Analysis in Several Variables
Math 140C—Fall 2006

Bernard Russo

December 2, 2006

Contents

1 Friday September 22—Course information; Schwarz inequality (Assignment 1)

1.1 Course Information

- Course: Mathematics 140C MWF 1:00–1:50 ET 204 FALL 2006
  Webpage for the course: *www.math.uci.edu/∼brusso*

- Prerequisite: Math 140AB. Rigorous study of differentiation and integration of real-valued functions of one real variable. All of this can be found in the six chapters of the recent text for 140AB, namely, *Elementary Analysis:The Theory of the Calculus*, by Kenneth A. Ross. This includes the set of real numbers and the completeness axiom; sequences of real numbers, continuity, uniform continuity, sequences and series of functions, differentiation and integration up to the fundamental theorem of calculus.

- Instructor: Bernard Russo MSTB 263 Office Hours M 2:30-3:30 W 10:30-11:30 and by appointment (a good time for short questions is right after class just outside the classroom; appointments can be arranged by email—*brusso@uci.edu*)

- Discussion section: TuTh 1:00–1:50 HICF 100M

- Teaching Assistant: TBA

- Homework: There will be approximately 35 to 40 assignments with about one week's notice before the due date. Most, but not all of these assignments will be from the textbook (Buck).

- Grading: The in-class exams are "closed book and notes." Homework and take home midterm are "open book and notes".

First midterm (in class)	October 20 (Friday of week 4)	20 percent
Second midterm (take home)	November 17 (Friday of week 8)	20 percent
Final Exam (in class)	December 6 (Wednesday)	40 percent
Homework	approximately 35-40 assignments	20 percent

- Holidays: November 10, 23, and 24

- Text: R. C. Buck, Advanced Calculus

- Material to be Covered. (Page numbers refer to the text Buck)

  Schwarz inequality Theorem 1, page 13 (1 lecture)

  topology §1.5 pp 28–33: open, closed, boundary, interior, exterior, closure, neighborhood, cluster point (5 lectures)

  compactness §1.8 pp 64–67: Heine-Borel and Bolzano-Weierstrass properties (Theorems 25,26,27, page 65) (3 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) (3 lectures)

differentiation (of functions) §3.3: Implies continuity, characterization by approximation (Corollary, page 129 and Theorem 8, page 131) (2 lectures)

integration §4.2: Integrability of continuous functions (Theorems 1,4 on pages 169,176) (5 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) (9 lectures)

1.2 Schwarz 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 Theorem 1.1 below) can be phrased compactly as

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

Theorem 1.1 (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: Let $Q := \alpha p - \beta q$ where α and β are unspecified real numbers. From $|Q|^2 \geq 0$ we obtain

$$\alpha^2|p|^2 + \beta^2|q|^2 - 2\alpha\beta p \cdot q \geq 0 \text{ for all } \alpha, \beta \in \mathbf{R}.$$

Choosing $\alpha = |q|$ and $\beta\|p|$, we have $2|p||q|p \cdot q \leq 2|q|^2|p|^2$ from which the theorem follows. □

Assignment 1 (Due September 29)

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 #1,2,5,6]

THINKING OUTSIDE THE BOX

- If x_j and y_j are infinite sequences, then

$$\sum_{j=1}^{\infty} x_j y_j \leq \left(\sum_{j=1}^{\infty} x_j^2\right)^{1/2} \left(\sum_{j=1}^{\infty} y_j^2\right)^{1/2},$$

provided the series on the left converges.

- If f and g are continuous functions on a closed interval $[a, b]$, then from the Schwarz inequality applied to Riemann sums

$$\begin{aligned}\sum_{j=1}^{n} f(t_j)g(t_j)(x_j - x_{j-1}) &= \sum_{j=1}^{n} f(t_j)(x_j - x_{j-1})^{1/2} g(t_j)(x_j - x_{j-1})1/2 \\ &\leq \left(\sum_{j=1}^{n} f(t_j)^2(x_j - x_{j-1})\right)^{1/2} \left(\sum_{j=1}^{n} g(t_j)^2(x_j - x_{j-1})\right)^{1/2},\end{aligned}$$

you get the Schwarz inequality for functions $\int_a^b f(x)g(x)\,dx \leq \left(\int_a^b f(x)^2\,dx\right)^{1/2} \left(\int g(x)^2\,dx\right)^{1/2}$.

- If you define $f \cdot g = \int_a^b f(x)g(x)\,dx$, then $f \cdot g$ has the same properties as the scalar product $p \cdot q$ and the proof above of Theorem 1.1 applies word for word to give an alternate proof of the Schwarz inequality for functions.

2 Monday, September 25—The triangle inequality and open sets (Assignment 2)

2.1 The triangle inequality

Here is an important consequences of the Schwarz inequality.

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

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

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 2.2 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 2.3 *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 2.4 ((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). □

Assignment 2 (Due October 6)

- Buck [§1.4 page 27 #3,15,16]
- Buck [§1.5 page 36 #1,5,9,13]

3 Wednesday September 27—More on open sets; closed sets, boundary, and closure (Assignments 3,4,5,6)

3.1 The final word on open sets (just kidding!)

Assignment 3 (Due October 6) Fix $p \in \mathbf{R}^n$. Show that $\{q \in \mathbf{R}^n : |q - p| > 2\}$ is an open set.

The next assignment outlines another proof of Schwarz's inequality, by asking you to prove Young's inequality which is as follows. The two corollaries (including Hölder's inequality) are proved for you. You get the Schwarz inequality from Hölder's inequality by taking $p = 2$.

Theorem 3.1 (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 $a, b \in [0, \infty)$,*

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

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

Corollary 3.2 *For $p \in (1, \infty)$, and $a, b \in [0, \infty)$,*

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

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

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

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

Corollary 3.3 (Hölder Inequality) *Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be real numbers and let $p \in (1, \infty)$. Then with $q := p/(p-1)$,*

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

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

Assignment 4 (Due October 13)
Give a rigorous proof of Theorem 3.1. 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.

Remark 3.4 *Every open set in $\mathbf{R}^n$ is the union of (not necessarily disjoint) open balls.*

Assignment 5 (Due October 13)

- Show that in $\mathbf{R}^1$, the open balls can be assumed to be disjoint
- Show that every open set in $\mathbf{R}^n$ is the union of a countable collection of open balls. (Hint: The answer is somewhere in the minutes for my 140C class of Fall 2005)

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.

3.2 Closed sets

Definition 3.5 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 3.6 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 2.4 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.

3.3 Boundary and closure

Definition 3.7 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 $S \cup \operatorname{bdy} S$.

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 3.8 ((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 6 (Due October 6) Prove the following assertions:

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

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

4 Friday, September 29—more on closed sets; cluster points (Assignment 7)

4.1 More on closed sets and boundary

We already mentioned the next proposition last time.

Proposition 4.1 ((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 4.2 (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[1]

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

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

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

Proposition 4.3 (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:

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

4.2 Cluster points

Definition 4.4 p is a *cluster point* of S if every ball with center p meets S in infinitely many points, that is, for every $\delta > 0$, the set $B(p,\delta) \cap S$ contains infinitely many points. We denote the set of cluster points of a set S by $\text{cl}\, S$.

Remark 4.5 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).

[1] Be sure to check this carefully

Proposition 4.6 ((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 $p_1, \dots, p_m$. Since these points are in S and $p \notin S$, if we set

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

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 4.6. □

Assignment 7 (Due October 6) [Buck §1.5 page 36 #2,6,10,11]

4.3 Proof of Proposition 3.8 ((vii) on page 32 of Buck)

Proof of Proposition 3.8[2]

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 3.8. □

[2] This proof was not done in class. Please make sure you read and understand it

5 Monday October 2—Compactness I (Assignment 8)

5.1 Bolzano-Weierstrass and Heine-Borel properties

Definition 5.1 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 $\mathcal{G}$ is a collection of open sets and if $S \subset \cup_{G \in \mathcal{G}} G$, then there is a finite subset $G_1, \ldots, G_N$ of $\mathcal{G}$ such that $S \subset G_1 \cup G_2 \cup \cdots \cup G_N$.

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 will be stated in a proposition below.

Definition 5.2 Let S be any subset of $\mathbf{R}^n$. We say S is *compact* if it satisfies HB.

Assignment 8 (Due October 13) 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—it is included here for comparison purposes only) If S satisfies HB, then S is a closed set.

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

5.2 HB implies closed

Proposition 5.3 *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 $\delta_q := |p-q|/2$. 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[3] 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.

6 Wednesday October 4—Compact sets II

6.1 Equivalence of HB and BW

Proposition 6.1 *Let* S *be any subset of* $\mathbf{R}^n$*. Then* S *satisfies BW if and only if it satisfies HB.*

Proof:

Step 1: HB$\Rightarrow$ BW.

Proof: Let T be an infinite sequence in S, and suppose that T has no cluster point in S. We seek a contradiction, which will then complete the proof of Step 1.

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.

Step 2: 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 $p \in B(p, r_p) \subset G_p$.

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

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 $S \subset \cup_{j=1}^{\infty} G_{p_j}$.

Step 3: BW$\Rightarrow$ HB.

Proof: Assume that S satisfies BW. By step 2, it suffices to prove that any countable open cover of S can be reduced to a finite subcover.

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[4] $p_n \notin \{p_1, \ldots, p_{n-1}\}$ 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 infinite 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 3 is proved and this completes the proof of Proposition 6.1. □

7 Friday October 6—Compactness III

7.1 Corrected proof of Step 3 of Proposition 6.1

Let $S \subset G_1 \cup G_2 \cup \cdots$. Since $S \not\subset G_1$, choose $p_1 \in S - G_1$. Choose n_1 such that $p_1 \in G_{n_1} - (G_1 \cup \cdots \cup G_{n_1 - 1})$. Since $S \not\subset G_1 \cup \cdots \cup G_{n_1}$, choose $p_2 \in S - (G_1 \cup \cdots \cup G_{n_1})$ and choose n_2 such that $p_2 \in G_{n_2} - (G_1 \cup \cdots \cup G_{n_2 - 1})$. Continuing in this way we obtain a sequence of distinct points $T := \{p_k\}_{k=1}^{\infty} \subset S$ and a subsequence $n_1 < n_2 < \cdots$ such that for each $k \ge 1$,

$$p_k \in [S - (G_1 \cup \cdots \cup G_{n_{k-1}})] \cap [G_{n_k} - (G_1 \cup \cdots \cup G_{n_k - 1})]. \tag{4}$$

By BW, there is a point $p \in \operatorname{cl} T \cap S$. Choose k_0 such that $p \in G_{k_0}$ and then choose $\delta > 0$ such that $B(p, \delta) \subset G_{k_0}$. Since $B(p, \delta) \cap T$ is infinite, there exists $m > k_0$ such that $p_m \in B(p, \delta)$ and thus $p_m \in G_{k_0}$. But by (4), $p_m \notin G_1 \cup \cdots \cup G_{n_m - 1}$. This contradicts the fact that $p_m \in G_{k_0}$, since $k_0 < m \le n_m - 1$. □

[4] This does not seem to be true; however all you need to know is that the sequence $\{p_k\}_{k=1}^{\infty}$ is actually infinite. See subsection7.1

7.2 Closed and Bounded implies Compact

Theorem 7.1 *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 7.1.

Lemma 7.2 (Bolzano-Weierstrass theorem in R) *Every bounded sequence of real numbers has a convergent subsequence.*

Lemma 7.3 *Every subsequence of a convergent sequence of real numbers converges to the same limit as the sequence.*

Proof of Theorem 7.1:
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 7.2 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 7.2 again) and call this limit x_2. By Lemma 7.3, the first coordinates of T_2 also converge to the previous x_1.

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

$$\lim_{k\to\infty} x_j^{(k)} = x_j \quad 1 \le j \le n. \tag{5}$$

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 (5), 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 7.1. □

8 Monday October 9—Continuity I (Assignment 9)

8.1 Overview

Here is a preview of our next topic: **continuous functions**. There are two major 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[5]: 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 forms for $n = 1$.

- (Ross, Theorem 18.1, p. 95) 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, which will be proved below.)

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

- (Ross, Theorem 19.2, p. 103) A continuous real valued function on a closed and bounded interval in **R** is uniformly continuous on that interval. (Application: a continuous function on a closed interval in **R** is Riemann integrable)

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. At least one of these will be proved below.

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.

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 (p. 89), 9 (p. 90) and 12 (p. 92) can be skipped (we need Theorem 8 later, but we can wait and prove it later). Theorems 14 (p. 93), 15 (p. 94), 16 (p. 95) involve connectedness and we shall skip them now.

8.2 Continuous functions—continuous image of a compact set

Definition 8.1 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[6]

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

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

It is important to realize that this lengthy definition can be put in the compact [7] 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 9 (Due October 27) [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 8.2 *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 shallo show that $f(D)$ satisfies the HB property. By Lemma **??**, we only need to deal with *countable* open covers. We shall use the fact that D satisfies the HB property (for *arbitrary* covers!).

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[8]:

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

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. $\square$

[7] no pun intended

[8] the redundant cover!

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).

9 Wednesday October 11—Continuity II (Assignment 10)

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

Definition 9.1 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:

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

Then

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

This proves the following:

Theorem 9.2 (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

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

Theorem 9.3 (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

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

that is, the sequence is bounded. □

9.2 Continuity and limits of sequences

Theorem 9.4 (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

$$f[D \cap B(p_0, \delta)] \subset B(f(p_0), \epsilon). \tag{6}$$

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

$$p_k \in B(p_0, \delta) \text{ for } k > N. \tag{7}$$

Putting together (6) and (7) results in $f(p_k) \in B(f(p_0), \epsilon)$ for $k > N$. □

Remark 9.5 Theorem 2 on page 74 of Buck is an important converse to Theorem 9.4. We omit the details of the straightforward (indirect) proof.

Theorem 9.6 (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 trivial from Theorem 8.2: $f(D)$ is compact, hence bounded. □

Lemma 9.7 *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 9.8 (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 9.6, 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

$$\beta - \frac{1}{k} \leq f(p_k) \leq \beta. \tag{8}$$

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 9.7, there is a subsequence p_{k_j} such that $\lim_{j\to\infty} p_{k_j} = p_0$. In particular, from (8), 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 20) [Buck, §1.6 page 54 #1, 2, 3, 4, 32, 35]

10 Friday October 13—Continuity III (Assignment 11)

10.1 Uniform continuity

Definition 10.1 (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 10.2 (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 $\delta = \min_{1 \le j \le m}\{\delta_{p_j}/2\}$. 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. □

There are non-trivial uniformly continuous functions on non-compact sets.

(A) $f(p) = |p|$ is uniformly continuous from $\mathbf{R}^n$ to $\mathbf{R}$ by the "backwards triangle inequality": $||p| - |q|| \le |p - q|$.

(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 is uniformly continuous from $\mathbf{R}^n$ to $\mathbf{R}$ by the Schwarz inequality: $|p \cdot q - p' \cdot q| = |(p - p') \cdot q| \le |p - p'||q|$.

Assignment 11 (Due October 27) [Buck, §2.3 page 88 #1–7]

11 Monday October 16—Differentiability implies continuity I (Assignment 12)

11.1 Motivation—one variable

Let's begin by recalling the mean value theorem in one variable. (See Theorem 29.3, page 163 of Ross) We shall use Lemma 11.1 (a result in one dimension) in the proof of Theorem 11.3 below (a theorem in $n \ge 1$ dimensions).

Lemma 11.1 (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. (For example, see Ross, page 160: The function f defined by $f(0) = 0$ and $f(x) = x^2 \sin(1/x)$ for $x \ne 0$ 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 11.1 and Theorem 11.2. This is one difference between these two one-dimensional results, and the n-dimensional theorem Theorem 11.3.

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

Theorem 11.2 (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

$$\lim_{x\to c} f(x)=f(c)+f'(c)\cdot 0=f(c).\Box$$

11.2 Partial Differentiation

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 $p_0=(x_1^0,\dots,x_n^0)\in D$, the *partial derivatives* at p_0 are defined by

$$D_1f(p_0)=\lim_{x_1\to x_1^0}\frac{f(x_1,x_2^0,\dots,x_n^0)-f(x_1^0,x_2^0,\dots,x_n^0)}{x_1-x_1^0}=\left.\frac{d}{dx_1}\right|_{x_1=x_1^0} f(x_1,x_2^0,x_3^0,\dots,x_n^0),$$

$$D_2f(p_0)=\lim_{x_2\to x_2^0}\frac{f(x_1^0,x_2,x_3^0,\dots,x_n^0)-f(x_1^0,x_2^0,\dots,x_n^0)}{x_2-x_2^0}=\left.\frac{d}{dx_2}\right|_{x_2=x_2^0} f(x_1^0,x_2,x_3^0,\dots,x_n^0),$$

and so forth, until

$$D_nf(p_0)=\lim_{x_n\to x_n^0}\frac{f(x_1^0,\dots,x_{n-1}^0,x_n)-f(x_1^0,x_2^0,\dots,x_n^0)}{x_n-x_n^0}=\left.\frac{d}{dx_n}\right|_{x_n=x_n^0} f(x_1^0,\dots,x_{n-1}^0,x_n).$$

Some common notations for this are

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

You can also write (if you prefer)

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

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

11.3 Differentiability (+ continuity) implies continuity

We want to prove an analog of Theorem 11.2 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 12) 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 11.3 (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 $D_1 f, \ldots, D_n f$ exist and are continuous at all points of D,. Then f is continuous on 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[9].

Assignment 12 (Due October 27) [Buck, §3.3 page 134 #4,5]

12 Wednesday October 18—Proof of Theorem 11.3

Proof of theorem 11.3: Fix $p_0 \in D$ and let $p \in B(p_0, r) \subset D$ for some $r > 0$.

> Outline: 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
>
> $$f(p) - f(p_0) = f_1(q_1)(x_1 - x_1^0) + f_2(q_2)(x_2 - x_2^0) + \cdots + f_n(q_n)(x_n - x_n^0) \quad (9)$$
>
> 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 $|p - p_0| < \delta$.

Now for the details. 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 $p = (x, y, z)$.

[9] this is a little white lie, see Problem 5 in the next assignment

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

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

(Question: what does d depend on?)

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

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

(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 (9).

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 (9) 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 $p \in D$. □

13 Friday October 20—First Midterm

Problem 1 (20 points) *Prove rigorously that the set* $S = \{0, 1, 1/2, 2/3, 3/4, \ldots\}$ *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* $S \times S := \{(x, y) : x, y \in S\}$ *a closed subset of* $\mathbf{R}^2$*? (Yes or no, no proof required).*

Problem 2 (20 points) *Find* bdy S, int S, *and all cluster points of* S *if*

$$S = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 \leq 4, 1 < x < 2\} \cup \{(x, 0) : 0 \leq x \leq 1\}$$

Just write down your answer, no proof is required.

Problem 3 (20 points) *Show that every point in the boundary of a set is the limit of a sequence of points from the set.*

Problem 4 (20 points) *If* A *and* B *are compact subsets of* $\mathbf{R}$*, show that* $A \times B$ *is a compact subset of* $\mathbf{R}^2$.

Problem 5 (20 points) *Using the fact that the closure of a set is the smallest closed set containing the set, show that for any two sets* A *and* B,

$$\overline{A \cup B} = \overline{A} \cup \overline{B}.$$

14 Monday October 23—More on closed sets and closure (Assignments 13,14)

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.)

- $\mathbf{R}^n - S$ is an open set
- $S = \overline{S}$
- $\text{cl}\, S \subset S$
- $\text{bdy}\, S \subset S$ (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.)

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

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*

$$\overline{S} = \{\lim_{k\to\infty} p_k : \{p_k\} \subset S, \lim_k p_k \text{ exists }\}. \tag{10}$$

Proof: Suppose first that $p = \lim_k p_k$ for some sequence p_k from S. If $p \notin \overline{S} = \text{bdy}\, S \cup S$, then $p \notin S$ and $p \notin \text{bdy}\, S$. Thus there exists $\delta > 0$ such that at least one of $B(p, \delta) \cap S$ or $B(p, \delta) \cap (\mathbf{R}^n \setminus S)$ 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 (10) is contained in the closure of S.

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

Corollary 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.*

Remark 14.3 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: (Corrected October 31) The function $f(x, y, z) := xy - z$ is continuous. The set S is defined by the inequality $f(p) > 0$. If $p_0 \in S$, that is, $f(p_0) > 0$, then by the continuity of f at p_0, there exists $\delta > 0$ such that $f[S \cap B(p_0, \delta)] \subset B(f(p_0), |f(p_0)|/2)$. It follows that $B(p_0, \delta) \subset S$, proving that S is open. □

(NOTE: This shows you how to prove the first general statement above)

Assignment 13 (Due November 3 Hint: This was done in class)

(A) Give a proof of Problem 3(c) on page 37 of Buck using Corollary 14.2.

(B) Use Corollary 14.2 to prove that if A and B are closed sets in $\mathbf{R}$, then $A \times B$ is a closed set in $\mathbf{R}^2$.

Assignment 14 (Due November 3) Let S be any subset of $\mathbf{R}^n$. Using only the definitions of cluster point and boundary, prove the following statements.

- $\text{cl}(\text{cl}\, S) \subset \text{cl}\, S$ (Hint: See the solution to Problem 8(b) on the first midterm for Math 140C, Fall 2005, which is at the top of page 26 of the minutes)
- $\mathbf{R}^n - \text{cl}\, S$ is open
- $\text{bdy}\,(\text{cl}\, S) \subset \text{cl}\, S$
- $\{\lim_k p_k : p_k \in \text{cl}\, S\} \subset \text{cl}\, S$

15 Wednesday October 25—Differential as a Linear approximation (the case of functions)

15.1 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 15.1 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. [10]

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

$$C^1(D) \supset C^2(D) \supset \cdots \supset C^k(D) \supset C^{k+1}(D) \supset \cdots \tag{11}$$

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

$$C^0(I) \supset C^1(I) \supset C^2(I) \supset \cdots \supset C^k(I) \supset C^{k+1}(I) \supset \cdots \tag{12}$$

Notice that (12) has an extra inclusion at the beginning, namely $C^0(I) \supset C^1(I)$, due to Theorem 11.2. We have shown in Theorem 11.3 that (11) has an extra inclusion too, namely $C^0(D) \supset C^1(D)$. (Question: how do these two extra inclusion relations differ from eachother?)

15.2 Linear Approximation

Let's examine the equation (9). 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 (9) can be rewritten as a dot product of vectors:

$$f(p) - f(p_0) = (f_1(q_1), f_2(q_2), \cdots, f_n(q_n)) \cdot (x_1 - x_1^0, x_2 - x_2^0, \cdots, x_n - x_n^0), \tag{13}$$

[10] In [Buck, Definition 1,page 128], the definition of C^k requires that f be continuous. By Theorem 11.3, Buck's definition of C^k and our Definition 15.1 are equivalent

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 (13).

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!

   > Informal exercise: 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 15.2 If $f : S \to \mathbf{R}$ is defined on an open set $S \subset \mathbf{R}^n$, the *total derivative* of f at $p \in S$ is the vector $\mathbf{D}f(p) = (D_1 f(p), D_2 f(p), \cdots, D_n f(p))$. Of course $\mathbf{D}f$ is defined only at those points of D where all first order partial derivatives of f exist.

For $n = 3$, $\mathbf{D}f$ is the *gradient* of f: $\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z})$ and for $n = 1$, $\mathbf{D}f = f'$, a numerical function.

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

$$\lim_{x \to c} \frac{|f(x) - f(c) - f'(c)(x - c)|}{|x - c|} = 0. \tag{14}$$

The following is the analog, for functions of several variables, of (14). 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 (15) 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 15.3 (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

$$\frac{|f(p) - f(p_0) - \nabla f(p_0) \cdot (p - p_0)|}{|p - p_0|} < \epsilon \text{ whenever } p \in B(p_0, \delta) \cap D. \tag{15}$$

Proof: Let $R := f(p) - f(p_0) - \nabla f(p_0) \cdot (p - p_0)$. By (13) (which is the main point in the proof of Theorem 11.3), $f(p) - f(p_0) = V \cdot (p - p_0)$, 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:

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

that is

$$\frac{|R|}{|p - p_0|} \leq |V - \nabla f(p_0)|, \tag{16}$$

and if you write out the coordinates of $V - \nabla f(p_0)$ you will see that $|V - \nabla f(p_0)|$ approaches zero, and hence by (16) $|R|/|p - p_0|$, that approaches zero as p approaches p_0. s 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

$$|V - \nabla f(p_0)|^2 = (f_1(q_1) - f_1(p_0))^2 + (f_2(q_2) - f_2(p_0))^2 + \cdots + (f_n(q_n) - f_n(p_0))^2. \tag{17}$$

Since each f_j is continuous and since

$$\begin{aligned} |q_j - p_0|^2 &= |((x_1, x_2, \ldots, x_{j-1}, c_j, x_{j+1}^0, \ldots, x_n^0) - (x_1^0, \ldots, x_n^0)|^2 \\ &= \sum_{k=1}^{j-1} (x_k - x_k^0)^2 + (c_j - x_j^0)^2 \leq |p - p_0|^2 \end{aligned}$$

for each j, we see from (16) and (17) that (15) holds.

16 Friday October 27—Transformations (Assignments 15,16)

16.1 Transformations

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

Definition 16.1 A *transformation* is any function $T : D \to \mathbf{R}^m$, where $D \subset \mathbf{R}^n$.

Here, $m \geq 1$ and $n \geq 1$, so this includes the special case of a function f considered up to now (that is, $m = 1, n$ arbitrary). Every transformation gives rise to *coordinate functions* as follows: if $p = (x_1, \cdots, x_n) \in D$, 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. Thus

$$T(p) = (f_1(p), \cdots, f_m(p)),$$

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 theorem.

The following is the analog of Definition 8.1

Definition 16.2 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

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

Notice that if $f : D \to \mathbf{R}$ is a function which is of class C^1 on a subset $D \subset \mathbf{R}^n$, then $\mathbf{D}f$ 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).

The following is the analog of Definition 10.1

Definition 16.3 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 $|p - q| < \delta$.

Theorem 16.4 *Let* $T : D \to \mathbf{R}^m$ *be a transformation on a subset* D *of* $\mathbf{R}^n$ *with coordinate functions* $f_1, \ldots, f_m$.

(a) (Buck,p. 80,#17 and p. 334,#6) *Prove that* T *is continuous at* p_0 *if and only if each coordinate function* f_j, $1 \le j \le m$, *is continuous at* p_0.

(b) *If* 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*

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

(c) (Theorem 4 on page 333 of Buck) *If* T *is continuous on* D, *and* D *is a compact subset of* $\mathbf{R}^n$, *then* $T(D)$ *is a compact subset of* $\mathbf{R}^m$.

(d) (Extreme values Theorem) *If* T *is continuous and* D *is compact, then there exist* $p_0, q_0 \in D$ *such that* $|T(q_0)| \le |T(p)| \le |T(p_0)|$ *for every* $p \in D$.

(e) (Uniform Continuity) *If* T *is continuous and* D *is compact, then* T *is uniformly continuous on* D.

Assignment 15 (Due November 3) Prove Theorem 16.4.

Assignment 16 (Due November 3) Show that a linear transformation (see [Buck, Section 7.3]) is uniformly continuous. (Hint: Use [Buck, Theorem 8,page 338])

17 Monday October 30—Uniqueness of the differential (Assignment 17)

17.1 The case of functions

Let us begin with the simple case of a real-valued function f of one variable.

Remark 17.1 *Let* f *be a real-valued function defined on an open interval containing the real number* c.

(a) *There is at most one number* L *satisfying*

$$\lim_{x\to c} \frac{f(x) - f(c) - L(x - c)}{x - c} = 0. \tag{18}$$

(b) *If such a number* L *exists, then* f *is differentiable at* c *and* $L = f'(c)$.

Proof: The most straightforward proof of (b) is to set $R = \frac{f(x)-f(c)-L(x-c)}{x-c}$ and write $\frac{f(x)-f(c)}{x-c} = L + R$ and then let $x \to c$:

$$\lim_{x\to c} \frac{f(x)-f(c)}{x-c} = \lim_{x\to c} L + \lim_{x\to c} R = L. \tag{19}$$

Although (b) implies (a), we shall give a proof of (a) which works in more general situations. Let L and L' both satisfy (18). Then,

$$\begin{aligned}(L-L'))(x-c) &= [-(f(x)-f(c)) + L(x-c)] - (f(x)-f(c)) \\ &+ [f(x)-f(c)-L'(x-c)] + (f(x)-f(c))\end{aligned}$$

so that $|(L-L')(x-c)| \le 2\epsilon|x-c|$ for $|x-c| < \delta = \delta(\epsilon, c)$. This shows that $|L-L'| \le 2\epsilon$. Since ϵ is arbitrary, $L = L'$, proving (a). As noted above, from (19), (b) follows. □

A corresponding result[11] can be proved for real-valued functions defined on subsets of $\mathbf{R}^n$.

Proposition 17.2 (Theorem 9 on page 132 of Buck) *If $f : D \to \mathbf{R}$, where $D \subset \mathbf{R}^n$, if $p_0 \in D$, $u = (u_1, \ldots, u_n) \in \mathbf{R}^n$, and if*

$$\lim_{p\to p_0} \frac{f(p) - f(p_0) - u\cdot(p-p_0)}{|p-p_0|} = 0, \tag{20}$$

then

(a) *There is at most one vector u satisfying (20).*

(b) *If such a vector u exists, then $D_j f(p_0)$ exists for $1 \le j \le n$ and $u = \mathbf{D}f(p_0)$.*

Proof:

Let u_1 and u_2 both be vectors satisfying (20). For $j = 1, 2$,

$$u_j \cdot (p-p_0) = [u_j\cdot(p-p_0) - f(p) + f(p_0)] + [f(p) - f(p_0)].$$

Thus

$$\frac{|(u_1-u_2)\cdot(p-p_0)|}{|p-p_0|} \le 2\epsilon \text{ for } |p-p_0| < \delta = \delta(\epsilon, p_0). \tag{21}$$

Let us rewrite (21) as

$$\frac{|(u_1-u_2)\cdot q|}{|q|} \le 2\epsilon \text{ for } |q| < \delta = \delta(\epsilon, p_0). \tag{22}$$

Now let r be an arbitrary non-zero vector in $\mathbf{R}^n$ and put $q = \delta r/2|r|$ in (22) to get $|(u_1-u_2)\cdot(r/|r|)| \le 2\epsilon$ for every $\epsilon > 0$ and every $r \in \mathbf{R}^n$. Thus $u_1 - u_2$ is orthogonal to every unit vector, hence $u_1 - u_2 = 0$. This proves (a).

[11] Buck assumes that f is continuous, but this does not seem to be needed. The assumption (20) is very strong

The limit (20) exists at p approaches p_0 in any manner. So let $p = p_0 + te_j$ where $t \in \mathbf{R}$ and e_j is the vector with 1 in the j^{th} coordinate and zeros elsewhere. From (20),

$$\lim_{t \to 0} \frac{f(p_0 + te_j) - f(p_0) - u \cdot (te_j)}{|t|} = 0.$$

As $t \to 0^+$, this implies

$$\lim_{t \to 0^+} \frac{f(p_0 + te_j) - f(p_0) - tu_j}{t} = 0. \tag{23}$$

As $t \to 0^-$, this implies

$$\lim_{t \to 0^-} \frac{f(p_0 - |t|e_j) - f(p_0) + |t|u_j}{|t|} = 0. \tag{24}$$

Assignment 17 (Due November 10) From (23) and (24), show that $D_j f(p_0)$ exists. Hint: Use the property $\mathbf{D}_{-e_j} f = -\mathbf{D}_{e_j} f$ of directional derivatives (see Buck, page 126)

From (23) and Assignment 17, it now follows that $D_j f(p_0) = u_j$. $\square$

17.2 Coordinate free definition of derivative

Definition 17.3 (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

$$\lim_{p \to p_0} \frac{|T(p) - T(p_0) - L(p - p_0)|}{|p - p_0|} = 0. \tag{25}$$

We denote L by $T'(p_0)$ (this is justified by Proposition 17.4 below) and call it the *derivative* of T at p_0. (Other names for this are *differential, Frechét derivative, ...*; other notations are $dT|_{p_0}$, $DT(p_0), \ldots$)

Proposition 17.4 *Let T be a transformation defined on a subset A of $\mathbf{R}^n$ with $T(A) \subset \mathbf{R}^m$, and let $f_1, f_2, \ldots, f_m$ be the coordinate functions of T.*

(a) *For a fixed p_0, at most one linear transformation L can satisfy* (25). (This is the same as Exercise #10, page 352 in Buck)

(b) *If such a linear transformation L exists, then $D_j f_i(p_0)$ exists for all $1 \le j \le n$ and $1 \le i \le m$ and the matrix $[l_{ij}]$ of the linear transformation L is given by $l_{ij} = D_j f_i(p_0)$.*

Proof: For (a), we imitate the proof of Proposition 17.2.

Let L_1 and L_2 both be linear transformations satisfying (25). For $j = 1, 2$,

$$L_j(p - p_0) = [L_j(p - p_0) - T(p) + T(p_0)] + [T(p) - T(p_0)].$$

Thus

$$\frac{|(L_1 - L_2)(p - p_0)|}{|p - p_0|} \le 2\epsilon \text{ for } |p - p_0| < \delta = \delta(\epsilon, p_0). \tag{26}$$

Let us rewrite (26) as

$$\frac{|(L_1 - L_2)(q)|}{|q|} \le 2\epsilon \text{ for } |q| < \delta = \delta(\epsilon, p_0). \tag{27}$$

Now let r be an arbitrary non-zero vector in $\mathbf{R}^n$ and put $q = \delta r/2|r|$ in (27) to get $|(L_1 - L_2)(r/|r|)| \le 2\epsilon$ for every $\epsilon > 0$ and every $r \in \mathbf{R}^n$. Thus $L_1 - L_2$ vanishes on every unit vector, and since $L_1 - L_2$ is a linear transformation, it vanishes on every vector. □

We now prove (b). Let $U_1, \ldots, U_m$ be the rows of the matrix representing L. From

$$\begin{aligned} L(p - p_0) &= \begin{bmatrix} l_{11} & \ldots & l_{1n} \\ l_{21} & \ldots & l_{2n} \\ \ldots & \ldots & \ldots \\ l_{m1} & \ldots & l_{mn} \end{bmatrix} \times \begin{bmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \\ \ldots \\ x_n - x_n^0 \end{bmatrix} \\ &= (U_1 \cdot (p - p_0), \cdots, U_m \cdot (p - p_0)), \end{aligned}$$

it follows that

$$\frac{f_i(p) - f_i(p_0) - U_i \cdot (p - p_0)}{|p - p_0|} \le \frac{|T(p) - T(p_0) - L(p - p_0)}{|p - p_0|} \to 0.$$

Hence, by Proposition 17.2, $Djf_i(p_0)$ exists and $U_i = \mathbf{D}f_i(p_0)$. □

18 Wednesday November 1—Existence of the differential

Even though the word "exists" appeared in the previous lecture, that lecture was about the uniqueness of derivatives (or differentials). Today, we discuss existence. But there is no free lunch. We have to pay a price, so it is more appropriate to describe "existence of derivatives" as "sufficient conditions for the existence of derivatives."

18.1 The case of functions

Let us begin with the simple case of a real-valued function f of one variable. The following Remark was proved in Remark 17.1.

Remark 18.1 *Let f be a real-valued function defined on an open interval containing the real number c. If $L := \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$ exists, then $\lim_{x \to c} \frac{f(x) - f(c) - L(x - c)}{x - c} = 0$.*

A corresponding result is true for real-valued functions defined on subsets of $\mathbf{R}^n$ (see Theorem 15.3, which is Theorem 8 on page 131 of Buck).

18.2 Differential as a linear approximation—the case of transformations

Our next main result is the analog for transformations of (15) in Theorem 15.3. First we need to define the replacement for the total derivative.

Definition 18.2 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 *differential* of T at $p \in D$ is the m by n matrix

$$dT|_p = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(p) & \cdots & \frac{\partial f_1}{\partial x_n}(p) \\ \frac{\partial f_2}{\partial x_1}(p) & \cdots & \frac{\partial f_2}{\partial x_n}(p) \\ \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial x_1}(p) & \cdots & \frac{\partial f_m}{\partial x_n}(p) \end{bmatrix}.$$

Of course $dT|_p$ 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

$$dT|_p = [\tfrac{\partial f_i}{\partial x_j}(p)]_{1\le i\le m, 1\le j\le n} = [D_j f_i(p)]_{1\le i\le m, 1\le j\le n}$$

We shall use $\times$ to denote matrix multiplication. Thus, for example, if q is any (row) vector in $\mathbf{R}^n$, $dT|_p \times q^t$ 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 for example, we will have $m = n$, and it will be very important to consider the *Jacobian* of T, which is defined to be $\det dT|_p$.

At this point it is necessary to include the following obvious definition.

Definition 18.3 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.

Remark 18.4 A transformation of class C^1 is continuous. For if $T = (f_1, \ldots, f_m)$ then by Theorem 11.3, each f_i is continuous and by (a) of Theorem 16.4, T is continuous.

Theorem 18.5 (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*[12]*, for any $p_0 \in D$,*

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

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

$$\frac{|T(p) - T(p_0) - (dT|_{p_0}) \times (p - p_0)^t|}{|p - p_0|} < \epsilon \text{ whenever } p \in B(p_0, \delta) \cap D. \qquad (28)$$

[12]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.

19 Friday November 3—More on existence of differentials (Assignments 18,19)

19.1 Proof of Theorem 18.5

Proof of Theorem 18.5: Let $T = (f_1, \cdots, f_m)$. By Theorem 15.3, for each $1 \le i \le m$

$$\frac{|f_i(p) - f_i(p_0) - \mathbf{D}f_i(p_0) \cdot (p - p_0)|}{|p - p_0|} \to 0 \text{ as } p \to p_0. \tag{29}$$

We have

$$\begin{aligned}(dT|_{p_0}) \times (p - p_0)^t &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(p_0) & \dots & \frac{\partial f_1}{\partial x_n}(p_0) \\ \frac{\partial f_2}{\partial x_1}(p_0) & \dots & \frac{\partial f_2}{\partial x_n}(p_0) \\ \dots & \dots & \dots \\ \frac{\partial f_m}{\partial x_1}(p_0) & \dots & \frac{\partial f_m}{\partial x_n}(p_0) \end{bmatrix} \times \begin{bmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \\ \dots \\ x_n - x_n^0 \end{bmatrix} \\ &= (\mathbf{D}f_1(p_0) \cdot (p - p_0), \cdots, (\mathbf{D}f_m(p_0) \cdot (p - p_0)).\end{aligned}$$

Thus

$$\left(\frac{|T(p) - T(p_0) - dT|_{p_0}) \times (p - p_0)^t|}{|p - p_0|}\right)^2 = \left(\sum_{i=1}^m \frac{(f_i(p) - f_i(p_0) - \mathbf{D}f_i(p_0) \cdot (p - p_0))^2}{|p - p_0|^2}\right)$$

the theorem follows from (29). □

Assignment 18 (Due November 10) [Buck page 351 #1,2,7,8]

19.2 Two questions on differentials

Recall the following two results.

Proposition 17.4 T differentiable at p implies $D_j f_i(p)$ exists for every i, j.

Remark 18.4 T of class C^1 on an open set D implies T continuous on D.

It is obvious in one dimension (take $f(x) = |x|$) that the converse to the second result is false. Here are the two questions.

1. Is it true that T differentiable implies T continuous?
2. Is the converse to the first result above true, that is, if $D_j f_i(p)$ exists for all p and all i, j, does it follow that T is differentiable?

The answer to the first question is YES and this will be an important tool in the proof of the chain rule below. The answer to the second question is NO as shown by the following example (Problem #4 on page 135 of Buck is another example of a non-differentiable transformation (in this case, a function) which has all first order partial derivatives existing.

EXAMPLE: Let $f(x, y) = (x^3 - y^3)/(x^2 + y^2)$ for $(x, y) \ne (0, 0)$ and $f(0, 0) = 0$. Then

- $D_1 f$ and $D_2 f$ exist everywhere (This is obvious for $(x, y) \neq (0, 0)$ and needs to be checked for $(0, 0)$).
- f is not differentiable at $(0, 0)$, that is, there are no vectors $u = (u_1, u_2)$ such that
$$\lim_{(x,y)\to(0,0)} \frac{f(x, y) - (u_1, u_2) \cdot (x, y)}{|(x, y)|} = 0.$$

Assignment 19 (Due November 17) Show that the function of Problem 4 on page 135 of Buck, namely, $f(x, y) = xy/(x^2 + y^2)$ for $(x, y) \neq (0, 0)$ and $f(0, 0) = 0$ is not differentiable at $(0, 0)$.

20 Monday November 6—Chain rule for transformations

20.1 Composition of transformations; statement of chain rule

Definition 20.1 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[13] 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, $ST \neq TS$.

Proposition 20.2 (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.*

Proof: If $p_k \to p_0$ in A, then by Theorem 16.4(b), $S(p_k) \to S(p_0)$ in B. Again by Theorem 16.4(b), $T(S(p_k)) \to T(S(p_0))$ in $\mathbf{R}^k$. Thus, by the converse of Theorem 16.4(b) (which is true but was not stated earlier), TS is continuous at p_0. □

Theorem 20.3 (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*

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

[13] 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

20.2 Proof of the one-dimensional chain rule

We first recall 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[14], but not by vectors.

We denote the composition of functions $f, g : \mathbf{R} \to \mathbf{R}$ 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 20.4 (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*

$$(f \circ g)'(a) = f'(g(a))\, g'(a). \tag{30}$$

The usual false proof of Theorem 20.4 is as follows. As long as $g(x) \neq g(a)$,

$$\frac{f \circ g(x) - f \circ g(a)}{x - a} = \frac{f \circ g(x) - f \circ g(a)}{g(x) - g(a)} \cdot \frac{g(x) - g(a)}{x - a}. \tag{31}$$

Since g is continuous at a, $g(x) \to g(a)$ as $x \to a$, and so

$$\lim_{x \to a} \frac{f \circ g(x) - f \circ g(a)}{x - a} = f'(g(a) \cdot g'(a).$$

The problem with this argument is that there is no guarantee that $g(x)$ is not a constant in some open interval containing a, so we would be dividing by zero, a NO NO. We can try to salvage the argument as follows. There are two possibilities.

- g is a constant (necessarily $g(a)$) in some interval containg a. In this case, $g'(a) = 0$ and since $f \circ g$ is also a constant in the interval, $(f \circ g)'(a) = 0$ as well, so (30) holds in this case.

- In the contrary case, for each $k \geq 1$, there exists $x_k \in (a - 1/k, a + 1/k)$ with $g(x_k) \neq g(a)$. Since $x_k \to a$ and g is continuous, $g(x_k) \to g(a)$, so by (31),

$$\frac{f \circ g(x_k) - f \circ g(a)}{x_k - a} \to (f \circ g)'(a) g'(a).$$

[14]Except for zero

Note that this does not prove Theorem 20.4, since it assumes that $(f \circ g)'(a)$ exists.

Proof of Theorem 20.4:

Step 1. $g'(a)$ exists Hence $\forall \epsilon' > 0, \exists \delta' > 0$ such that

$$|g(x) - g(a) - g'(a)(x-a)| < \epsilon'|x-a| \quad \text{if } |x-a| < \delta'. \tag{32}$$

Step 2. $f'(g(a))$ exists Hence $\forall \epsilon'' > 0, \exists \delta'' > 0$ such that

$$|f(y) - f(g(a)) - f'(g(a))(y - g(a))| < \epsilon''|y - g(a)| \quad \text{if } |y - g(a)| < \delta''. \tag{33}$$

Step 3. Wish to show $\forall \epsilon > 0, \exists \delta > 0$ such that

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

Step 4. g is continuous at a Hence $\exists \delta_c > 0$ such that

$$|g(x) - g(a)| < \delta'' \text{ if } |x-a| < \delta_c. \tag{35}$$

Step 5. Substitute Step 4 in Step 2 Using (35), replace y in (33) by $g(x)$ to obtain

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

Step 6. Rewrite Step 1 Set $\eta(x) := g(x) - g(a) - g'(a)(x-a)$ so that

$$g(x) - g(a) = g'(a)(x-a) + \eta(x) \tag{37}$$

and by (32),

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

Step 7. Substitute Step 6 in Step 5 Putting (37) into (36) (in two places!) and setting

$$A(x) := f(g(x)) - f(g(a)) - f'(g(a))[g'(a)(x-a) + \eta(x)] \tag{39}$$

we obtain from (36)

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

Step 8. Prove Step 3 Given ϵ, choose ϵ' and ϵ'' such that

$$\epsilon''\epsilon' + \epsilon''|g'(a)| + |f'(g(a))|\epsilon' < \epsilon. \tag{41}$$

Then choose δ' and δ'' as in Steps 1 and 2, and set $\delta = \min(\delta_c, \delta')$. If $|x-a| < \delta$, we have,

$$\begin{aligned}
&|f(g(x)) - f(g(a)) - f'(g(a))g'(a)(x-a)| \\
=\ & |A(x) + f'(g(a))\eta(x)| \qquad \text{(by (39))} \\
\le\ & |A(x)| + |f'(g(a))\eta(x)| \\
\le\ & \epsilon''|g'(a)||x-a| + \epsilon''|\eta(x)| + |f'(g(a))||\eta(x)| \qquad \text{(by (40))} \\
\le\ & [\epsilon''|g'(a)| + \epsilon''\epsilon' + |f'(g(a))|\epsilon']|x-a| \qquad \text{(by (38))} \\
<\ & \epsilon|x-a| \qquad \text{(by (41)}.
\end{aligned}$$

This proves (34). □

21 Wednesday November 8—Proof of the chain rule for transformations (Assignments 20, 21)

21.1 Two lemmas

To make life simpler, we shall isolate two lemmas, which are themselves of independent interest. We first met Lemma 21.1 in Assignment 16.

Lemma 21.1 (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 21.2 (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 21.1.) Thus T is continuous at p_0. $\square$

21.2 Proof of the chain rule

We are now ready to prove the chain rule for composition of transformations. 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 20.4.

In the proof of Theorem 20.3 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 20.3 (Chain Rule):

Step 1. $T'(p_0)$ exists Hence $\forall \epsilon' > 0, \exists \delta' > 0$ such that

$$|T(p) - T(p_0) - T'(p_0)(p - p_0)| < \epsilon'|p - p_0| \quad \text{if } |p - p_0| < \delta'. \tag{42}$$

Step 2. $S'(T(p_0))$ exists Hence $\forall \epsilon'' > 0, \exists \delta'' > 0$ such that

$$|S(q) - S(T(p_0)) - S'(T(p_0))(q - T(p_0))| < \epsilon''|q - T(p_0)| \quad \text{if } |q - T(p_0)| < \delta''. \tag{43}$$

Step 3. Wish to show $\forall \epsilon > 0, \exists \delta > 0$ such that

$$|S \circ T(p) - S \circ T(p_0) - S'(T(p_0)) \circ T'(p_0)(p - p_0)| < \epsilon|p - p_0| \quad \text{if } |p - p_0| < \delta. \tag{44}$$

Step 4. T is continuous at a (by Lemma 21.2) Hence $\exists \delta_c > 0$ such that

$$|T(p) - T(p_0)| < \delta'' \text{ if } |p - p_0| < \delta_c. \tag{45}$$

Step 5. Substitute Step 4 in Step 2 Using (45), we may replace q in (43) by $T(p)$ to obtain

$$|S(T(p)) - S(T(p_0)) - S'(T(p_0))(T(p) - T(p_0))| < \epsilon''|T(p) - T(p_0)| \quad \text{if } |p - p_0| < \delta_c. \tag{46}$$

Step 6. Rewrite Step 1 Set $\eta(p) := T(p) - T(p_0) - T'(p_0)(p - p_0)$ so that

$$T(p) - T(p_0) = T'(p_0)(p - p_0) + \eta(p) \tag{47}$$

and by (42),

$$|\eta(p)| < \epsilon'|p - p_0| \text{ if } |p - p_0| < \delta. \tag{48}$$

Step 7. Substitute Step 6 in Step 5 Substitute (47) into (46) (in two places!) and set

$$A(p) := S(T(p)) - S(T(p_0)) - S'(T(p_0))[T'(p_0)(p - p_0) + \eta(p)] \tag{49}$$

to obtain from (46)

$$|A(p)| < \epsilon''|T'(p_0)(p - p_0) + \eta(p)| \quad \text{if } |p - p_0| < \delta. \tag{50}$$

Step 8. Prove Step 3 Given ϵ, choose ϵ' and ϵ'' such that

$$\epsilon''\epsilon' + \epsilon'' C_1 + \epsilon' C_2 < \epsilon, \tag{51}$$

where by Lemma 21.1, C_1 and C_2 are chosen so that

$$|T'(p_0)(p)| \leq C_1|p| \quad (p \in \mathbf{R}^n) \text{ and } |S'(T(p_0))(q)| \leq C_2|q| \quad (q \in \mathbf{R}^m).$$

Then choose δ' and δ'' as in Steps 1 and 2, and set $\delta = \min(\delta_c, \delta')$.

If $|p - p_0| < \delta$, we have,

$$\begin{aligned}
&|S(T(p)) - S(T(p_0)) - S'(T(p_0) \circ T'(p_0)(p - p_0)| \\
= \;& |A(p) + S'(T(p_0))\eta(p)| && (\text{ by } (49)) \\
\leq \;& |A(p)| + |S'(T(p_0))\eta(p)| \\
\leq \;& \epsilon''|T'(p_0)(p - p_0)| + \epsilon''|\eta(p)| + |S'(T(p_0))\eta(p)| && (\text{by } (50)) \\
\leq \;& \epsilon'' C_1|p - p_0| + \epsilon''\epsilon'|p - p_0| + C_2\epsilon'|p - p_0| && (\text{by } (48)\ ) \\
< \;& \epsilon|p - p_0| && (\text{by } (51))
\end{aligned}$$

This proves (44). □

The following is the "coordinatized" version of the chain rule. Notice that there is nothing to prove, given Theorem 20.3.

Corollary 21.3 *Let T be a transformation which is differentiable on an open set D, and let S be a transformation which is differentiable on an open set containing $T(D)$. Then $S \circ T$ is differentiable on D, and if $p \in D$, then*

$$dS \circ T|_p = dS|_{T(p)} \times dT_p.$$

As an illustration of the power of Corollary 21.3, we have the following theorem from [Buck,section 3.4].

Theorem 21.4 (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,*

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

Assignment 20 (Due November 17)

(a) Use Theorem 20.3 to prove Theorem 21.4

(b) Let $F(x, y) = f(g(x, y), h(x, y))$, where $f : \mathbf{R}^2 \to \mathbf{R}$, $g : \mathbf{R}^2 \to \mathbf{R}$, and $h : \mathbf{R}^2 \to \mathbf{R}$ are differentiable. Use Theorem 20.3 to prove that F is differentiable and

$$D_1F(x, y) = D_1f(g(x, y), h(x, y))D_1g(x, y) + D_2f(g(x, y), h(x, y))D_1h(x, y)$$

and

$$D_2F(x, y) = D_1f(g(x, y), h(x, y))D_2g(x, y) + D_2f(g(x, y), h(x, y))D_2h(x, y).$$

Assignment 21 (Due November 17) [Buck page 145 #1,2]

22 Friday November 10—holiday (Veteran's Day)

23 Monday November 13—How to use the chain rule; mean value theorems

23.1 An application of the chain rule

The following is one of the seven examples in [Buck; pp. 137–145] illustrating the "baby chain rule" (Theorem 21.4).

EXAMPLE: Let $w = f(x, u, v)$, $u = g(x, v, y)$ and $v = h(x, y)$ where $f : \mathbf{R}^3 \to \mathbf{R}$, $g : \mathbf{R}^3 \to \mathbf{R}$, and $h : \mathbf{R}^2 \to \mathbf{R}$. Then $w = F(x, y) = f(x, g(x, h(x, y), y), h(x, y))$ and it is required to find $D_1 f$ and $D_2 f$.

Write this as $F = f \circ T$ where $T : \mathbf{R}^2 \to \mathbf{R}^3$ is given by $T(x, y) = (x, G(x, y), h(x, y))$ and $G(x, y) = g(x, h(x, y), y)$. By the chain rule

$$F'(x, y) = f'(T(x, y)) \circ T'(x, y),$$

which is the same as

$$[F_x, F_y] = [D_1 f, D_2 f, D_3 f] \times \begin{bmatrix} 1 & 0 \\ G_x & G_y \\ h_x & h_y \end{bmatrix}$$

Write $G(x, y) = g(x, h(x, y), y)$ as $G = g \circ S$ where $S : \mathbf{R}^2 \to \mathbf{R}^3$ is given by $S(x, y) = (x, h(x, y), y)$. By the chain rule,

$$G'(x, y) = g'(S(x, y)) \circ S'(x, y),$$

which is the same as

$$[G_x, G_y] = [D_1 g, D_2 g, D_3 g] \times \begin{bmatrix} 1 & 0 \\ h_x & h_y \\ 0 & 1 \end{bmatrix}$$

Carrying out the two matrix multiplications, we obtain

- $G_x = D_1 g + (D_2 g) h_x$
- $G_y = (D_2 g) h_y + D_3 g$

and

- $F_x = D_1 f + (D_2 f) G_x + (D_3 f) h_x = D_1 f + D_2 f (D_1 g + (D_2 g) h_x) + (D_3 f) h_x$
- $F_y = (D_2 f) G_y + (D_3 f) h_y = D_2 f ((D_2 g) h_y + D_3 g) + (D_3 f) h_y$

This agrees with the result on page 138 of Buck, which in the "taboo" notation for partial derivatives states that

- $\frac{\partial w}{\partial x} = f_1 + f_2 g_1 + f_3 h_1 + f_2 g_2 h_1$
- $\frac{\partial w}{\partial y} = f_2 g_3 + f_3 h_2 + f_2 g_2 h_2$ □

23.2 Mean Value Theorems

Up to now we have used the mean value theorem in one variable (Theorem 11.1). But we mentioned the mean value theorem in several variables above (see the first question at the beginning of the lecture for October 25), 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 18.5 (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 24.1 below.

We note that the version for functions (Theorem 23.1), nicknamed the "Little Mean Value Theorem" will be used in the proof of the version for transformations (Theorem 23.2), nicknamed the "Big Mean Value Theorem". Also, the "Baby Chain Rule" (Theorem 21.4) is needed in the proof of the "Little Mean Value Theorem"[15].

Theorem 23.1 ("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*[16] *segment $L := \{tp_2 + (1-t)p_1 : 0 \le t \le 1\}$ connecting p_1 and p_2 such that*

$$f(p_2) - f(p_1) = \mathbf{D}f(p^*) \cdot (p_2 - p_1).$$

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 $\phi'(\lambda) = (p_2 - p_1)^t$, $\forall \lambda \in [0, 1]$.

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

$$f(p_2) - f(p_1) = F'(\lambda_0) \tag{52}$$

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

Letting $p^* = \phi(\lambda_0)$ we get by the "coordinatized" chain rule (Corollary 21.3),

$$F'(\lambda_0) = f'(\phi(\lambda_0)) \times \phi'(\lambda_0) = \mathbf{D}f(\phi(\lambda_0)) \times (p_2 - p_1)^t = \mathbf{D}f(p^*) \cdot (p_2 - p_1). \tag{53}$$

Compare (52) and (53). □

Theorem 23.2 ("Big" Mean Value Theorem, Theorem 12,page 350 of Buck) *Let $T = (f_1, \ldots, f_m) : D \to \mathbf{R}^m$ 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 $L := \{tp' + (1-t)p'' : 0 \le t \le 1\}$ is a subset of D. Then there exist points $p^*_1, \ldots, p^*_m \in L$ such that*[17]

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

[15] We have a little and big mean value theorem. Question: what is the "tiny mean value theorem"?

[16] Note that this line segment is a subset of $B(p_0, r)$

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

where M is the matrix $(D_j f_i(p_i^*))_{1\le i\le m, 1\le j\le n}$, that is[18],[19],

$$M = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(p_1^*) & \cdots & \frac{\partial f_1}{\partial x_n}(p_1^*) \\ \frac{\partial f_2}{\partial x_1}(p_2^*) & \cdots & \frac{\partial f_2}{\partial x_n}(p_2^*) \\ \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial x_1}(p_m^*) & \cdots & \frac{\partial f_m}{\partial x_n}(p_m^*) \end{bmatrix}.$$

Proof: Apply the Little mean value theorem (Theorem 23.1) to each $f_i : D \to \mathbf{R}$ to get points $p_i^* \in L$ such that

$$f_i(p'') - f_i(p') = \mathbf{D}f_i(p_i^*)\cdot(p''-p') \quad (1 \le i \le m). \tag{54}$$

Now write down the coordinates of the vector $T(p'')-T(p')$, thinking of it as a column vector, and use (54):

$$\begin{aligned} T(p'')-T(p') &= (f_1(p''),\cdots,f_m(p''))^t - (f_1(p'),\cdots,f_m(p'))^t \\ &= (f_1(p'')-f_1(p'),\cdots,f_m(p'')-f_m(p'))^t \\ &= (\mathbf{D}f_1(p_1^*)\cdot(p''-p'),\cdots,\mathbf{D}f_m(p_m^*)\cdot(p''-p'))^t. \end{aligned}$$

On the other hand, writing $p' = (x_1',\ldots,x_n')$ and $p'' = (x_1'',\ldots,x_n'')$,

$$\begin{aligned} M \times (p''-p')^t &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(p_1^*) & \cdots & \frac{\partial f_1}{\partial x_n}(p_1^*) \\ \frac{\partial f_2}{\partial x_1}(p_2^*) & \cdots & \frac{\partial f_2}{\partial x_n}(p_2^*) \\ \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial x_1}(p_m^*) & \cdots & \frac{\partial f_m}{\partial x_n}(p_m^*) \end{bmatrix} \times \begin{bmatrix} x_1''-x_1' \\ x_2''-x_2' \\ \cdots \\ x_n''-x_n' \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{D}f_1(p_1^*) \\ \cdots \\ \mathbf{D}f_m(P_m^*) \end{bmatrix} \times \begin{bmatrix} x_1''-x_1' \\ \cdots \\ x_n''-x_n' \end{bmatrix} = \begin{bmatrix} \mathbf{D}f_1(p_1^*)\cdot(p''-p') \\ \cdots \\ \mathbf{D}f_m(p_m^*)\cdot(p''-p') \end{bmatrix}. \end{aligned}$$

Now compare the last two displayed equations. □

Assignment 22 (Due November 22) [Buck page 154 #18] (Look in the index of Buck to find the definitions of *convex* and *Lipschitz condition*)

24 Wednesday November 15—Applications of Big Mean Value Theorem; local invertibility (Assignment 23)

24.1 Alternate proof of linear approximation for differentiable transformations

For no particularly good reason, we now give an alternate proof to the approximation property of the Jacobian matrix (Theorem 18.5).

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

[19] Note that $M = (\mathbf{D}f_1(p_1^*),\cdots,\mathbf{D}f_m(p_m^*))^t$

Second Proof of Theorem 18.5: By the Big mean value theorem (Theorem 23.2), $T(p)-T(p_0) = L^* \times (p-p_0)^t$ where $L^* := (D_j f_i(p_i^*))$. Look at the matrix entries of $L^* - T'(p_0) = (a_{ij})$; they are $a_{ij} = D_j f_i(p_i^*) - D_j f_i(p_0)$. By Lemma 21.1, for all column vectors $q \in \mathbf{R}^n$,

$$|(L^* - T'(p_0)) \times q| \le \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2} |q|.$$

Since $T(p) - T(p_0) - T'(p_0) \times (p-p_0)^t = (L^* - T'(p_0)) \times (p-p_0)^t$, we have,

$$\begin{aligned} \frac{|T(p) - T(p_0) - T'(p_0) \times (p-p_0)^t|}{|p-p_0|} &\le \frac{|(L^* - T'(p_0)) \times (p-p_0)^t|}{|p-p_0|} \\ &\le \frac{\left(\sum_{i,j} (a_{ij})^2\right)^{1/2} |p-p_0|}{|p-p_0|} \\ &= \left(\sum_{i,j} (D_j f_i(p_i^*) - D_j f_i(p_0))^2 \right)^{1/2} \\ &\to 0 \text{ as } p \to p_0, \end{aligned}$$

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

24.2 The local invertibility theorem

The following simple one-dimensional illustration gives the flavor of the statement and proof of the local invertibility theorem, Theorem 24.1. Let $f : D \to \mathbf{R}$ be differentiable on an open set $D \subset \mathbf{R}$ and suppose that $f'(x) \ne 0$ 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'',

$$f(x'') - f(x') = f'(\tilde{x})(x'' - x'). \tag{55}$$

If $f(x'') = f(x')$, then since $f'(\tilde{x}) \ne 0$, (55) implies $x'' = x'$.

Theorem 24.1 (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*[20]

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

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 $B(p_0, \delta)$.

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

Proof: Consider the open[21] 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

$$F(p_1, \ldots, p_n) = \det[D_j f_i(p_i)] \text{ for } (p_1, \ldots, p_n) \in \Omega.$$

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[22].

We note next that the value of F at a special point of Ω of the form $(p, \ldots, p)$ is given by $F(p, \ldots, p) = \det[D_j f_i(p)] = \det T'(p)$ and so for every $p \in D$, $F(p, \ldots, p) \neq 0$.

It follows from the last two paragraphs[23] that, given a point of D, let's call it p_0 now, there is a $\delta > 0$ such that $B(p_0, \delta) \subset D$ and

$$F(p_1, \ldots, p_n) \neq 0 \text{ for every } (p_1, \ldots, p_n) \in B(p_0, \delta) \times \cdots \times B(p_0, \delta). \tag{56}$$

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

To prove this claim, we use the Mean value theorem for transformations, Theorem 23.2. 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 $M = [D_j f_i(p_i^*)]$,

$$T(p'') - T(p') = M \times (p'' - p')^t. \tag{57}$$

Now $\det M = F(p_1^*, \ldots, p_n^*) \neq 0$ by (56), so M is non-singular. Since we are assuming $T(p'') = T(p')$, (57) shows $p'' - p' = 0$. $\square$

Assignment 23 (Due November 22) [Buck page 361 #11] (The answer to the question is NO. Look at the hint in Buck to construct a proof)

25 Friday November 17—Implicit Function Theorem I (Assignment 24)

25.1 Motivation

In much of *analysis*, the linear functions are the easiest to work with[24]. 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

$$F(x_1, \ldots, x_n) = \sum_{j=1}^{n} a_j x_j.$$

[21] 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 $B((p_1, \ldots, p_n), \delta) \subset D \times \cdots \times D$

[22] Note that F is "jointly" continuous (which is what is needed), and not just separately continuous

[23] Persistence of sign for continuous functions

[24] This is not necessarily the case for *algebra*

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 we can take $r = 1 - |x_0|$, and set $\phi(x) = +\sqrt{1 - x^2}$ for $x \in (x_0 - r, x_0 + r)$. The only problem arises when $|x_0| = 1$, that is $y_0 = 0$, which is precisely where $\frac{\partial F}{\partial y}$ vanishes. Another solution is obtained by taking $\phi(x) = -\sqrt{1 - x^2}$. Before we leave this example, let's note that we can interchange the roles of the variables x and y and obtain a function $x = \psi(y)$ satisfying, among other things $(\psi(y))^2 + y^2 - 1 = 0$.

Let's now consider a third example, which is not so easy (in fact, 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 24 (Due December 1) *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 derivative of T:

$$T'(x,y) = \begin{pmatrix} \frac{\partial G}{\partial x}(x,y) & \frac{\partial G}{\partial y}(x,y) \\ \frac{\partial F}{\partial x}(x,y) & \frac{\partial F}{\partial y}(x,y) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{\partial F}{\partial x}(x,y) & \frac{\partial F}{\partial y}(x,y) \end{pmatrix}.$$

It follows that $\det T'(x,y) = \frac{\partial F}{\partial y}(x,y)$.

25.2 Implicit function theorems

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

Theorem 25.1 (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 25.1, 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[25] 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[26]
- Question: Can we calculate $\phi'(x)$ by implicit differentiation and the chain rule?[27]

The proof of the general Implicit Function Theorem (see the statement below), and in particular of Theorem 25.1, depends on three main results about transformations of class C^1. These are

- Local Invertibility Theorem
- Open Mapping Theorem
- Inverse Function Theorem

The local invertibility theorem has already been stated and proved as Theorem 24.1. We shall state and use the other two theorems here and prove them in later lectures.

[25]This already says a lot! If you stop here you got a bargain.
[26]This statement implies the previous two statements
[27]Yes, but it is not entirely satisfactory because the answer is in terms of $\phi(x)$

Theorem 25.2 (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*

$$\det T'(p) \neq 0 \text{ for all } p \in D.$$

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

Theorem 25.3 (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*

$$\det T'(p) \neq 0 \text{ for all } p \in D.^{28}$$

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*

$$(T^{-1})'(T(p)) = [T'(p)]^{-1} \text{ for all } p \in D. \tag{58}$$

Proof of Theorem 25.1: Define a transformation $T = (G, F)$ by setting $G(x, y) = x$. Let p_0 denote (x_0, y_0). Since $T'(p_0)$ is invertible, by the "local invertibility theorem" (Theorem 24.1), T is locally one-to-one at p_0. That is, there is a ball B with center p_0 (contained in D since D is open) 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 $T'(p)$ is invertible for every p in this smaller ball[29]. 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 25.2), $T(B')$ 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[30]

$$(u, v) = T \circ T^{-1}(u, v) = T(T^{-1}(u, v)) = T(g(u, v), h(u, v)) = (g(u, v), F(g(u, v), h(u, v))$$

for all $(u, v) \in B((x_0, 0), r)$. In particular, $u = g(u, v)$ and

$$v = F(g(u, v), h(u, v)) = F(u, h(u, v)). \tag{59}$$

Substitute for (u, v), any point of the form $(x, 0) \in B((x_0, 0), r)$. From (59), 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 "inverse function theorem" (Theorem 25.3), T^{-1} is of class C^1, hence h is of class C^1, and hence ϕ is of class C^1 on $(x_0 - r, x_0 + r)$. This completes the proof.□

[28] so that $(T'(p))^{-1}$ exists

[29] What is the reason for this?

[30] We also 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 $(x, y) \in B'$, but this is of no use to us

26 Monday November 20—Implicit Function Theorem II (Assignment 25)

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.

Theorem 26.1 (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 26.2 *Let $F : D \to \mathbf{R}$ be of class C^1 on an open set $D \subset \mathbf{R}^n$, let $(x_1^0, \ldots, x_n^0)$ be a point of D, and suppose that*

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

Then there exists $r > 0$ and a function

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

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

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

and

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

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

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

> Let $F : D \to \mathbf{R}$ be of class C^1 on an open set $D \subset \mathbf{R}^n$, let $p_0 = (x_1^0, \ldots, x_n^0)$ be a point of D, and suppose that $F(p_0) = 0$ and $\frac{\partial F}{\partial x_k}(p_0) \neq 0$ for some k. Let $p_0^{(k)} = (x_1^0, \ldots, x_{k-1}^0, x_{k+1}^0, \ldots, x_n^0)$. Then there exists $r > 0$ and a function $\phi : B(p_0^{(k)}, r) \to \mathbf{R}$ of class C^1 on $B(p_0^{(k)}, r) \subset \mathbf{R}^{n-1}$, such that, with $p = (x_1, \ldots, x_n)$ and $p^{(k)} = (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n)$, we have $\phi(p_0^{(k)}) = x_k^0$ and $F(x_1, \ldots, x_{k-1}, \phi(p^{(k)}), x_{k+1}, \ldots, x_n) = 0$ 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 $x_1, \ldots, x_n$ 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.

Definition 26.3 For m functions $\phi_1, \ldots, \phi_m$ defined on a subset D of $\mathbf{R}^n$ (with $m \leq n$) and any choice of m variables $x_{i_1}, \ldots, x_{i_m}$ from $x_1, \ldots, x_n$ we define the *Jacobian* of $\phi_1, \ldots, \phi_m$ with respect to $x_{i_1}, \ldots, x_{i_m}$ to be the m by m matrix-valued function

$$\frac{\partial(\phi_1,\ldots,\phi_m)}{\partial(x_{i_1},\ldots,x_{i_m})} = \begin{bmatrix} \frac{\partial \phi_1}{\partial i_1} & \cdots & \frac{\partial \phi_1}{\partial i_m} \\ \cdot & \cdots & \cdot \\ \frac{\partial \phi_m}{\partial i_1} & \cdots & \frac{\partial \phi_m}{\partial i_m} \end{bmatrix}$$

Theorem 26.4 (Theorem 18,page 364 of Buck) *Let $F : D \to \mathbf{R}$ and $G : D \to \mathbf{R}$ be of class C^1 on an open set $D \subset \mathbf{R}^5$, let $(x_0, y_0, z_0, u_0, v_0) \in D$, and suppose that $F(x_0, y_0, z_0, u_0, v_0) = G(x_0, y_0, z_0, u_0, v_0) = 0$ and*

$$\frac{\partial(F,G)}{\partial(u,v)}(x_0, y_0, z_0, u_0, v_0) \neq 0.$$

Then there exists a $r > 0$ and functions $\phi : B((x_0, y_0, z_0), r) \to \mathbf{R}$ and $\psi : B((x_0, y_0, z_0), r) \to \mathbf{R}$ of class C^1 on $B((x_0, y_0, z_0), r) \subset \mathbf{R}^3$ such that $\phi(x_0, y_0, z_0) = u_0$, $\psi(x_0, y_0, z_0) = v_0$ and $F(x, y, z, \phi(x, y, z), \psi(x, y, z)) = 0$ and $G(x, y, z, \phi(x, y, z), \psi(x, y, z)) = 0$ for all $(x, y) \in B((x_0, y_0, z_0), r)$.

Proof: Define a transformation $T : D \to \mathbf{R}^5$ by setting

$$T(x, y, z, u, v) = (x, y, z, F(x, y, z, u, v), G(x, y, z, u, v)).$$

Let p_0 denote $(x_0, y_0, z_0, u_0, v_0)$. Then $T(p_0) = (x_0, y_0, z_0, 0, 0)$ and for $p \in D$,

$$T'(p) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & F_u(p) & F_v(p) \\ 0 & 0 & 0 & G_u(p) & G_v(p) \end{bmatrix}$$

so that $\det T'(p) = \frac{\partial(F,G)}{\partial(u,v)}(p)$, and since $\det T'(p_0) \neq 0$, T has a local inverse at p_0, that is, there exists $\delta > 0$ such that with $B = B(p_0, \delta)$, T is a one-to-one map of B onto $T(B)$ with inverse $T^{-1} : T(B) \to B$. As $T(B)$ is an open set, we can choose $r > 0$ such that $B((x_0, y_0, z_0, 0, 0), r) \subset T(B)$. Let the coordinate functions of T^{-1} be $(\alpha, \beta, \gamma, f, g)$ so that for all $\mathbf{t} = (t_1, t_2, t_3, t_4, t_5) \in T(B)$,

$$\begin{aligned} (t_1, t_2, t_3, t_4, t_5) &= T \circ T^{-1}(t_1, t_2, t_3, t_4, t_5) = T(\alpha(\mathbf{t}), \beta(\mathbf{t}), \gamma(\mathbf{t}), f(\mathbf{t}), g(\mathbf{t})) \\ &= (\alpha(\mathbf{t}), \beta(\mathbf{t}), \gamma(\mathbf{t}), F(\alpha(\mathbf{t}), \beta(\mathbf{t}), \gamma(\mathbf{t})), f(\mathbf{t}), g(\mathbf{t})), G(\alpha(\mathbf{t}), \beta(\mathbf{t}), \gamma(\mathbf{t}), f(\mathbf{t}), g(\mathbf{t}))). \end{aligned}$$

This shows that for all $\mathbf{t} \in T(B)$, $\alpha(\mathbf{t}) = t_1$, $\beta(\mathbf{t}) = t_2$, $\gamma(\mathbf{t}) = t_3$ and

$$F(t_1, t_2, t_3, f(t_1, t_2, t_3, 0, 0), g(t_1, t_2, t_3, 0, 0)) = 0$$

and

$$G(t_1, t_2, t_3, f(t_1, t_2, t_3, 0, 0), g(t_1, t_2, t_3, 0, 0)) = 0.$$

We now see that if we define $\phi(t_1,t_2,t_3) = f(t_1,t_2,t_3,0,0)$ and $\psi(t_1,t_2,t_3) = g(t_1,t_2,t_3,0,0)$ for $(t_1,t_2,t_3) \in B((x_0,y_0,z_0),r)$, the theorem is proved. $\square$

The general form of Theorem 26.4 is now clear: Given m functions $\phi_1,\ldots,\phi_m$ defined on an open set $D \subset \mathbf{R}^n$ with $m \le n$, if $\phi_i(p_0) = 0$ for each $1 \le i \le m$ and some point $p_0 \in D$, and if

$$\tfrac{\partial(\phi_1,\ldots,\phi_m)}{\partial(x_{i_1},\ldots,x_{i_m})}(p_0) \neq 0,$$

then you can solve the system of equations $\phi_i(p) = 0$, $1 \le i \le m$ for each of the variables $x_{i_1},\ldots,x_{i_m}$ in terms of the other variables, and in a continuously differentiable way.

Assignment 25 (Due December 1) [Buck page 366 #2,5,9,11]

27 Wednesday November 22—Proof of Open Mapping Theorem (Assignment 26)

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 $1 \le j \le n$. Stated another way, $\mathbf{D}f(p_0) = 0$.*

Proof: The meaning of "local minimum" is that there exists a ball $B(p_0,r) \subset D$ such that $f(p) \ge f(p_0)$ for all $p \in B(p_0,r)$. By definition,

$$D_j f(p_0) = \lim_{t\to 0} \frac{f(p_0 + te_j) - f(p_0)}{t}. \tag{60}$$

In (60), 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 $D_j f(p_0) \ge 0$, whereas if we let t approach zero through negative values, we get $D_j f(p_0) \le 0$. Thus $D_j f(p_0) = 0$. $\square$

We shall also use the following fact about compact sets.

Assignment 26 (Due December 1) *Prove that if K is a compact set in $\mathbf{R}^n$ and $q \notin K$, then*

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

Proof of the open mapping theorem, Theorem 25.2: Let $q_0 \in T(D)$. Choose a point $p_0 \in D$ such that $q_0 = T(p_0)$. By Theorem 24.1, 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 $N := \{p \in D : |p - p_0| \le \delta\} \subset D$. 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 26, $d := \inf\{|q_0 - q| : q \in T(C)\} > 0$.

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:

$$\forall p \in N, \quad |T(p) - q_1|^2 \geq |T(p^*) - q_1|. \tag{61}$$

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

To prove[31] 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 $p \in C$,

$$|T(p) - q_1| \geq |T(p) - q_0| - |q_0 - q_1| \geq d - d/3 = 2d/3. \tag{62}$$

Note that $p_0 \in N$, $T(p_0) = q_0$, and $|q_0 - q_1| < d/3$. Suppose now that $p^* \in C$. Then we would have on the one hand, by (62), $|T(p^*) - q_1| \geq 2d/3$, and on the other hand, by (61), $|T(p^*) - q_1| \leq |T(p_0) - q_1| < d/3$, 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 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 $q_1 \in T(D)$, proving CLAIM 1 and the theorem. □

[31] We still need to prove claim 1

28 Friday November 24—holiday (Thanksgiving)

29 Monday November 27—Proof of Inverse Mapping 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, whose (indirect) proof is straightforward.

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.

Let us also note that if a transformation preserves convergent sequences, then it is continuous. (Same proof as [Buck, Theorem 2,page 74].)

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 $\lim_{k\to\infty} T(p_k) = T(p)$. By the remark preceding the theorem, 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 25.3 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*

$$(f^{-1})'(f(x_0)) = \frac{1}{f'(x_0)}.$$

Proof of the inverse function theorem, Theorem 25.3: Since T is of class C^1, by Theorem 18.5, (considering $T(p)$ and $T(p_0)$ as column vectors)

$$T(p) - T(p_0) = T'(p_0) \times (p - p_0)^t + R(p) \tag{63}$$

where

$$\lim_{p \to p_0} \frac{|R(p)|}{|p - p_0|} = 0. \tag{64}$$

By assumption $\det T'(p_0) \neq 0$ so $T'(p_0)$ is non-singular. Multiplying (63) (on the left) by $[T'(p_0)]^{-1}$, you get

$$[T'(p_0)]^{-1}(T(p) - T(p_0)) = (p - p_0)^t + [T'(p_0)]^{-1}(R(p)). \tag{65}$$

Let us now denote by q and q_0, the column vectors which are the images of p^t and p_0^t under T; that is $q = T(p^t)$ and $q_0 = T(p_0^t)$, so that $p^t = T^{-1}(q), p_0^t = T^{-1}(q_0)$. Then by (65),

$$T^{-1}(q) - T^{-1}(q_0) = (p - p_0)^t = [T'(p_0)]^{-1}(T(p) - T(p_0)) - [T'(p_0)]^{-1}(R(p)),$$

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

$$T^{-1}(q) - T^{-1}(q_0) - [T'(p_0)]^{-1}(T(p) - T(p_0)) = -[T'(p_0)]^{-1}(R(p)). \tag{66}$$

If we can show that the right hand side of (66) satisfies

$$\lim_{q \to q_0} \frac{|[T'(p_0)]^{-1}(R(p))|}{|q - q_0|} = 0, \tag{67}$$

then (66) will say that (58) is true. So we need to prove (67).

First recall that by Lemma 21.1 there is a constant M such that $|[T'(p_0)]^{-1}(u)| \leq M|u|$ for all $u \in \mathbf{R}^n$. Therefore,

$$\frac{|[T'(p_0)]^{-1}(R(p))|}{|q - q_0|} \leq \frac{M|R(p)|}{|q - q_0|}. \tag{68}$$

By (65), $(p - p_0)^t = [T'(p_0)]^{-1}(T(p) - T(p_0)) - [T'(p_0)]^{-1}(R(p))$ so

$$|p - p_0| \leq M|q - q_0| + M|R(p)|, \tag{69}$$

and by (64),

$$|R(p)| \leq \epsilon|p - p_0| \text{ for } |p - p_0| < \delta \quad (\delta \text{ depending on } \epsilon). \tag{70}$$

Therefore, (69) becomes

$$|p - p_0| \leq M|q - q_0| + M\epsilon|p - p_0|,$$

or,

$$(1 - \epsilon M)|p - p_0| \leq M|q - q_0|,$$

that is,

$$|p - p_0| \le \frac{M}{1 - \epsilon M}|q - q_0| \text{ for } |p - p_0| < \delta. \tag{71}$$

Taking recriprocals in (71) you get

$$\frac{1}{|q - q_0|} \le \frac{M}{1 - \epsilon M}\,\frac{1}{|p - p_0|} \text{ for } |p - p_0| < \delta. \tag{72}$$

Now by (68),(72), and (70), we have, for $|p - p_0| < \delta$,

$$\frac{|[T'(p_0)]^{-1}(R(p))|}{|q - q_0|} \le M|R(p)|\frac{M}{|p - p_0|(1 - \epsilon M)} \le \frac{\epsilon M^2}{1 - \epsilon M}.$$

The quantity

$$\frac{\epsilon M^2}{1 - \epsilon M}$$

is "just as good" as ϵ (since it goes to zero as ϵ does). Therefore (67) 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 $T'(p)$ are continuous functions by assumption and therefore the entries of the inverse matrix $T'(p)^{-1}$ are continuous functions (Why?). By (58) then, the entries of $(T^{-1})'(T(p))$ are continuous functions of $q = T(p)$. □

30 Wednesday November 29—Mixed Partials Theorem

30.1 Mixed Partials Theorem—weak version

Theorem 30.1 (Theorem 11,page 189 of Buck) *Let f be of class C^2 on a closed rectangle $R \subset \mathbf{R}^2$ with vertices $P_1 = (a_1, b_1), Q_1 = (a_2, b_1), P_2 = (a_2, b_2), Q_2 = (a_1, b_2)$ with $a_1 < a_2$ and $b_1 < b_2$. Then*

$$\int_R\int D_1(D_2 f)(x, y)\,dxdy = \int_R\int D_2(D_1 f)(x, y)\,dxdy = f(P_1) - f(Q_1) + f(P_2) - f(Q_2).$$

Proof: In this proof, let f_{12} denote $D_2(D_1 f)$ and f_{21} denote $D_1(D_2 f)$. Then, since f_{12}, $D_1 f$ and $D_2 f$ are continuous, by the Fundamental Theorem of Calculus,

$$\begin{aligned}\int_R\int f_{12}(x, y)\,dxdy &= \int_{a_1}^{a_2}\left[\int_{b_1}^{b_2} \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\,dy\right]dx \\ &= \int_{a_1}^{a_2}\left[\frac{\partial f}{\partial x}(x, b_2) - \frac{\partial f}{\partial x}(x, b_1)\right]dx \\ &= f(a_2, b_2) - f(a_1, b_2) - f(a_2, b_1) + f(a_1, b_1).\end{aligned}$$

Similarly, since f_{21} is continuous, by the Fundamental Theorem of Calculus,

$$\begin{aligned}\int_R\int f_{21}(x,y)\,dxdy &= \int_{b_1}^{b_2}\left[\int_{a_1}^{a_2}\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)\,dx\right]\,dy\\ &= \int_{b_1}^{b_2}\left[\frac{\partial f}{\partial y}(a_2,y)-\frac{\partial f}{\partial y}(a_1,y)\right]\,dy\\ &= f(a_2,b_2)-f(a_2,b_1)-f(a_1,b_2)+f(a_1,b_1).\square\end{aligned}$$

Corollary 30.2 (Corollary, page 190 of Buck) *If f is of class C^2 on an open set $D \subset \mathbf{R}^2$, then $D_1(D_2f) = D_2(D_1f)$ on D.*

Proof: By the theorem, $\int_R\int(f_{12} - f_{21})(x, y)\,dxdy = 0$ for every closed rectangle $R \subset D$. Since $f_{12} - f_{21}$ is continuous on D, this implies that $f_{12} - f_{21}$ is zero everywhere on D. $\square$

30.2 Mixed Partials Theorem—strong version

Theorem 30.3 (Tom M. Apostol, Mathematical Analysis 1957, page 121) *Let D_1f, D_2f, and $D_2(D_1f)$ exist and be continuous in a neighborhood of (x_0, y_0) in $\mathbf{R}^2$. Then $D_1(D_2f)(x_0, y_0)$ exists and equals $D_2(D_1f)(x_0, y_0)$.*

Proof: For notation's sake, let $N := D_2f(x_0+h, y_0)-D_2f(x_0, y_0)$. We need to prove that $f_{21}(x_0, y_0) = \lim_{h\to 0} N/h$ exists and equals $f_{12}(x_0, y_0)$.

Write

$$N = \lim_{k\to 0}\frac{f(x_0+h,y_0+k)-f(x_0+h,y_0)}{k} - \lim_{k\to 0}\frac{f(x_0,y_0+k)-f(x_0,y_0)}{k}.$$

Fix k and introduce $g_k(t) = f(x_0 + t, y_0 + k) - f(x_0 + t, y_0)$ so that by the Mean Value Theorem,

$$N = \lim_{k\to 0}\frac{g_k(h)-g_k(0)}{k} = \lim_{k\to 0}\frac{hg_k'(\overline{h})}{k}$$

where $\overline{h}$ lies between 0 and h (and depends on k).

By the chain rule, $g_k'(t) = D_1f(x_0+t, y_0+k)-D_1f(x_0+t, y_0)$ so that by the Mean Value Theorem again,

$$\frac{N}{h} = \lim_{k\to 0}\frac{D_1f(x_0+\overline{h},y_0+k)-D_1f(x_0+\overline{h},y_0)}{k} = \lim_{k\to 0} f_{12}(x_0+\overline{h},\overline{y}),$$

where $\overline{y}$ lies between y_0 and $y_0 + k$ (and depends on $\overline{h}$).

We remind ourselves that we now need to prove that

$$\lim_{h\to 0}\lim_{k\to 0} f_{12}(x_0+\overline{h},\overline{y}) \text{ exists, and equals } f_{12}(x_0,y_0).$$

Digression: As $k \to 0$, then $\overline{y} \to y_0$ but the dependence of $\overline{h}$ on k is unknown. If $\overline{h}$ were in fact independent of k, then we would have

$$\lim_{k\to 0} f_{12}(z_0 + \overline{h}, \overline{y}) = f_{12}(x_0 + \overline{h}, y_0),$$

since f_{12} is continuous in a neighborhood of (x_0, y_0). Again, by the continuity of f_{12} at (x_0, y_0), we would have

$$\lim_{h\to 0} f_{12}(x_0 + \overline{h}, y_0) = f_{12}(x_0, y_0),$$

which would complete the proof of the theorem. However, we do not know that $\overline{h}$ is independent of k, so this argument is invalid.

Let us now return to the proof of Theorem 30.3. Let $F(h) = \lim_{k\to 0} f_{12}(x_0 + \overline{h}, \overline{y})$ (recall that $\overline{h}$ and $\overline{y}$ depend on k as well as on h. Once again, we need to show that

$$\lim_{h\to 0} F(h) \text{ exists and equals } f_{12}(x_0, y_0). \tag{73}$$

Let $\epsilon > 0$ and choose $\delta > 0$ such that

$$|f_{12}(x, y) - f_{12}(x_0, y_0)| < \epsilon/2 \text{ if } (x, y) \in B((x_0, y_0), \delta).$$

Note that if $|h| < \delta/2$ and $|k| < \delta/2$, then $(x_0 + \overline{h}, \overline{y}) \in B((x_0, y_0), \delta)$. Now keep h fixed with $|h| < \delta/2$. Then

$$|f_{12}(x_0 + \overline{h}, \overline{y}) - f_{12}(x_0, y_0)| < \epsilon/2 \text{ if } |k| < \delta/2.$$

Now let $k \to 0$ to get

$$|F(h) - f_{12}(x_0, y_0)| < \epsilon/2 \text{ if } |h| < \delta/2.$$

This proves (73) and the theorem. □

Remark 30.4 Note that it follows that $D_1(D_2 f)(x, y)$ exists and equals $D_2(D_1 f)(x, y)$ for every (x, y) in the given neighborhood of (x_0, y_0).

31 Friday December 1—Extensions of (uniformly) continuous functions; Course summary

31.1 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 9.4. 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[32] to $p \in \overline{S}$, then p_k is a Cauchy sequence, and by uniform continuity of f, $f(p_k)$ 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.

[32] Such a sequence exists by Theorem 14.1

31.2 The extension theorem

This subsection is devoted to the answer to the third question raised above. Let us state this as a theorem.

Theorem 31.1 *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 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: Here are the basic ideas of the proof.

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.

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

$$|\tilde{f}(p) - \tilde{f}(q)| \leq |\tilde{f}(p) - f(p_k)| + |f(p_k) - f(q_j)| + |f(q_j) - f(q)|. \tag{74}$$

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 (74)).

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 (74), $|f(p_k) - f(q_j)| < \epsilon/3$ if $|p_k - q_j| < \delta_1$.

Now note that (again by the triangle inequality)

$$|p_k - q_j| \leq |p_k - p| + |p - q| + |q - q_j|. \tag{75}$$

Thus, if we define $\delta := \delta_1/2$, then from (75), 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 (74), where k, j are chosen so that $k > N_1, j > N_2$ and $1/k + 1/j < \delta_1$. □

31.3 Course summary—from Buck (and the Minutes)

1.3 Schwarz inequality—Theorem 1,p.13 (Th.1.1); Triangle inequality—Cor. p.14 (Cor.2.1)

1.5 topology—open,closed,interior,boundary,closure,cluster point

1.6 sequences—characterization of closure: Theorem 5,p.30 (Th.14.1)

1.8 compactness—Bolzano Weierstrass,Heine Borel, Theorems 25,26,27, p.65 (Th.7.1)

2.2 continuity—sequential criteria,Theorem 1,p.73 (Th.9.4),Theorem 2,p.74

2.3 uniform continuity—on compact sets, Theorem 6,p.84 (Th.10.2)

2.4 extreme values—Theorem 10,p.90 (Th.9.6), Theorem 11, p.91 (Th.9.8), Theorem 13,p.93 (Th.8.2)

2.6 extension—Theorem 24,p.109 (Th.31.1)

3.3 differentiable implies continuous:Cor.,p.129 (Th. 11.3), existence (= linear approximation), Theorem 8,p.131 (Th.15.3),uniqueness, Theorem 9,p.132 (Prop.17.2)

3.4 baby chain rule—Theorem 14,p.136 (Th.21.4)

3.5 little mean value theorem—Theorem 16,p.151 (Th.23.1)

4.3 mixed partials theorem—Cor. p.190 (Th.30.1 and Th.30.3)

7.2 transformations—continuity, Theorem 3,p.333 (Th.16.4),compactness, Theorem 4,p.333 (Th.16.4)

7.3 linear transformation-uniform continuity of them, Theorem 8,p.338 (Lemma 21.1)

7.4 coordinate free derivative—existence (=linear approximation) Theorem 10, p.344 (Th.18.5), chain rule, Theorem 11, p.346 (Th.20.3), big mean value theorem, Theorem 12,p.350 (Th.23.2), uniqueness, Exercise 10,p.352, (Prop.17.4)

7.5 inverse functions—automatic continuity of inverse Theorem 13,p.353 (Th.29.2), local invertibility Theorem 14,p.355 (Th.24.1), open mapping Theorem 15,p.356 (Th.25.2), inverse function Theorem 16,p.358 (Th.25.3)

7.6 implicit functions—implicit function theorems, Theorem 17,p.363 (Ths.25.1,26.1,26.2), Theorem 18,p.364 (Th.26.4)