \subset S$ such that $a \leq a'' < x < y < b'' < b$. In either case, $y \in S$.

Theorem 3.6 *Every open set in* $\mathbf{R}^1$ *is the union of a countable collection of disjoint open intervals. (*This decomposition is unique but we shall ignore this fact—enough is enough!)

Proof: Let S be the given open set and let $S_n := S \cap (-n, n)$. Then $S = \cup_1^\infty S_n$ and each S_n is the union of a countable collection of disjoint open intervals. The existence follows from this. □

4 Friday September 30—Closed sets

Here are the first two propositions on page 32 of Buck. The proofs are written out in detail in Buck on pages 32–34.

(i) If A and B are open sets, then so are $A \cap B$ and $A \cup B$.

(ii) If $\{A_\alpha : \alpha \in I\}$ is an arbitrary family of open sets, then $\cup_{\alpha \in I} A_\alpha$ is an open set.

4.1 Closed sets

Definition 4.1 A subset S of $\mathbf{R}^n$ is said to be a *closed* set if its complement $\mathbf{R}^n \setminus S$ is an open set.

Remark 4.2 The second part of Assignment 3 shows that the set $\{q \in \mathbf{R}^n : |q-p| \leq r\}$ is a closed set for any $p \in \mathbf{R}^n$ and $r > 0$. Needless to say, we call such a set a "closed ball".

In order to facilitate the study of closed sets, we recall De Morgan's laws. If $\{A_\alpha : \alpha \in I\}$ is an arbitrary family of sets, then

$$\mathbf{R}^n \setminus \cup_{\alpha \in I} A_\alpha = \cap_{\alpha \in I} (\mathbf{R}^n \setminus A_\alpha)$$

and

$$\mathbf{R}^n \setminus \cap_{\alpha \in I} A_\alpha = \cup_{\alpha \in I} (\mathbf{R}^n \setminus A_\alpha).$$

Using De Morgan's laws we obtain immediately from (i) and (ii) the following propositions ((iii) and (iv)) on page 32 of Buck. From the definition of closed set, (v) is obvious, and (vi) has already been proved in Proposition 2.5 above.

(iii) If A and B are closed sets, then so are $A \cap B$ and $A \cup B$.

(iv) If $\{A_\alpha : \alpha \in I\}$ is an arbitrary family of closed sets, then $\cap_{\alpha \in I} A_\alpha$ is a closed set.

(v) A set is open if and only if its complement is closed.

4.2 Boundary and closure

Definition 4.3 Let $S \subset \mathbf{R}^n$ and let $p \in \mathbf{R}^n$. We say that p is a *boundary point* of S if every ball with center p meets both S and its complement $\mathbf{R}^n \setminus S$, that is, for every $\delta > 0$, $B(p, \delta) \cap S \neq \emptyset$ and $B(p, \delta) \cap (\mathbf{R}^n \setminus S) \neq \emptyset$. The *boundary* of S, denoted by $\operatorname{bdy} S$, is the set of all boundary points of S. The *closure* of S, notation $\overline{S}$ is defined to be $S \cup \operatorname{bdy} S$.

Here are some examples in $\mathbf{R}^1$:

S	$(a,b]$	$[a,b]$	(a,b)	$\{5+1/n\}_{n=1}^{\infty}$	$(a,b)\cap\mathbf{Q}$
bdy S	$\{a,b\}$	$\{a,b\}$	$\{a,b\}$	$\{5\}\cup\{5+1/n\}_{n=1}^{\infty}$	$[a,b]$
$\overline{S}$	$[a,b]$	$[a,b]$	$[a,b]$	$\{5\}\cup\{5+1/n\}_{n=1}^{\infty}$	$[a,b]$

The following proposition is the analog for closed sets of (vi) on page 32 of Buck. It will be proved in the next lecture.

Proposition 4.4 ((vii) on p.32 of Buck) *Let S be any subset of $\mathbf{R}^n$. Then $\overline{S}$ is the smallest closed set containing S.* (you know what this means.)

Assignment 4 (Due October 7) Prove the following assertions:

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

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

5 Monday October 3—More on closed sets

5.1 Proof of Proposition 4.4 ((vii) on page 32 of Buck)

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

Proof: We have to prove that the complement $\mathbf{R}^n\setminus\overline{S}$ is an open set, so let $q\in\mathbf{R}^n\setminus\overline{S}$. We must find a ball $B(q,\delta)\subset\mathbf{R}^n\setminus\overline{S}$. Since $q\notin\overline{S}=S\cup\operatorname{bdy}S$, $q\notin S$ and $q\notin\operatorname{bdy}S$. The latter implies that there is a $\delta>0$ such that either $B(q,\delta)\cap S=\emptyset$ or $B(q,\delta)\cap(\mathbf{R}^n\setminus S)=\emptyset$. The point q belongs to the latter set, so for sure $B(q,\delta)\cap S=\emptyset$, that is, $B(q,\delta)\subset\mathbf{R}^n\setminus S$. We complete the proof of Step 1 by showing that in fact $B(q,\delta)\subset\mathbf{R}^n\setminus\overline{S}$. If this were not true, there would be a point $q'\in B(q,\delta)\cap\overline{S}$. Since $B(q,\delta)\subset\mathbf{R}^n\setminus S$, in fact we have $q'\in B(q,\delta)\cap\operatorname{bdy}S$. 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, $B(q',\epsilon)\cap S\neq\emptyset$, 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 $\overline{S}=S\cup\operatorname{bdy}S$, and we are given that $S\subset F$, we have to show only that $\operatorname{bdy}S\subset F$. Suppose that $p\in\operatorname{bdy}S$ and $p\notin F$. If we arrive at some contradiction, we will be done. Since F is closed, $\mathbf{R}^n\setminus F$ is open, so there exists $\delta>0$ such that $B(p,\delta)\subset\mathbf{R}^n\setminus F$, that is, $B(p,\delta)\cap F=\emptyset$. By the definition of boundary point, $B(p,\delta)\cap S\neq\emptyset$. 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 Proposition 4.4. □

5.2 Cluster points

Definition 5.1 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 $\operatorname{cl}S$.

Remark 5.2 Although it is hard to believe, the point $p \in \mathbf{R}^n$ is a cluster point of $S \subset \mathbf{R}^n$ 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).

Proposition 5.3 ((ix) on p.32 of Buck) *Let S be any subset of $\mathbf{R}^n$. Then S is a closed set if and only if every cluster point of S belongs to S.*

Proof:
Step1: 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 $p \notin S$, then since $\mathbf{R}^n \setminus S$ is open, there exists a ball $B(p,\delta) \subset \mathbf{R}^n \setminus S$, that is, $B(p,\delta) \cap S = \emptyset$. But $B(p,\delta) \cap S$ is an infinite set, contradiction, so step 1 is proved.

Step 2: 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 $\mathbf{R}^n \setminus S$ is open. Let $p \in \mathbf{R}^n \setminus S$, 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, $B(p,\delta') \subset \mathbf{R}^n \setminus S$. Thus $\mathbf{R}^n \setminus S$ is open, and S is closed. Step 2 is proved.

Steps 1 and 2 constitute a proof of Proposition 5.3. □

Assignment 5 (Due October 14) [Buck §1.5 page 36 #6,10,11]

6 Wednesday October 5, 2005—Bolzano-Weierstrass and Heine-Borel properties

Definition 6.1 Let S be any subset of $\mathbf{R}^n$.

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 $S \subset \cup_{n=1}^{\infty} G_n$, then for some $N \geq 1$, $S \subset \cup_{n=1}^{N} G_n$.

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 Buck [Theorem 21,p. 62].
- $[0, 1]$ satisfies HB. This is [Theorem 24,p.65] in Buck..

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

7 Friday October 7,2005—Compact sets

Proposition 7.1 *Let S be any subset of $\mathbf{R}^n$. Then S satisfies BW if and only if it satisfies HB.*

Proof:

Step 1: BW$\Rightarrow$ HB.

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

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

For each n there is thus a point $p_n \in S$ such that $p_n \notin \{p_1, \ldots, p_{n-1}\}$ and

$$p_n \notin G_k \text{ for } 1 \le k \le n. \tag{3}$$

Because S satisfies BW, there is a cluster point, say p of the infinte sequence $T = \{p_1, p_2, \ldots,\}$ 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 $B(p, \delta) \subset G_{k_0}$. 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.

Step 2: HB$\Rightarrow$ BW.

Let $T = \{p_1, p_2, \ldots\} \subset S$ 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

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

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

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

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 7.1.

Definition 7.2 Let S be any subset of $\mathbf{R}^n$. We say S is *compact* if it satisfies BW or HB.

Assignment 6 (Due October 14) 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 in Buck as [§1.8 page 69 #1,2]

8 Monday October 10, 2005—Characterization of compact sets

8.1 Two remarks on the property HB

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 8.1 *Every compact set in* $\mathbf{R}^n$ *is closed.*

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

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

By HB, there exist finitely many points $q_1, \ldots, q_m \in S$ such that $S \subset \cup_{j=1}^m B(q_j, \delta_{q_j})$. Then $V := \cap_{j=1}^m B(p, \delta_{q_j})$ is an open set[1] 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 $\cup_{j=1}^m B(q_j, \delta_{q_j})$, and hence from S, that is, $V \subset \mathbf{R}^n \setminus S$. Thus S is closed. This completes the proof.

[1] because it is a finite intersection!! (this is the beauty of the Heine-Borel property)

Remark 8.2 *In the proof of Proposition 7.1 it wasn't shown yet that BW⇒ HB, only that BW implies that every countable cover of S by open sets could be reduced to a finite subcover. On the other hand, the proof of HB⇒ BW uses the full strength of the property HB, namely that an arbitrary (that is, possibly uncountable) open cover of S could be reduced to a finite subcover. So to complete the proof of Proposition 7.1, we need the following lemma, whose proof is left for you to think about.*

Lemma 8.3 *Every open cover of any set $S \subset \mathbf{R}^n$ can be reduced to a countable cover of S.*

8.2 Another characterization of compactness

We now come to a major theorem. [2]

Theorem 8.4 *Let S be any subset of $\mathbf{R}^n$. 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 $\mathbf{R}^n$). 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 $\mathbf{R}^n$). You will find the Heine-Borel theorem in Buck as Theorem 24 on page 65 (for $n = 1$) and Theorem 25 on page 65 of Buck 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 8.4.

Lemma 8.5 (Bolzano-Weierstrass theorem in R) *Every bounded sequence of real numbers has a convergent subsequence.*

Lemma 8.6 *Every subsequence of a convergent sequence of real numbers converges to the same limit as the sequence.*

Proof of Theorem 8.4:
Since S is bounded, there is a ball $B(0, M)$ with $S \subset B(0, M)$. Obviously

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

Now let $T = \{p_1, p_2, \ldots\} \subset S$ 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 8.5 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 8.5 again) and call this limit x_2. By Lemma 8.6, the first coordinates of T_2 also converge to the previous x_1.

[2] This makes today a very important day in your life

Continuing in this way, you obtain subsequences

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

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 $p = (x_1, \ldots, x_n) \in \mathbf{R}^n$.

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

$$s_k = (x_1^{(k)}, \ldots, x_n^{(k)}) \quad k = 1, 2, \ldots,$$

so that

$$\lim_{k\to\infty} x_j^{(k)} = x_j \quad 1 \le j \le n. \tag{4}$$

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

$$|p - s_k| < \delta \text{ for infinitely many } k.$$

By (4), there exist N_j $(1 \le j \le n)$ such that

$$|x_j - x_j^{(k)}| < \delta/\sqrt{n} \text{ for } k \ge N_j.$$

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

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

This completes the proof of Theorem 8.4. □

9 Wednesday October 12, 2005—More on boundary and closed sets (Second Midterm Alert)

Proposition 9.1 (Part of (viii) on page 32 of Buck) *For any subset S of $\mathbf{R}^n$, its boundary bdy S is a closed set.*

Proof: Just note that for any set S, we have the decomposition[3]

$$\mathbf{R}^n = \operatorname{int} S \cup \operatorname{bdy} S \cup \operatorname{int}(\mathbf{R}^n \setminus S)$$

of Euclidean space $\mathbf{R}^n$ into three mutually disjoint subsets. It follows that $\operatorname{bdy} S = \mathbf{R}^n \setminus (\operatorname{int} S \cup \operatorname{int}(\mathbf{R}^n \setminus S))$ is the complement of an open set. □

Note that $\operatorname{bdy} S = \operatorname{bdy} \mathbf{R}^n \setminus S$, for any set $S \subset \mathbf{R}^n$.

[3]Be sure to check this carefully

MIDTERM ALERT NUMBER 2

The first midterm, (WHICH IS CLOSED BOOK AND NOTES!) will take place on Friday October 21 and will cover sections 1.5 (pages 28–33 only) and 1.8 of Buck (pages 64–65 only). Of particular interest are the 10 statements (Propositions) on page 32. You should understand each step in the proofs of these propositions.

Assignment 7 (Due October 21—no penalty for turning it in on October 24, so you can write it up elegantly)

- Buck page 36, #1,3,4,11 (any two of these four)
- Buck page 36, #7,8,12 (any one of these)
- Buck page 36, #9, Buck page 69, #3 (both of these) (For #3, see the hint at the end of the book)

You will of course be responsible for all of these problems on the midterm.

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 may not have time to study the important topic of connectedness in this course[4]

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

- BW$\Rightarrow$HB
- HB$\Rightarrow$BW
- BW$\Rightarrow$ closed
- HB$\Rightarrow$ closed
- BW$\Rightarrow$ bounded
- HB$\Rightarrow$ bounded
- closed and bounded $\Rightarrow$BW
- closed and bounded $\Rightarrow$HB[5]

You may ignore Theorems 27,28,29,30 on pages 65–69 of Buck. We shall not discuss them[6]. 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 September 30 and October 7 and the homework you will turn in on October 14 and October 21 (or October 24).

[4] You can do this on your own—I will provide notes later

[5] we did not discuss this one—this is included in the proof of Theorem 25 in [Buck, p.67] Of course the result follows from the preceding fact since BW$\Rightarrow$ HB

[6] however, you are in a good position to understand them

10 Friday October 14, 2005—Even More on boundary and closed sets; proof of Lemma 8.3

10.1 Two remarks on closed sets and boundary

Proposition 10.1 ((iii) and (iv) on p.32 of Buck)

(a) *If A and B are closed subset of $\mathbf{R}^n$, 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.*

(c) *If $\{A_\alpha : \alpha \in \Lambda\}$ is a family of closed sets, then $\cap_{\alpha \in \Lambda}$ is closed.*

First proof: use De Morgan's law:

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

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, $p \in S$.

The same proofs work for (c). □

Proposition 10.2 (Another part of (viii) on p.32 of Buck) *For any subset S of $\mathbf{R}^n$,*

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

Proof:

$$\begin{aligned} \overline{S} \cap \overline{(\mathbf{R}^n \setminus S)} &= (S \cup \mathrm{bdy}\, S) \cap ((\mathbf{R}^n \setminus S) \cup \mathrm{bdy}\, (\mathbf{R}^n \setminus S)) \\ &= (S \cup \mathrm{bdy}\, S) \cap ((\mathbf{R}^n \setminus S) \cup \mathrm{bdy}\, S) \\ &= \mathrm{bdy}\, S. \end{aligned}$$

10.2 Proof of Lemma 8.3

The lemma states: Every open cover of *any* set $S \subset \mathbf{R}^n$ can be reduced to a countable cover of S.

Proof: Let S be covered by a family $\mathcal{G}$ of open sets. For each $p \in S$ choose a set $G_p \in \mathcal{G}$ containing p. Since G_p is open, choose an open ball $B(p, \delta_p) \subset G_p$. Since $\mathbf{Q}$ is dense in $\mathbf{R}$, we can find a rational number $r_p \in (0, \delta_p)$, hence $p \in B(p, r_p) \subset G_p$. Again, since $\mathbf{Q}$ is dense in $\mathbf{R}$, we can find a vector q_p with rational coordinates such that $q_p \in B(p, r_p/2)$. By the triangle inequality, $B(q_p, r_p/2) \subset B(p, r_p)$ (Check this!), so for each $p \in S$, we have $p \in B(q_p, r_p/2) \subset G_p$. The collection $\{B(q_p, r_p/2) : p \in S\}$ is countable, so we can enumerate it as $\{B(q_{p_j}, r_{p_j}/2)\}_{j=1}^{\infty}$, where $\{p_j\}$ is a sequence of points in S. For each $j = 1, 2, \ldots$ pick the corresponding $G_{p_j} \in \mathcal{G}$. Then $S \subset \cup_{j=1}^{\infty} G_{p_j}$, proving the lemma. □

11 Monday October 17, 2005—Continuous functions

11.1 Overview

Here is a preview of our next topic: **continuous functions**. There are only two main 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 Buck[7]: The continuous real valued image of a compact subset of $\mathbf{R}^n$ is a compact subset of $\mathbf{R}$.
- Theorem 6 on page 84 of Buck: A continuous real valued function on a compact subset of $\mathbf{R}^n$ 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 $f(\alpha) \le f(x) \le f(\beta)$ for every $x \in [a, b]$. (This is stated for functions defined on compact subsets of $\mathbf{R}^n$ as Theorem 10 on page 90 and Theorem 11 on page 91 of Buck)
- A continuous real valued function on a closed and bounded interval in $\mathbf{R}$ is uniformly continuous on that interval.

Here is a description of the first five theorems of Chapter 2 of Buck

Theorems 1,2 page 73-74 These 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 right now.

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 Buck to give a proof of the extreme value theorem ([Theorem 11,page 91] of Buck), 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.

[7] Do not read the proof of Theorem 13 in Buck, we will present a better one

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

In [Buck, Section 2.4] Theorems 10 and 11 follow easily from Theorem 13, as we will 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 may have to skip them now.

11.2 Continuous functions—continuous image of a compact set

Definition 11.1 Let $f : D \to \mathbf{R}$ be a function, where D is any subset of $\mathbf{R}^n$, and let $p_0 \in D$. We say that f is *continuous at* p_0 if

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

such that[8]

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

It is important to realize that this lengthy definition can be put in the compact [9] form

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

Here, we are using the notation

$$f(A) := \{f(p) : p \in A\} \text{ if } A \subset D.$$

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 gives a "global" definition of continuity.

Assignment 8 (Due October 28) [Buck, §2.2 page 80 #1 or 2,3 or 4,7 or 8,12 or 13,14 or 17] You are to hand in 5 problems, one from each of these 5 pairs. You will of course be responsible for all of the problems.

Theorem 11.2 *The continuous image of a compact set is compact. In other words, if $f : D \to \mathbf{R}$ is a continuous function on D, and D is a compact subset of $\mathbf{R}^n$, then $f(D)$ is a compact subset of $\mathbf{R}$.*

Proof: We choose[10] to show that $f(D)$ satisfies the HB property. By Lemma 8.3, we only need to deal with *countable* open covers. We shall use the fact that D satisfies the HB property (for *arbitrary* covers!).

[8] δ depends in general on p_0 as well as on ϵ

[9] no pun intended

[10] how many choices are there?

Let

$$f(D) \subset \cup_{k=1}^{\infty} G_k$$

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]:

$$D \subset \cup_{p \in D} B(p, \delta_p).$$

Since D is compact, the HB property tells us there are a finite number of points $p_1, \ldots, p_m$ say, such that

$$D \subset \cup_{j=1}^{m} B(p_j, \delta_{p_j}).$$

It follows that $D = \cup_{j=1}^{m} [B(p_j, \delta_{p_j}) \cap D]$, and therefore that

$$f(D) = \cup_{j=1}^{m} f[B(p_j, \delta_{p_j}) \cap D] \subset \cup_{j=1}^{m} B(f(p_j), \epsilon_{p_j}) \subset \cup_{j=1}^{m} G_{p_j}.$$

We have reduced the given (countable) cover to a finite subcover, so the proof is complete. □

An alternate proof would show that if S satisfies BW, then $f(S)$ satisfied BW, as follows. Let $\{\alpha_n\}_{n=1}^{\infty}$ be an infinite sequence in $f(S)$, which we may assume without loss of generality, consists of distinct points. For each n, choose a point $p_n \in S$ such that $f(p_n) = \alpha_n$. Since f is a function (well-defined!), $\{p_n\}_{n=1}^{\infty}$ is an infinite sequence in S so there exists a vector $p \in S$ which is a cluster point of $\{p_n\}_{n=1}^{\infty}$. Now verify that $f(p)$ is a cluster point of $\{\alpha_n\}_{n=1}^{\infty}$ (details omitted).

Assignment 9 (Due November 4) [Buck, §2.3 page 88 #1,3 or 4,5 or 6,7]

Remark 11.3 Whenever a set in $\mathbf{R}^n$ is defined by inequalities (or equalities) involving continuous functions, the set is open if all inequalities are strict ($>$ or $<$), and closed if all inequalities are not strict ($\leq$ or $\geq$ or $=$). Also, the boundary is obtained by changing one or more of the inequalites to $=$. As an example, here is a proof of the fact that the set $S = \{(x, y, z) \in \mathbf{R}^3 : xy > z\}$ is open in $\mathbf{R}^3$ (Problem 3(c) on page 37 of Buck).

Proof: Let $p_0 = (x_0, y_0, z_0) \in S$. We must find $\delta > 0$ such that $|p - p_0| < \delta$ implies $p \in S$. For any δ we note that if $p = (x, y, z)$ and $|p - p_0| < \delta$, then $(x - x_0)^2 \leq |p - p_0|^2 < \delta^2$ so that $|x - x_0| < \delta$ and similarly $|y = y_0| < \delta$ and $|z - z_0| < \delta$. Rewriting these last three inequalities as $x_0 - \delta < x < x_0 + \delta$, $y_0 - \delta < y < y_0 + \delta$, and $z_0 - \delta < z < z_0 + \delta$ implies $xy - z > (x_0 - \delta)(y_0 - \delta) - (z_0 + \delta) = x_0 y_0 - \delta y_0 - \delta x_0 + \delta^2 - z_0 - \delta = x_0 y_0 - z_0 + \delta(\delta - y_0 - x_0 - 1)$, which is strictly positive for sufficiently small δ. This proves that $B(p_0, \delta) \subset S$ for some $\delta > 0$. □

[11] the redundant cover!

12 Wednesday October 19, 2005—Continuity in terms of sequences

12.1 Limits of sequences of points in $\mathbf{R}^n$

Definition 12.1 Let $\{p_k\}_{k=1}^{\infty} \subset \mathbf{R}^n$ be a subset indexed by the natural numbers, and let $p \in \mathbf{R}^n$. 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 $k \to \infty$, or just plain $p_k \to p$.

Introduce coordinates of the points p_k and p:

$$p = (x_1, \dots, x_n) \text{ and } p_k = (x_1^{(k)}, \dots, x_n^{(k)}).$$

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 12.2 (Theorem 7 on page 42 of Buck) *Let $\{p_k\}_{k=1}^{\infty} \subset \mathbf{R}^n$ be a sequence, and let $p \in \mathbf{R}^n$. Then*

$$\lim_{k\to\infty} p_k = p,$$

if and only if

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

Theorem 12.3 (Theorem 3 on page 40 of Buck) *A convergent sequence in $\mathbf{R}^n$ 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

$$M = \max\{1 + |p|, |p_1|, \dots, |p_N|\},$$

that is, the sequence is bounded. □

Theorem 12.3 raises the following question. Does the set of points of a convergent sequence constitute a compact set, that is, is it closed. The answer is easily seen to be no. However, an illuminating informal exercise would be to prove that the set consisting of the points of convergent sequence together with its limit is a compact set. This exercise becomes even more instructive if you proved it in three ways, using successively, BW, HB, and CB (closed and bounded).

12.2 Continuity and limits of sequences

Theorem 12.4 (Theorem 1 on page 73 of Buck) *Let $f : D \to \mathbf{R}$, where $D \subset \mathbf{R}^n$, 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

$$f[D \cap B(p_0, \delta)] \subset B(f(p_0), \epsilon). \tag{5}$$

Since $p_k \to p_0$, and since $\delta > 0$, there exists N such that

$$p_k \in B(p_0, \delta) \text{ for } k > N. \tag{6}$$

Putting together (5) and (6) results in $f(p_k) \in B(f(p_0), \epsilon)$ for $k > N$. □

Remark 12.5
- Theorem 2 on page 74 of Buck is an important converse to Theorem 12.4. I suggest you read this theorem as we will not cover it in lecture.
- At this point you are in a position to give another proof of Theorem 11.2 above using the property BW (at both ends). We did this in class but I strongly suggest that you do this again for yourself as an informal exercise. The following lemma, which we shall use in the extreme value theorem (Theorem 12.8 below) may be helpful in that informal exercise.

Lemma 12.6 *For any subset $S \subset \mathbf{R}^n$, 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 $p_k \notin \{p_1, \ldots, p_{k-1}\}$. 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. □

Theorem 12.7 (Theorem 10 on page 90 of Buck) *A continuous function on a compact set is bounded. That is, if $f : D \to \mathbf{R}$ is continuous on $D \subset \mathbf{R}^n$ and D is compact, then f is a bounded function on D.*

Proof: This is now trivial, since by Theorem 11.2, $f(D)$ is compact, hence bounded. (Note that Theorem 11.2 does not depend on Theorem 12.7, so it is OK to use it in the proof).

Theorem 12.8 (Theorem 11 on page 91 of Buck, Extreme values Theorem) *A continuous function f on a compact set $D \subset \mathbf{R}^n$ assumes its maximum and its minimum at some points of D.*

Proof: By Theorem 12.7, f is bounded, that is $f(D)$ is a bounded subset of $\mathbf{R}$. Let

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

so that $\beta \in \mathbf{R}$. 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 12.6, 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. □

Assignment 10 (Due October 28) [Buck, §1.6 page 54 #1 or 2,3 or 4,31 or 33,32 or 35]

13 Friday October 21, 2005—First Midterm

Do all problems. However, there is a choice in one of them, number 8

Problem 1 (12 points) *Prove rigorously that the set $\mathbf{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ of integers is a closed subset of $\mathbf{R}^1$. Is it a closed subset of $\mathbf{R}^2$? (Yes or no, no proof required for this part of the question). Is*

$$S := \{(m, k) : m \in \mathbf{N}, k \in \mathbf{Z}\}$$

a closed subset of $\mathbf{R}^2$? (Yes or no, no proof required).

Problem 2 (12 points) *Find $bdy\,S$, $int\,S$, and all cluster points of S if*

$$S = \{(x, y) \in \mathbf{R}^2 : 1 \leq x^2 + y^2 < 2\} \cup \{(x, 0) : 0 < x \leq 1/2\}$$

Just write down your answer, no proof is required.

Problem 3 (5 points) *Prove or disprove: For every $S \subset \mathbf{R}^n$, $S \setminus int\,S = bdy\,S$.*

Problem 4 (25 points) *Let A be a bounded subset and B a closed subset of $\mathbf{R}^n$ 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) *$A \cap (\mathbf{R}^n \setminus B)$ is bounded*

Problem 5 (12 points) *Prove that a compact set is bounded. You may use BW or HB.*

Problem 6 (12 points) *Let S be a closed subset of $\mathbf{R}^n$, that is $\mathbf{R}^n \setminus S$ is an open set. Prove*

(A) *$bdy\, S \subset S$.*

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

Problem 7 (12 points) *Let A be any subset of a compact set S.*

(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 $\overline{A}$, the closure of A, is a compact set.*

Problem 8 (10 points) *Let S be an arbitrary subset of $\mathbf{R}^n$. Do only one of (A) or (B), not both.*

(A) *Let $p \in \mathbf{R}^n$ 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 $cl\, S$ be the set of cluster points of S. Prove that $cl\, S$ is a closed set. Hint: Use the fact that open balls are open sets to show that $cl(cl\, S) \subset cl\, S$*

14 Monday October 24, 2005—More on closure; Uniform continuity

14.1 A discussion of closed sets and closure

A closed set was originally defined to be a set whose complement is an open set and the closure of a set was originally defined to be the union of the set and its boundary. These definitions are not always workable so it is desirable to note that the following five statements are all equivalent to a set S being closed and can therefore serve as the definition of closed set. (The last one has not been discussed before and is proved in the next subsection. I stated it as an equality in class; however, it is also correct as stated here.)

- $\mathbf{R}^n - S$ is an open set
- $S = \overline{S}$
- $\operatorname{cl} S \subset S$
- $\operatorname{bdy} S \subset S$ (I failed to mention this one in class!)
- $\{\lim_k p_k : \{p_k\}_{k=1}^{\infty} \subset S, \text{ the limit exists}\} \subset S$

Besides being defined as the union of the set and its boundary points, the closure of a set has also been shown to be equivalent to several other statements, listed below. (The last one is proved in the next subsection.)

- $\overline{S} = S \cup \operatorname{bdy} S$
- $\overline{S}$ is the smallest closed set containing S
- $\overline{S}$ is the intersection of all closed sets containing S
- $\overline{S} = \operatorname{int} S \cup \operatorname{bdy} S$
- $\overline{S} = \{\lim_{k\to\infty} p_k : \{p_k\} \subset S, \lim_k p_k \text{ exists }\}$

14.2 A characterization of closed sets in terms of convergent sequences

Theorem 14.1 (Theorem 5 on page 40 of Buck) *Let S be any subset of $\mathbf{R}^n$. 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 $p \notin \overline{S} = \operatorname{bdy} S \cup S$, then $p \notin S$ and $p \notin \operatorname{bdy} S$. Thus there exists $\delta > 0$ such that at least one of $B(p,\delta) \cap S$ or $B(p,\delta) \cap (\mathbf{R}^n \setminus S)$ 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 $p \in \operatorname{bdy} S$, 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 14.2 (Corollary 2 on page 41 of Buck) *A set S is closed if and only if it contains the limit of each convergent sequence of points from S.*

14.3 Cauchy sequences

The concept of Cauchy sequence is needed in Assignment 11.

Definition 14.3 (Definition 6 on page 52 of Buck) A sequence p_k of points in $\mathbf{R}^n$ 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 14.4 (Corollary on page 63 and exercise 32 on page 56 of Buck) *A sequence in $\mathbf{R}^n$ is convergent if and only if it is a Cauchy sequence.*

14.4 Uniform continuity

Definition 14.5 (Definition 2 on page 82 of Buck) A function $f : E \to \mathbf{R}$, where $E \subset \mathbf{R}^n$, 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 $|p - q| < \delta$.

A function which is uniformly continuous on a set S is certainly continuous at every point of S, that is, is continuous on S. However, a function continuous on a set S need not be uniformly continuous on S. There are exceptions, as in the next theorem.

Theorem 14.6 (Theorem 6 on page 84 of Buck) *A function which is continuous on a compact set D is uniformly continuous on D.*

Proof: First an outline:

- Given ϵ, use $\epsilon/2$ to get a "continuity ball" $B(p, \delta_p$ for every $p \in S$
- Use $\delta_p/2$ to get a "covering ball" for every $p \in S$
- Use HB to get a finite number of covering balls and pick δ to be the smallest of their radii
- Use the triangle inequality to get the uniform continuity

Now the details. 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,

$$D \subset \cup_{p \in D} B(p, \delta_p/2).$$

We can refer to $B(p, \delta_p/2)$ as a "covering ball". By compactness, we have $D \subset \cup_{j=1}^{m} B(p_j, \delta_{p_j}/2)$. 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 $|x - y| < \delta \le \delta_{p_j}/2$ we have $|y - p_j| \le |y - x| + |x - p_j| < \delta + \delta_{p_j}/2 \le \delta_{p_j}$. In other words, x and y both belong to the same continuity ball $B(p_j, \delta_{p_j})$. Thus

$$|f(x) - f(y)| \le |f(x) - f(p_j)| + |f(p_j) - f(y)| < \epsilon/2 + \epsilon/2 = \epsilon.$$

The proof is complete. □

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

Assignment 11 (Due November 4) In (A) and (B), show that f and g are uniformly continuous on $\mathbf{R}^n$, where

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

(B) $g(p) = x_1y_1 + \cdots + x_ny_n$ where $p = (x_1, \ldots, x_n) \in \mathbf{R}^n$ is a variable point and $y_1, \ldots, y_n \in \mathbf{R}$ are fixed.

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

15 Wednesday October 26,2005—Discussion of First Midterm

15.1 Statistics

- Mean =49
- Median =41
- tentative letter grade 91-100=A, 85-90=A-, 80-84=B+, 72-80=B, 65-71=B-,55-65=C+,45-54=C,40-44=C-,35-39=D+, 30-34=D,25-29=D-,0-24=F
- mean on each problem: #1 6.72=56%, #2 5.40=45%, #3 3.68=74%, #4 13.64=55%, #5 6.88=57%, #6 7.40=62%, #7 3.24=27%, #8 2.68=27%

15.2 Answers to the problems

Problem 1 **(a)** $\mathbf{R} - \mathbf{Z} = \cup_{n \in \mathbf{Z}}(n, n+1)$ is a union of open sets. Alternatively, given $x \in \mathbf{R} - \mathbf{Z}$, pick $n \in \mathbf{Z}$ with $x \in (n, n+1)$ and define $\delta = \min\{x - n, n + 1 - x\}$. Then $(x - \delta, x + \delta) \subset \mathbf{R} - \mathbf{Z}$.

(b) yes

(c) yes

Problem 2 **(a)** bdy $S = \{(x, 0) : 0 \le x \le 1/2\} \cup \{(x, y) : x^2 + y^2 = 1\} \cup \{(x, y) : x^2 + y^2 = 2\}$

(b) $\operatorname{int} S = \{(x, y) : 1 < x^2 + y^2 < 2\}$

(c) $\operatorname{cl} S = \{(x, 0) : 0 \leq x \leq 1/2\} \cup \{(x, y) : 1 \leq x^2 + y^2 \leq 2\}$

Problem 3 False. Counterexample: $S = \{(x, y) : x^2 + y^2 < 1\}$

Problem 4 **(a)** True; $A \cap B$ is a subset of A

(b) False; $A = (0, 1),\ B = [1/2, \infty)$

(c) False; $A = (0, 1),\ B = [1/2, 3]$

(d) False; $A = (0, 1),\ B = [1/2, 3]$

(e) True; $A \cap (\mathbf{R}^n - B)$ is a subset of A

Problem 5

- Using BW: If S is compact and not bounded, then for every $n \geq 1$ there exists $p_n \in S$ with $|p_n| > n$. By BW there exists $p \in (\operatorname{cl} \{p_n\}_{n=1}^{\infty}) \cap S$ so that for any $\delta > 0$, $B(p, \delta) \cap \{p_n\}$ is infinite, say $B(p, \delta) \cap \{p_n\} = \{p_{n_k}\}_{k=1}^{\infty}$. Then $n_k < |p_{n_k}| \leq |p_{n_k} - p| + |p| \leq \delta + |p|$, a contradiction since $\lim_k n_k = \infty$.
- Using HB: The collection $\mathcal{G} = \{\mathcal{B}(\prime, \backslash) : \backslash = \infty, \in, \ldots\}$ is an open cover of any set $S \subset \mathbf{R}^n$. If S is compact then for some $N \geq 1$, $S \subset \cup_{n=1}^{N} B(0, n) = B(0, N)$, that is, S is bounded.

Problem 6 **(a)** Let $p \in \operatorname{bdy} S$ and suppose $p \notin S$. Then $\exists B(p, \delta) \subset \mathbf{R}^n - S$, so that then $B(p, \delta) \cap S = \emptyset$, contradicting the fact that $p \in \operatorname{bdy} S$

(b) $\overline{S} = \operatorname{bdy} S \cup S \subset S$ by (a). But $\overline{S}$ is the smallest closed set containing S. In particular, $S \subset \overline{S}$. Hence $S = \overline{S}$

Problem 7 **(a)**

- Using HB: Let $A \subset \cup_{k=1}^{\infty} G_k$ where G_k is a sequence of open sets. Then $S \subset (\mathbf{R}^n - A) \cup \cup G_k$ is an open cover of S , so that $\exists N \geq 1$ with $S \subset (\mathbf{R}^n - A) \cup \cup_{k=1}^{N} G_k$. Then $A \subset \cup_{k=1}^{N} G_k$.
- Using BW: Let $\{p_n\}$ be an infinite sequence in A. Since $A \subset S$ by the BW property for S, $\exists p \in \operatorname{cl}\{p_n\} \cap S$. Since $\{p_n\} \subset A$, $p \in \operatorname{cl} A$ and since A is closed ($\operatorname{cl} A \subset A$), $p \in A$. Thus A satisfies BW
- Using CB: A is given to be closed, and S is closed because it is compact. Since $A \subset S$, A is also bounded, so A is closed and bounded, therefore compact.

(b) S is closed because it is compact. So S is a closed set containing A. Since $\overline{A}$ is the smallest closed set containing A, we have $\overline{A} \subset S$. Then by (a), $\overline{A}$, being closed, is compact.

Problem 8 **(a)** We need to show that for every $\delta > 0$, the set $B(p, \delta) \cap S$ is infinite. Given δ, pick $q_1 \neq p$ and $q_1 \in B(p, \delta) \cap S$ and set $\delta_1 = |p - q_1|$. Then pick $q_2 \neq p$, and $q_2 \in B(p, \delta_1) \cap S$ and make sure that $q_1 \neq q_2$. Continuing in this was we obtain a sequence of distinct points $\{q_i, q_2, \ldots\} \subset B(p, \delta) \cap S$

(b) Let's show (directly) that $\mathrm{cl}(\mathrm{cl}\, S) \subset \mathrm{cl}\, S$. Take $p \in \mathrm{cl}(\mathrm{cl}\, S)$ so that $B(p,\delta) \cap \mathrm{cl}\, S$ contains a point $q \neq p$. Choose δ_1 such that $B(q,\delta_1) \subset B(p,\delta)$. But $B(q,\delta_1) \cap S$ is infinite, so $B(p,\delta) \cap S$ is infinite, proving that $p \in \mathrm{cl}\, S$.

15.3 An assignment related to Problem 8

Assignment 12 (Due November 4) Give three other proofs that the set $\mathrm{cl}\, S$ of cluster points of an arbitrary set S is a closed set, more precisely,

- Show $\mathbf{R}^n - \mathrm{cl}\, S$ is open
- Show $\mathrm{bdy}\,(\mathrm{cl}\, S) \subset \mathrm{cl}\, S$
- Show $\{\lim_k p_k : p_k \in \mathrm{cl}\, S\} \subset \mathrm{cl}\, S$

16 Friday October 28, 2005—A uniformly continuous function extends (continuously!) to the closure of its domain

16.1 Motivation and statement of the problem

There are two main applications of uniform continuity. In the theory of Riemann integration the fact that a continuous function on a close rectangle in $\mathbf{R}^2$ is integrable follows very readily 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 $\mathbf{R}^n$ and let $f : S \to \mathbf{R}$ 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 $p \in S$? Let me repeat this: given a continuous function f on S, does there exist a continuous function $\tilde{f}$ on $\overline{S}$ such that $\tilde{f}_{|S} = f$?

We know already that the answer is no, as the example $f(x) = 1/x$ on $S = (0,1) \subset \mathbf{R}$ 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 14.1 and 12.4. 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 first.

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[12] to $p \in \overline{S}$, then p_k is a Cauchy sequence, and by uniform continuity of f, Assignment 11(C) tells us that $f(p_k)$ is a Cauchy sequence in $\mathbf{R}$. 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 $p \in \overline{S}$. 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 γ. Since every subsequence of this sequence must also converge to γ, it follows that $\alpha = \gamma$ and $\beta = \gamma$, so $\alpha = \beta$, as required. The second question is answered affirmatively.

17 Monday October 31, 2005—The extension theorem

This section is devoted to the answer to the third question raised in the last lecture. Let us state this as a theorem.

Theorem 17.1 *Let $f : S \to \mathbf{R}$ be a uniformly continuous function defined on a subset S of $\mathbf{R}^n$. Define a function $\tilde{f} : \overline{S} \to \mathbf{R}$ 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[13] 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.

[12] Such a sequence exists by Theorem 14.1

[13] The proof will show that actually $\tilde{f}$ is uniformly continuous on $\overline{S}$

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:

$$|\tilde{f}(p) - \tilde{f}(q)| \leq |\tilde{f}(p) - f(p_k)| + |f(p_k) - f(q_j)| + |f(q_j) - f(q)|. \tag{9}$$

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)

$$|p_k - q_j| \leq |p_k - p| + |p - q| + |q - q_j|. \tag{10}$$

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 δ_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 $1/k + 1/j < \delta_1$. □

Assignment 13 (Due November 14)

(A) Let $S \subset \mathbf{R}^n$ be a bounded set and let $f : S \to \mathbf{R}$ 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 $f : \mathbf{R} \to \mathbf{R}$ 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 $\mathbf{R}$.

18 Wednesday November 2, 2005—Differentiability implies continuity for functions

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

Let's begin by recalling the mean value theorem in one variable. We shall use Lemma 18.1 (a result in one dimension) in the proof of Theorem 18.3 below (a theorem in $n \geq 1$ dimensions).

Lemma 18.1 (Mean Value Theorem in one variable) *If $f : (a,b) \to \mathbf{R}$ 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. (See the textbook for 140AB by Ross, page 160. The function f defined by $f(0) = 0$ and $f(x) = x^2 \sin(1/x)$ for $x \neq 0$ is differentiable for every real x, but the derivative f' is not continuous at $x = 0$.) However, only the existence of a derivative, not the continuity of the derivative, is required in Lemma 18.1 and Theorem 18.2. This is one difference between these two one-dimensional results, and the n-dimensional theorem Theorem 18.3.

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

Theorem 18.2 (Differentiability implies continuity—one variable) *If $f : (a,b) \to \mathbf{R}$ is differentiable at a point c in (a,b), then f is continuous at c. In particular, if f is differentiable on all of (a,b) then it is continuous on (a,b).*

Proof: If $f : (a,b) \to \mathbf{R}$ 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 $f : D \to \mathbf{R}$ defined on open subsets D of $\mathbf{R}^n$. 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

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

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

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

$$D_j f(p_0) = f_j(p_0) = \tfrac{\partial f}{\partial x_j}(p_0).$$

You can also write (if you prefer)

$$\frac{\partial f}{\partial x_j}(p_0) = \lim_{t\to 0} \frac{f(x_1^0, \cdots, x_{j-1}^0, x_j + t, x_{j+1}^0, \cdots, x_n^0) - f(x_1^0, x_2^0, \ldots, x_n^0)}{t}.$$

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

We want to prove an analog of Theorem 18.2 for functions of n variables. We will see that it differs both in statement and difficulty of proof from the case $n = 1$. The following example (Problem 4 on page 135 and part of Assignment 14) indicates a striking difference between one variable and two variables.

Let $f(x, y) = xy/(x^2 + y^2)$ for $(x, y) \in \mathbf{R}^2 - \{(0, 0)\}$ and $f(0, 0) = 0$. Then

- $D_1 f(0, 0)$ and $D_2 f(0, 0)$ exist
- f is not continuous at $(0, 0)$
- $D_1 f$ and $D_2 f$ are not continuous at $(0, 0)$

Theorem 18.3 (Corollary on page 129 of Buck) *Let $f : D \to \mathbf{R}$ be defined on an open subset D of $\mathbf{R}^n$, 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 $p_0 = (x_1^0, \ldots, x_n^0)$ 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
>
> $$f(p) - f(p_0) = f_1(q_1)(x_1 - x_1^0) + f_2(q_2)(x_2 - x_2^0) + \cdots + f_n(q_n)(x_n - x_n^0) \quad (11)$$
>
> 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 $p = (x, y, z)$.

Step 1 Let $p_1 = (x, y_0, z_0)$. Then by the mean value theorem in one variable

$$f(p_1) - f(p_0) = \frac{\partial f}{\partial x}(c, y_0, z_0)(x - x_0) \text{ for some } c \text{ between } x \text{ and } x_0.$$

(Question: what does c depend on?)

Step 2 Let $p_2 = (x, y, z_0)$. Then by the mean value theorem in one variable

$$f(p_2) - f(p_1) = \tfrac{\partial f}{\partial y}(x, d, z_0)(y - y_0) \text{ for some } d \text{ between } y \text{ and } y_0.$$

(Question: what does d depend on?)

Step 3 Let $p_3 = (x, y, z)$ $(= p)$. Then by the mean value theorem in one variable

$$f(p) - f(p_2) = \tfrac{\partial f}{\partial z}(x, y, e)(z - z_0) \text{ for some } e \text{ between } z \text{ and } z_0.$$

(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

$$\begin{aligned} f(p) - f(p_0) &= [f(p_1) - f(p_0)] + [f(p_2) - f(p_1)] + [f(p) - f(p_2)] \\ &= f_1(q_1)(x - x_0) + f_2(q_2)(y - y_0) + f_3(q_3)(z - z_0). \end{aligned}$$

This proves (11).

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 (11) 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 $p \in D$. □

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[14].

Assignment 14 (Due November 14) [Buck, §3.3 page 134 #4,5,11]

19 Friday November 4, 2005—Differential as a Linear approximation (the case of functions)

Let's examine the equation (11). 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 (11) can be rewritten as a dot product of vectors:

$$f(p) - f(p_0) = (f_1(q_1), f_2(q_2), \cdots, f_n(q_n)) \cdot (x_1 - x_1^0, x_2 - x_2^0, \cdots, x_n - x_n^0), \quad (12)$$

or, $f(p) - f(p_0) = V \cdot (p - p_0)$, where V is the vector $V = (f_1(q_1), f_2(q_2), \cdots, f_n(q_n))$. 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 (12).

[14] 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 [Buck, 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 15)

Assignment 15 (Due November 14) 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 19.1 If $f : D \to \mathbf{R}$ is defined on an open set $D \subset \mathbf{R}^n$, the *gradient* of f at $p \in D$ is the vector $\nabla f(p) = (D_1 f(p), D_2 f(p), \cdots, D_n f(p))$. Of course ∇f is defined only at those points of D where all first order partial derivatives of f exist.

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 $c \in (a, b) \subset \mathbf{R}$ with derivative $f'(c)$, then

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

This is the same as

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

The following is the analog, for functions of several variables, of (13). It says that a C^1-function can be approximated, in some sense, by an essentially linear function, namely the function $T(p) := f(p_0) + \nabla f(p_0) \cdot (p - p_0)$. Note that (14) is much stronger than the obvious statement that $|f(p) - f(p_0) - \nabla f(p_0) \cdot (p - p_0)| \to 0$ as $p \to p_0$, which follows from the continuity of f at p_0.

Theorem 19.2 (Theorem 8 on page 131 of Buck) *Let f be of class C^1 on an open set $D \subset \mathbf{R}^n$. 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

$$\frac{|f(p) - f(p_0) - \nabla f(p_0) \cdot (p - p_0)|}{|p - p_0|} < \epsilon \text{ whenever } p \in B(p_0, \delta) \cap D. \tag{14}$$

Proof: Let $R := f(p) - f(p_0) - \nabla f(p_0)\cdot(p-p_0)$. By (12) (which is the main point in the proof of Theorem 18.3), $f(p)-f(p_0) = V\cdot(p-p_0)$, where V is the vector $V = (f_1(q_1), f_2(q_2), \cdots, f_n(q_n))$. 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:

$$|R| = |[V-\nabla f(p_0)]\cdot[p-p_0]| \le |V-\nabla f(p_0)||(p-p_0)|,$$

that is

$$\frac{|R|}{|p-p_0|} \le |V-\nabla f(p_0)|, \tag{15}$$

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 (15) $|R|/|p-p_0|$, approaches zero as p approaches p_0.

Here are the details:

$$\begin{aligned} V-\nabla f(p_0) &= [f_1(q_1), f_2(q_2), \cdots, f_n(q_n)] - [f_1(p_0), f_2(p_0), \cdots, f_n(p_0)] \\ &= [f_1(q_1)-f_1(p_0), f_2(q_2)-f_2(p_0), \cdots, f_n(q_n)-f_n(p_0)], \end{aligned}$$

so that

$$|V-\nabla f(p_0)|^2 = (f_1(q_1)-f_1(p_0))^2 + (f_2(q_2)-f_2(p_0))^2 + \cdots + (f_n(q_n)-f_n(p_0))^2. \tag{16}$$

Since each f_j is continuous and since

$$\begin{aligned} |q_j - p_0|^2 &= |((x_1, x_2, \ldots, x_{j-1}, c_j, x^0_{j+1}, \ldots, x^0_n) - (x^0_1, \ldots s, x^0_n)|^2 \\ &= \sum_{k=1}^{j-1}(x_k - x^0_k)^2 + (c_j - x^0_j)^2 \le |p-p_0| \end{aligned}$$

for each j, we see from (15) and (16) that (14) holds.

20 Monday November 7, 2005—Higher derivatives; Transformations

20.1 Higher order partial derivatives

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 $\mathbf{R}^n$, 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,

$$D_i(D_j f) = (f_j)_i = f_{ji} = \tfrac{\partial}{\partial x_i}(\tfrac{\partial f}{\partial x_j}) = \frac{\partial^2 f}{\partial x_i \partial x_j},$$

and if $i = j$,

$$D_j^2 f = D_j(D_j f) = (f_j)_j = f_{jj} = \tfrac{\partial}{\partial x_j}(\tfrac{\partial f}{\partial x_j}) = \frac{\partial^2 f}{\partial x_j^2}.$$

Definition 20.1 Let k be any positive integer, $k = 1, 2, \ldots$. A function f defined on an open set D in $\mathbf{R}^n$ 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. [15]

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:

$$D_j f \ (1 \le j \le n), \qquad D_m(D_i f) \ (1 \le i \le n, 1 \le m \le n).$$

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 $\mathbf{R}$, 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 18.2. We have shown in Theorem 18.3 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 eachother?)

20.2 Transformations

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

Definition 20.2 A *transformation* is any function $T : D \to \mathbf{R}^m$, where $D \subset \mathbf{R}^n$.

Here, $m \ge 1$ and $n \ge 1$, 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 $p = (x_1, \cdots, x_n) \in D$, and $T(p) = (y_1, \cdots, y_m) \in \mathbf{R}^m$, then each y_j is a function of $p = (x_1, \ldots, x_n)$, which we can denote by f_j or f^j [16]. Thus

$$T(p) = (f^1(p), \cdots, f^m(p)),$$

[15] In [Buck, Definition 1,page 128], the definition of C^k requires that f be continuous. By Theorem 18.3, Buck's definition of C^k and our Definition 20.1 are equivalent

[16] 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 $f^j : D \to \mathbf{R}$ is a function of n variables $x_1, \ldots, x_n$.

Transformations are the subject of [Buck, Chapter 7] and their geometric properties are discussed in [Buck, 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 16 to 23.[17]

The following is the analog of Definition 11.1

Definition 20.3 Let $T : D \to \mathbf{R}^m$ be a transformation, where D is any subset of $\mathbf{R}^n$, and let $p_0 \in D$. We say that T 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

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

Notice that if $f : D \to \mathbf{R}$ is a function which is of class C^1 on a subset $D \subset \mathbf{R}^n$, the ∇f is an example of a transformation. 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 $T : D \to \mathbf{R}^m$, such as continuity and differentiability (suitably defined).

Assignment 16 (Due November 21) Let $T(p) = (f^1(p), \cdots, f^m(p))$ 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 11.2.

Theorem 20.4 (Theorem 4 on page 333 of Buck) *The continuous image of a compact set is compact. In other words, if $T : D \to \mathbf{R}^m$ is a continuous transformation on D, and D is a compact subset of $\mathbf{R}^n$, then $T(D)$ is a compact subset of $\mathbf{R}^m$.*

Assignment 17 (Due November 21) Prove Theorem 20.4.

The following is the analog of Theorem 12.4.

[17] Don't worry, not all of these assignments will be handed in

Theorem 20.5 *Let $T : D \to \mathbf{R}^m$, where $D \subset \mathbf{R}^n$, 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 18 (Due November 21) Prove Theorem 20.5.

Assignment 19 (Due November 21) State and prove an analog of the Extreme values theorem, Theorem 12.4. (Hint: Since $\mathbf{R}^m$ has no order structure, you have to express the theorem in terms of $|T(p)|$.)

The following is the analog of Definition 14.5

Definition 20.6 A transformation : $T : E \to \mathbf{R}^m$, where $E \subset \mathbf{R}^n$, 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 14.6.

Theorem 20.7 *A transformation which is continuous on a compact set D is uniformly continuous on D.*

Assignment 20 (Due November 21) Prove Theorem 20.7.

Assignment 21 (Due November 21) Show that a linear transformation (see [Buck, Section 7.3]) is uniformly continuous. (Hint: Use [Buck, Theorem 8,page 338])

The following is the analog of Theorem 17.1

Theorem 20.8 *Let $T : D \to \mathbf{R}^m$ be a uniformly continuous transformation defined on a subset D of $\mathbf{R}^n$. Define a transformation $\tilde{T} : \overline{D} \to \mathbf{R}^m$ by*

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

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 22 (Due November 21) Prove Theorem 20.8.

Assignment 23 (Due November 21) Let $D \subset \mathbf{R}^n$ be a bounded set and let $T : D \to \mathbf{R}^m$ be a continuous transformation. Prove that T has a continuous extension to $\overline{D}$ if and only if T is uniformly continuous on D.

21 Wednesday November 9, 2005—Approximation by the differential—the case of transformations

Our next main result is the analog for transformations of (14) in Theorem 19.2. First we need to define the replacement for the gradient.

Definition 21.1 If $T : D \to \mathbf{R}^m$ is defined on an open set $D \subset \mathbf{R}^n$, with coordinate functions $f^1, \ldots, f^m$, the *Jacobian matrix* of T at $p \in D$ is the m by n matrix

$$J_T(p) = \left[\begin{array}{ccc} \frac{\partial f^1}{\partial x_1}(p) & \ldots & \frac{\partial f^1}{\partial x_n}(p) \\ \frac{\partial f^2}{\partial x_1}(p) & \ldots & \frac{\partial f^2}{\partial x_n}(p) \\ \ldots & \ldots & \ldots \\ \frac{\partial f^m}{\partial x_1}(p) & \ldots & \frac{\partial f^m}{\partial x_n}(p) \end{array} \right].$$

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

$$J_T(p) = [\tfrac{\partial f^i}{\partial x_j}(p)]_{1\le i\le m, 1\le j\le n} = [D_j f^i(p)]_{1\le i\le m, 1\le j\le n}$$

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

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 $\det J_T(p)$.[18]

At this point it is necessary to include the following obvious definition. A transformation $T = (f^1, \ldots, f^m)$ is said to be of *class* C^k on an open set $D \subset \mathbf{R}^n$ for a fixed integer $k \ge 1$, if each of its coordinate functions f^i is of class C^k on D.

Assignment 24 (Due November 21) Prove that a transformation of class C^1 is continuous.

Theorem 21.2 (Theorem 10 on page 344 of Buck) *Let $T : D \to \mathbf{R}^m$ be a transformation of class C^1 on an open set $D \subset \mathbf{R}^n$. Then*[19]*, for any $p_0 \in D$,*

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

[18]Be careful. Some authors (including Buck) define the Jacobian to be what I am calling Jacobian determinant. Others, like me, who are sensible, distinguish between the two definitions: Jacobian matrix and Jacobian determinant

[19]Strictly speaking, $T(p)$ and $T(p_0)$ are row vectors and $J_T(p_0) \times (p - p_0)^t$ is a column vector, so to be perfectly truthful this should be written as $\lim_{p\to p_0} \frac{|T(p)^t - T(p_0)^t - J_T(p_0)\times(p-p_0)^t|}{|p-p_0|} = 0$. However, we won't do this as it makes the notation cumbersome and it is clear that we are talking about vectors, and it doesn't matter if we call them row vectors or column vectors.

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

$$\frac{|T(p) - T(p_0) - J_T(p_0) \times (p - p_0)^t|}{|p - p_0|} < \epsilon \text{ whenever } p \in B(p_0, \delta) \cap D. \tag{19}$$

Proof: Let $T = (f^1, \cdots, f^m)$. By Theorem 19.2, for each $1 \leq i \leq m$

$$\frac{|f^i(p) - f^i(p_0) - \nabla f^i(p_0) \cdot (p - p_0)|}{|p - p_0|} \to 0 \text{ as } p \to p_0. \tag{20}$$

Using the notation $R_i(p) = f^i(p) - f^i(p_0) - \nabla f^i(p_0) \cdot (p - p_0)$, (20) becomes

$$\frac{|R_i(p)|}{|p - p_0|} \to 0, \tag{21}$$

and we have

$$\begin{aligned} T(p) - T(p_0) &= (f^1(p) - f^1(p_0), \cdots, f^m(p) - f^m(p_0)) \\ &= (\nabla f^1(p_0) \cdot (p - p_0), \cdots, \nabla f^m(p_0) \cdot (p - p_0)) + (R_1(p), \cdots, R_m(p)) \\ &= (\sum_{j=1}^n \tfrac{\partial f^1}{\partial x_j}(p_0)(x_j - x_j^0), \cdots, \sum_{j=1}^n \tfrac{\partial f^m}{\partial x_j}(p_0)(x_j - x_j^0)) + (R_1(p), \cdots, R_m(p)). \end{aligned}$$

On the other hand,

$$\begin{aligned} J_T(p_0) \times (p - p_0)^t &= \begin{bmatrix} \frac{\partial f^1}{\partial x_1}(p_0) & \cdots & \frac{\partial f^1}{\partial x_n}(p_0) \\ \frac{\partial f^2}{\partial x_1}(p_0) & \cdots & \frac{\partial f^2}{\partial x_n}(p_0) \\ \cdots & \cdots & \cdots \\ \frac{\partial f^m}{\partial x_1}(p_0) & \cdots & \frac{\partial f^m}{\partial x_n}(p_0) \end{bmatrix} \times \begin{bmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \\ \cdots \\ x_n - x_n^0 \end{bmatrix} \\ &= (\sum_{j=1}^n \tfrac{\partial f^1}{\partial x_j}(p_0)(x_j - x_j^0), \cdots, \sum_{j=1}^n \tfrac{\partial f^m}{\partial x_j}(p_0)(x_j - x_j^0))^t. \end{aligned}$$

Now let us subtract the last two equations. We get

$$T(p) - T(p_0) - J_T(p_0) \times (p - p_0)^t = (R_1(p), \cdots, R_m(p)).$$

Now use (21) to obtain

$$\frac{|T(p) - T(p_0) - J_T(p_0) \times (p - p_0)^t|}{|p - p_0|} = \left(\sum_{i=1}^m \frac{R_i(p)^2}{|p - p_0|^2} \right)^{1/2} \to 0$$

as $p \to p_0$. □

22 Friday November 11, 2005—Holiday

(Veteran's day)

23 Monday November 14, 2005—Chain rule I. The one-dimensional case

We begin our 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-dimension 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 23.1 (One-dimensional chain rule) *Let g be a real valued function defined on an open interval containing $a \in \mathbf{R}$ 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

$$|g(x) - g(a) - g'(a)(x-a)| < \epsilon'|x-a| \quad \text{if } |x-a| < \delta'. \tag{22}$$

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

$$|f(y) - f(g(a)) - f'(g(a))(y - g(a))| < \epsilon''|y - g(a)| \quad \text{if } |y - g(a)| < \delta''. \tag{23}$$

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

$$|f(g(x)) - f(g(a)) - f'(g(a))g'(a)(x-a)| < \epsilon|x-a| \quad \text{if } |x-a| < \delta. \tag{24}$$

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

$$|g(x) - g(a)| < \delta'' \text{ if } |x-a| < \delta_c. \tag{25}$$

Using (25), we may replace y in (23) by $g(x)$ to obtain

$$|f(g(x)) - f(g(a)) - f'(g(a))(g(x) - g(a))| < \epsilon''|g(x) - g(a)| \quad \text{if } |x-a| < \delta_c. \tag{26}$$

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

$$g(x) - g(a) = g'(a)(x-a) + \eta(x) \tag{27}$$

and by (22),

$$|\eta(x)| < \epsilon'|x-a| \text{ if } |x-a| < \delta. \tag{28}$$

Now substitute (27) into (26) (in two places!) and set

$$A := f(g(x)) - f(g(a)) - f'(g(a))[g'(a)(x-a) + \eta(x)] \tag{29}$$

to obtain from (26)

$$|A| < \epsilon''|g'(a)(x-a) + \eta(x)| \quad \text{if } |x-a| < \delta. \tag{30}$$

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

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

$$\begin{aligned}
&= |A + f'(g(a))\eta(x)| && \text{(by (29))} \\
&\leq |A| + |f'(g(a))\eta(x)| \\
&\leq \epsilon''|g'(a)||x-a| + \epsilon''|\eta(x)| + |f'(g(a))||\eta(x)| && \text{(by (30))} \\
&\leq [\epsilon''|g'(a)| + \epsilon''\epsilon' + |f'(g(a))|\epsilon']|x-a| && \text{(by (28))} \\
&< \epsilon|x-a|,
\end{aligned}$$

the last step provided we simply choose ϵ' and ϵ'' so that $[\epsilon''|g'(a)|+\epsilon''\epsilon'+|f'(g(a))|\epsilon'] < \epsilon$. This proves (24). □

24 Wednesday November 16, 2005—Coordinate-free definition of derivative

24.1 Composition of transformations

We now consider composition of transformations and the chain rule in arbitrary dimensions.

Definition 24.1 Let T be a transformation defined on a subset A of $\mathbf{R}^n$ with $T(A) \subset \mathbf{R}^m$. Suppose that S is a transformation defined on a subset C of $\mathbf{R}^m$ with $S(C) \subset \mathbf{R}^k$. We suppose that $C \subset T(A)$. Under these circumstances, the *composition* of S and T is the transformation $S \circ T$ (also denoted[20] simply by ST) defined by

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

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 $TS\,(x,y,z) = T(S(x,y,z)) = T(x-y, yz) = ((x-y)yz, 2(x-y), -yz)$. Note that in this case, $ST \neq TS$.

[20] 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

Theorem 24.2 (Theorem 3, page 333 of Buck) *If $S : A \to \mathbf{R}^m$ is a transformation which is continuous at a point $p_0 \in A \subset \mathbf{R}^n$, and $T : B \to \mathbf{R}^k$ is a transformation which is continuous at the point $S(p_0) \in B \subset \mathbf{R}^m$, then the composition $T \circ S : A \to \mathbf{R}^k$ is continuous at the point p_0.*

Assignment 25 (Due November 28) Prove Theorem 24.2.

24.2 Coordinate free definition of derivative

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

Definition 24.3 (Coordinate-free definition of derivative) Let T be a transformation defined on a subset A of $\mathbf{R}^n$ with $T(A) \subset \mathbf{R}^m$. We say that T is *differentiable* at $p_0 \in A$ if there exists a linear transformation $L : \mathbf{R}^n \to \mathbf{R}^m$, such that

$$\lim_{p \to p_0} \frac{|T(p) - T(p_0) - L(p - p_0)|}{|p - p_0|} = 0. \tag{31}$$

We denote L by $T'(p_0)$ (this is justified by Assignment 26) 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 26 (Due November 28) *Prove that, for a fixed p_0, at most one linear transformation L can satisfy* (31). (This is the same as Exercise #10, page 352 in Buck)

Since at most one linear transformation can satisfy (31), the notation $T'(p_0)$ is justified, that is, T' is a function (single valued, or well-defined) with domain $\{p \in A : T \text{ is differentiable at } p\}$, which has its values in the set of all linear transformations from $\mathbf{R}^n$ to $\mathbf{R}^m$.

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

Remark 24.4 *If $m = 1$ and $n = 1$, then a transformation T is just a function $f : A \to \mathbf{R}$, where $A \subset \mathbf{R}$. 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 $L : \mathbf{R} \to \mathbf{R}$ given by $L(x) = f'(x_0)x$.* (What is the justification for this?)

Remark 24.5 *If $m = 1$ and $n \geq 1$, then a transformation T is just a function $f : A \to \mathbf{R}$, where this time $A \subset \mathbf{R}^n$. 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 $L : \mathbf{R}^n \to \mathbf{R}$ given by $L(p) = \nabla f(p_0) \cdot p$.* (What is the justification for this?)

Remark 24.6 *If $m \geq 1$ and $n \geq 1$, then a transformation T is just a function $T : A \to \mathbf{R}^m$, where $A \subset \mathbf{R}^n$. 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 $L : \mathbf{R}^n \to \mathbf{R}^m$ given*[21] *by $L(p) = J_T(p_0) \times p^t$.* (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 24.6, 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 27 (Due November 28) *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 $t \in \mathbf{R}$ and $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0), \ldots,$ $e_n = (0, \ldots, 0, 1)$).

24.3 What the second midterm will cover

Assignments: Assignments 8-24

Text pages covered: [22] 72-74, 81-85, 89-93, 109-110, 127-131, 328-333, 341-344

Lecture Material: [23]

- Continuous real valued functions, continuous image of a compact set (Oct 17)
- Continuous functions and sequences, extreme value theorem (Oct 19)
- Closure of a set, continuous functions on compact sets are uniformly continuous (Oct 24)
- Extension theorem for functions (Oct 28 and 31)
- Differentiability implies continuity for functions (Nov 2)
- Linear approximation (Nov 4)
- Properties of transformations (Nov 7)
- Differentiability implies continuity for transformations, linear approximation for transformations (Nov 9)

[21] Recall that p^t is the transpose of the row vector p and that strictly speaking $L(p)$ is a row vector and $J_T(p_0) \times p^t$ is a column vector. As stated in an earlier footnote, we shall ignore this notational inconsistency since it does not cause any confusion

[22] I suggest you rely on these notes rather than on these pages of the text

[23] The second midterm is a take home midterm which will focus on the first four of these items; the last four items are covered in the Assignments which are due on November 21

25 Friday November 18, 2005—Chain rule II. The general case; applications

25.1 Proof of the chain rule

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 23.1.

Theorem 25.1 (Chain Rule,Theorem 11, page 346 of Buck) *Let $T : D \to \mathbf{R}^m$ be a transformation which is differentiable on an open set $D \subset \mathbf{R}^n$, and let $S : E \to \mathbf{R}^k$ be a differentiable transformation on an open subset E of $\mathbf{R}^m$ containing $T(D)$. Then $S \circ T$ is differentiable on D, and if $p \in D$, then*

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

To make life simpler, we shall isolate two lemmas, which are themselves of independent interest. We first met Lemma 25.2 in Assignment 21.

Lemma 25.2 (Theorem 8,page 338 of Buck) *A linear transformation L from $\mathbf{R}^n$ to $\mathbf{R}^m$ is continuous. In fact, L is uniformly continuous and there is a constant C such that $|L(p)| \le C|p|$ for every $p \in \mathbf{R}^n$. More precisely, if L is given by an $m \times n$ matrix $A := [a_{ij}]_{1\le i\le m, 1\le j\le n}$ as follows:*

$$L(\sum_{j=1}^{n} x_j e_j) = \sum_{j=1}^{n} x_j L(e_j) \text{ where } Le_j = A \times e_j^t = \sum_{i=1}^{m} a_{ij} e_i$$

and $e_1 = (1, 0, \ldots, 0),\ e_2 = (0, 1, 0, \ldots, 0), \ldots,\ e_n = (0, \ldots, 0, 1)$ is the usual basis for $\mathbf{R}^n$ (and $e_1, \ldots, e_m$ is the usual basis for $\mathbf{R}^m$!), then

$$|L(p)| \le (\sum_i \sum_j a_{ij}^2)^{1/2} |p|.$$

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

$$L(p) = \sum_j x_j \sum_i a_{ij} e_i = \sum_i (\sum_j x_j a_{ij}) e_i,$$

so, by the Schwarz inequality,

$$|L(p)|^2 = \sum_i |\sum_j x_j a_{ij}|^2 \le \sum_i (\sum_j x_j^2)(\sum_j a_{ij}^2) = (\sum_i \sum_j a_{ij}^2)|p|^2. \ \square$$

Lemma 25.3 (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

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

Writing this "horizontally", you get

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

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

$$\begin{aligned} |T(p)-T(p_0)| &\le |T(p)-T(p_0)-T'(p_0)(p-p_0)| + |T'(p_0)(p-p_0)| \\ &\le 365|p-p_0| + C|p-p_0| = (365+C)|p-p_0|. \end{aligned}$$

(The constant C comes from Lemma 25.2.) Thus T is continuous at p_0. □

Question: What is the difference between Lemma 25.3 and Assignment 24.

In the proof of Theorem 25.1 which follows, the names of the characters were changed to protect the innocent. Any similarity with any characters, living or dead, is purely intentional.

Proof of Theorem 25.1 (Chain Rule): Let $p_0 \in D$. Since T is differentiable at p_0, $\forall \epsilon' > 0, \exists \delta' > 0$ such that

$$|T(p)-T(p_0)-T'(p_0)(p-p_0)| < \epsilon'|p-p_0| \quad \text{if } |p-p_0| < \delta'. \tag{32}$$

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

$$|S(q)-S(T(p_0))-S'(T(p_0))(q-T(p_0))| < \epsilon''|q-T(p_0)| \quad \text{if } |q-T(p_0)| < \delta''. \tag{33}$$

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

$$|S\circ T(p)-S\circ T(p_0)-S'(T(p_0))\circ T'(p_0)(p-p_0)| < \epsilon|p-p_0| \quad \text{if } |p-p_0| < \delta. \tag{34}$$

By Lemma 25.3, T is continuous at p_0, so $\exists \delta_c > 0$ such that

$$|T(p)-T(p_0)| < \delta'' \text{ if } |p-p_0| < \delta_c. \tag{35}$$

Using (35), we may replace q in (33) by $T(p)$ to obtain

$$|S(T(p))-S(T(p_0))-S'(T(p_0))(T(p)-T(p_0))| < \epsilon''|T(p)-T(p_0)| \quad \text{if } |p-p_0| < \delta_c. \tag{36}$$

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

$$T(p)-T(p_0) = T'(p_0)(p-p_0) + \eta(p) \tag{37}$$

and by (32),

$$|\eta(p)| < \epsilon'|p - p_0| \text{ if } |p - p_0| < \delta. \tag{38}$$

Now substitute (37) into (36) (in two places!) and set

$$A(p) := S(T(p)) - S(T(p_0)) - S'(T(p_0))[T'(p_0)(p - p_0) + \eta(p)] \tag{39}$$

to obtain from (36)

$$|A(p)| < \epsilon''|T'(p_0)(p - p_0) = \eta(p)| \quad \text{if } |p - p_0| < \delta. \tag{40}$$

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

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

$$\begin{aligned}
&= |A(p) + S'(T(p_0))\eta(p)| && \text{(by (39))} \\
&\leq |A(p)| + |S'(T(p_0))\eta(p)| \\
&\leq \epsilon''|T'(p_0)(p - p_0)| + \epsilon''|\eta(p)| + |S'(T(p_0))\eta(p)| && \text{(by (40))} \\
&\leq \epsilon'' C_1|p - p_0| + \epsilon''\epsilon'|p - p_0| + C_2\epsilon'|p - p_0| && \text{(by (38) and Lemma 25.3)} \\
&< \epsilon|p - p_0|,
\end{aligned}$$

the last step provided we simply choose ϵ' and ϵ'' so that $[\epsilon'' C_1 + \epsilon''\epsilon' + C_2\epsilon'] < \epsilon$. This proves (34). □

The power of Theorem 25.1 is that by setting $m = n = k = 1$ you get the one-dimensional chain rule (Theorem 23.1), and by setting $m = k = 1$ and leaving $n \geq 1$ you subsume the discussion of the chain rule in [Buck,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 $1 \leq i \leq m, 1 \leq j \leq k, 1 \leq r \leq k$,

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

Since

$$\begin{aligned}
S \circ T(p) &= S(T(p)) = S(f^1(p), \ldots, f^m(p)) \\
&= (g^1(f^1(p), \ldots, f^m(p)), \ldots, g^k(f^1(p), \ldots, f^m(p))),
\end{aligned}$$

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 28 (Due November 28) 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.

25.2 Baby chain rule

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 25.1, Remark 24.6, and Assignment 28.

Corollary 25.4 *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 25.4, we prove the following theorem from [Buck,section 3.4].

Theorem 25.5 (Baby chain rule,Theorem 14, page 136 of Buck) *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 $t_0 \in \mathbf{R}$, 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 $t_0 \in \mathbf{R}$, and for t in that interval,*

$$F'(t) = \tfrac{\partial f}{\partial x}(g(t),h(t))g'(t) + \tfrac{\partial f}{\partial y}(g(t),h(t))h'(t).$$

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 25.4,

$$\begin{aligned} F'(t) &= J_F(t) = J_{S\circ T}(t) = J_S(T(t)) \times J_T(t) \\ &= [\tfrac{\partial f}{\partial x}(g(t),h(t)) \ \tfrac{\partial f}{\partial y}(g(t),h(t))] \times \begin{bmatrix} g'(t) \\ h'(t) \end{bmatrix} \\ &= \tfrac{\partial f}{\partial x}(g(t),h(t))g'(t) + \tfrac{\partial f}{\partial y}(g(t),h(t))h'(t). \ \Box \end{aligned}$$

Assignment 29 (Due November 28) *Let $F(x,y) = f(g(x,y), h(x,y))$, where $g : \mathbf{R}^2 \to \mathbf{R}$ and $h : \mathbf{R}^2 \to \mathbf{R}$. Use Corollary 25.4 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 25.4 with the method on page 137 of Buck.)

Assignment 30 (Due November 28) [Buck, §7.4 page 351, #2,5,(7 or 8)](three problems)

26 Monday November 21, 2005—Mean value theorems; local invertibility

26.1 Mean Value Theorems

Up to now we have used the mean value theorem in one variable (Theorem 18.1). But we mentioned the mean value theorem in several variables above (see the first question at the beginning of the lecture for November 4), 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 21.2 (linear approximation for transformations). 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 (Buck, Theorem 14, page 355)—see Theorem 26.3 below.

We note that the version for functions (Theorem 26.1), nicknamed the "Little Mean Value Theorem" will be used in the proof of the version for transformations (Theorem 26.2), nicknamed the "Big Mean Value Theorem". Also, the "Baby Chain Rule" (Theorem 25.5) is needed in the proof of the "Little Mean Value Theorem"[24].

Theorem 26.1 ("Little" Mean Value Theorem, Theorem 16,page 151 of Buck) *Let $f : B(p_0, r) \to \mathbf{R}$ be of class C^1 on a ball $B(p_0, r) \subset \mathbf{R}^n$. Then for any two points $p_1, p_2 \in B(p_0, r)$, there is another point p^* on the line[25] segment $L := \{tp_2 + (1-t)p_1 : 0 \le t \le 1\}$ connecting p_1 and p_2 such that*

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

Proof: Define a function $F : [0, 1] \to \mathbf{R}$ by

$$F(\lambda) = f(\lambda p_2 + (1 - \lambda)p_1).$$

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

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

$$f(p_2) - f(p_1) = F'(\lambda_0) \tag{41}$$

for some $\lambda_0 \in (0, 1)$.

Letting $p^* = \phi(\lambda_0)$ we get by the "coordinatized" chain rule (Corollary 25.4),

$$F'(\lambda_0) = \nabla f(\phi(\lambda_0)) \times J_\phi(\lambda_0) = \nabla f(\phi(\lambda_0)) \times (p_2 - p_1)^t = \nabla f(p^*) \cdot (p_2 - p_1). \tag{42}$$

Compare (41) and (42). □

[24] We have a little and big mean value theorem. Question: what is the "tiny mean value theorem"?
[25] Note that this line segment is a subset of $B(p_0, r)$

Theorem 26.2 ("Big" Mean Value Theorem, Theorem 12,page 350 of Buck) *Let $T : D \to \mathbf{R}^m$ be a transformation of class C^1 on an open set $D \subset \mathbf{R}^n$. 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*[26]

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

where M is the matrix $(D_j f^i(p_i^*))_{1\leq i\leq m, 1\leq j\leq n}$, that is[27][28]*,*

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

Proof: Apply the Little mean value theorem (Theorem 26.1) to each $f^i : D \to \mathbf{R}$ to get points $p_i^* \in L$ such that

$$f^i(p'') - f^i(p') = \nabla f^i(p_i^*) \cdot (p'' - p') \quad (1 \leq i \leq m). \tag{43}$$

Now write down the coordinates of the vector $T(p'') - T(p')$, thinking of it as a column vector, and use (43):

$$\begin{aligned} T(p'') - T(p') &= (f^1(p''), \cdots, f^m(p''))^t - (f^1(p'), \cdots, f^m(p'))^t \\ &= (f^1(p'') - f^1(p'), \cdots, f^m(p'') - f^m(p'))^t \\ &= (\nabla f^1(p_1^*) \cdot (p'' - p'), \cdots, \nabla f^m(p_m^*) \cdot (p'' - p'))^t. \end{aligned}$$

On the other hand,

$$\begin{aligned} M \times (p'' - p')^t &= \begin{bmatrix} \frac{\partial f^1}{\partial x_1}(p_1^*) & \ldots & \frac{\partial f^1}{\partial x_n}(p_1^*) \\ \frac{\partial f^2}{\partial x_1}(p_2^*) & \ldots & \frac{\partial f^2}{\partial x_n}(p_2^*) \\ \ldots & \ldots & \ldots \\ \frac{\partial f^m}{\partial x_1}(p_m^*) & \ldots & \frac{\partial f^m}{\partial x_n}(p_m^*) \end{bmatrix} \times \begin{bmatrix} p_1'' - p_1' \\ p_2'' - p_2' \\ \cdots \\ p_n'' - p_n' \end{bmatrix} \\ &= \begin{bmatrix} \nabla f^1(p_1^*) \\ \cdots \\ \nabla f^m(P_m^*) \end{bmatrix} \times \begin{bmatrix} p_1'' - p_1' \\ \cdots \\ p_n'' - p_n' \end{bmatrix} = \begin{bmatrix} \nabla f^1(p_1^*) \cdot (p'' - p') \\ \cdots \\ \nabla f^m(p_m^*) \cdot (p'' - p') \end{bmatrix}. \end{aligned}$$

Now compare the last two displayed equations. □

For no particularly good reason, we now give an alternate proof to the approximation property of the Jacobian matrix (Theorem 21.2).

Second Proof of Theorem 21.2: By the Big mean value theorem (Theorem 26.2), $T(p) - T(p_0) = L^* \times (p - p_0)^t$ where $L^* := (D_j f^i(p_i^*))$. Look at the matrix

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

[27] How does M differ from the Jacobian matrix of T?

[28] Note that $M = (\nabla f^1(p_1^*), \cdots, \nabla f^m(p_m^*))^t$

entries of $L^* - J_T(p_0) = (a_{ij})$; they are $a_{ij} = D_j f^i(p_i^*) - D_j f^i(p_0)$. By Lemma 25.2, for all column vectors $q \in \mathbf{R}^n$,

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

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,

$$\begin{aligned}
\frac{|T(p) - T(p_0) - J_T(p_0) \times (p - p_0)^t|}{|p - p_0|} &\leq \frac{|(L^* - J_T(p_0)) \times (p - p_0)^t|}{|p - p_0|} \\
&\leq \frac{\left(\sum_{i,j} (a_{ij})^2\right)^{1/2} |p - p_0|}{|p - p_0|} \\
&= \left(\sum_{i,j} (D_j f^i(p_i^*) - D_j f^i(p_0))^2 \right)^{1/2} \\
&\to 0 \text{ as } p \to p_0,
\end{aligned}$$

because, as $p \to p_0$, each $p_i^* \to p_0$ and T is of class C^1. □

26.2 The local invertibility theorem

The following simple one-dimensional illustration gives the flavor of the statement and proof of the local invertibility theorem, Theorem 26.3. Let $f : D \to \mathbf{R}$ be differentiable on an open set $D \subset \mathbf{R}$ 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'',

$$f(x'') - f(x') = f'(\tilde{x})(x'' - x'). \tag{44}$$

If $f(x'') = f(x')$, then since $f'(\tilde{x}) \neq 0$, (44) implies $x'' = x'$.

Theorem 26.3 (Local invertibility,Theorem 14,page 355 of Buck) *Let $T : D \to \mathbf{R}^n$ be a transformation of class C^1 defined on an open set $D \subset \mathbf{R}^n$ and suppose that*[29]

$$\det J_T(p) \neq 0 \textit{ for all } p \in D.$$

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[30] set $\Omega := D \times \cdots \times D \subset \mathbf{R}^n \times \cdots \times \mathbf{R}^n$. The set Ω is a subset of $\mathbf{R}^{n^2}$. Here is the trick: define a function $F : \Omega \to \mathbf{R}$ by

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

[29]note that $J_T(p)$ is an n by n matrix, so its determinant makes sense

[30]If $(p_1, \ldots, p_n) \in D \times \cdots \times D$, let $B(p_j, \delta_j) \subset D$ and let $\delta := \min\{\delta_1, \ldots, \delta_n\}$. Then $B((p_1, \ldots, p_n), \delta) \subset D \times \cdots \times D$

We note first that F is a continuous function on Ω 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 Ω of the form $(p, \ldots, p)$ is given by $F(p, \ldots, p) = \det[D_j f^i(p)] = \det J_T(p)$ 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

$$F(p_1, \ldots, p_n) \neq 0 \text{ for every } (p_1, \ldots, p_n) \in B(p_0, \delta) \times \cdots \times B(p_0, \delta). \tag{45}$$

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

To prove this claim, we use the Mean value theorem for transformations, Theorem 26.2. 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 $M = [D_j f^i(p_i^*)]$,

$$T(p'') - T(p') = M \times (p'' - p')^t. \tag{46}$$

Now $\det M = F(p_1^*, \ldots, p_n^*) \neq 0$ by (45), so M is non-singular. Since we are assuming $T(p'') = T(p')$, (46) shows $p'' - p' = 0$. □

27 Tuesday November 22, 2005—Open Mapping Theorem

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

Lemma 27.1 (Theorem 11,page 133 of Buck) *Let $f : D \to \mathbf{R}$ be of class C^1 on an open set $D \subset \mathbf{R}^n$ 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, $\nabla f(p_0) = 0$.*

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,

$$D_j f(p_0) = \lim_{t \to 0} \frac{f(p_0 + te_j) - f(p_0)}{t}. \tag{47}$$

In (47), the numerator is non-negative whenever $p_0 + te_j \in B(p_0, r)$. Thus if we let t approach zero through positive values, we get $D_j f(p_0) \geq 0$, whereas if we let t approach zero through negative values, we get $D_j f(p_0) \leq 0$. Thus $D_j f(p_0) = 0$. □

We shall also use the following fact about compact sets.

Assignment 31 (Due December 2) *Prove that if K is a compact set in $\mathbf{R}^n$ and $q \notin K$, then*

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

Theorem 27.2 (Open mapping,Theorem 15,page 356 of Buck) *Let $T : D \to \mathbf{R}^n$ be a transformation of class C^1 defined on an open set $D \subset \mathbf{R}^n$ and suppose that*

$$\det J_T(p) \neq 0 \textit{ for all } p \in D.$$

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

Proof: Let $q_0 \in T(D)$. Choose a point $p_0 \in D$ such that $q_0 = T(p_0)$. By Theorem 26.3, 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 31, $d := \inf\{|q_0 - q| : q \in T(C)\} > 0$.

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 ϕ 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:

$$\forall p \in N, \quad |T(p) - q_1|^2 \geq |T(p^*) - q_1|^2. \tag{48}$$

CLAIM 2: $p^* \in \operatorname{int} N$, 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 $p \in C$,

$$|T(p) - q_1| \geq |T(p) - q_0| - |q_0 - q_1| \geq d - d/3 = 2d/3. \tag{49}$$

Note that $T(p_0) = q_0$, and $|q_0 - q_1| < d/3$. Suppose now that $p^* \in C$. Then we would have on the one hand, by (49), $|T(p^*) - q_1| \geq 2d/3$, and on the other hand, by (48), $|T(p^*) - q_1| \leq |T(p_0) - q_1| < d/3$, a contradiction, proving claim 2.

By claim 2, p^* is an interior point of N so that by Lemma 27.1, $D_j\phi(p^*) = 0$ for $1 \leq j \leq n$.

We now need to write down some explicit formulas for the function ϕ. 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. □

28 Friday November 25, 2005—Holiday

(Thanksgiving)

29 Monday November 28, 2005—Inverse Function Theorem

29.1 Automatic continuity of the inverse

The special case of Theorem 29.2 below, in which $m = n = 1$ and D is a compact interval, is proved in [Ross 18.4,18.6]. Before stating and proving Theorem 29.2, let's state a very simple and very useful lemma.

Lemma 29.1 *A sequence of points in $\mathbf{R}^n$ converges to a point $p \in \mathbf{R}^n$ if and only if every subsequence of the given sequence has a subsequence which converges to p.*

Assignment 32 (Due December 7—the day of the final exam) Prove Lemma 29.1.

Assignment 33 (Due December 7—the day of the final exam) *If a transformation preserves convergent sequences, then it is continuous.* (Same proof as [Buck, Theorem 2,page 74].)

The following theorem in the case of functions was a problem on the take-home midterm.

Theorem 29.2 (Automatic continuity of inverse,Theorem 13,page 353 of Buck) *Let $T : D \to \mathbf{R}^m$ be a continuous one-to-one transformation defined on a compact set $D \subset \mathbf{R}^n$. 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 33 all we need to do is prove $\lim_{k\to\infty} p_k = p$. For this we shall use Lemma 29.1. So let p_{k_j} be a subsequence of p_k. By the BW property there is a further subsequence $p_{k_{j_l}}$ and a point $q \in D$ such that

$$\lim_{l\to\infty} p_{k_{j_l}} = q.$$

Since T is continuous, $\lim_{l\to\infty} T(p_{k_{j_l}}) = T(q)$. But $T(p_{k_{j_l}})$ is a subsequence of $T(p_k)$ so $T(p_{k_{j_l}}) \to T(p)$. Thus $T(p) = T(q)$ and since T is one-to-one, $p = q$. By Lemma 29.1, $\lim_k p_k = p$. □

29.2 The inverse function theorem

The inverse function theorem (Theorem 29.4 below) is the n-dimensional analog of the following result in one-variable which we state here for comparison purposes.

Theorem 29.3 (Theorem 29.9,page 165 of Ross) *Let f be a one-to-one continuous function on an open interval $I \subset \mathbf{R}$ 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 29.4 (Inverse Function Theorem, Theorem 16,page 358 of Buck) *Let $T : D \to \mathbf{R}^n$ be a transformation of class C^1 defined on an open set $D \subset \mathbf{R}^n$ and suppose that*

$$\det J_T(p) \neq 0 \text{ for all } p \in D.\text{[31]}$$

Suppose also that T is one-to-one on D. Then the inverse T^{-1} (which exists and is defined on the open subset $T(D) \subset \mathbf{R}^n$) *is of class C^1 on $T(D)$ and*

$$J_{T^{-1}}(T(p)) = [J_T(p)]^{-1} \text{ for all } p \in D. \tag{50}$$

Proof: Since T is of class C^1, by Theorem 21.2, (considering $T(p)$ and $T(p_0)$ as column vectors)

$$T(p) - T(p_0) = J_T(p_0) \times (p - p_0)^t + R(p) \tag{51}$$

where

$$\lim_{p\to p_0} \frac{|R(p)|}{|p - p_0|} = 0. \tag{52}$$

By assumption $\det J_T(p_0) \neq 0$ so $J_T(p_0)$ is non-singular. Multiplying (51) (on the left) by $[J_T(p_0)]^{-1}$, you get

$$[J_T(p_0)]^{-1}(T(p) - T(p_0)) = (p - p_0)^t + [J_T(p_0)]^{-1}(R(p)). \tag{53}$$

[31] so that $J_T(p)^{-1}$ exists

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 (53),

$$T^{-1}(q) - T^{-1}(q_0) = (p - p_0)^t = [J_T(p_0)]^{-1}(T(p) - T(p_0)) - [J_T(p_0)]^{-1}(R(p)),$$

that is (eliminating the middle person $(p - p_0)^t$),

$$T^{-1}(q) - T^{-1}(q_0) - [J_T(p_0)]^{-1}(T(p) - T(p_0)) = -[J_T(p_0)]^{-1}(R(p)). \tag{54}$$

If we can show that the right hand side of (54) satisfies

$$\lim_{q \to q_0} \frac{|[J_T(p_0)]^{-1}(R(p))|}{|q - q_0|} = 0, \tag{55}$$

then (54) will say that (50) is true. So we need to prove (55).

First recall that by Lemma 25.2 there is a constant M such that $|[J_T(p_0)]^{-1}(u)| \leq M|u|$ for all $u \in \mathbf{R}^n$. Therefore,

$$\frac{|[J_T(p_0)]^{-1}(R(p))|}{|q - q_0|} \leq \frac{M|R(p)|}{|q - q_0|}. \tag{56}$$

By (53), $(p - p_0)^t = [J_T(p_0)]^{-1}(T(p) - T(p_0)) - [J_T(p_0)]^{-1}(R(p))$ so

$$|p - p_0| \leq M|q - q_0| + M|R(p)|, \tag{57}$$

and by (52),

$$|R(p)| \leq \epsilon|p - p_0| \text{ for } |p - p_0| < \delta \quad (\delta \text{ depending on } \epsilon). \tag{58}$$

Therefore, (57) 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,

$$|p - p_0| \leq \frac{M}{1 - \epsilon M}|q - q_0| \text{ for } |p - p_0| < \delta. \tag{59}$$

Taking recriprocals in (59) you get

$$\frac{1}{|q - q_0|} \leq \frac{M}{1 - \epsilon M} \, \frac{1}{|p - p_0|} \text{ for } |p - p_0| < \delta. \tag{60}$$

Now by (56),(60), and (58), 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 ϵ (since it goes to zero as ϵ does). Therefore (55) holds. Note that we have used the fact that T^{-1} is continuous (Theorem 29.2). 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 (50) then, the entries of $J_{T^{-1}}(T(p))$ are continuous functions of $q = T(p)$. □

Assignment 34 (Due December 7) [Buck, §7.5,page 361,#11,14]

30 Wednesday November 30, 2005—Implicit Function Theorem

30.1 Motivation

In much of *analysis*, the linear functions are the easiest to work with[32]. Let $F : \mathbf{R}^n \to \mathbf{R}$ 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 $(x,y) \in \mathbf{R}^2$ so that $F : D \to \mathbf{R}$ where $D = \mathbf{R}^2$. Note that $\frac{\partial F}{\partial y}(x,y) = 2y$.

Suppose that $(x_0, y_0) \in \mathbf{R}^2$ 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 ϕ, 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 ϕ 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 $\phi(x) =$

[32] This is not necessarily the case for *linear algebra*

$+\sqrt{1-x^2}$ 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 $F(x, y) = x + 2y + x^2y^5 - 8$, for $(x, y) \in \mathbf{R}^2$. 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 ϕ 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 $x + 2y + x^2y^5 - 8 = 0$ 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 ϕ 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 continous.

Let's return to our second example, that is, $F(x, y) = x^2 + y^2 - 1$ for $(x, y) \in \mathbf{R}^2$ so that $F : D \to \mathbf{R}$ where $D = \mathbf{R}^2$. Of course F is a function. Let's construct a related *transformation* $T_F : D \to \mathbf{R}^2$ 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 35 (Due December 7) *Show that, for $F = x^2 + y^2 - 1$, $T = T_F$ is not one-to-one on $D = \mathbf{R}^2$ and $T(\mathbf{R}^2)$ is not an open subset of $\mathbf{R}^2$.*

Suppose again that $(x_0, y_0) \in \mathbf{R}^2$ 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

$$\det J_T(x, y) = \frac{\partial F}{\partial y}(x, y).$$

30.2 Implicit function theorems

Since we have just introduced most of the ideas in its proof, it seems appropriate now to state the implicit function theorem.

Theorem 30.1 (Theorem 17,page 363 of Buck,"downgraded" to two variables) *Let $F : D \to \mathbf{R}$ be of class C^1 on an open set $D \subset \mathbf{R}^2$, 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 $\phi : (x_0 - r, x_0 + r) \to \mathbf{R}$ 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 30.1, let's reiterate exactly all that it says.

- There is (theoretically!) a function ϕ, such that for each x close enough to x_0, $y = \phi(x)$ is a solution[33] of the equation $F(x, y) = 0$
- As a function of x, ϕ is continuous
- Actually, ϕ is differentiable
- Actually, the derivative of ϕ is a continuous function[34]
- Question: Can we calculate $\phi'(x)$ by implicit differentiation and the chain rule?[35]

Proof of Theorem 30.1: 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" (Theorem 26.3), 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 $T^{-1} : T(B) \to B$. 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[36]. 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" (Theorem 27.2), $T(B')$ 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

$$v = F(g(u, v), h(u, v)) = F(u, h(u, v)). \tag{61}$$

Substitute for (u, v), any point of the form $(x, 0) \in B((x_0, 0), r)$. From (61), we have

$$0 = F(x, h(x, 0)) \text{ for all } |x - x_0| < r.$$

Thus, if we define $\phi(x) = h(x, 0)$ for $|x - x_0| < r$, we have the desired function ϕ. Note that by the chain rule, $\phi'(x) = \frac{\partial h}{\partial x}(x, 0)$ so that ϕ is of class C^1 on $(x_0 - r, x_0 + r)$. This completes the proof. □

[33] This already says a lot! If you stop here you got a bargain.
[34] This statement implies the previous two statements
[35] Yes, but it is not entirely satisfactory because the answer is in terms of $\phi(x)$
[36] What is the reason for this?

We now state a version of the implicit function theorem in 3 variables. We refer to Buck 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 30.2 (Theorem 17,page 363 of Buck—three variables) *Let $F : D \to \mathbf{R}$ be of class C^1 on an open set $D \subset \mathbf{R}^3$, let $(x_0, y_0, z_0) \in D$, and suppose that $F(x_0, y_0, z_0) = 0$ and $\frac{\partial F}{\partial z}(x_0, y_0, z_0) \neq 0$. Then there exists a $r > 0$ and a function $\phi : B((x_0, y_0), r) \to \mathbf{R}$ 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.

Theorem 30.3 *Let $F : D \to \mathbf{R}$ be of class C^1 on an open set $D \subset \mathbf{R}^n$, let $(x_1^0, \dots, x_n^0)$ be a point of D, and suppose that*

$$F(x_1^0, x_2^0, \dots, x_n^0) = 0 \text{ and for some } k,\ \frac{\partial F}{\partial x_k}(x_1^0, x_2^0, \dots, x_n^0) \neq 0.$$

Then there exists $r > 0$ and a function

$$\phi : B((x_1^0, \dots, x_{k-1}^0, x_{k+1}^0, \dots, x_n^0), r) \to \mathbf{R}$$

of class C^1 on $B((x_1^0, \dots, x_{k-1}^0, x_{k+1}^0, \dots, x_n^0), r) \subset \mathbf{R}^{n-1}$, such that

$$\phi(x_1^0, \dots, x_{k-1}^0, x_{k+1}^0, \dots, x_n^0) = x_k^0$$

and

$$F(x_1, \dots, x_{k-1}, \phi(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n), x_{k+1}, \dots, x_n) = 0$$

for all $(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n) \in B((x_1^0, \dots, x_{k-1}^0, x_{k+1}^0, \dots, x_n^0), r)$.

If we introduce a little notation we can make the last theorem easier to read.

> Let $F : D \to \mathbf{R}$ be of class C^1 on an open set $D \subset \mathbf{R}^n$, let $p_0 = (x_1^0, \dots, x_n^0)$ 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 $p_0^{(k)} = (x_1^0, \dots, x_{k-1}^0, x_{k+1}^0, \dots, x_n^0)$. Then there exists $r > 0$ and a function $\phi : B(p_0^{(k)}, r) \to \mathbf{R}$ of class C^1 on $B(p_0^{(k)}, r) \subset \mathbf{R}^{n-1}$, such that, with $p = (x_1, \dots, x_n)$ and $p^{(k)} = (x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$, we have $\phi(p_0^{(k)}) = x_k^0$ and $F(x_1, \dots, x_{k-1}, \phi(p^{(k)}), x_{k+1}, \dots, x_n) = 0$ 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 $x_1, \dots, x_n$ can be solved in terms of the remaining variables. The situation is described in [Buck, Theorem 18,page 364], and the discussion on page 366 of Buck.

We now present some examples in the form of exercises.

Assignment 36 (Due December 7) *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* $\mathbf{R}^3$*:* $x_0^2+y_0^2+z_0^2 = 1$*, that is,* $F(x_0,y_0,z_0) = 0$*. "Prove" that*[37] $z = \phi(x,y) := \sqrt{1-x^2-y^2}$ *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 37 (Due December 7) *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 38 (Due December 7) *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 39 (Due December 7) *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

$$z = \phi(x,y) \text{ for } (x,y) \text{ close to } (0,\pi/2),$$

as well as

$$x = \psi(y,z) \text{ for } (y,z) \text{ close to } (\pi/2,\pi),$$

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 40 (Due December 7) [Buck, §7.6,page 366,#1,2,5]

[37] Don't laugh, you need to assume that $x^2+y^2 < 1$

31 Friday December 2, 2005—Review of course

31.1 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

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

7.6 implicit functions—implicit function theorems, Theorems 17,18

31.2 The four theorems on transformations

We proved these four theorems in class on November 21,22,30. 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 is the climax of this course, and is a very useful result in almost any branch of analysis. Even in one variable, the inverse function theorem is not so easy. We recalled the statement (but not the proof) of the one-variable result below for motivation (see Theorem 29.3).

Here is a summary of the four theorems. We presented them in a slightly different order from that of [Buck, §7.6]. We shall give each of these theorems a "nickname".

“Automatic continuity of inverse”	Theorem 29.2 ([Buck, Theorem 13,page 353])
Hypothesis	Conclusion
T continuous,one-to-one on compact $D \subset \mathbf{R}^n$	T^{-1} is continuous

“Local invertibility”	Theorem 26.3 ([Buck, Theorem 14,page 355])
Hypothesis	Conclusion
T is of class C^1 on open $D \subset \mathbf{R}^n$ and $\det J_T(p_0) \neq 0$	T is locally one-to-one at p_0

“Open mapping”	Theorem 27.2 ([Buck, Theorem 15,page 356])
Hypothesis	Conclusion
T is of class C^1 on open $D \subset \mathbf{R}^n$ and $\det J_T(p) \neq 0$ for all $p \in D$	$T(D)$ is an open set

“Inverse function”	Theorem 29.4 (Buck, [Theorem 16,page 358])
Hypothesis	Conclusion
T is of class C^1 on open $D \subset \mathbf{R}^n$ and $\det J_T(p) \neq 0$ for all $p \in D$ T is globally one-to-one on D	T^{-1} is of class C^1 on $T(D)$ and $(T^{-1})'(T(p)) = (T'(p))^{-1}$ and $J_{T^{-1}}(T(p)) = (J_T(p))^{-1}$

31.3 Functions vs. Transformations

thing to be differentiated	function $f : \mathbf{R} \to \mathbf{R}$	function $f : \mathbf{R}^n \to \mathbf{R}$	transformation $T : \mathbf{R}^n \to \mathbf{R}^m$
notations for derivative	$f'(a)$	$\frac{\partial f}{\partial x_j}(p)$,$D_j f(p)$, $\nabla f(p)$	$T'(p)$, $J_T(p)$
differentiability implies continuity	[Ross, 28.2]	[Buck, Cor,p129]	Lemma 25.3 of these notes
approximation	(13) p.32 of these notes	[Buck, Thm8,p131]	[Buck, Thm10,p344]
algebra of continuity and differentiation	[Ross, 17.4,28.3]	[Buck, Thm4,p77]	just a vector space
chain rule	[Ross, 17.5,28.4]	[Buck, Thm14,p136]	[Buck, Thm11,p346]
critical points	[Ross, 29.1]	[Buck, Thm11,p133]	doesn’t make sense
Rolle’s theorem	[Ross, 29.2]		
Mean value theorem	[Ross, 29.3]	[Buck, Thm16,p151]	[Ross, Thm12,p350]
Inverse function theorem	[Ross, 29.9]	doesn’t make sense	[Buck, Thm16,p358]
local invertibility	p.49 of these notes		[Buck, Thm14,p355]
automatic continuity of inverse	[Ross, 18.4,18.6]		[Buck, Thm13,p353]
open mapping theorem			[Buck, Thm15,p356]
implicit function theorem	doesn’t make sense	[Buck, Thm17,18,p363-4]	

32 Wednesday December 7, 2005—Final Examination 1:30-3:30 pm