ANALYSIS 2

SIDDHANT CHAUDHARY

ABSTRACT. These are my classnotes for Analysis 2. These notes mainly cover differentiation in general Euclidean spaces.

Proof of Cauchy-Schwarz Inequality

Here, I will mention a proof of the Cauchy-Schwarz inequality, which is one of my favourite proofs that I have seen for this particular inequality. We will prove both the real and the complex versions of it. The good thing about this proof is that it is general, and it is based on a simple geometric idea (that the length of the hypotenuse is the greatest in a right triangle).

First, to prove the real version, we will use the idea of a *positive definite scalar product*. Suppose V is a vector space over $\mathbb{R}$. A *scalar product* is a mapping from V^2 to $\mathbb{R}$, denoted by $\langle \cdot, \cdot \rangle$ satisfying the following conditions:

(1) $\langle v, w \rangle = \langle w, v \rangle$, for all $v, w \in V$.
(2) $\langle v, u + w \rangle = \langle v, u \rangle + \langle v, w \rangle$, for all $u, v, w \in V$.
(3) $\langle cv, w \rangle = c\langle v, w \rangle$ for all $c \in \mathbb{R}$ and $v, w \in V$.

In addition, the scalar product is said to be *positive definite*, if, in addition, $\langle v, v \rangle > 0$ if $v \neq O$ and $\langle O, O \rangle = 0$. In this case, we define the *norm* of a vector v to be $N(v) = \sqrt{\langle v, v \rangle}$. One consequence of this definition is the *Pythagoras theorem*, which can be easily proved by using the properties of the scalar product: if $\langle u, v \rangle = 0$ (orthogonality), then

$$[N(u + v)]^2 = [N(u)]^2 + [N(v)]^2$$

Let's now prove the inequality:

Theorem 0.1. For all $u, v \in V$, we have

$$|\langle u, v \rangle| \leq N(u)N(v)$$

Proof: Without loss of generality we assume that both u, v are non-zero, because otherwise the inequality is trivial.

First, suppose the inequality is true when v is a *unit* vector. Then, let $u, v \in V$, and hence $\dfrac{v}{N(v)}$ is a unit vector. So, we have

$$\left| \left\langle u, \frac{v}{N(v)} \right\rangle \right| \leq N(u)$$

and hence the desired inequality is obtained by multiplying both sides by $N(v)$.

Now, suppose e is a unit vector, and let $u \in V$. Then, it can be checked that $u_1 = \langle u, e \rangle e$ and $u_2 = u - \langle u, e \rangle e$ are orthogonal. Hence, by pythagoras theorem, we have

$$[N(u)]^2 = [N(u_1)]^2 + [N(u_2)]^2$$

Date: January 2020.

and hence

$$[N(u)]^2 \geq [N(u_1)]^2$$

and taking square roots on both sides, the inequality is obtained.

Observe that equality follows if and only if $N(u_2) = 0$, which means that

$$u = \langle u, e \rangle e$$

which can be written as $u = \lambda e$, for some $\lambda \in \mathbb{R}$. Similarly, it can be checked that equality in Cauchy-Schwarz follows if an only if

$$u = \lambda v$$

for some $\lambda \in K$.

Now, if $V = \mathbb{R}^n$, and we consider the usual dot product, it is a positive definite scalar product, and hence the real version of the Cauchy-Schwarz inequality is proven, i.e

$$|x_1 y_1 + ... + x_n y_n| \leq \sqrt{x_1^2 + ... + x_n^2}\sqrt{y_1^2 + ... + y_n^2}$$

If we now consider a vector space V over $\mathbb{C}$ with a positive-definite Hermitian product, then the same proof as above works in that case as well. Now, for $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \mathbb{C}^n$, we define their product as

$$\langle x, y \rangle = x_1 \overline{y_1} + ... + x_n \overline{y_n}$$

and it can be easily verified that this is a positive definite Hermitian product. So, the corresponding Cauchy-Schwarz inequality follows, and we have

$$|x_1 \overline{y_1} + ... + x_n \overline{y_n}| \leq \sqrt{|x_1|^2 + ... + |x_n|^2}\sqrt{|y_1|^2 + ... + |y_n|^2}$$

1. Some special metric spaces

In this section, we will study a special kind of vector space, called a *normed* vector space. Let K be the real or the complex field. A map $|| \cdot || : V \to [0, \infty)$, where V is a vector space over K, is said to be a *norm* if the following conditions hold:

(1) $||x|| = 0$ if and only if $x = 0$.
(2) $||cx|| = |c| ||x||$, for all $c \in K$.
(3) $||x_1 + x_2|| \leq ||x_1|| + ||x_2||$

Note that if V is a normed space, then we define the metric (distance) on V by

$$d(x, y) = ||x - y||$$

and it is easily observed that the properties of a metric space are satisfied, and hence V is a metric space as well.

Let's look at some specific kind of norms:

1.1. **Euclidean Norm:** If our vector space if $\mathbb{R}^n$ over $\mathbb{R}$ or $\mathbb{C}^n$ over $\mathbb{C}$, we define

$$||(x_1, ..., x_n)|| = \sqrt{|x_1|^2 + |x_2|^2 + ... + |x_n|^2}$$

It is easily seen that in this case, the first two properties of the norm are satisfied. For the third property, the Cauchy-Schwarz Inequality will give an easy proof.

1.2. l_p **norms:** For $p \in [1, \infty)$, and $K = \mathbb{R}^n, \mathbb{C}^n$, define the l_p norm as

$$||(x_1, ..., x_n)||_p = \left\{ \sum_{k=1}^{n} |x_i|^p \right\}^{\frac{1}{p}}$$

l_2 is the usual Euclidean norm. In the next section, we will show that l_p actually satisfies the norm axioms.

1.3. l^∞ **norm:** On $\mathbb{R}^n$ or $\mathbb{C}^n$, the l^∞ norm is defined as:

$$||(x_1, ..., x_n)||_\infty = \max_{1 \le i \le n} |x_i|$$

For a fixed vector in $\mathbb{R}^n$ or $\mathbb{C}^n$, we can show that $l_p \to l^\infty$ as $p \to \infty$. So, let $x = (x_1, ..., x_n)$ in K^n. Let $C = \max_{1 \le i \le n} |x_i|$. Then, observe that

$$||x||_p = \left\{ \sum_{k=1}^{n} |x_i|^p \right\}^{\frac{1}{p}} = C \left\{ \sum_{k=1}^{n} \frac{|x_i|^p}{C^p} \right\}^{\frac{1}{p}}$$

Define $f : [1, \infty) \to \mathbb{R}$ by

$$f(p) = \left\{ \sum_{k=1}^{n} \frac{|x_i|^p}{C^p} \right\}^{\frac{1}{p}}$$

Then, f is a positive continuous function on its domain. Also, by the definition of C, observe that

$$1 \le \sum_{k=1}^{n} \frac{|x_i|^p}{C^p} \le n$$

which means

$$1 \le \left\{ \sum_{k=1}^{n} \frac{|x_i|^p}{C^p} \right\}^{\frac{1}{p}} \le n^{\frac{1}{p}}$$

and hence we get

$$0 \le \log(f(p)) \le \frac{\log n}{p}$$

and hence as $p \to \infty$, $\log(f(p)) \to 0$, which means that $f(p) \to 1$. Hence, we have

$$\lim_{p \to \infty} ||x||_p = C$$

which shows that $l_p \to l^\infty$ as $p \to \infty$.

Now let's look at some normed function spaces:

1.4. L^p **spaces:** For $p \in [1, \infty)$, let

$$L^p := \left\{ f : [0,1] \to \mathbb{R} : \left(\int_0^1 |f|^p \right)^{\frac{1}{p}} < \infty \right\}$$

It turns out that for $f \in L^p$, the norm defined as

$$||f||_p = \left(\int_0^1 |f|^p \right)^{\frac{1}{p}}$$

satisfies the norm axioms.

Proving that l_p is a Norm

This section will involve three primary inequalities: Young's Inequality, Holder's Inequality and Minkowski's inequality. The first two are used to prove Minkowski, which is basically the fact that l_p is a norm.

First, let's start with a little definition. If $p \in (1\infty)$, then the number $q \in (1\infty)$ which satisfies

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

is called the *Holder conjugate* of p.

Theorem 1.1. *Young's Inequality*: Suppose $u, v \geq 0$, and $p, q \in (1\infty)$ be Holder conjugates. We then have

$$uv \leq \frac{u^p}{p} + \frac{v^q}{q}$$

Proof: Define the function

$$f(u) = \frac{u^p}{p} + \frac{v^q}{q} - uv$$

for $u \in [0, \infty)$, where v, p, q are fixed. Observe that

$$f'(u) = u^{p-1} - v$$

and that

$$f''(u) = (p-1)u^{p-2} > 0$$

which implies that the point where $f'(u) = 0$ is a global minimum of f. This point is

$$u = v^{\frac{1}{p-1}}$$

and it can be easily seen that $f(v^{\frac{1}{p-1}}) = 0$, and hence the inequality follows.

Theorem 1.2. Holder's Inequality: Suppose $x, y \in K^n$, where $K \in \{\mathbb{R}, \mathbb{C}\}$. Let $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$, and suppose p, q are Holder conjugates. Then we have

$$\left|\sum_{i=1}^{n} |x_i y_i|\right| \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |y_i|^q\right)^{\frac{1}{q}}$$

Note: There is an integral version of this inequality which has a very similar proof. I have written the proof in the Advanced Analysis pdf. The proof there can be imitated in this case as well.

Given the Holder inequality, we can finally prove that l_p is a norm, which is also called the Minkowski inequality:

Theorem 1.3. Suppose $x = (x_1, ...,_n)$ and $y = (y_1, ..., y_n)$ are in $\mathbb{R}^n$ or $\mathbb{C}^n$. Then,

$$||x + y||_p \leq ||x||_p + ||y||_p$$

Proof: If $p = 1$, then the inequality is just the triangle inequality for $\mathbb{R}$. So, suppose $p > 1$, and hence it as a holder conjugate q. We want to show that

$$\left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{\frac{1}{p}}$$

Consider the term on the left. We split that term as follows:

$$\sum_{i=1}^{n}|x_i+y_i|^p=\sum_{i=1}^{n}|x_i+y_i||x_i+y_i|^{p-1}\le\sum_{i=1}^{n}|x_i||x_i+y_i|^{p-1}+\sum_{i=1}^{n}|y_i||x_i+y_i|^{p-1}$$

and applying the Holder inequality on the last two sums, we get that

$$\begin{aligned}\sum_{i=1}^{n}|x_i+y_i|^p&\le\left[\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}\right]\left(\sum_{i=1}^{n}|x_i+y_i|^{(p-1)q}\right)^{\frac{1}{q}}\\&=\left[\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}\right]\left(\sum_{i=1}^{n}|x_i+y_i|^{p}\right)^{\frac{1}{q}}\end{aligned}$$

and from here we get the desired inequality.

2. Topology and Equivalent Norms

We have already done topology on general metric spaces, and since every norm is nothing but a metric space, we don't have to repeat everything again. However, let us introduce one important idea.

Suppose X is a K-vector space ($K\in\{\mathbb{R},\mathbb{C}\}$), and let $||\cdot||$ and $||\cdot||'$ be two norms on X. We say that these norms are *equivalent* if there are constants $c,C>0$ such that

$$c||x||\le||x||'\le C||x||$$

for all $x\in X$. It is not hard to see that this is an equivalence relation on the set of norms on X.

One useful consequence of equivalent norms is the following: $S\subset X$ is an open set in $(X,||\cdot||)$ if and only if it is an open set in $(X,||\cdot||')$ (and the same statement holds for closed sets). A sequence $\{x_n\}$ in $(X,||\cdot||)$ converges to a point $x\in X$ if and only if it converges to x in $(X,||\cdot||')$. So in some sense, both the norms induce the same topology on X, and many ideas like convergence and continuity remain invariant under equivalent norms. (I am not proving this fact here, but it is not hard to prove either)

Let's now look at some important aspects of compactness for normed spaces (the definition of compactness is the same as we did before).

We already know that compactness is an *intrinsic* property, not an embedding property. This basically means that if K is a subset of a space X, then K is a compact subset of X if and only if K is itself a compact metric space. We also know that closed subsets of compact spaces are compact.

Let's now define *boundedness* in a metric space. A subset K of X is said to be *bounded* if there is some $x\in X$ and $\epsilon>0$ such that

$$K\subset B(x,\epsilon)$$

An important observation regarding boundedness is that it remains invariant under equivalent norms.

Let's now prove one direction of the Heine-Borel theorem for metric spaces:

Theorem 2.1. Suppose X is a metric space, and let K be a compact subset of X. Then, K is closed and bounded.

Proof: That it is closed is already proven before. To show that it is bounded, fix $x \in X$, and consider the set $M = \{B(x,n) | n \in \mathbb{N}\}$. Then, M is an open cover of K, and so it has a finite subcover. So, it follows that K is bounded.

Before proving that *closed and bounded boxes*(which will be defined later) in $\mathbb{R}^n$ are compact, let us prove an important regarding product spaces:

Theorem 2.2. Identify K^n as $K^r \times K^d$, where $r + d = n$. Let $A \times B$ be a subset of $K^r \times K^d$. Then, $A \times B$ is open in K^n if and only if A is open in K^r and B is open in K^d.

Proof: First, suppose A is open in K^r and B is open in K^d. Let $(x, y) \in A \times B$. Now, there is a $\delta > 0$ such that $B(x, \delta) \subset A$ in K^r and $B(y, \delta) \subset B$ in K^d. Consider $B((x, y), \delta)$. If $(a, b) \in B((x, y), \delta)$, we have

$$||(x, y) - (a, b)||_2 < \delta$$

which means that

$$||(x, y) - (a, b)||_2^2 < \delta^2$$

and hence

$$||x - a||_2^2 + ||y - b||_2^2 < \delta^2$$

which means that $a \in B(x, \delta)$ and $b \in B(y, \delta)$, and hence $(a, b) \in A \times B$, proving that $B((x, y), \delta) \in A \times B$, and hence $A \times B$ is open.

Conversely, suppose $A \times B$ is open in K^n. We will show that A is open in K^r, and similarly it will follow that B is open in K^d. Fix $z \in B$, and let $x \in A$. Consider the point $(x, z) \in A \times B$. There is some $\delta > 0$ such that

$$B((x, z), \delta) \in A \times B$$

Now, let $y \in K^r$ such that $||x - y||_2 \le \delta$. Consider the point $(y, z) \in K^r \times K^d$. We have

$$||(x, z) - (y, z)||_2^2 = ||x - y||_2^2 < \delta^2$$

and hence

$$||(x, z) - (y, z)||_2 < \delta$$

which implies that $(y, z) \in A \times B$, and hence $y \in A$. So, A is open.

Using the above theorem and induction, we can now prove that closed and bounded *boxes* in $\mathbb{R}^n$ are compact. A *box* in $\mathbb{R}^n$ is defined a set of the form

$$I_1 \times I_2 \times ... \times I_n$$

where each I_i is a closed and bounded interval in $\mathbb{R}$.

Theorem 2.3. Closed and bounded boxes in $\mathbb{R}^n$ are compact.

Proof: Let $K = I_1 \times ... \times I_n$ be a closed and bounded box in $\mathbb{R}^n$. Suppose $\{Q_\alpha\}$ is an open cover of K in $\mathbb{R}^n$. Suppose $Q_\alpha = J_1 \times J_2 \times ... \times J_n$. By the above theorem, each J_i is an open set in $\mathbb{R}$. Let $B_{Q_\alpha} = J_1$ and let $C_{Q_\alpha} = J_2 \times ... \times J_n$, so that B_{Q_α} is an open set in $\mathbb{R}$.

Let $K' = I_2 \times ... \times I_n$, and hence K' is a closed and bounded box in $\mathbb{R}^{n-1}$. Fix $t \in I_1$. Let β_t be the set of all $\{Q_\alpha\}$ such that $(\{t\} \times K') \cap Q_\alpha$ is non-empty. Observe that

$$\beta_t = \{Q_\alpha | \ t \in B_{Q_\alpha}\}$$

Now, the set $\{C_{Q_\alpha} | \ Q_\alpha \in \beta_t\}$ is an open cover of K' in $\mathbb{R}^{n-1}$. By our induction hypothesis, there must be a finite subcover. Let F_t be this finite subcover of K' in $\mathbb{R}^{n-1}$.

As a corollary to the above theorem, the **Heine-Borel** theorem in $\mathbb{R}^n$ follows, which says that any closed and bounded set in $\mathbb{R}^n$ is compact, because any such set is a subset of some closed and bounded box, and the fact that closed subsets of compact sets are compact.

Using the Heine-Borel theorem, we can prove that all norms in $\mathbb{R}^n$ are equivalent:

Theorem 2.4. If $||\cdot||$ is any norm in $\mathbb{R}^n$, then $||\cdot||$ is equivalent to $||\cdot||_1$. Hence, all norms in $\mathbb{R}^n$ are equivalent.

Proof: Let $||\cdot||$ be any norm in $\mathbb{R}^n$. Consider the function $f : (\mathbb{R}^n, ||\cdot||_1) \to [0, \infty)$ defined as

$$f(x) = ||x||$$

Let us show that this function is continuous. So, let $x_0 \in \mathbb{R}^n$, and let $\{e_1, ..., e_n\}$ be the standard orthonormal basis of $\mathbb{R}^n$. Let $M = \max_{1\leq i\leq n} ||e_i||$. Suppose $x_0 = a_1e_1 + ... + a_ne_n$. For any $x = b_1e_1 + ... + b_ne_n \in \mathbb{R}^n$, we have

$$\begin{aligned} ||x - x_0|| &= \left|\left|\sum_{i=1}^{n}(a_i - b_i)e_i\right|\right| \\ &\leq \sum_{i=1}^{n} |a_i - b_i|||e_i|| \\ &\leq M\sum_{i=1}^{n} |a_i - b_i| \\ &= M||x - x_0||_1 \end{aligned}$$

and hence it follows that f is continuous on $\mathbb{R}^n$.

Now, consider the unit sphere S on the space $(\mathbb{R}^n, ||\cdot||_1)$. Being closed and bounded, it is compact. So, the function f attains a maximum and a minimum on the unit sphere, and let them be c, C respectively. So, we have that for any $u \in S$,

$$c \leq ||u|| \leq C$$

and hence it follows that $||\cdot||$ is equivalent to $||\cdot||_1$.

Thus, if V is any finite dimensional vector space over $\mathbb{R}$, then all norms on the space are equivalent.

3. Differentiation in Euclidean Spaces

First, let us introduce a norm on the space of linear transformations. Let $L(\mathbb{R}^n, \mathbb{R}^m)$ denote the space of linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$, where both the spaces have the Euclidean norm. We know that any $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ is continuous. So, consider the unit sphere $S := \{x \in \mathbb{R}^n | \, ||x|| = 1\}$. This unit sphere is closed and bounded, hence compact. Now, define

$$||T||_L = \sup_{x\in S} ||T(x)|| < \infty$$

Theorem 3.1. $||\cdot||_L$ is a norm.

Proof: First, if $T = O$, then $||T||_L = 0$. Conversely, if $||T||_L = 0$, then T is O on the unit sphere. Since T is a linear transformation, it is zero everywhere.

Second, let $\lambda \in \mathbb{R}$. So,

$$||\lambda T||_L = \sup_{x\in S} ||\lambda T(x)|| = |\lambda| \sup_{x\in S} ||T(x)|| = |\lambda|||T||_L$$

Finally, if $T_1, T_2 \in L(\mathbb{R}^n, \mathbb{R}^m)$,then we have

$$\begin{aligned}||T_1 + T_2||_L = \sup_{x\in S} ||T_1(x) + T_2(x)|| &\leq \sup_{x\in S} ||T_1(x)|| + ||T_2(x)|| \\ &\leq \sup_{x\in S} ||T_1(x)|| + \sup_{x\in S} ||T_2(x)|| \\ &= ||T_1||_L + ||T_2||_L\end{aligned}$$

and hence the proof is complete.

Let's prove some more properties of the operator norm:

Theorem 3.2. Suppose $A : V \to W$ is a linear map between finite dimensional vector spaces over $\mathbb{R}$. Suppose there is a real number M such that $||Ax|| \leq M||x||$ for all $x \in V$, then $||A|| \leq M$. In fact $||A||$ is the infimum of all such numbers M.

Proof: For any u in the unit sphere, we have

$$||Au|| \leq M$$

and hence by definition, we have

$$||A|| \leq M$$

It is clear that $||A||$ is the infimum of all such numbers.

Theorem 3.3. Suppose U, V, W are finite dimensional normed linear spaces over $\mathbb{R}$. Let $A : U \to V$ and $B : V \to W$ be linear transformations. Then,

$$||BA|| \leq ||B||\,||A||$$

Proof: Let $x \in S$, where S is the unit sphere in U. Then,

$$||BAx|| \leq ||B||||Ax|| \leq ||B||\,||A||$$

and hence taking the supremum over S, we get

$$||BA|| \leq ||B||\,||A||$$

Theorem 3.4. Suppose U, V, W are as in the previous theorem. Let $\{A_n\}$ be a sequence of linear maps from U to V, and let $\{B_n\}$ be a sequence of linear maps from V to W. Further, suppose that $A_n \to A$ and $B_n \to B$ as $n \to \infty$ (here, the limits are taken with respect to the operator norm). Then, $B_nA_n \to BA$ as $n \to \infty$. So, the map $P : L(U, V) \times L(V, W) \to L(U, W)$ given by $(S, T) \mapsto TS$ is continuous, where the space $L(U, V) \times L(V, W)$ is given any norm. (note that it is a finite dimensional space over $\mathbb{R}$)

Proof: Observe that

$$\begin{aligned}||B_nA_n - BA|| &= ||B_nA_n - B_nA + B_nA - BA|| \\ &\leq ||B_nA_n - B_nA|| + ||B_nA - BA|| \\ &\leq ||B_n||\,||A_n - A|| + ||B_n - B||\,||A||\end{aligned}$$

and the right hand side converges to 0 as $n \to \infty$. So, the proof is complete.

Infinite Series Representations for Operators: In this section, we will see how the notion of infinite series also makes sense in the space of operators. We define a series of operators to be *convergent* if the partial sums of the series converges to some quantity, the limits being with respect to the operator norm.

Theorem 3.5. Suppose V is a finite-dimensional vector space over $\mathbb{R}$. Suppose $A \in L(V, V)$ such that $||A|| < 1$. Then:

(1) $\sum_{i=0}^{\infty} A^i$ is convergent.
(2) $I - A$ is invertible. Moreover, we have

$$(I-A)^{-1} = \sum_{i=0}^{\infty} A^i$$

Proof: First, observe that the space $L(V,V)$ is isomorphic to $\mathbb{R}^{(\dim V)^2}$, and since the latter is complete, the former is also complete. So, any Cauchy-Sequence in $L(V,V)$ converges.

Now, suppose $\epsilon > 0$ is given. So, there is some $N \in \mathbb{N}$ such that

$$\sum_{i=n}^{m} ||A||^i < \epsilon$$

for all $n, m \geq N$. Also, by the properties of the operator norm, we have

$$\left|\left|\sum_{i=n}^{m} A^i\right|\right| \leq \sum_{i=n}^{m} ||A^i|| \leq \sum_{i=n}^{m} ||A||^i < \epsilon$$

and hence the partial sums of the series form a Cauchy-sequence, and hence the series is convergent.

To prove the second assertion, observe that

$$||S_n(I-A) - I|| = ||A^{n+1}||$$

and the right hand side tends to 0 as $n \to \infty$. So, it follows that

$$(I-A)^{-1} = \sum_{k=0}^{\infty} A^i$$

Theorem 3.6. The function $\det : L(K^n, K^n) \to K$ is a continuous map. $\mathrm{GL}(K^n)$ is an open subset of $L(K^n, K^n)$, and the function $\mathrm{Inv} : \mathrm{GL}(K^n) \to \mathrm{GL}(K^n)$ defined by

$$M \mapsto M^{-1}$$

is also a continuous map. Here, K is either $\mathbb{R}$ or $\mathbb{C}$, and $\mathrm{GL}(V)$ is the set of all invertible operators on V, where V is a finite dimensional vector space over K.

Proof: First, we identify $T \in L(K^n, K^n)$ by its matrix. Since the determinant is a polynomial in the entries of the matrix, it is a continuous function (here again, we identify the space $M_n(K)$ with K^{n^2}, and we use the fact that all norms are equivalent).

Now, to show that $\mathrm{GL}(K^n)$ is an open subset of $L(K^n, K^m)$, observe that $\det^{-1}(K - \{0\}) = \mathrm{GL}(K^n)$, and since $K - \{0\}$ is an open subset of K, it follows that $\mathrm{GL}(K^n)$ is an open subset of $L(K^n, K^m)$.

Finally, we show that Inv is a continuous function. Here, we will identify the space $\mathrm{GL}(K^n)$ as the space of all invertible matrices in $M_n(K)$, which we denote by $\mathrm{GL}_n(K)$. Observe that, if $M \in \mathrm{GL}_n(K)$, then

$$M^{-1} = \frac{1}{\det M}\mathrm{adj}(M)$$

where $\mathrm{adj}(M)$ is the adjoint of M, or the transpose of the cofactor matrix. We have already seen that det is continuous on $\mathrm{GL}_n(K)$. Now, each cofactor of the matrix is a polynomial in the entries, and hence continuous. Combining all this, we see that Inv is continuous.

Note: The exercises referred to in this section are taken from the book "Calculus on Manifolds" by Michael Spivak.

Here, we will reformulate the definition of the derivative a bit, so that it can also be applied to general Euclidean spaces.

A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at a point $a \in \mathbb{R}$ if there is a number $f'(a) \in \mathbb{R}$ such that

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

Define $\lambda : \mathbb{R} \to \mathbb{R}$ by λ is a linear map.

In general, consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$, where both spaces have been given the $||\cdot||_2$ norm. We say that f is differentiable at $a \in \mathbb{R}^n$, if there is a linear map $\lambda_a : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h\to 0} \frac{||f(a+h) - f(a) - \lambda_a(h)||_2}{||h||_2} = 0$$

Let us show the uniqueness of this linear transformation:

Theorem 3.7. Suppose f is differentiable at some point $a \in \mathbb{R}^n$. Then, the linear transformation λ_a is unique.

Proof: Suppose μ and λ are linear transformations such that

$$\lim_{h\to 0} \frac{||f(a+h) - f(a) - \lambda(h)||}{||h||} = \lim_{h\to 0} \frac{||f(a+h) - f(a) - \mu(h)||}{||h||} = 0$$

Observe that

$$\lim_{h\to 0} \frac{||\lambda(h) - \mu(h)||}{||h||} \leq \lim_{h\to 0} \frac{||f(a+h) - f(a) - \lambda(h)||}{||h||} + \lim_{h\to 0} \frac{||f(a+h) - f(a) - \mu(h)||}{||h||} = 0$$

Since λ and μ are linear maps, they agree at the origin. If $x \in \mathbb{R}^n$ such that $x \neq O$, we have

$$\lim_{t\to 0} \frac{||\lambda(tx) - \mu(tx)||}{||tx||} = 0$$

and hence $\lambda(x) = \mu(x)$.

The matrix associated to the linear map λ_a is called the *Jacobian matrix*.

4. Exercises on Page 17

2-1. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a point $a \in \mathbb{R}^n$. Then, there is a linear map λ such that

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

If $\epsilon > 0$ is given, then there is some $\delta > 0$ such that $h \in B(O, \delta)$ implies that

$$||f(a+h) - f(a) - \lambda(h)|| < \epsilon||h||$$

which implies that

$$||f(a+h) - f(a)|| < ||\lambda(h)|| + \epsilon||h||$$

As $h \to O$, the right hand side tends to 0 (because λ is a continuous map and $\lambda(O) = 0$) and hence the function is continuous at a.

2-4. Let f be the given function.

(a) Suppose $x \in \mathbb{R}^2$. If $x = O$, then $h(t) = 0$ for all $t \in \mathbb{R}$, and hence h is differentiable. So, let $x \neq 0$. Observe that

$$h(t) = \begin{cases} t|x|g\left(\dfrac{x}{|x|}\right), t \neq 0 \\ 0, \, t = 0 \end{cases}$$

It follows that h is differentiable at all points other than $x = 0$. At $x = 0$, we have

$$\lim_{t\to 0} \frac{h(t)}{t} = \lim_{t \to 0} |x|g\left(\frac{x}{|x|}\right) = |x|g\left(\frac{x}{|x|}\right)$$

and hence h is differentiable everywhere.

(b)

Let's now prove the **chain rule** of differentiation:

Theorem 4.1. Chain Rule: If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^n$, and if $g : \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $f(a)$, then $g \circ f$ is differentiable at a, and we have

$$D(g \circ f)(a) = Dg(f(a)) \circ Df(a)$$

Proof: Before proving this result, we use the following: if $\mu : \mathbb{R}^a \to \mathbb{R}^b$ is a linear transformation, define

$$||\mu||_L = \sup_{x \in S} ||\mu(x)||$$

where S is the unit sphere in $\mathbb{R}^a$. This value always exists because S is compact and μ is continuous, and this is a norm on the space of linear transformations. The following inequality (which is not difficult to prove) holds:

$$||\mu(h)||_2 \leq ||\mu||_L ||h||_2$$

for all $h \in \mathbb{R}^a$.

Now, let λ and μ be the derivatives of f and g at a and $f(a)$ respectively. Define

$$\begin{aligned} \phi(h) &= f(a+h) - f(a) - \lambda(h) \\ \epsilon(h) &= g(f(a)+h) - g(f(a)) - \mu(h) \\ \beta(h) &= g(f(a+h)) - g(f(a)) - \mu(\lambda(h)) \end{aligned}$$

Now, observe that

$$\begin{aligned} g(f(a+h)) - g(f(a)) - \mu(\lambda(h)) &= g(f(a) + \phi(h) + \lambda(h)) - g(f(a)) - \mu(\lambda(h)) \\ &= \epsilon(\phi(h) + \lambda(h)) + \mu(\phi(h) + \lambda(h)) - \mu(\lambda(h)) \\ &= \epsilon(\phi(h) + \lambda(h)) + \mu(\phi(h)) \end{aligned}$$

Now, we know that $\phi(h) + \lambda(h)$ converges to O as $h \to O$. Also, we know that

$$\frac{||\epsilon(h)||}{||h||} \to O$$

as $h \to O$. Let $\epsilon_0 > 0$ be any given number. So, there some $B(O, \delta_1)$ such that for all $h \in B(O, \delta_1)$, we have

$$||\epsilon(h)|| \leq \epsilon_0 ||h||$$

Now, let $B(O, \delta_2)$ be a ball such that for all $h \in B(O, \delta_2)$, we have that

$$||\phi(h) + \lambda(h)|| < \delta_1$$

which in turn implies that

$$||\epsilon(\phi(h)+\lambda(h))|| \leq \epsilon_0||\phi(h)+\lambda(h)|| \leq \epsilon_0||\phi(h)||+\epsilon_0||\lambda(h)||$$

So, for all $h \in B(O, \delta_2)$, we have

$$\begin{aligned}\frac{||\beta(h)||}{||h||} = \frac{||\epsilon(\phi(h)+\lambda(h))+\mu(\phi(h))||}{||h||} &\leq \frac{||\epsilon(\phi(h)+\lambda(h))||}{||h||}+\frac{||\mu(\phi(h))||}{||h||} \\ &\leq \frac{\epsilon_0||\phi(h)||}{||h||}+\frac{\epsilon_0||\lambda(h)||}{||h||}+\frac{||\mu(\phi(h))||}{||h||} \\ &\leq \frac{\epsilon_0||\phi(h)||}{||h||}+\frac{\epsilon_0||\lambda||_L||h||}{||h||}+\frac{||\mu||_L||\phi(h)||}{||h||} \\ &= \frac{\epsilon_0||\phi(h)||}{||h||}+\epsilon_0||\lambda||_L+\frac{||\mu||_L||\phi(h)||}{||h||}\end{aligned}$$

Now, as $h \to O$ the right hand side tends to $\epsilon_0||\lambda||_L$. But, since ϵ_0 was arbitrary, the claim follows.

Let's now prove some more basic properties of derivatives:

Theorem 4.2. The following properties hold:

(1) If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a constant function, then $Df(a) = O$ at every $a \in \mathbb{R}^n$, where O linear map.

(2) If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, then $Df(a) = f$ for all $a \in \mathbb{R}^n$.

(3) $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^n$ if and only if each co-ordinate function f_i is differentiable at a, and we have

$$Df(a) = (Df_1(a), ..., Df_m(a))$$

In terms of matrices, $Df(a)$ is the matrix, whose i^{th} row is $Df_i(a)$.

(4) If $s : \mathbb{R}^2 \to \mathbb{R}$ such that $s(x, y) = x + y$, then

$$Ds(a, b) = s$$

at all points $(a, b) \in \mathbb{R}^2$.

Proof: To prove (1), suppose f is a constant function. Let $a \in \mathbb{R}^n$. Then we have

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

and hence $Df(a) = O$, where O is the zero linear map.

To prove (2), suppose f is linear. Then, we have

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

and hence $Df(a) = f$.

Let's now prove (3). First, suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^n$, and let λ be the derivative. Let

$$\lambda = (\lambda_1, ..., \lambda_m)$$

where each λ_i is a linear map from $\mathbb{R}^n \to \mathbb{R}$. We will show that $Df_i(a) = \lambda_i$. We know that

$$\lim_{h \to O} \frac{||f(a+h)-f(a)-\lambda(h)||}{||h||} = 0$$

and hence for each i, we have

$$\begin{aligned}\lim_{h\to O}\frac{|f_i(a+h)-f_i(a)-\lambda_i(h)|}{||h||} &\leq \lim_{h\to O}\frac{||f(a+h)-f(a)-\lambda(h)||_\infty}{||h||}\\ &\leq \lim_{h\to O}\frac{||f(a+h)-f(a)-\lambda(h)||_2}{||h||}=0\end{aligned}$$

and hence it follows that $Df_i(a)=\lambda_i$. Similarly, the converse may be proved.

To prove (4), observe that s is a linear map. Hence the claim follows.

Let's prove the product and quotient rules as a corollary:

Theorem 4.3. Suppose $f, g : \mathbb{R}^n \to \mathbb{R}^m$ are differentiable at $a \in \mathbb{R}^n$. Then, we have

$$\begin{aligned}D(f+g)(a) &= Df(a)+Dg(a)\\ D(fg)(a) &= Df(a)g(a)+f(a)Dg(a)\\ D(f/g)(a) &= \frac{g(a)Df(a)-f(a)Dg(a)}{[g(a)]^2}\end{aligned}$$

Proof: The first equation is true because

$$\lim_{h\to O}\frac{||f(a+h)+g(a+h)-f(a)-g(a)-\lambda_1(h)-\lambda_2(h)||}{||h||}=0$$

as a consequence of the triangle inequality. In fact, this statement holds true for $f, g : \mathbb{R}^n \to \mathbb{R}^m$.

We now prove the second statement. Let $a \in \mathbb{R}^n$. Consider the function $F : \mathbb{R}^n \to \mathbb{R}$ given by $F(x) = h \circ f(x)$, where $h = x^2$. By the chain rule, we know that h is differentiable at a, and hence

$$DF(a) = Dh(f(a))Df(a) = 2f(a)Df(a)$$

We have the identity

$$4fg = (f+g)^2 - (f-g)^2$$

and hence

$$\begin{aligned}4D(fg)(a) &= 2(f+g)(a)D(f+g)(a) - 2(f-g)(a)D(f-g)(a)\\ &= 4f(a)Dg(a) + 4g(a)Df(a)\end{aligned}$$

To prove the last equality, suppose $g(a) \neq 0$. Consider the function $G : \mathbb{R}^n \to \mathbb{R}$ defined by $G(x) = h \circ g(x)$, where $h(x) = \dfrac{1}{x^2}$. By the chain rule, we have

$$DG(a) = \frac{-Dg(a)}{[g(a)]^2}$$

and applying the product rule with $\dfrac{f}{g}$, we get the desired equality.

Using the previous theorem and the chain rule, we can compute the derivatives of many functions from $\mathbb{R}^n \to \mathbb{R}$.

5. Exercises on Page 22

2-10. In this problem, we will compute the derivatives of some functions using the theorems we have thus far:

(a) $f(x, y, z) = x^y$. If $x \in \mathbb{R}^3$, then we have

$$f(x) = h_1 \circ (\pi_2 \log \pi_1)(x)$$

where $h_1(y) = e^y$ for $y \in \mathbb{R}$. First, let us compute $D(\pi_2 \log \pi_1)(x)$. By the product rule, we have

$$D(\pi_2 \log \pi_1)(x) = \pi_2(x)D(\log \pi_1)(x) + \log \pi_1(x)D(\pi_2)(x)$$

So, if $x = (x, y, z)$, we have

$$D(\pi_2 \log \pi_1)(x) = \begin{bmatrix} \frac{y}{x} & \log x & 0 \end{bmatrix}$$

where the right hand side is a 3×3 matrix.

Now, we have

$$Df(x) = x^y \begin{bmatrix} \frac{y}{x} & \log x & 0 \end{bmatrix}$$

(b) $f(x, y, z) = (x^y, z)$. We know that f is differentiable if and only if each component is differentiable. From part a, we know that the first component of f is differentiable. Now, consider

$$f_2(x, y, z) = z$$

This is nothing but the third projection mapping π_3. So, we have that, if $x = (x, y, z)$, then

$$Df(x) = (Df_1(x), Df_2(x))$$

and so the matrix of $Df(x)$ is given by

$$\begin{bmatrix} x^{y-1}y & x^y \log x & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(j) $f(x, y) = (\sin(xy), \sin(x \sin y), x^y)$ Here, we shall calculate the derivatives of the component functions separately, and that will give us the derivative of f.

The derivative of $f_1(x, y, z) = \sin(xy)$ is given by

$$Df_1(x, y, z) = \cos(xy)[x, y]$$

Similarly, we have

$$Df_2(x, y, z) = \cos(x \sin y)[\sin y, x \sin y]$$

and

$$Df_3(x, y, z) = x^y \begin{bmatrix} \frac{y}{x} & \log x \end{bmatrix}$$

and hence the matrix of $Df(x, y, z)$ has been found.

2-11. In this problem, we assume that $g : \mathbb{R} \to \mathbb{R}$ is continuous:

(a) $f(x, y) = \int_a^{x+y} g$. Let G be defined by

$$G(x) = \int_a^x g$$

So, observe that

$$f(x, y) = G \circ (\pi_1 + \pi_2)$$

and hence by the chain rule we see that

$$f'(x, y) = \begin{bmatrix} g(x + y) & g(x + y) \end{bmatrix}$$

(b) $f(x, y) = \int_a^{xy} g$. Again, by the same technique as above, we obtain

$$f'(x, y) = g(xy) \begin{bmatrix} y & x \end{bmatrix}$$

6. Partial Derivatives

Now, we will look at the connection between $Df(a)$ and the partial derivatives of f.

First, suppose $f : \mathbb{R}^n \to \mathbb{R}$. The i^{th} partial derivative of f at $a = (a_1, ..., a_i, ..., a_n) \in \mathbb{R}^n$, denoted by $D_i f(a)$, is the limit (if it exists)

$$\lim_{t\to 0} \frac{f(a_1, ..., a_i + t, ..., a_n) - f(a_1, ..., a_n)}{t} = \lim_{t\to 0} \frac{f(a + te_i) - f(a)}{t}$$

where e_i is the i^{th} member of the standard orthonormal basis of $\mathbb{R}^n$.

Now, we extend the notion of partial derivatives. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is function. Then, $f = (f_1, ..., f_m)$. We denote by $D_i f_j(a)$ the i^{th} partial derivative of f_j at a (if it exists).

We now show an important relationship between the derivative of a function and its partial derivatives:

Theorem 6.1. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a point $a \in \mathbb{R}^n$. Then, $D_j f_i(a)$ exists for each $1 \leq j \leq n$ and $1 \leq i \leq m$, and the matrix of $Df(a)$ is the $m \times n$ matrix $[D_j f_i(a)]$.

Proof: Suppose f is differentiable at a. Then, each component function f_i ($1 \leq i \leq m$) is also differentiable at a. Now, let $Df_i(a)$ be the derivative of f_i at a (which is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}$). Now, we have

$$\lim_{t\to O} \frac{|f_i(a + t) - f_i(a) - Df_i(a)(t)|}{||t||} = 0$$

Let e_j be the j^{th} member of the standard orthonormal basis of $\mathbb{R}^n$. Observe that

$$Df_i(a)(e_j) = [Df(a)]_{ij}$$

where $[Df(a)]$ is the matrix of $Df(a)$. Now, we have

$$\lim_{t\to 0} \frac{|f_i(a + e_j t) - f(a) - Df_i(a)(e_j t)|}{||te_j||} = 0$$

which implies that

$$\lim_{t\to 0} \left| \frac{f_i(a + e_j t) - f_i(a)}{t} - Df_i(a)(e_j) \right| = 0$$

and hence we have

$$D_j f_i(a) = Df_i(a)(e_j)$$

and hence the claim follows.

The converse of the above theorem is not true, but with an additional hypothesis it becomes true:

Theorem 6.2. Suppose $f : U \to \mathbb{R}^m$, where U is an open subset of $\mathbb{R}^n$. Suppose f is differentiable in U. Then, it is C^1 if and only if each partial derivative $D_j f_i$ is continuous in U. (a function is said to be C^1 if it is differentiable and the derivative is continuous)

Note: Such a function is called *continuously differentiable*.

Proof: First, suppose that f is differentiable on U. Then, we know that the matrix of $Df(x)$ at some point $x \in U$ is the Jacobian matrix $D_j f_i(x)$, which is a matrix in the space $M_{mn}(\mathbb{R})$. Now, $Df : U \to L(\mathbb{R}^m, \mathbb{R}^n)$, and we first identify $L(\mathbb{R}^n, \mathbb{R}^m)$ with the space $M_{mn}(\mathbb{R})$. So, Df is a continuous function from U to M_{nm}, where the norm on $M_{nm}(\mathbb{R})$ is again the operator norm. Finally, we identify M_{nm} with the space $\mathbb{R}^{nm}$. So, Df is a continuous map from U to $\mathbb{R}^{nm}$, where

$$Df(x) = (D_1 f_1(x), D_2 f_1(x), ..., D_1 f_2(x), ..., D_n f_m(x))$$

Finally, all norms on $\mathbb{R}^{nm}$ are equivalent, and hence it follows that each component is continuous, meaning that each partial derivative is continuous. The converse also has the same argument in its proof.

Remark: As seen in this proof, using vector space isomorphisms to identify one object with the other reduces a lot of work!

Now, we know by the previous theorem that if the function is C^1, then each partial derivative exists and is continuous. It actually turns out that the converse is also true:

Theorem 6.3. Suppose $f : U \to \mathbb{R}^m$ is a function such that each partial derivative $D_j f_i$ exists on U and is continuous on U. Then, f is differentiable on U.

Proof: To prove this statement, we will assume without loss of generality that $f : \mathbb{R}^n \to \mathbb{R}$, because we know that a function is differentiable if and only if each component is differentiable.

So, let $a = (a_1, ..., a_n) \in U$ be fixed, and let $h = (h_1, ..., h_n) \in \mathbb{R}^n$ such that $a+h \in U$. We claim that f is differentiable at a, and the derivative is the linear map $\lambda : \mathbb{R}^n \to \mathbb{R}$ given by

$$\lambda(h) = \langle \nabla f(a), h \rangle$$

Define $v_0 = a$, and recursively define

$$v_k = v_{k-1} + h_k e_k$$

Observe that we have $v_n = a + h$. So, we can write

$$f(a+h) - f(a) = f(v_n) - f(v_{n-1}) + ... + f(v_1) - f(v_0)$$

Moreover, observe that

$$f(v_k) - f(v_{k-1}) = h_k D_k f(\zeta_k)$$

for some ζ_k in the line segment $[v_{k-1}, v_k]$. It follows that

$$f(a+h) - f(a) = h_n D_n f(\zeta_n) + ... + h_1 D_1 f(\zeta_1) = \langle w, h \rangle$$

where we define

$$w = (D_1 f(\zeta_1), ..., D_n f(\zeta_n))$$

So, we have

$$\frac{||f(a+h) - f(a) - \lambda(h)||}{||h||} = \frac{||\langle w - \nabla f(a), h \rangle||}{||h||} \leq ||w - \nabla f(a)||$$

Since it is assumed that each partial derivative is continuous, we see that $w \to \nabla f(a)$ as $h \to O$. This completes the proof.

7. The Mean Value Theorems in $\mathbb{R}^n$

Here, we will see two versions of the mean value theorems in $\mathbb{R}^n$.

Theorem 7.1. MVT 1: Suppose $g : [a, b] \to \mathbb{R}^n$ is a function, continuous on $[a, b]$ and differentiable on (a, b). Then, there is some $c \in (a, b)$ such that

$$||g(b) - g(a)|| \leq (b - a)||g'(c)||$$

Note: Here, the default norm in $\mathbb{R}^n$ is the l_2 norm. Observe that we are identifying the derivative $g'(c)$ as a vector in $\mathbb{R}^n$, because the operator norm and the l_2 norm are the same in this case.

Proof: Put $z = g(b) - g(a)$. Now, we define $\phi : [a, b] \to \mathbb{R}$ by

$$\phi(t) = \langle g(t), z \rangle$$

Because the inner product is continuous and g is continuous, it follows that ϕ is continuous on $[a, b]$. Also, because the inner product is a linear function in the first variable and since g is differentiable on (a, b), it follows that ϕ is also differentiable on (a, b), and for $t \in (a, b)$ we have

$$\phi'(t) = \langle g'(t), z \rangle$$

Now, applying the mean value theorem to ϕ, we get

$$\phi(b) - \phi(a) = (b - a)\langle g'(c), z \rangle$$

for some $c \in (a, b)$. So,

$$||z||^2 = (b - a)|\langle g'(c), z \rangle|$$

and by the Cauchy-Schwarz inequality, we see that

$$||z||^2 \leq (b - a)||g'(c)||||z||$$

and from here, we get the desired inequality. Here is another version of the MVT:

Theorem 7.2. Suppose U is a convex open set in $\mathbb{R}^n$, and let $f : U \to \mathbb{R}^m$ be a differentiable function on U. Suppose for each point $x \in U$, it is true that $||f'(x)|| \leq M$, where the norm being used in the operator norm. Then, for any $a, b \in \mathbb{R}^n$, we have

$$||f(b) - f(a)|| \leq M||b - a||$$

Proof: Suppose $a, b \in U$. Since U is convex, every point of the form $a + t(b - a)$ for $t \in [0, 1]$ is in U. Define a function $g : [0, 1] \to \mathbb{R}^m$ by

$$g(t) = f(a + t(b - a))$$

So, g is differentiable in $(0, 1)$ and continuous in $[0, 1]$. Also, because $g = f \circ \gamma$, where

$$\gamma(t) = a + (b - a)t$$

we have that

$$g'(t) = f'(\gamma(t))(b - a)$$

where $(b - a)$ is interpreted as a column matrix.

Now, by the mean value theorem on g, there is some $c \in (0, 1)$ such that

$$||g(1) - g(0)|| = ||f(b) - f(a)|| \leq ||g'(c)|| = ||f'(\gamma(t))(b - a)|| \leq M||b - a||$$

by the inequality of operator norm. So, the proof is complete.

8. The Inverse and Implicit Function Theorems

These are perhaps the most impportant theorems when it comes to multivariable calculus. Let's start with the **inverse function theorem**. Suppose $f : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function. Suppose there is a point $a \in \mathbb{R}$ where the derivative is non-zero. Then, f is monotonic in some neighborhood of that point, and hence f is one-one in that neighborhood. It is also clear that the inverse function f^{-1} is differentiable in a neighborhood of the point $f(a)$. The same situation occurs in multivariable functions as well.

Theorem 8.1. *Inverse function theorem*: Let U be an open set in $\mathbb{R}^n$. Suppose $f : U \to \mathbb{R}^n$ is $\mathcal{C}^1$ on U, such that $f(0) = 0$ and $f'(0) = I$, where I is the identity mapping. Then, there is an open neighborhood V of 0 contained in U, and an open neighborhood W of 0 contained in $\mathbb{R}^n$, such that $f(V) = W$, and f is one-one on V. Also, the inverse function $f^{-1} : W \to V$ is $\mathcal{C}^1$, and for any $y \in W$,

$$(f^{-1})'(y) = [f'(f^{-1}(y))]^{-1}$$

Proof: First, we will show that f is one-one in some neighborhood of 0 contained in U. Since Df is assumed to be continuous on U, it follows that there is an open neighborhood Q (which can also assumed to be convex, as we can just take an open ball) containing 0 such that

$$||I - f'(x)|| \leq \frac{1}{2}$$

for all $x \in Q$ (this is the operator norm). Observe that, as we proved earlier, this means that $I - (I - f'(x)) = f'(x)$ is invertible for all $x \in Q$. Now, define

$$g(x) = f(x) - x$$

for $x \in U$. Then,

$$g'(x) = f'(x) - I$$

and for $x \in Q$, we have

$$||g'(x)|| = ||f'(x) - I|| \leq \frac{1}{2}$$

Applying the second version of the mean value theorem on g, we see that for any $x_1, x_2 \in Q$ (recall Q is convex), we have

$$||g(x_1) - g(x_2)|| \leq \frac{1}{2}||x_1 - x_2||$$

and hence

$$||x_1 - x_2|| - ||f(x_1) - f(x_2)|| \leq \frac{1}{2}||x_1 - x_2||$$

and hence

$$\frac{1}{2}||x_1 - x_2|| \leq ||f(x_1) - f(x_2)||$$

which implies that f is one-one on Q.

Now, take we take a closed ball $\overline{B}(0, r)$ contained in Q (possible because Q is open). Let $\partial\overline{B}$ be the boundary of this ball, which is a compact set. Now, $||f||$ attains a minimum on this boundary (because f is continuous). Let δ be this minimum. First, since f is one-one on $\overline{B}(0, r)$, and since $f(0) = 0$, we see that $\delta > 0$. We take an open ball $W = B'\left(0, \frac{\delta}{2}\right)$ around the origin, and claim that $B'(0, \frac{\delta}{2}) \subset f(B(0, r))$, i.e this ball is contained in this image. Let's try to prove this.

Let $y \in B'(0, \frac{\delta}{2})$. Then, consider the function $F : \overline{B}(0, r) \to \mathbb{R}$ be given by

$$F(x) = ||f(x) - y||^2 = \sum_{i=1}^{n} (f_i(x) - y_i)^2$$

Since each f_i is continuous (because f is continuous), it follows that F is continuous on $\overline{B}(0, r)$. Since $\overline{B}(0, r)$ is compact, F attains a minimum on $\overline{B}(0, r)$. Also, we claim that this minimum is in the interior of $\overline{B}(0, r)$, i.e in the open ball $B(0, r)$. To show this, observe that

$$||f(0) - y|| = ||y|| < \frac{\delta}{2}$$

and if $x \in \partial\overline{B}$, then we have

$$||f(x) - y|| \geq ||f(x)|| - ||y|| > \delta - \frac{\delta}{2} = \frac{\delta}{2}$$

So, let $x_0 \in B(0, r)$ be the point where F attains its minimum. Since F is differentiable in $\overline{B}(0, r)$, it follows that each partial derivative of F at x_0 vanishes. But, observe that for $1 \leq j \leq n$, we have

$$D_jF(x_0) = \sum_{i=1}^{n} 2D_jf_i(x_0)(f_i(x_0) - y_i) = 0$$

which implies that

$$\begin{bmatrix} f_1(x_0) - y_1 & ... & f_n(x_0) - y_n \end{bmatrix} \begin{bmatrix} D_jf_1(x_0) \\ ... \\ D_jf_n(x_0) \end{bmatrix} = 0$$

which implies that

$$\begin{bmatrix} f_1(x_0) - y_1 & ... & f_n(x_0) - y_n \end{bmatrix} Df(x_0) = 0$$

where the last equation is matrix multiplication. However, we have shown that $Df(x_0)$ is invertible, and hence it follows that

$$f(x_0) = y$$

and so the claim has been proven. Now, let $V = B(0, r) \cap f^{-1}(W)$, so that V is open. It is easy to see that $f(V) = W$, and f is one-one on V, because it is one-one on $B(0, r)$. So, the required open sets V and W have been found.

Now, consider $f^{-1} : W \to V$, which is obviously one-one and continuous, because

$$\frac{1}{2}||f^{-1}(x_1) - f^{-1}(x_2)|| \leq ||x_1 - x_2||$$

for all $x_1, x_2 \in W$, by the inequality we proved above in the proof.

We will now show that f^{-1} is differentiable on W, and infact it is $\mathcal{C}^1$ on W. So, let $y_0 \in W$, let $k \neq O \in W$ such that $y_0 + k \in W$ (possible because W is open). Let $x_0 = f^{-1}(y_0)$. Let $A = f'(x_0)$. We know that

$$\frac{||f(x_0 + h) - f(x_0) - Ah||}{||h||} \to 0$$

as $h \to 0$. Now, let $f^{-1}(y_0 + k) = x_0 + h$, so that $h \neq O$ (because f^{-1} is one-one) and that $h \to O$ as $k \to O$ (by the continuity of f^{-1}. Also, observe that h depends on k). Also, note that

$$||f^{-1}(y_0 + k) - f^{-1}(y_0)|| \leq 2||k||$$

implying that

$$\frac{||h||}{||k||} \leq 2$$

for all such $k \neq O$. Now, we have

$$\begin{aligned}\frac{||f^{-1}(y_0+k)-f^{-1}(y_0)-A^{-1}k||}{||k||} &= \frac{||h-A^{-1}k||}{||k||} \\ &= \frac{||A^{-1}(Ah-k)||}{||k||} \\ &= \frac{||A^{-1}(Ah-k)||}{||h||}\cdot\frac{||h||}{||k||} \\ &\leq 2\frac{||A^{-1}||||Ah-k||}{||h||}\end{aligned}$$

and the right hand side tends to 0 as $k \to O$. So, this shows that $Df^{-1}(y_0) = A^{-1}$, proving the claim that f^{-1} is differentiable.

Now, the derivative of f^{-1} is given by

$$Df^{-1} = \text{Inv} \circ Df$$

and since both Inv and Df are continuous, it follows that Df^{-1} is also continuous, proving that f^{-1} is $\mathcal{C}^1$.

Instead of looking at 0, we can generalise this to any point $a \in U$, and in fact we only want $f'(a)$ to be invertible:

Corollary: Let $f : U \to \mathbb{R}^n$ be a $\mathcal{C}^1$ function, where U is open in $\mathbb{R}^n$. Let a be a point in U, and suppose $f(a) = b$. Also, suppose that $f'(a)$ is invertible. Then, there is an open set V containing a in $\mathbb{R}^n$ and an open set W containing b in $\mathbb{R}^n$ such that $f(V) = W$, f is one-one on V, and the inverse function $f^{-1} : W \to V$ is $\mathcal{C}^1$ with

$$Df^{-1}(y) = [Df(f^{-1}(y))]^{-1}$$

Proof: First, let us assume that $f'(a) = I$. Consider the function $g : U \to \mathbb{R}^m$ defined by

$$g(x) = f(a + x) - b$$

and clearly, g is $\mathcal{C}^1$ with

$$g'(0) = f'(a + 0) = I$$

and g is defined in an open set containing 0. So, the hypothesis of the inverse function theorem apply to g, and it is not hard to see that the statement follows for f as well, where $0 \in U$ is replaced by a and $0 \in \mathbb{R}^n$ is replaced by b.

Next, suppose that $f'(a) = L$ is invertible. Define a new function $g : U \to \mathbb{R}^n$ by

$$g(x) = (L^{-1} \circ f)(x)$$

Then, the hypothesis of the inverse function theorem apply to g as well, and this time, $g'(a) = I$. So, the theorem is true for g, and hence it is also true for f because L^{-1} is an automorphism of $\mathbb{R}^n$.

Next, we will look at an extremely important corollary of the inverse function theorem, called the **Implicit Function theorem**. In simple words, this theorem says that a curve in $\mathbb{R}^m$, under some conditions, looks like the graph of a $\mathcal{C}^1$ function:

Theorem 8.2. Suppose $\phi : U \to \mathbb{R}^m$ is a $\mathcal{C}^1$ function, where U is open open subset of $\mathbb{R}^n \times \mathbb{R}^m$. Suppose the derivative of ϕ at a point (x_0, y_0) is given by the block matrix

$$D\phi(x,y) = \begin{bmatrix} \dfrac{\partial\phi}{\partial x}_{(x_0,y_0)} & \dfrac{\partial\phi}{\partial y}_{(x_0,y_0)} \end{bmatrix}$$

Let $(a, b) \in U$ ($a \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$) such that

$$\phi(a, b) = c$$

for some $c \in \mathbb{R}^m$. Further, assume that the map

$$\frac{\partial \phi}{\partial y}_{(a,b)}$$

at the point (a, b) is invertible. Then, there is an open neighborhood W of a in $\mathbb{R}^n$, and a unique $\mathcal{C}^1$ map $f : W \to \mathbb{R}^m$ such that for all $x \in W$, we have that

$$(x, f(x)) \in U$$

and

$$\phi(x, f(x)) = c$$

meaning that f is an implicit solution to the equation $\phi(x, y) = c$.

Proof: First, we define a new function $\psi : U \to \mathbb{R}^n \times \mathbb{R}^m$ given by

$$\psi(x, y) = (x, \phi(x, y))$$

Then, the derivative of ψ at the point (a, b) is given by the block matrix

$$D\psi(a, b) = \begin{bmatrix} I & 0 \\ \dfrac{\partial \phi}{\partial x} & \dfrac{\partial \phi}{\partial y} \end{bmatrix}$$

where I is the identity $n \times n$ matrix. Clearly, we can see that $D\psi(a, b)$ is invertible at the point (a, b), and hence the inverse function theorem is applicable to ψ at the point (a, b). So, there is a open neighborhood U_0 containing (a, b), and an open neighborhood V containing $\psi(a, b) = (a, c)$, such that $\psi : U_0 \to V$ is bijective, and the inverse function $\psi^{-1} : V \to U_0$ is $\mathcal{C}^1$.

Now, V is a open neighborhood of (a, c) in $\mathbb{R}^n \times \mathbb{R}^m$. Decompose V to

$$V = W \times W'$$

where $W \subset \mathbb{R}^n$ is an open neighborhood of a, and $W' \subset \mathbb{R}^m$ is an open neighborhood of c. We can write ψ^{-1} as

$$\psi^{-1} = (I, F)$$

where F is the component function from $W \times W'$ to $\mathbb{R}^m$. Since ψ^{-1} is $\mathcal{C}^1$, it follows that F is also $\mathcal{C}^1$. Finally, we define $f : W \to \mathbb{R}^m$ by

$$f(x) = F(x, c)$$

so that f is also $\mathcal{C}^1$ on W. Also, observe that

$$(x, f(x)) = \psi^{-1}(x, c) \in U_0$$

and clearly

$$\phi(x, f(x)) = c$$

and hence the required f has been found. (Uniqueness of f to be proven.)

9. Problems

Here are some problems I practiced for the mid-sem exam. I will also mention the source of the problems.

The following problems are from the book *Calculus on Manifolds* by M. Spivak.

2-12. Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ is bilinear.

In this problem, $||\cdot||$ will denote the standard euclidean norm.

(a) We will show that

$$\lim_{(h,k)\to 0} \frac{||f(h,k)||}{||(h,k)||} = 0$$

To show this, suppose $\{e_1, ..., e_n\}$ is a basis of $\mathbb{R}^n$, and suppose $\{v_1, ..., v_m\}$ is a basis of $\mathbb{R}^m$. Let

$$M = \max 1 \leq i \leq n, 1 \leq j \leq m ||f(e_i, v_j)||$$

Now, observe that for any $(h,k) \in \mathbb{R}^n \times \mathbb{R}^m$, we have

$$\begin{aligned} ||f(h,k)|| = ||f(a_1e_1 + ... + a_ne_n, b_1v_1 + ...b_mv_m)|| &= \left|\left|\sum_{i,j} a_ib_jf(e_i, v_j)\right|\right| \\ &\leq M\sum_{i,j} |a_ib_j| \\ &= M||h||_1||k||_1 \end{aligned}$$

where $||\cdot||_1$ is the l_1 norm. Since all norms on $\mathbb{R}^m$ are equivalent, there is some $C > 0$ such that

$$\frac{||k||_1}{||k||} \leq C$$

for all $k \in \mathbb{R}^m$. So, we have

$$\frac{||f(h,k)||}{||(h,k))||} \leq \frac{M||h||_1||k||_1}{||(h,k)||} \leq \frac{M||h||_1||k||_1}{||k||} \leq M||h||_1C$$

and hence as $(h,k) \to O$, $h \to O$, which means that the right hand side above goes to 0. This proves the claim.

(b) Next, let $(a,b) \in \mathbb{R}^n \times \mathbb{R}^m$, and let $P_{(a,b)}$ be the linear map from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^p$ given by

$$P_{(a,b)}(x,y) = f(a,y) + f(x,b)$$

We will show that $Df(a,b) = P_{(a,b)}$. Observe that

$$\frac{||f(a+h, b+k) - f(a,b) - P_{(a,b)}(h,k)||}{||(h,k)||} = \frac{||f(h,k)||}{||(h,k)||}$$

because f is bilinear, and hence the claim is true.

2-13. Consider the standard inner product on $\mathbb{R}^n$, and call it IP.

(a) Let $(a,b) \in \mathbb{R}^n \times \mathbb{R}^n$. By the previous problem, we have that

$$D(IP)(a,b)(x,y) = IP(a,y) + IP(x,b)$$

2-14. Let E_i, $i = 1, ..., k$ be Euclidean spaces of various dimensions. Let $f : E_1 \times ... \times E_k$ be a multilinear function.

(a) Suppose $i \neq j$, and let $h = (h_1, ..., h_k)$, where each $h_i \in E_i$. We have

$$\lim_{h\to O} \frac{||f(a_1, ..., h_i, h_j, ..., a_k)||}{||h||} = 0$$

where again $||\cdot||$ is the default Euclidean norm.

To show this, let $g : E_i \times E_j \to \mathbb{R}^p$ defined by

$$g(x,y) = f(a_1, ..., x, ..., y, ..., a_k)$$

then it is easy to see that g is a bilinear function. So, we have that

$$\lim_{(h_i,h_j)\to O} \frac{||g(h_i,h_j)||}{||(h_i,h_j)||} = 0$$

Now, observe that

$$\frac{||f(a_1,...,h_i,...,h_j,...,a_k)||}{||h||} = \frac{||g(h_i,h_j)||}{||h||} \leq \frac{||g(h_i,h_j)||}{||(h_i,h_j)||}$$

and thus the claim has been proven.

2-36. Let $A \subset \mathbb{R}^n$ be an open set and $f : A \to \mathbb{R}^n$ be a $\mathcal{C}^1$ one-one function such that $\det f'(x) \neq 0$ for all x.

We first show that $f(A)$ is open. So, let $y \in f(A)$, and let $x = f^{-1}(y)$. Since $f'(x)$ is invertible by our hypothesis, we apply the inverse function theorem to f at x. So, there is an open set $U \subset A$ containing x, and an open set $W \in\subset \mathbb{R}^n$ containing y, such that $f(U) = W$, and f is obviously one-one on U. Since W is open and $y \in W$, there is an open ball $B(y,r)$ for some $r > 0$ such that $B(y,r) \subset W = f(U) \subset f(A)$, which proves that $f(A)$ is open. It is clear that $f^{-1} : f(A) \to A$ is $\mathcal{C}^1$.

Now, let B be any open set in A. We will show that $f(B)$ is open. If $f^{-1} = \phi$, then observe that

$$\phi^{-1}(B) = f(B)$$

because f is one-one. Since B is open and ϕ is continuous, it follows that $f(B)$ is also open in $\mathbb{R}^n$.

2-37. **(a)** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a $\mathcal{C}^1$ function. We will show that f is *not* one-one.

To prove this, we consider two cases. First, suppose $\dfrac{\partial f}{\partial y}$ is non-zero at all points $(x,y) \in \mathbb{R}^2$. Take $(x_0,y_0) \in \mathbb{R}$, and let $f(x_0,y_0) = c$. Applying the implicit function theorem to f at this point, we see that there is an open set $U \subset \mathbb{R}$ containing x_0 and a unique function $g : U \to \mathbb{R}$ such that

$$f(q, g(q)) = c$$

for all $q \in W$. In that case, f is clearly not one-one.

Now, suppose $\dfrac{\partial f}{\partial y} = 0$ at all points. Then, for any (x_1,y) and $(x_2,y) \in \mathbb{R}^2$ such that $x_1 \neq x_2$, we see that

$$f(x_1,y) = f(x_2,y)$$

by the mean value theorem, and hence f is not one-one again. So, in any case, f is not one-one.

(b) We now generalise the above argument to functions $f : \mathbb{R}^n \to \mathbb{R}^m$. Suppose f is $\mathcal{C}^1$. Consider the first component function f_1 of f. Then, f is also $\mathbb{C}^1$. Then, $f_1 : \mathbb{R}^n \to \mathbb{R}$.

10. Directional Derivatives

Suppose $f : U \to \mathbb{R}^m$ is a map, where U is an open subset of $\mathbb{R}^n$. Let $p \in U$, and let $u \in \mathbb{R}^n$ be a unit vector. The directional derivative $D_u f(p)$ is defined as the vector

$$D_u f(p) = \lim_{h\to 0} \frac{f(p+uh) - f(p)}{h}$$

if the limit exists. The following result shows that the directional derivative exists if the map is differentiable at a point:

Theorem 10.1. Let f be as above, and suppose $Df(p)$ exists. Then, every directional derivative of f at p exists, and for any unit vector u we have

$$D_u f(p) = f'(p)(u)$$

where $f'(p)(u)$ denotes the value of the linear map $f'(p)$ at u.

Proof: We know that

$$\lim_{h \to O} \frac{||f(p+h) - f(p) - f'(p)(h)||}{||h||} = 0$$

So, we have that

$$\lim_{t \to 0} \frac{||f(p+tu) - f(p) - f'(p)(tu)||}{||tu||} = 0$$

which implies that

$$\lim_{t \to 0} \left|\left| \frac{f(p+tu) - f(p)}{t} - f'(p)(u) \right|\right| = 0$$

proving the claim.

As a special case, if $f : U \to \mathbb{R}$, then the directional derivative is given by

$$D_u f(p) = \langle \nabla f(p), u \rangle$$

Here are a few more calculations. In this special case, observe that the last equation gives

$$|D_u f(p)| \leq ||\nabla f(p)||$$

for any unit vector u. Moreover, if $\nabla f(p) \neq O$ and we put $u = \dfrac{\nabla f(p)}{||\nabla f(p)||}$, then we get

$$|D_u f(p)| = ||\nabla f(p)||$$

This is interpreted as follows: if the gradient at a point is non-zero, it points in the direction where the rate of change of f is maximum.

A point about directional derivatives: even if they exist in all directions, the function need not be differentiable, and counterexamples are not hard to find.

11. Mixed Partial Derivatives

In this section, the primary question of interest will be the following: does the order of differentiation matter? In particular, when are equations of the form

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

justified?

We will first consider a special case. Suppose $f : U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^2$. Let us begin with a lemma, which is a 2D version of the mean value theorem. Before that, let us introduce a notation. Let h, k be real numbers such that the product of the open intervals I_1 with endpoints a and $a+h$, and I_2 with endpoints b and $b+k$ is contained inside U, i.e $Q(h,k) = I_1 \times I_2 \subset U$. For these h and k, define

$$\Delta(Q, f) = f(a+h, b+k) - f(a, b+k) - f(a+h, b) + f(a, b)$$

Theorem 11.1. 2D MVT: Suppose f is defined as above, and in addition suppose $D_1 f$ and $D_{21} f$ exist at all points of U (note that existence of D_2 is not needed here). If Q is defined as above, then there is some $(s, t) \in I_1 \times I_2$ such that

$$\Delta(Q, f) = hk D_{21} f(s, t)$$

Proof: We know that

$$\Delta(Q, f) = f(a+h, b+k) - f(a, b+k) - f(a+h, b) + f(a, b)$$

For $x \in I_1$, define

$$u(x) = f(x, b+k) - f(x, b)$$

then we have that

$$\Delta(Q, f) = u(a+h) - u(a)$$

Now, u is clearly differentiable on I_1 and continuous on the closed interval. So by mean value theorem applied to u, there is some $s \in I_1$ such that

$$u(a+h) - u(a) = hu'(s) = h[D_1 f(s, b+k) - D_1 f(s, b)]$$

and again applying mean value theorem to D_1 (keeping the first variable constant), we see that there is some $t \in I_2$ such that

$$h[D_1 f(s, b+k) - D_1 f(s, b)] = hkD_{21} f(s, t)$$

and this proves the claim.

We are now in position to prove the equality of mixed partial derivatives under some conditions:

Theorem 11.2. Suppose $f : U \to \mathbb{R}$ (U is open in $\mathbb{R}$) is a map such that $D_1 f$, $D_2 f$ and $D_{21} f$ all exist on U, and such that $D_{21} f$ is continuous on U. Then, for all points $x \in U$, $D_{12} f(x)$ exists, and

$$D_{12} f(x) = D_{21} f(x)$$

so that both the mixed partials exist and are continuous.

Proof: Let $(x, y) \in U$, and choose (h, k) small enough such that $(x+h, y+k) \in U$, and $I_1 \times I_2 \in U$, where I_1 is the interval with endpoints x and $x+h$, and similarly for I_2. Since $D_{21} f$ is assumed to be continuous, we have that

$$|D_{21}(p, q) - D_{21}(x, y)| < \epsilon$$

given any $\epsilon > 0$, where (p, q) is close enough to (x, y). So, if h, k are small enough, we see that

$$|D_{21}(s, t) - D_{21}(x, y)| = \left| \frac{\Delta(Q, f)}{hk} - D_{21}(x, y) \right| < \epsilon$$

which means that

$$\left| \frac{f(x+h, y+k) - f(x, y+k) - f(x+h, y) + f(x, y)}{hk} - D_{21} f(x, y) \right| < \epsilon$$

Letting $k \to 0$ we see that

$$\left| \frac{D_2 f(x+h, y) - D_2 f(x, y)}{h} - D_{21} f(x, y) \right| < \epsilon$$

and the last inequality shows that

$$D_{12} f(x, y) = D_{21} f(x, y)$$

As a corollary, if $f : U \to \mathbb{R}$ is $\mathcal{C}^2$, then the mixed partials are equal at all points of consideration. We can extend this corollary to higher dimensions as follows:

Theorem 11.3. Suppose $f : U \to \mathbb{R}^m$ is a $\mathcal{C}^k$ function, where U is an open subset of $\mathbb{R}^n$. Let $(i_1, ..., i_k) \in \{1, ..., n\}^k$. Then,

$$D_{i_1 i_2 \ldots i_k} f(p) = D_{j_1 j_2 \ldots j_k} f(p)$$

where $(j_1, j_2, ..., j_k)$ is any permutation of the tuple $(i_1, ..., i_k)$ and $p \in U$.

Proof: The idea is to consider only two variables at a time, and use the fact that any permutation is a product of transpositions.

12. A Projection Theorem

Here is a little useful theorem that I is used as an intuitive idea behind why the implicit function theorem is true. You can find an application of this in one of my Analysis-2 homework assignments.

Here is the theorem:

Theorem 12.1. Consider the space K^n (where $K \in \{\mathbb{R}, \mathbb{C}\}$) with the usual inner product (so that the inner product is positive definite). Identify $K^n = K^d \times K^m$, where $d + m = n$. Let $\pi : K^n \to K^d$ and $\pi' : K^n \to K^m$ be the usual projections. Let H be a subspace of K^n of dimension d. Then, $\pi(H) = K^d$ if and only if $\pi'(H^\perp) = K^m$. (Draw a picture in $\mathbb{R}^2$ to understand this theorem).

Proof: Proving one direction is enough because we can symmetrically apply the same argument to the other direction.

So, suppose $\pi(H) = K^d$. Observe that it is enough to show that the kernel of $\pi' : H^\perp \to K^m$ is zero (because $H^\perp$ has dimension m). So, suppose $v \in H^\perp$ such that $\pi'(v) = 0$. So, $v = (x, 0)$, for some $x \in K^d$ (here $0 \in K^m$). Since $\pi : H \to K^d$ is surjective, there is some $h \in H$ such that $\pi(h) = v$. So, let $h = (x, w)$, where $w \in K^m$. Also, observe that

$$(x, 0) \cdot (x, w) = 0$$

and this implies that $x \cdot x = 0$, showing that $x = 0$, and hence $v = 0$. This completes the proof.

13. Quadratic Forms and Quadric Hypersurfaces

Classically, a *quadratic form* (over $\mathbb{R}$, but the field can be anything) is just a two-degree homogeneous polynomial. For example, in three dimensions it looks like

$$Q(x_1, x_2, x_3) = Ax^2 + By^2 + Cz^2 + 2Dxy + 2Eyz + 2Fzx$$

In modern language, quadratic forms are determined by symmetric matrices, i.e a quadratic form is a function of the form

$$Q(x) = x^t Ax$$

where A is a symmetric matrix.

A *quadric hypersurface* in $\mathbb{R}^n$ is the locus of points satisfying the equation

$$\Phi(x_1, ..., x_n) = 0$$

where $\Phi \in \mathbb{R}[x_1, ..., x_n]$ is a degree two polynomial. For instance, ellipses and hyperbolas are quadric hypersurfaces in 2 dimensions.

Changing a quadric equation to "diagonal form": Since matrices of quadratic forms are symmetric, they are diagonalisable by an orthogonal transformations. We will use this fact to convert a quadric equation to diagonal form.

Now, suppose $\Phi : \mathbb{R}^n \to \mathbb{R}$ is any degree two polynomial. So we can write

$$\Phi(x) = Q(x) + P(x) + r$$

where Q is a quadratic form, P is a linear homogeneous polynomial, and r is a constant. Let Q be the matrix of the quadratic form, and hence Q is diagonalisable by an orthogonal matrix. So, suppose

$$\Gamma^t Q \Gamma = D$$

where D is a diagonal matrix. Let $x^* = \Gamma^{-1}x = \Gamma^t x$ where we view Γ as a change of coordinates matrix. So, we have

$$\Phi(x) = (x^*)^t D(x^*) + P(\Gamma x^*) + r$$

Now, if we set $P^*(x^*) = P(\Gamma x^*)$, then P^* will also be linear homogeneous.

So, we have

$$\Phi(x) = \sum_{i=1}^{n} \lambda_i (x_i')^2 + \sum_{i=1}^{n} \mu_i x_i' + r$$

where λ_i are the diagonal entries of D, and μ_i are constants and $x^* = (x_1', ..., x_n')$. So, we have written the polynomial Φ in "diagonal" form. So if we have a quadric hypersurface of the form $\Phi = c$, this corresponds to the equation

$$\sum_{i=1}^{n} \lambda_i (x_i')^2 + \sum_{i=1}^{n} \mu_i x_i' + r = c$$

and by completing squares and shifting the origin, this equation may be written

$$\sum_{i=1}^{n} \nu_i y_i^2 = \rho$$

where ν_i and ρ are some constants. Some examples of this are given in one of my homeworks.

The advantage of writing surfaces in this form is that it is easier to work with simpler equations. Also, we can determine what type of surface an equation represents if we know its standard form.

14. Taylor's Theorem in Several Variables

In this section, we will try to derive a form of Taylor's theorem for general differentiable functions.

Suppose $f : U \to \mathbb{R}$ is $\mathcal{C}^m$ on U, and let $a \in U$. Just like in one variable, we would like to approximate $f(a+h)$ (for small h) by a multivariate polynomial in h. Let's see how we can do this.

Since U is open, there is a closed ball $\overline{B}$ around the origin such that $a + \overline{B}$ is contained in U. Let $h \in \overline{B}$ and fix this h, and consider the point $a+h$ and a. The idea is to define a function on the line segment between a and $a+h$, which is a differentiable function in one variable, and apply Taylor's theorem to that function.

So, consider the set $a + th$ for $t \in [0, 1]$, which is obviously contained in $a + \overline{B}$ (because balls are convex). Consider the function $g : [0, 1] \to \mathbb{R}$ given by

$$g(t) = f(a + th)$$

Now, $g \in \mathcal{C}^m$ on $[0, 1]$. Also, convince yourself that for $k \in \{1, ..., m\}$, the k^{th} derivative of g at some point $t \in [0, 1]$ is given by

$$g^{(k)}(t) = \sum_{i_1, i_2, ..., i_k} D_{i_1,...,i_k} f(a + th) h_{i_1} ... h_{i_k}$$

where the sum ranges over all subsets of $\{1, ..., n\}^k$.

Now, observe that if $(j_1, ..., j_n)$ is any permutation of $(i_1, ..., i_n)$, then the mixed partials $D_{j_1...j_n}f$ and $D_{i_1...i_n}f$ are equal, since $f \in \mathcal{C}^k$ for $k \in \{1, ..., m\}$. Now, in a subset of $\{1, ..., n\}^k$, suppose 1 occurs s_1 times, 2 occurs s_2 times, and s_i occurs s_i times, so that

$$s_1 + ... + s_n = k$$

and the number of permutation of this tuple is

$$\frac{k!}{s_1!....s_n!}$$

and hence we have

$$g^{(k)}(t) = \sum_{s_1+...+s_n=k} \frac{k!}{s_1!...s_n!} D_1^{s_1}...D_n^{s_n} f(a+th) h_1^{s_1}...h_n^{s_n}$$

where the notation $D_1^{s_1}....D_n^{s_n} f(a+th)$ means the mixed partial $D_{111...1222...2...nnn...n}f(a+th)$, where i occurs s_i times in the subscript.

Now, we have that $g(0) = f(a)$ and $g(1) = f(a+h)$. Applying Taylor's theorem to g with the two points being 1 and 0, we see that there is some $\theta \in (0, 1)$ such that

$$g(1) = \sum_{k=0}^{m-1} \frac{g^{(k)}(0)}{k!} + \frac{g^{(m)}(\theta)}{m!}$$

and hence we have

$$\begin{aligned} f(a+h) &= \sum_{k=0}^{m-1} \sum_{s_1+...+s_n=k} \frac{D_1^{s_1}...D_n^{s_n} f(a) h_1^{s_1}...h_n^{s_n}}{s_1!...s_n!} + \sum_{s_1+...+s_n=m} \frac{D_1^{s_1}...D_n^{s_n} f(a+\theta h) h_1^{s_1}...h_n^{s_n}}{s_1!...s_n!} \\ &= \sum_{k=0}^{m-1} \sum_{s_1+...+s_n=k} \frac{D_1^{s_1}...D_n^{s_n} f(a) h_1^{s_1}...h_n^{s_n}}{s_1!...s_n!} + r(h) \end{aligned}$$

and this is Taylor's theorem for f.

Now, since f is $\mathcal{C}^m$, it follows that $D_1^{s_1}...D_n^{s_n} f$ is continuous on U, in particular continuous on $a + \overline{B}$. So, it must be bounded on $a + \overline{B}$ (because it is compact). So, there is some constant M such that

$$\frac{1}{s_1!...s_n!} D_1^{s_1}...D_n^{s_n} f(a+\theta h) \leq M$$

for every possible $s_1, ..., s_n$. Hence, we see that

$$||r(h)|| \leq KM||h||^m$$

where K is the total number of solutions to $s_1 + ... + s_n = m$. The last inequality also proves that $\dfrac{||r(h)||}{||h||^{m-1}} \to 0$ as $h \to O$. So, we have proven the following theorem:

Theorem 14.1. Suppose $f : U \to \mathbb{R}$ is a $\mathcal{C}^m$ function, where U is an open subset of $\mathbb{R}^n$, and let $a \in U$. Then, there is an open ball B around the origin such that $a + \overline{B} \subset U$, and for any $h \in \overline{B}$, we have that

$$f(a+h) = \sum_{k=0}^{m-1} \sum_{s_1+...+s_n=k} \frac{D_1^{s_1}...D_n^{s_n} f(a) h_1^{s_1}...h_n^{s_n}}{s_1!...s_n!} + r(h)$$

where $r : \overline{B} \to \mathbb{R}$ is given by

$$\sum_{s_1+...+s_n=m} \frac{D_1^{s_1}...D_n^{s_n} f(a+\theta h) h_1^{s_1}...h_n^{s_n}}{s_1!...s_n!}$$

for some $\theta \in (0,1)$. Moreover, $||r(h)||/||h||^{m-1} \to 0$ as $h \to O$.

The Hessian: Let $f : U \to \mathbb{R}$ be any function, where U is an open subset of $\mathbb{R}^n$. Let $a \in U$ such that both f' and f'' exist at a. The *Hessian matrix* of f at a is the matrix whose ij^{th} entry is $D_{ij}f(a)$, i.e it is the matrix of mixed partial derivatives. If f is $\mathcal{C}^2$, then observe that Hf is a symmetric matrix.

We now apply Taylor's theorem to a function which will relate to its Hessian matrix:

Theorem 14.2. Suppose $f : U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^n$, and suppose f is $\mathcal{C}^3$ on U. Let $a \in U$. So, there is an open ball B centered at the origin such that $a + \overline{B}$ is contained in U, and for any $h \in \overline{B}$, we have

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} h^t Hf(a) h + r(h)$$

where $||r(h)||/||h||^2 \to 0$ as $h \to O$.

Proof: We just apply Taylor's theorem here. That gives us the proof.

15. Local Maxima and Minima

Let $f : U \to \mathbb{R}$ be a $\mathcal{C}^1$ map, where U is an open subset of $\mathbb{R}^n$. Let $a \in U$ be a point such that $f'(a) = 0$. Then, a is said to be a *critical* point of f.

Now, suppose $a = (a_1, ..., a_n)$ is a point of local extremum for f. It is then clear that every directional derivative $D_u f(a) = 0$ (from one variable calculus). In particular, we have that $D_i f(a) = 0$ for every i, and hence $Df(a) = 0$. So, points of local extrema are critical points of f. The converse is not even true in one dimension.

Now, suppose the function f above is $\mathcal{C}^2$. A point $a \in U$ is said to be a *non-degenerate critical point* if a is a critical point of f, and the Hessian $Hf(a)$ is invertible. The index $\sigma(a)$ of a non-degenerate critical point is the number of negative eigenvalues of $Hf(a)$.

Clearly, if f is $\mathcal{C}^2$, then Hf is symmetric at all points of U. So, $Hf(a)$ will be positive definite if and only if its index is 0, and negative definite if and only if its index is n. A non-degenerate critical point is said to be a *saddle point* if $0 < \sigma(a) < n$, i.e the Hessian $Hf(a)$ is indefinite.

We now prove a test to check extremums:

Theorem 15.1. Suppose $f : U \to \mathbb{R}$ is a $\mathcal{C}^3$ function and let a be a non-degenerate critical point of f. Let H be the Hessian $Hf(a)$. Then:

(1) a is a point of local minimum if and only if H is positive definite.
(2) a is a point of local maximum if and only if H is negative definite.

Before proving this, observe that a will be a sadle point if and only if H is indefinite.
Proof: Since f is $\mathcal{C}^3$, we know that there is some closed ball $\overline{B}$ such that $a + \overline{B} \subset U$, and also for $h \in \overline{B}$ we have

$$\begin{aligned} f(a+h) &= f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} h^t H h + r(h) \\ &= f(a) + \frac{1}{2} h^t H h + r(h) \end{aligned}$$

and also $||r(h)||/||h||^2 \to 0$ as $h \to O$. Observe that H is a symmetric matrix, and hence it is positive definite (or negative definite) if and only if all eigenvalues are positive (or negative). We only prove part (1), the proof of part (2) being similar.

First, suppose H is positive definite. Observe that we have

$$f(a+h) - f(a) = \frac{1}{2}h^t Hh + r(h)$$

where h is small enough.

First, observe that $h^t Hh$ is a $\mathcal{C}^\infty$ map from $\mathbb{R}^n$ (a quadratic form), in particular it is continuous, and hence attains a minimum on the unit sphere S in $\mathbb{R}^n$. Let

$$\mu = \min_{x \in S} x^t Hx$$

and since H is positive definite, it follows that $\mu > 0$. Also, for any non-zero vector $h \in \mathbb{R}^n$, we know that $h/||h|| \in S$, and hence this means that

$$h^t Hh/||h||^2 \geq \mu$$

for any non-zero h. Now, let h be so small such that

$$\frac{|r(h)|}{||h||^2} < \frac{\mu}{2}$$

and hence we have that

$$\frac{h^t Hh + r(h)}{||h||^2} > 0$$

which means that $f(a+h) - f(a) > 0$ for small enough h, and hence a is a point of local minimum.

Conversely, suppose a is a point of local minimum. Let λ be an eigenvalue of H, and let u be a unit eigenvector. By one variable calculus, we see that $D_u^2 f(a) > 0$. Also, convince yourself that

$$D_u^2 f(a) = u^t Hu$$

but we have that

$$u^t Hu = \lambda u^t u = \lambda$$

and hence $\lambda > 0$. So, all eigenvalues of H are positive, and hence H is positive definite.

Lagrange Multipliers: Here, we will look at the Lagrange Multiplier method of optimization. Suppose U is an open subset of $\mathbb{R}^n$, and let $h, f : U \to \mathbb{R}$ be two $\mathcal{C}^1$ functions. Let $c \in h(U)$, and consider the level set $M = \{x \in U : h(x) = c\}$. We wish to maximise f subject to the constraint $h = c$, i.e we wish to maximise f on the level set M.

In all of this section, we assume that h' does not vanish at any point of U, and hence h has a tangent space at every point of M (this is easy to see, because the derivative will be a non-zero row, so h' will have full rank). Moreover, from problems **6.** and **7.** in HW-6, we know that if $f|_M$ has a local extremum at some $x \in M$, then $\nabla f(x)$ lies in the orthogonal space of the tangent space T_x. Also, there is a unique scalar $\lambda \in \mathbb{R}$ such that $\nabla f(x) = \lambda \nabla h(x)$. This scalar is called a *Lagrange Multiplier*. We'll have a look at two examples to see how to use these:

Example 1:

Example 2: Now, we will look at the case when we have multiple constraints. In the above discussion we had only one h, now we will see how to use Lagrange multipliers if we have multiple such hs. This is given by the following theorems:

Theorem 15.2. Let $d + m = n$, where d and m are non-negative integers. Suppose $h : U \to \mathbb{R}^m$ and $f : U \to \mathbb{R}$ are $\mathcal{C}^1$ functions (here, the function h will represent m constraints). Let $c \in h(U)$, and let $M = h^{-1}(c)$ be a level set. Suppose the rank of $h'(x)$ at every $x \in M$ is m (i.e, $h'(x)$ has full rank, and hence the tangent space at each point exists). Suppose f has a point of extremum on M. Then, $\nabla f(x)$ lies is the orthogonal space of T_x, the tangent space at x. In particular, there are scalars $\lambda_1, ..., \lambda_m$ such that

$$\nabla f(x) = \sum_{i=1}^{m} \lambda_i \nabla h_i(x)$$

Proof: Suppose f has a point of extremum on M, and let that point be a. Let the tangent space at the point a be denoted by T_a. From a previous homework, we know that T_a is spanned by velocity vectors of $\mathcal{C}^1$ paths taking values in M and passing through a (as an exercise, try to prove this. But you can always look at the proof in one of my homeworks).

Now, let $\gamma : (a_1, b_1) \to M$ be such a $\mathcal{C}^1$ path, and let $g(\theta) = a$ where $\theta \in (a_1, b_1)$ (so that the path passes through a). If we consider the function $g : (a, b) \to \mathbb{R}$ given by $g(t) = f(\gamma(t))$, then observe that g has a point of extremum at θ, and by one variable calculus we see that $g'(a) = 0$ (g is differentiable because it a a composition of two differentiable functions). Also, it is easy to see that

$$g'(t) = \langle \nabla f(a), \gamma'(\theta) \rangle$$

and this implies that $\nabla f(a)$ is orthogonal to $\gamma'(\theta)$. So, this implies that $\nabla f(a) \in N_a$, where N_a is the orthogonal complement of T_a.

Finally, it is not hard to see that N_a is spanned by the vectors $\nabla h_i(a)$, for $1 \leq i \leq m$. So, it follows that

$$\nabla f(a) = \sum_{i=1}^{m} \lambda_i \nabla h_i(a)$$

for some scalars $\lambda_i \in \mathbb{R}$. This completes the proof.

Note: In the above proof, I have used some facts blatantly, but all these facts are proven in my homeworks. Check them out if you want.

Remark: Note that in the above proof, $m \leq n$, and hence h represents at most n constraints. This is reasonable, because more than n equations in n variables may not even have solutions (like in the case of linear equations).