Math E-23a - Linear Algebra and Real Analysis I

Proofs

Jimmy Tran[1]

Fall 2023

[1]Originally scribed by Chris Watson. Adapted with permission.

Preface

This is the first half of an integrated treatment of linear algebra, real analysis, and multivariable calculus. By combining these disciplines into one course, we show important relations between each, which allows us to use results from one topic to gain deeper understanding of other topics. We cover matrices, eigenvectors, dot and cross products, limits, continuity, and differentiability, all in multiple dimensions, with an introduction to manifolds. This course covers both mathematical proofs as well as applications.
These notes serve as a reference for proofs that students in the course are expected to know how to recreate and teach to others.

Historical Note. *Math E–23a* began as a direct cross-listing of Harvard College's accelerated *Math 23a*. For many years Extension- and College-registered students sat in the same lecture hall, turned in the same problem sets, and took the same exams—the only difference was the catalog number.

Contents

Proof 0.1

Suppose that a and b are two elements of a field F. Prove the following:

(a) $\forall a \in F, 0a = 0$

(b) $\forall a, b \in F$, if $ab = 0$, then either $a = 0$ or $b = 0$

(c) $\forall a \in F$, the additive inverse of a is unique

Proof

Part (a)

$$
\begin{aligned}
0 + 0 &= 0 && \text{(Additive identity)} \\
(0 + 0)a &= 0a && \\
0a + 0a &= 0a && \text{(Distributivity)} \\
(0a + 0a) + (-0a) &= 0a + (-0a) && \text{(Existence of additive inverse)} \\
0a + (0a + (-0a)) &= 0a + (-0a) && \text{(Associativity)} \\
0a + 0 &= 0 && \text{(Additive inverse of } 0a\text{)} \\
0a &= 0 && \text{(Additive identity)}
\end{aligned}
$$

■

Part (b)

WLOG, assume $a \neq 0$. Then $\exists a^{-1}$ such that $a^{-1}a = 1$. So,

$$
\begin{aligned}
ab &= 0 && \\
\Rightarrow a^{-1}(ab) &= a^{-1}0 && \\
\Rightarrow (a^{-1}a)b &= 0 && \text{(Associativity (LHS) and Proof 0.1a (RHS))} \\
\Rightarrow 1b &= 0 && \text{(Multiplicative inverse)} \\
\Rightarrow b &= 0 && \text{(Multiplicative identity)}
\end{aligned}
$$

■

Part (c)

Assume x_1 and x_2 are both additive inverses of some $a \in F$. Then $a + x_1 = 0$ and $a + x_2 = 0$, and they equal each other.

$$
\begin{aligned}
a + x_1 &= a + x_2 && \\
\Rightarrow (x_1 + a) + x_1 &= (x_1 + a) + x_2 && \\
\Rightarrow 0 + x_1 &= 0 + x_2 && \text{(Definition of additive inverse)} \\
\Rightarrow x_1 &= x_2 && \text{(Additive identity)}
\end{aligned}
$$

■

Proof 1.1

A is a $n \times m$ matrix, B is $m \times p$, and C is $p \times r$. Using summation notation, prove that matrix multiplication is associative: that $(AB)C = A(BC)$.

Proof

By definition, $(AB)_{ik} = \sum_{j=1}^{m} a_{ij}b_{jk}$. Also, $(BC)_{jq} = \sum_{k=1}^{p} b_{jk}c_{kq}$.
We calculate both products:

$$((AB)C)_{iq} = \sum_{k=1}^{p} (AB)_{ik}c_{kq} = \sum_{k=1}^{p} \left(\sum_{j=1}^{m} a_{ij}b_{jk} \right) c_{kq} = \sum_{k=1}^{p} \sum_{j=1}^{m} a_{ij}b_{jk}c_{kq}$$

$$(A(BC))_{iq} = \sum_{j=1}^{m} a_{ij}(BC)_{jq} = \sum_{j=1}^{m} a_{ij} \left(\sum_{k=1}^{p} b_{jk}c_{kq} \right) = \sum_{j=1}^{m} \sum_{k=1}^{p} a_{ij}b_{jk}c_{kq}$$

Because multiplication and addition (of the matrix elements) are commutative and satisfy distributivity, we can interchange the order of summation. Therefore, the products are equal.
■

Proof 1.2

Suppose that linear transformation $T : F^n \to F^m$ is represented by the $m \times n$ matrix $\left[T\right]$.

(i) Suppose the matrix $\left[T\right]$ is invertible. Prove that the linear transformation T is one-to-one and onto, hence invertible.

(ii) Suppose that the linear transformation T is invertible. Prove that its inverse S is linear.

(iii) Given that S is the inverse of T, and that both are linear, prove that the matrix of S is the inverse of the matrix of T; that is, prove that $\left[S\right] = \left[T\right]^{-1}$.

Proof

Part (a)

Let $y \in F^m$ arbitrary. Since $\left[T\right]$ is invertible, $\exists x \in F^n$ such that $\left[T\right]^{-1} y = x$. In particular, we can left-multiply by $\left[T\right]$ so that $\left[T\right] \left[T\right]^{-1} y = \left[T\right] x$, which simplifies to $y = T(x)$. Since y was arbitrary in the codomain, this shows that T is surjective.

Assume $x_1, x_2 \in F^n$, and $T(x_1) = T(x_2)$. But $T(x_1) = \left[T\right] x_1 = \left[T\right] x_2$. We can left-multiply by its inverse, so that $\left[T\right]^{-1} \left[T\right] x_1 = \left[T\right]^{-1} \left[T\right] x_2$. This can be simplified as $x_1 = x_2$, so T is injective.

Both results together show that T is invertible. ■

Part (b)

We want to show that $S(ay_1 + by_2) = aS(y_1) + bS(y_2)$.

Since T is surjective, we can let $y_1 = T(x_1)$ and $y_2 = T(x_2)$ for some $y_1, y_2 \in F^m$, and let $y = ay_1 + by_2$ for $a, b \in F$.

Then, since T is linear, $S(ay_1 + by_2) = S(aT(x_1) + bT(x_2)) = S(T(ax_1 + bx_2))$.

Since S is the inverse of T, $(S \circ T)(x) = x$ for any $x \in F^n$, so $S(T(ax_1 + bx_2)) = ax_1 + bx_2$. Because S is an inverse, we know that $S(y_1) = S(T(x_1)) = x_1$ and similarly $S(y_2) = x_2$, so $ax_1 + bx_2 = aS(y_1) + bS(y_2)$. Therefore, S is linear. ■

Part (c)

Since S is the inverse of T, that means $(S \circ T)(x) = S(T(x)) = x$ for all $x \in F^n$. $S(T(x)) = S(\left[T\right] x) = \left[S\right] \left[T\right] x = x$. By definition, then, $\left[S\right] \left[T\right] = I_n$.

Also, $(T \circ S)(y) = T(S(y)) = y$ for all $y \in F^m$. So $T(S(y)) = T(\left[S\right] y) = \left[T\right] \left[S\right] y = y$. Therefore, $\left[T\right] \left[S\right] = I_m$.

Together these show that $\left[S\right]$ is both a left- and right-inverse of $\left[T\right]$, which means $\left[S\right] = \left[T\right]^{-1}$.
■

Proof 2.1

(i) Show that for any vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$, $|\vec{v} \cdot \vec{w}| \leq |\vec{v}|\,|\vec{w}|$.

(ii) Prove that $|\vec{x} + \vec{y}| \leq |\vec{x}| + |\vec{y}|$

Proof

Part (a)

Define a quadratic function of the variable $t \in \mathbb{R}$ by

$$f(t) = |t\vec{v} - \vec{w}|^2 = (t\vec{v} - \vec{w}) \cdot (t\vec{v} - \vec{w})$$

Since $f(t)$ is the square of a length of a vector, it cannot be negative, so the quadratic equation $f(t) = 0$ does not have two real roots. So its discriminant must satisfy $b^2 - 4ac \leq 0$.

We can use the distributivity of the dot product to get

$$\begin{aligned} f(t) &= (t\vec{v} - \vec{w}) \cdot (t\vec{v} - \vec{w}) \\ &= t^2\vec{v} \cdot \vec{v} - 2t\vec{v} \cdot \vec{w} + \vec{w} \cdot \vec{w} \\ &= |\vec{v}|^2\, t^2 - 2\vec{v} \cdot \vec{w}t + |\vec{w}|^2 \end{aligned}$$

So $a = |\vec{v}|^2$, $b = -2\vec{v} \cdot \vec{w}$, and $c = |\vec{w}|^2$. Then

$$\begin{gathered} b^2 - 4ac \leq 0 \\ \Rightarrow 4(\vec{v} \cdot \vec{w})^2 - 4\,|\vec{v}|^2\,|\vec{w}|^2 \leq 0 \\ \Rightarrow 4(\vec{v} \cdot \vec{w})^2 \leq 4\,|\vec{v}|^2\,|\vec{w}|^2 \end{gathered}$$

If $\vec{v} \cdot \vec{w} = 0$, the result is trivial. If $\vec{v} \cdot \vec{w} \neq 0$, we divide by 4 and take the square root of both sides and get $|\vec{v} \cdot \vec{w}| \leq |\vec{v}|\,|\vec{w}|$. ■

Part (b)

$$\begin{aligned} |\vec{x} + \vec{y}|^2 &= (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y}) \\ &= (\vec{x} + \vec{y}) \cdot \vec{x} + (\vec{x} + \vec{y}) \cdot \vec{y} && \text{(Distributivity of the dot product)} \\ &\leq |\vec{x} + \vec{y}|\,|\vec{x}| + |\vec{x} + \vec{y}|\,|\vec{y}| && \text{(Cauchy-Schwarz)} \end{aligned}$$

If $|\vec{x} + \vec{y}| = 0$, the inequality is trivially true. Otherwise, we can divide by the common factor $|\vec{x} + \vec{y}|$ and get

$$|\vec{x} + \vec{y}| \leq |\vec{x}| + |\vec{y}|$$

■

Proof 3.1

Prove that a maximal set of linearly independent vectors for a subspace of $\mathbb{R}^n$ is also a minimal spanning set for that subspace.

Proof

We need to show that the set of vectors is a spanning set and that it is minimal.

Let our subspace be $V \subseteq \mathbb{R}^n$, and the basis vectors $\vec{v}_1, \ldots, \vec{v}_k$. Since our set is *maximal*, if we add any other vector $\vec{w} \in V$ to the basis set, the resulting set is linearly dependent:

$$a_1\vec{v}_1 + \cdots + a_k\vec{v}_k + b\vec{w} = \vec{0}$$

with not all of the a_i's and b equal to 0 (i.e., at least one must be nonzero).

(i) If $b = 0$, then since $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is a linearly independent set, we must have $a_i = 0$ for all i, which is a contradiction.

(ii) If $b \neq 0$, then we can rewrite the above equation as $\vec{w} = -\frac{1}{b}\left(a_1\vec{v}_1 + \cdots + a_k\vec{v}_k\right)$. Since $\vec{w} \in V$ was arbitrary, the $\vec{v}_i$'s span V.

To show that the vectors are *minimal*, WLOG remove $\vec{v}_k$, and assume that $\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$ still spans V. Then, since $\vec{v}_k \in V$, we can write it as

$$\begin{aligned} \vec{v}_k &= a_1\vec{v}_1 + \cdots + a_{k-1}\vec{v}_{k-1} \\ \Rightarrow \vec{0} &= a_1\vec{v}_1 + \cdots + a_{k-1}\vec{v}_{k-1} - 1 \cdot \vec{v}_k \end{aligned}$$

However, since clearly not all the coefficients are 0, this implies that $\{\vec{v}_1, \ldots, \vec{v}_k\}$ are linearly *dependent*, which is a contradiction. Therefore, the set of vectors must be minimal. ■

Proof 3.2

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation.

(i) Define kernel and image.

(ii) Prove that $\ker T$ and $\operatorname{Img} T$ are subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively.

(iii) Briefly explain (without proof) how you can construct a basis for $\ker T$ and for $\operatorname{Img} T$ by row reducing the matrix $[T]$.

(iv) Prove that $\dim(\ker T) + \dim(\operatorname{Img} T) = n$.

Proof

Part (a)

$$\ker T = \left\{\vec{x} \in \mathbb{R}^n : T(\vec{x}) = \vec{0}\right\}$$

$$\operatorname{Img} T = \{\vec{y} \in \mathbb{R}^m : \exists \vec{x} \in \mathbb{R}^n \text{ such that } \vec{y} = T(\vec{x})\}$$

Part (b)

For the image, let $\vec{y}_1, \vec{y}_2 \in \operatorname{Img} T$, and $\vec{y} = a\vec{y}_1 + b\vec{y}_2$. We can write $\vec{y}_1 = T(\vec{x}_1)$ and $\vec{y}_2 = T(\vec{x}_2)$ for some $\vec{x}_1, \vec{x}_2 \in \mathbb{R}^n$. So $\vec{y} = aT(\vec{x}_1) + bT(\vec{x}_2)$. By linearity, this means that $\vec{y} = T(a\vec{x}_1 + b\vec{x}_2)$, which shows that a linear combination of elements of the image is also in the image. In other words, that $\operatorname{Img} T \subseteq \mathbb{R}^m$.

For the kernel, let $\vec{x}_1, \vec{x}_2 \in \ker T$, and $\vec{x} = a\vec{x}_1 + b\vec{x}_2$. Then $T(\vec{x}) = T(a\vec{x}_1 + b\vec{x}_2)$. By linearity, then, $T(\vec{x}) = aT(\vec{x}_1) + bT(\vec{x}_2) = \vec{0} + \vec{0} = \vec{0}$. So $\vec{x} \in \ker T$ and $\ker T \subseteq \mathbb{R}^n$. ■

Part (c)

The image of T is the set of all linear combinations of the columns of $\left[T\right]$. Since some columns may be redundant, we row reduce $\left[T\right]$ and "retain" only the pivotal columns. A basis for the image will then be the pivotal columns of the *original* matrix $\left[T\right]$ (not the row-reduced matrix).

For the kernel, we create a basis vector for each of the *non-pivotal* columns of the row-reduced matrix. For each vector/non-pivotal column, insert a 1 into the entry for that non-pivotal column, 0s for all other non-pivotal columns, and solve for the remaining entries. These vectors will all be linearly independent, because the entry for each non-pivotal column will have a 1 in only one basis vector, and 0 in the remaining basis vectors.

Part (d)

Since a basis for the image consists of all pivotal columns, say it is r, then $\dim \operatorname{Img} T = r$. A basis for the kernel will consist of all non-pivotal columns, or $\dim \ker T = n - r$. Thus $\dim \ker T + \dim \operatorname{Img} T = n - r + r = n$. ■

Proof 4.1

If $\vec{v}_1, \dots, \vec{v}_k$ are eigenvectors of $A : \mathbb{R}^n \to \mathbb{R}^n$ with distinct eigenvalues $\lambda_1, \dots, \lambda_k$, then they are linearly independent.

Proof

Assume for contradiction that $\vec{v}_1, \dots, \vec{v}_k$ are linearly *dependent*. Then there is a first $\vec{v}_j$ that is a linear combination of its predecessors:

$$\vec{v}_j = a_1\vec{v}_1 + \cdots + a_{j-1}\vec{v}_{j-1} \tag{1}$$

where not all of the $a_1, \dots, a_{j-1} \in \mathbb{R}$ are equal to 0.

First, we will multiply both sides of Equation (1) by A, and substitute since they are all eigenvectors:

$$A\vec{v}_j = a_1 A\vec{v}_1 + \cdots + a_{j-1}A\vec{v}_{j-1} \tag{2}$$

$$\Rightarrow \lambda_j\vec{v}_j = a_1\lambda_1\vec{v}_1 + \cdots + a_{j-1}\lambda_{j-1}\vec{v}_{j-1} \tag{3}$$

We can also multiply both sides of Equation (1) by λ_j:

$$\lambda_j\vec{v}_j = a_1\lambda_j\vec{v}_1 + \cdots + a_{j-1}\lambda_j\vec{v}_{j-1} \tag{4}$$

Now, we subtract Equation (3) from Equation (4). Since we are given that all the eigenvectors are distinct, and not all of the a_i's are 0, then the following shows that $\vec{v}_1, \dots, \vec{v}_{j-1}$ are linearly *dependent*:

$$\vec{0} = a_1(\underbrace{\lambda_j - \lambda_1}_{\neq 0})\vec{v}_1 + \cdots + a_{j-1}(\underbrace{\lambda_j - \lambda_{j-1}}_{\neq 0})\vec{v}_{j-1}$$

However, we assumed that $\vec{v}_j$ was the first linearly dependent vector, a contradiction. So the set of eigenvectors $\{\vec{v}_1, \dots, \vec{v}_k\}$ must be linearly independent. ■

Proof 4.2

For a real $n \times n$ matrix A, each $p_i(t)$ is simple and has real roots if and only if there is a basis of $\mathbb{R}^n$ consisting of the eigenvectors of A (an *eigenbasis*).

Proof

(⇒)

First calculate $\vec{e}_1, A\vec{e}_1, A^2\vec{e}_1, \ldots, A^m\vec{e}_1$, where $A^m\vec{e}_1$ is the first linearly dependent vector. Now consider the subspace $E_1 = \operatorname{span}\left\{\vec{e}_1, A\vec{e}_1, \ldots, A^{m-1}\vec{e}_1\right\}$. We know that $\dim E_1 = m$; $p(A)\vec{e}_1 = \vec{0}$ where, by assumption, $p(t)$ has simple, real roots; $p(t)$ is of degree m; and therefore $p(t)$ will have m distinct roots.

So we will find m eigenvectors $\vec{v}_1, \ldots, \vec{v}_m$ with distinct eigenvalues $\lambda_1, \ldots, \lambda_m$; these eigenvectors will be linearly independent (*Proof 4.1*). By our algorithm, each $\vec{v}_i$ will be calculated as $\vec{v}_i = q(A)\vec{e}_1$; that is, each $\vec{v}_i$ is a linear combination of $\vec{e}_1, A\vec{e}_1, \ldots, A^{m-1}\vec{e}_1$. This means that all the $\vec{v}_1, \ldots, \vec{v}_m \in E_1$. Since there are m linearly independent eigenvectors, and $\dim E_1 = m$, the $\vec{v}_1, \ldots, \vec{v}_m$ form a basis for E_1. Furthermore, since $\vec{e}_1 \in E_1$, then $\vec{e}_1 \in \operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_m\}$.

We repeat this process for starting vectors $\vec{e}_2, \ldots, \vec{e}_n$; in each step, we will find some eigenvectors (possibly some repeated). We then take the *union* of all eigenvectors. In addition, each $\vec{e}_i$ will be in the span of the set of all eigenvectors we found. Therefore, the set of all eigenvectors found form a basis for $\mathbb{R}^n$.

(⇐)

Assume for contradiction that $\exists \vec{w} \in \mathbb{R}^n$ for which $p(t)$ has a repeated root λ_*. Then we can write the polynomial as

$$p(t) = (t - \lambda_*)^2(t - \lambda_1)(t - \lambda_2)\cdots(t - \lambda_k)$$

From this polynomial, we will find eigenvectors $\vec{v}_*, \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$.

Consider our starting vector $\vec{w}$; since we assume that the eigenvectors form a basis for $\mathbb{R}^n$, we can write it as

$$\vec{w} = a_*\vec{v}_* + a_1\vec{v}_1 + \cdots + a_k\vec{v}_k.$$

We also know that $p(A)\vec{w} = \vec{0}$. So we can see that each factor $(A - \lambda_i I)$ makes the term $(A - \lambda_i I)a_i\vec{v}_i = \vec{0}$ (since the $\vec{v}_i$'s are eigenvectors). This includes $(A - \lambda_* I)^2 a_*\vec{v}_* = \vec{0}$; it follows that also $(A - \lambda_* I)a_*\vec{v}_* = \vec{0}$ (i.e., we don't need the additional factor of $(A - \lambda_* I)$ to make the term equal to 0). This means there is another polynomial

$$p_*(t) = (t - \lambda_*)(t - \lambda_1)\cdots(t - \lambda_k)$$

with the property that $p_*(A)\vec{w} = \vec{0}$. However, the degree of $p_*(t)$ is one less than the degree of $p(t)$; but we assumed that $p(t)$ was the lowest degree polynomial for which $p(A)\vec{w} = \vec{0}$. This is because in our algorithm, we start with the vector $\vec{w}$, then calculate $A\vec{w}, A^2\vec{w}$, and so on, and we stop as soon as we find a linear dependence (i.e., we would have first found a polynomial with only simple roots). Since this is a contradiction, then there is not a polynomial with repeated roots. ■

Proof 5.1

(i) Define "countably infinite".

(ii) Prove that the set of positive rational numbers is countably infinite.

(iii) Prove that the set of real numbers in the interval $[0, 1]$, as represented by infinite decimals, is not countable.

Proof

Part (a)

A set A is countably infinite if and only if there exists a *bijection* $f : \mathbb{N} \to A$.

Part (b)

We can organize $\mathbb{Q}^+$ in a two-dimensional table, with the numerator along the columns and denominator along the rows.

	1	2	3	4	$\cdots$
1	1/1	2/1	3/1	4/1	$\cdots$
2	1/2	2/2	3/2	4/2	$\cdots$
3	1/3	2/3	$\cdots$		
4	1/4	$\cdots$			
$\vdots$					

We can then give each of the elements an index, starting from the top-left, and moving diagonally. We skip any "repeats"; e.g., 2/2 is the same as 1/1. So the elements could be listed as $\{1/1,\ 2/1,\ 1/2,\ 1/3,\ 3/1,\ 4/1,\ 3/2,\ 2/3,\ 1/4,\ \ldots\}$. Therefore, since each unique rational number has a unique index in $\mathbb{N}$, there is a bijection $f : \mathbb{N} \to \mathbb{Q}^+$. ∎

Part (c)

Assume for contradiction that there is some bijection $f : \mathbb{N} \to \mathbb{R}_{[0,1]}$. We can list the elements:

$$\begin{aligned} r_1 &= 0.14375\ldots \\ r_2 &= 0.25980\ldots \\ r_3 &= 0.37214\ldots \\ r_4 &= 0.12359\ldots \\ &\vdots \end{aligned}$$

We can create a number x that is not in the list by choosing the kth digit of r_k, for all k (in red above). We will then change that digit $j \mapsto 9 - j$; so the first digit of r_1 is 1, which becomes 8. Then $x = 0.8474\ldots$. However, for all k, the kth digit of x is different from the kth digit of r_k (by construction), so $x \neq r_k$ for all k, and the list is not surjective. Therefore, $\mathbb{R}_{[0,1]}$ is not countable. ∎

Proof 5.2 (The Archimedean Property of $\mathbb{R}$)

Use the Completeness Axiom of $\mathbb{R}$ to prove that for any two positive real numbers a and b, there exists a positive integer n such that $na > b$.

Proof

Assume for contradiction that

$$\exists a, b \in \mathbb{R}^+ : \forall n > 0,\ \ na \leq b.$$

Consider the set $S = \{x \in \mathbb{R} : x = na,\ \ n \in \mathbb{N}\}$; that is, the set S consists of multiples of a. We know that S is nonempty, because $a \in S$. We also know that S is bounded above, because b is an upper bound by assumption: $na \leq b$.

So by the *Completeness Axiom*, S has a *least upper bound* (or *supremum*). We will call this $y = \sup S$. Since $a > 0$, then $y - a < y$, so $y - a$ is *not* an upper bound for S. That means $\exists x \in S : x > y - a$; we can write this as

$$\begin{aligned} na &> y - a \\ \Rightarrow (n+1)a &> y \end{aligned}$$

Clearly, $(n+1)a \in S$ (it is a multiple of a); however, y was supposed to be an upper bound for S, yet we've found an element of S that is bigger. Since this is a contradiction, the Archimedean Property of $\mathbb{R}$ must be true. ■

Proof 5.3

Suppose that $s_n \neq 0$ for all n and that $s = \lim s_n > 0$.

Prove that $\exists N$ such that $\forall n > N$, $s_n > s/2$ and that $\frac{1}{s_n}$ converges to $\frac{1}{s}$.

Proof

Part (a)

Set $\varepsilon_1 = \frac{s}{2}$. Because $\lim_{n\to\infty} s_n = s$, $\exists N_1 : \forall n > N_1$, $|s_n - s| < \frac{s}{2}$. We can rewrite the last inequality as

$$\begin{aligned} -\frac{s}{2} &< s_n - s < \frac{s}{2} \\ \Rightarrow s - \frac{s}{2} &< s_n < s + \frac{s}{2} \\ \Rightarrow \boxed{\frac{s}{2} < s_n} &< \frac{3s}{2} \end{aligned}$$

So we have shown that $s_n > s/2$.

Part (b)

$\forall \varepsilon > 0$, choose N_2 such that $\forall n > N_2$, $|s - s_n| < \varepsilon \frac{s^2}{2}$.

Then $\forall n > \max\{N_1, N_2\}$,

$$\begin{aligned} \left| \frac{1}{s_n} - \frac{1}{s} \right| &= \left| \frac{s - s_n}{s_n s} \right| \\ &= \frac{|s - s_n|}{s_n s} \\ &< \frac{\varepsilon \frac{s^2}{2}}{\frac{s}{2} s} \\ &= \varepsilon \end{aligned}$$

So, $\lim \frac{1}{s_n} = \frac{1}{s}$. ■

Proof 5.4

Suppose that $\lim s_n = +\infty$ and $\lim t_n > 0$. Prove that $\lim s_n t_n = +\infty$.

Proof

Since (t_n) converges, let $\lim t_n = t$. Then, $\forall \varepsilon > 0, \exists N : \forall n > N,\ |t_n - t| < \varepsilon$.

Let $\varepsilon = \frac{t}{2}$; then $\exists N_t : \forall n > N_t, |t_n - t| < \frac{t}{2}$, which we can rewrite as (see *Proof 5.3a*)

$$t_n > \frac{t}{2}.$$

Since (s_n) *diverges*, $\forall M > 0, \exists N : \forall n > N, s_n > M$.

Choose some N_s such that $\forall n > N_s$, $s_n > \frac{2M}{t}$.
Then $\forall n > \max\{N_s, N_t\}$, both $s_n > \frac{2M}{t}$ and $t_n > \frac{t}{2}$, so

$$\begin{aligned} s_n t_n &> \frac{2M}{t} \cdot \frac{t}{2} \\ &= M \end{aligned}$$ ■

Proof 6.1

A Cauchy sequence is defined as a sequence where

$$\forall \varepsilon > 0,\ \exists N : \forall m, n > N \implies |s_n - s_m| < \varepsilon$$

(i) Prove that any Cauchy sequence is bounded.
(ii) Prove that any convergent sequence is Cauchy.
(iii) Prove that any Cauchy sequence of real numbers is convergent.

Proof

Part (a)

Set $\varepsilon = 1$; then since (s_n) is Cauchy, $\exists N : \forall m, n > N,\ |s_n - s_m| < 1$.

If we set $m = N + 1$, this becomes $|s_n - s_{N+1}| < 1$.

Then $\forall n > N$,

$$\begin{aligned} &|s_n| = |s_n - s_m + s_m| && \\ \Rightarrow\ &|s_n| \leq |s_n - s_m| + |s_m| && \text{(Triangle inequality)} \\ \Rightarrow\ &|s_n| \leq 1 + |s_{N+1}| && \text{(Substitute } m = N+1) \end{aligned}$$

Set $M = \max\{|s_1|, |s_2|, \ldots, |s_N|, |s_{N+1}| + 1\}$. Then, $\forall n$, $|s_n| \leq M$. Therefore, (s_n) is bounded. ■

Part (b)

Assume (s_n) is convergent, where $\lim s_n = s$.

Fix $\varepsilon > 0$, and choose some N so that $\forall n > N,\ |s_n - s| < \varepsilon/2$.

Then $\forall m, n > N$,

$$\begin{aligned} |s_n - s_m| &= |s_n - s + s - s_m| && \\ &\leq |s_n - s| + |s - s_m| && \text{(Triangle inequality)} \\ &< \varepsilon/2 + \varepsilon/2 && \\ &= \varepsilon && \end{aligned}$$

Therefore, (s_n) is Cauchy. ■

Part (c)

Assume for contradiction that there is a sequence (s_n) that is Cauchy but is *not* convergent. Then $\liminf s_n \neq \limsup s_n$ (see Topic 6 from lecture), and neither limit equals $+\infty$.

We can choose some $\varepsilon > 0$ with the property $\limsup s_n - \liminf s_n = 3\varepsilon$.

By definition, $\forall N, \exists n > N : s_n > \limsup s_n - \varepsilon$, and $\forall N, \exists m > N : s_m < \liminf s_n + \varepsilon$. Then

$$\begin{aligned} s_n - s_m &> (\limsup s_n - \varepsilon) - (\liminf s_n + \varepsilon) \\ &= \limsup s_n - \liminf s_n - 2\varepsilon \\ &= \varepsilon \end{aligned}$$

So $\exists \varepsilon > 0 : \forall N,\ \exists m, n > N$ for which $|s_n - s_m| > \varepsilon$, a contradiction. So (s_n) is convergent. ■

Proof 6.2: Bolzano-Weierstrass

(i) Prove that any bounded increasing sequence converges. (You may assume without additional proof the corresponding result, that any bounded decreasing sequence converges.)

(ii) Prove that every sequence (s_n) has a monotonic subsequence.

(iii) Prove the Bolzano-Weierstrass Theorem: every bounded sequence has a convergent subsequence.

Proof

Part (a)

By the *Completeness Axiom*, any bounded sequence of numbers has a supremum: $\sup s_k = s$. So $\forall k,\ s_k \leq s$.

Consider any $\varepsilon > 0$; then $s - \varepsilon < s$, so $s - \varepsilon$ is not an upper bound. Then $\exists N : s - \varepsilon < s_N \leq s$.

Since (s_k) is increasing, $\forall k > N,\ s_k \geq s_N$, so

$$s - \varepsilon < s_N \leq s_k \leq s$$

This means that $|s_k - s| < \varepsilon$, so (s_k) converges. ■

Part (b)

Define a *dominant term* of a sequence as a term that is greater than all terms after it. (s_n) has either infinitely many or finitely many dominant terms.

- If there are infinitely many, then the dominant terms form a *decreasing* subsequence.
- If there are finitely many, $\exists s_N$ that is the last dominant term. This means that s_{N+1} is *not* a dominant term; we pick this as the first element of our subsequence. Similarly, there is some element after s_{N+1} that is greater; we pick this as the next element of the subsequence. Since that term was not dominant, we keep doing this. Therefore, we have constructed an *increasing* subsequence.

Therefore, (s_n) has a monotonic subsequence. ■

Part (c)

We've shown that every (bounded) sequence has a monotonic subsequence. Furthermore, the monotonic subsequence of a bounded sequence is itself bounded. We also proved that every monotonic bounded subsequence converges. Therefore, every bounded sequence has a subsequence that converges. ■

Proof 6.3

(i) Define "Hausdorff space", and prove that in a Hausdorff space the limit of a sequence is unique.

(ii) Prove that $\mathbb{R}^n$, with the topology defined by open balls, is a Hausdorff space.

Proof

Part (a)

A *Hausdorff space* is a set with the following topology. Given any two distinct elements a and b, we can find open sets A and B with $a \in A,\ b \in B$, and $A \cap B = \emptyset$.

Suppose that $(s_n) \to a$, and assume A, B are open sets such that $A \cap B = \emptyset$.

Since A is an open set with $a \in A$, $\exists N : \forall n > N,\ s_n \in A$. Since A and B are disjoint, then all $s_n \notin B$. But B is an open set with $b \in B$. So s_n does not converge to b.

Therefore, if a and b are distinct, a sequence cannot converge to both a and b in a Hausdorff space.

$\blacksquare$

Part (b)

Let $\vec{a} \neq \vec{b}$; then $\left|\vec{a} - \vec{b}\right| = \varepsilon > 0$. Let $r = \varepsilon/4$; and let $A = B_r(\vec{a})$ and $B = B_r(\vec{b})$ be open balls.

Assume for contradiction that $A \cap B \neq \emptyset$; so $\exists \vec{x} \in \mathbb{R}^n : \vec{x} \in A \cap B$. Then since $\vec{x} \in A$, $|\vec{x} - \vec{a}| < \varepsilon/4$; similarly, $\vec{x} \in B \implies \left|\vec{x} - \vec{b}\right| < \varepsilon/4$.

Then

$$\begin{aligned}
\left|\vec{a} - \vec{b}\right| &= \left|\vec{a} - \vec{x} + \vec{x} - \vec{b}\right| \\
&\leq |\vec{a} - \vec{x}| + \left|\vec{x} - \vec{b}\right| \\
&< \varepsilon/4 + \varepsilon/4 \\
&= \varepsilon/2
\end{aligned}$$

But we assumed that $\left|\vec{a} - \vec{b}\right| = \varepsilon$, and since $\varepsilon > 0$, $\varepsilon < \varepsilon/2$ is a contradiction. Therefore, $\mathbb{R}^n$ is a Hausdorff space. $\blacksquare$

Proof 7.1

(i) If function $\mathbf{f}$ is continuous, every sequence is good.
Given that function $\mathbf{f} : \mathbb{R}^k \to \mathbb{R}^m$ is continuous at $\vec{x}_0$, prove that every sequence such that $\vec{x}_n \to \vec{x}_0$ is a "good sequence" in the sense that $\mathbf{f}(\vec{x}_n)$ converges to $\mathbf{f}(\vec{x}_0)$.

(ii) If function $\mathbf{f}$ is discontinuous, there exists a bad sequence.
Given that function $\mathbf{f} : \mathbb{R}^k \to \mathbb{R}^m$ is discontinuous at $\vec{x}_0$, show how to construct a "bad sequence" such that $\vec{x}_n \to \vec{x}_0$ but $\mathbf{f}(\vec{x}_n)$ does not converge to $\mathbf{f}(\vec{x}_0)$.

Proof

Part (a)

We are given that $\mathbf{f} : \mathbb{R}^k \to \mathbb{R}^m$ is continuous at $\vec{x_0}$; i.e.,

$$\forall \varepsilon > 0,\ \exists \delta > 0 : \forall \vec{x} \in \mathbb{R}^k,\ |\vec{x} - \vec{x}_0| < \delta \implies |\mathbf{f}(\vec{x}) - \mathbf{f}(\vec{x}_0| < \varepsilon \tag{1}$$

Consider any sequence $\vec{x}_n \in \mathbb{R}^k$ such that $(\vec{x}_n)$ converges to $\vec{x}_0$.

Then by the definition of convergence, $\forall \delta > 0,\ \exists N : \forall n > N,\ |\vec{x}_n - \vec{x}_0| < \delta$.

By Equation (1), then $|\mathbf{f}(\vec{x}_n) - \mathbf{f}(\vec{x}_0)| < \varepsilon$.

So $\forall \varepsilon > 0,\ \exists N : \forall n > N,\ |\mathbf{f}(\vec{x}_n) - \mathbf{f}(\vec{x}_0)| < \varepsilon$; in other words, $\mathbf{f}(\vec{x}_n) \to \mathbf{f}(\vec{x}_0)$, and every sequence is a good sequence. ■

Part (b)

If $\mathbf{f}$ is discontinuous,

$$\exists \varepsilon > 0 : \forall \delta > 0,\ \exists \vec{x} \in \mathbb{R}^k : |\vec{x} - \vec{x}_0| < \delta \wedge |\mathbf{f}(\vec{x}) - \mathbf{f}(\vec{x}_0)| \geq \varepsilon$$

We will construct a bad sequence:

$\forall n \in \mathbb{N}$, set $\delta_n = \frac{1}{n}$. Then $\exists \vec{x}_n \in \mathbb{R}^k : |\vec{x}_n - \vec{x}_0| < \delta_n \wedge |\mathbf{f}(\vec{x}_n) - \mathbf{f}(\vec{x}_0)| \geq \varepsilon$.

Since $|\vec{x}_n - \vec{x}_0| < \frac{1}{n}$, we can let $N > \frac{1}{\varepsilon}$, so $(\vec{x}_n)$ converges to $\vec{x}_0$. But since $|\mathbf{f}(\vec{x}_n) - \mathbf{f}(\vec{x}_0)| \geq \varepsilon$, $(\mathbf{f}(\vec{x}_n))$ does *not* converge to $\mathbf{f}(\vec{x}_0)$. Therefore, we have found a bad sequence. ■

Proof 7.2: The intermediate value theorem

Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous on the interval $[a, b]$ (with $a < b$), and $f(a) < y < f(b)$.

Prove that there exists at least one $x \in [a, b]$ such that $f(x) = y$.

Use Ross's "no bad sequence" definition of continuity, not the epsilon-delta definition. Constructing the appropriate sequences requires some care.

Proof

Define a set $S = \{x \in [a, b] \mid f(x) < y\}$. S is not empty, because $f(a) < y$, so $a \in S$. And S is bounded, because $y < f(b)$, so $b \notin S$. So, by the Completeness Axiom, S has a supremum: $x_0 = \sup S$.

We want to prove that $f(x_0) = y$.

($\leq$)

Note that $\forall n \in \mathbb{N},\ x_0 - \frac{1}{n}$ is not an upper bound of S. So $\exists s_n \in S : x_0 - \frac{1}{n} < s_n \leq x_0$. This means that (s_n) converges to x_0 (by the *Squeeze Lemma*).

Since f is continuous, every sequence is a good sequence; then since $(s_n) \to x_0$, $f(s_n) \to f(x_0)$, or $\lim f(s_n) = f(x_0)$. Furthermore, $\forall n,\ s_n \in S$; so by definition every $f(s_n) < y$ and therefore $\lim f(s_n) = \boxed{f(x_0) \leq y}$.

($\geq$)

Now we construct another sequence $(t_n) \notin S$. For all $n \in \mathbb{N}$, we define $t_n = \min\left\{x_0 + \frac{1}{n}, b\right\}$. First, $x_0 \leq b$ (because x_0 is the *least upper bound*), and clearly $x_0 < x_0 + \frac{1}{n}$ for all n, so $x_0 \leq \min\left\{x_0 + \frac{1}{n}, b\right\}$; this means $x_0 \leq t_n$. Furthermore, by definition $t_n \leq x_0 + \frac{1}{n}$. This gives us

$$x_0 \leq t_n \leq x_0 + \frac{1}{n}$$

Then by the *Squeeze Lemma*, $\lim_{n\to\infty} t_n = x_0$. Since f is continuous, (t_n) is a good sequence, so $\lim f(t_n) = f(x_0)$.

By definition, $\forall n,\ t_n \notin S$, so $f(t_n) \geq y$; therefore, $\lim f(t_n) = \boxed{f(x_0) \geq y}$.

Since we've shown that $f(x_0) \leq y \leq f(x_0)$, then $y = f(x_0)$. ■

Proof 7.3

Using the Bolzano-Weierstrass theorem, prove that if function $f : \mathbb{R} \to \mathbb{R}$ is continuous on the closed interval $[a, b]$, then f is uniformly continuous on $[a, b]$.

Proof

Prove the contrapositive: if f is not uniformly continuous on $[a, b]$, then f is not continuous for some $x_0 \in [a, b]$.

If f is not uniformly continuous, then

$$\exists \varepsilon > 0 : \forall \delta > 0,\ \exists x, y \in [a, b] : |x - y| < \delta \wedge |f(x) - f(y)| \geq \varepsilon$$

Set $\delta = \frac{1}{n}$. Then $\exists x_n, y_n \in [a, b] : |x_n - y_n| < \frac{1}{n} \wedge |f(x_n) - f(y_n)| \geq \varepsilon$.

These sequences (x_n) and (y_n) might not converge, but they are bounded on $[a, b]$.

By the Bolzano-Weierstrass theorem, we can construct a convergent subsequence (x_{n_k}) such that $\lim x_{n_k} = x_0 \in [a, b]$ (since a closed interval contains all its limit points).

We can also show that $(y_{n_k}) \to x_0$:

$$|y_{n_k} - x_0| \leq \underbrace{|y_{n_k} - x_{n_k}|}_{< \frac{1}{n_k}} + \underbrace{|x_{n_k} - x_0|}_{x_{n_k} \to x_0}$$

so both terms can be made arbitrarily small. We can then conclude that $\lim y_{n_k} = x_0$.

But what we seek from the earlier statement (by assumption) is that $|f(x_{n_k}) - f(y_{n_k})| \geq \varepsilon$, so both $f(x_{n_k})$ and $f(y_{n_k})$ cannot both converge to the same value $f(x_0)$. So at least one of (x_{n_k}) or (y_{n_k}) is a *bad sequence*.

Therefore, f is not continuous at $x_0 \in [a, b]$. ■

Proof 7.4

Prove that if $f : \mathbb{R} \to \mathbb{R}$ is uniformly continuous on a set S and (s_n) is a Cauchy sequence in S, then $(f(s_n))$ is a Cauchy sequence.

Proof

Let $\varepsilon > 0$. Since f is uniformly continuous,

$$\exists \delta > 0 : \forall x, y \in S,\ |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon$$

It is also true that

$$\forall s_n, s_m \in S,\ |s_n - s_m| < \delta \implies |f(s_n) - f(s_m)| < \varepsilon$$

Since (s_n) is Cauchy, $\exists N : \forall n, m > N,\ |s_n - s_m| < \delta$. Then also $|f(s_n) - f(s_m)| < \varepsilon$, so $f(s_n)$ is also Cauchy. ■

Proof 8.1: Extreme Value Theorem

Let C be a compact subset of $\mathbb{R}^n$. Let $f : C \to \mathbb{R}$ be a continuous function (a real-valued function). Then f has a supremum M and $\exists \mathbf{b} \in C$ (a maximum) where $f(\mathbf{b}) = M$.

Proof

Part (a)

Assume for contradiction that there is no supremum M.

Create a sequence $(\vec{x}_n) \in C$ such that $f(\vec{x}_1) > 1,\ f(\vec{x}_2) > 2,\ \dots,\ f(\vec{x}_n) > n,\ \dots$.

Since C is compact, by Bolzano-Weierstrass it has a convergent subsequence $(\vec{x}_{n_k}) \in C$, with $(\vec{x}_{n_k}) \to \vec{a} \in C$.

By construction, $f(\vec{x}_{n_k}) > n_k$ for all k, so $(f(\vec{x}_{n_k}))$ does not converge. However, since $(\vec{x}_{n_k}) \to \vec{a}$, this is a *bad sequence*. This contradicts our assumption that f was continuous, so f must have a supremum.

Part (b)

Create a new sequence $(\vec{x}_n) \in C$ such that $f(\vec{x}_n) > M - 1/n$. This sequence, by Bolzano-Weierstrass, has a convergent subsequence $(\vec{x}_{n_k}) \in C$ that converges to some $\vec{b} \in C$. Since f is continuous, then $f(\vec{x}_{n_k}) \to f(\vec{b})$.

By construction, $f(\vec{x}_{n_k}) > M - \frac{1}{n_k}$, and also $f(\vec{x}_{n_k}) \leq M$. Together, this shows that $(f(\vec{x}_{n_k})) \to M$. Since limits are unique ($\mathbb{R}^n$ is a *Hausdorff space*), we must have that $f(\vec{b}) = M$. ■

Proof 8.2: Heine-Borel Theorem

For a compact set, every open cover contains a finite subcover.

If $X \subset \mathbb{R}^n$ is compact (closed and bounded) and $\mathcal{U}$ is any open cover of X, then $\mathcal{U}$ contains a finite subcover: there exist finitely many open sets $U_1, U_2, \ldots, U_m \in \mathcal{U}$ such that

$$X \subset \bigcup_{i=1}^{m} U_i$$

Proof

Assume for contradiction that *no* such finite subcover exists.

Since X is bounded, first divide it into N (closed) unit squares. Since there is no finite subcover, at least one of the closed squares S_0 must require infinitely many of the open sets U_i to cover it.

Next divide S_0 into 4 smaller squares; at least 1 of these smaller squares S_1 will require infinitely many of the U_i to cover it.

Continue the process, which creates a sequence of nested compact sets:

$$S_0 \supset S_1 \supset S_2 \supset \cdots$$

By the *nested compact set theorem*, the infinite intersection is not empty:

$$\bigcap_{k=0}^{\infty} S_k \neq \emptyset$$

So it contains some point $\vec{x} \in X$; since $\vec{x}$ is in our original set, there must exist some U that covers $\vec{x}$. And because U is open, it contains some ball around $\vec{x}$ with $r > 0 : B_r(\vec{x}) \subset U$.

Since there's an open ball around $\vec{x}$ in U, at some point one of the squares S_j must be in the open ball: $B_r(\vec{x}) \supset S_j$. Therefore $S_j \subset U$, and since it is a nested sequence, $\forall i > j,\ S_i \subset U$.

But this is a contradiction, because we said each square S_k in the sequence required *infinitely many* open sets to cover it, while we have found a single open set U that covers S_j.

Therefore, there must be a finite subcover of our open cover. ■

Proof 9.1: Chain Rule

Assume the following:

- Function f is differentiable at a
- Function g is differentiable at $f(a)$
- There is an open interval J containing a on which f is defined and $f(x) \neq f(a)$ for $x \neq a$ (without this restriction, you need the messy Case 2 on page 229).
- Function g is defined on the open interval $I = f(J)$, which contains $f(a)$.

Using the sequential definition of a limit, prove that the composite function $g \circ f$ is differentiable at a and that

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

Proof

Consider any sequence $(x_n) \in J$ for which $(x_n) \to a$. Since f is differentiable, $f'(a) = \lim_{x \to a} \frac{f(x)-f(a)}{x-a}$. Also, the limit of sequences also converges:

$$\left(\frac{f(x_n) - f(a)}{x_n - a} \right) \to f'(a)$$

Plugging (x_n) into f gives us a new sequence, which we can call $y_n = f(x_n)$. We know as well that $(y_n) \to f(a)$ because f is continuous (and every sequence is good).

Consider the difference quotient; we multiply by 1 and rearrange:

$$\frac{g(f(x_n)) - g(f(a))}{x_n - a} = \frac{g(y_n) - g(f(a))}{x_n - a} \cdot \frac{y_n - f(a)}{y_n - f(a)} = \frac{g(y_n) - g(f(a))}{y_n - f(a)} \cdot \frac{f(x_n) - f(a)}{x_n - a}$$

Since g is differentiable at $y = f(a)$, then for any sequence $(y_n) \to f(a)$ the first term above converges to $g'(f(a))$. The second term is just $f'(a)$.

Altogether, since the limit of a product is the product of the limits, we have

$$\lim_{n \to \infty} \frac{g(f(x_n)) - g(f(a))}{x_n - a} = g'(f(a)) \cdot f'(a)$$

$\blacksquare$

Proof 9.2: Rolle's Theorem and Mean Value Theorem

(i) Prove Rolle's Theorem: if f is a continuous function on $[a, b]$ that is differentiable on (a, b) and satisfies $f(a) = f(b)$, then there exists at least one x in (a, b) such that $f'(x) = 0$.

(ii) Using Rolle's Theorem, prove the Mean Value Theorem: if f is a continuous function on $[a, b]$ that is differentiable on (a, b), then there exists at least one x in (a, b) such that

$$f'(x) = \frac{f(b) - f(a)}{b - a}$$

Proof

Part (a)

Since f is continuous on the closed and bounded interval $[a, b]$, by the *Extreme Value Theorem*, f has a minimum (x_0) and a maximum (y_0) on $[a, b]$. So $\forall x \in [a, b],\ f(x_0) \leq f(x) \leq f(y_0)$.

- If the minimum and maximum occur at the endpoints a and b, then since $f(a) = f(b)$ (by assumption), f is constant. That is, $f'(x) = 0$ everywhere on (a, b).
- Otherwise, at least one of the min or max occurs at the interior of (a, b). By a previous proof (*Topic 8* in *Module 9*), then $f'(x) = 0$ there.

Part (b)

Define $L(x)$ to be the line connecting the endpoints. So by the point-slope formula of a line, we have

$$L(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

Also define $g(x) = f(x) - L(x)$; then $g(a) = g(b) = 0$. Since $L(x)$ is continuous and differentiable (it is just a first-order, linear function) and $f(x)$ is continuous and differentiable (by assumption), then $g(x)$ is also differentiable on (a, b) and continuous on $[a, b]$. Therefore, $g(x)$ satisfies the conditions for Rolle's Theorem: $\exists x_0 \in (a, b) : g'(x_0) = 0$. Then,

$$\begin{aligned} g'(x_0) = f'(x_0) - L'(x_0) &= 0 \\ \Rightarrow f'(x_0) &= L'(x_0) \\ &= \frac{f(b) - f(a)}{b - a} \end{aligned} \qquad \blacksquare$$

Proof 9.3: Derivative of an inverse function

Suppose that f is a one-to-one continuous function on open interval I (either strictly increasing or strictly decreasing). Let open interval $J = f(I)$. Define the inverse function $g : J \to I$ such that

$g \circ f(x) = x$ for $x \in I$

$f \circ g(y) = y$ for $y \in J$

And define $y_0 = f(x_0)$.

Take it as proved that g is continuous at y_0.

Prove that, if f is differentiable at x_0 and $f'(x_0) \neq 0$, then

$$\lim_{y \to y_0} \frac{g(y) - g(y_0)}{y - y_0} = \frac{1}{f'(x_0)}$$

Proof

Consider any sequence $(y_n) \in J$ for which $(y_n) \to y_0$. Since g is continuous, then $(g(y_n)) \to g(y_0)$ as well.

Since g is the inverse function of f, then $f(x) = y$ and $g(y) = x$, so $(x_n) \to x_0$ as well. By assumption, f is differentiable at x_0, so

$$\lim_{n \to \infty} \frac{f(x_n) - f(x_0)}{x_n - x_0} = f'(x_0)$$

For sequences, we've shown that $\lim \frac{1}{s_n} = \frac{1}{\lim s_n}$, so we can rewrite the above as

$$\lim_{n \to \infty} \frac{x_n - x_0}{f(x_n) - f(x_0)} = \frac{1}{f'(x_0)}$$

We can then substitute into the above equation since g is the inverse of f:

$$\lim_{n \to \infty} \frac{g(y_n) - g(y_0)}{y_n - y_0} = \frac{1}{f'(x_0)}$$

Since this is true for any arbitrary $(y_n) \to y_0$, the function limit exists and is

$$g'(y_0) = \lim_{y \to y_0} \frac{g(y) - g(y_0)}{y - y_0} = \frac{1}{f'(x_0)}$$

■

Proof 10.1: The product rule in $\mathbb{R}^n$

Let $U \subset \mathbb{R}^n$ be an open set, and let f and g be functions from U to $\mathbb{R}$. Prove that if f and g are differentiable at $\mathbf{a}$ then so is fg, and that

$$[\mathbf{D}(fg)(\mathbf{a})] = f(\mathbf{a})[\mathbf{D}g(\mathbf{a})] + g(\mathbf{a})[\mathbf{D}f(\mathbf{a})]$$

Proof

The remainder is

$$r(\vec{h}) = f(\mathbf{a}+\vec{h})g(\mathbf{a}+\vec{h}) - f(\mathbf{a})g(\mathbf{a}) - f(\mathbf{a})[\mathbf{D}g(\mathbf{a})]\vec{h} - g(\mathbf{a})[\mathbf{D}f(\mathbf{a})]\vec{h}$$

We will rewrite this as $r(\vec{h}) = r_1(\vec{h}) + r_2(\vec{h}) + r_3(\vec{h})$ by adding and subtracting some terms.

$$\begin{aligned}
r_1(\vec{h}) &= f(\mathbf{a}+\vec{h})g(\mathbf{a}+\vec{h}) - f(\mathbf{a})g(\mathbf{a}+\vec{h}) - g(\mathbf{a}+\vec{h})[\mathbf{D}f(\mathbf{a})]\vec{h} \\
r_2(\vec{h}) &= f(\mathbf{a})g(\mathbf{a}+\vec{h}) - f(\mathbf{a})g(\mathbf{a}) - f(\mathbf{a})[\mathbf{D}g(\mathbf{a})]\vec{h} \\
r_3(\vec{h}) &= g(\mathbf{a}+\vec{h})[\mathbf{D}f(\mathbf{a})]\vec{h} - g(\mathbf{a})[\mathbf{D}f(\mathbf{a})]\vec{h}
\end{aligned}$$

We will look at the limits separately:

$$\lim_{\vec{h}\to\vec{0}} \frac{r_1(\vec{h})}{|\vec{h}|} = \lim_{\vec{h}\to\vec{0}} \underbrace{\frac{f(\mathbf{a}+\vec{h}) - f(\mathbf{a}) - [\mathbf{D}f(\mathbf{a})]\vec{h}}{|\vec{h}|}}_{\substack{\to 0 \text{ because} \\ f \text{ is differentiable}}} \underbrace{g(\mathbf{a}+\vec{h})}_{\substack{\to g(\mathbf{a}) \\ \text{(bounded)}}} = 0$$

$$\lim_{\vec{h}\to\vec{0}} \frac{r_2(\vec{h})}{|\vec{h}|} = \lim_{\vec{h}\to\vec{0}} \underbrace{\frac{g(\mathbf{a}+\vec{h}) - g(\mathbf{a}) - [\mathbf{D}g(\mathbf{a})]\vec{h}}{|\vec{h}|}}_{\substack{\to 0 \text{ because} \\ g \text{ is differentiable}}} \underbrace{f(\mathbf{a})}_{\text{constant}} = 0$$

$$\lim_{\vec{h}\to\vec{0}} \frac{r_3(\vec{h})}{|\vec{h}|} = \lim_{\vec{h}\to\vec{0}} \underbrace{\left(g(\mathbf{a}+\vec{h}) - g(\mathbf{a})\right)}_{\substack{\to 0 \text{ because} \\ g \text{ is continuous}}} \underbrace{[\mathbf{D}f(\mathbf{a})]}_{\substack{\text{fixed} \\ \text{matrix}}} \underbrace{\frac{\vec{h}}{|\vec{h}|}}_{\substack{\text{unit} \\ \text{vector}}} = 0$$

The final limit equals 0 because the latter two terms together are bounded.

Since the limit of each of the remainders divided by the length of $\vec{h}$ equals 0, the limit of their sum equals 0:

$$\lim_{\vec{h}\to\vec{0}} \frac{r(\vec{h})}{|\vec{h}|} = 0$$

Therefore,

$$[\mathbf{D}(fg)(\mathbf{a})] = f(\mathbf{a})[\mathbf{D}g(\mathbf{a})] + g(\mathbf{a})[\mathbf{D}f(\mathbf{a})] \qquad \blacksquare$$

Proof 10.2

Using the mean value theorem, prove that if a function $f : \mathbb{R}^2 \to \mathbb{R}$ has partial derivatives $D_1 f$ and $D_2 f$ that are continuous at $\mathbf{a}$, it is differentiable at $\mathbf{a}$ and its derivative is the Jacobian matrix $\begin{bmatrix} D_1 f(\mathbf{a}) & D_2 f(\mathbf{a}) \end{bmatrix}$.

Proof

Let $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ and $\vec{h} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$. So $\mathbf{a} + \vec{h} = \begin{pmatrix} a_1 + h_1 \\ a_2 + h_2 \end{pmatrix}$.

If $\begin{bmatrix} D_1 f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} & D_2 f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \end{bmatrix}$ is the derivative, then the remainder is

$$\begin{aligned} r\left(\vec{h}\right) = r \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} &= f(\mathbf{a} + \vec{h}) - f(\mathbf{a}) - \begin{bmatrix} D_1 f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} & D_2 f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \\ &= f \begin{pmatrix} a_1 + h_1 \\ a_2 + h_2 \end{pmatrix} - f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} - D_1 f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} h_1 - D_2 f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} h_2 \end{aligned}$$

We will split the remainder into two terms (after adding and subtracting a new term):

$$\begin{aligned} r_1\left(\vec{h}\right) &= f \begin{pmatrix} a_1 + h_1 \\ a_2 + h_2 \end{pmatrix} - f \begin{pmatrix} a_1 \\ a_2 + h_2 \end{pmatrix} - D_1 f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} h_1 \\ r_2\left(\vec{h}\right) &= f \begin{pmatrix} a_1 \\ a_2 + h_2 \end{pmatrix} - f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} - D_2 f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} h_2 \end{aligned}$$

For $r_1(\vec{h})$, apply the *Mean Value Theorem* (since it only varies in its first component):

$$\exists b_1 \in (a_1, a_1 + h_1) : D_1 f \begin{pmatrix} b_1 \\ a_2 + h_2 \end{pmatrix} h_1 = f \begin{pmatrix} a_1 + h_1 \\ a_2 + h_2 \end{pmatrix} - f \begin{pmatrix} a_1 \\ a_2 + h_2 \end{pmatrix}$$

And for $r_2(\vec{h})$:

$$\exists b_2 \in (a_2, a_2 + h_2) : D_2 f \begin{pmatrix} a_1 \\ b_2 \end{pmatrix} h_2 = f \begin{pmatrix} a_1 \\ a_2 + h_2 \end{pmatrix} - f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

We will substitute these values into the remainder terms and take the limit.

As $\vec{h} \to \vec{0}$, since $b_1 \in (a_1, a_1 + h_1)$, then $b_1 \to a_1$ and $(a_2 + h_2) \to a_2$.

And since $D_1 f$ is continuous, all sequences are good. So $\left(D_1 f \begin{pmatrix} b_1 \\ a_2 + h_2 \end{pmatrix} \right) \to D_1 f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$.

Then

$$\lim_{\vec{h} \to \vec{0}} \frac{r_1\left(\vec{h}\right)}{\left|\vec{h}\right|} = \lim_{\vec{h} \to \vec{0}} \underbrace{\left(D_1 f \begin{pmatrix} b_1 \\ a_2 + h_2 \end{pmatrix} - D_1 f \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \right)}_{\to 0} \underbrace{\frac{h_1}{\left|\vec{h}\right|}}_{\text{bounded}} = 0$$

The argument for $r_2(\vec{h})$ is the same, so

$$\lim_{\vec{h} \to \vec{0}} \frac{r(\vec{h})}{\left|\vec{h}\right|} = 0$$

and the function is differentiable. ■

Proof 11.1: The Implicit Function Theorem

Let W be an open subset of $\mathbb{R}^n$, and let $\mathbf{F} : W \to \mathbb{R}^{n-k}$ be a C^1 mapping such that $\mathbf{F}(\mathbf{c}) = \mathbf{0}$. Assume that $\big[\mathbf{DF}(\mathbf{c})\big]$ is onto.

(i) Prove that the n variables can be ordered so that the first $n-k$ columns of $\big[\mathbf{DF}(\mathbf{c})\big]$ are linearly independent and that $\big[\mathbf{DF}(\mathbf{c})\big] = \big[A \mid B\big]$ where A is an invertible $(n-k)\times(n-k)$ matrix.

Set $\mathbf{c} = \begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix}$, where $\mathbf{a}$ are the $n-k$ passive variables and $\mathbf{b}$ are the k active variables. Let $\mathbf{g}$ be the "implicit function" from a neighborhood of $\mathbf{b}$ to a neighborhood of $\mathbf{a}$ such that $\mathbf{g}(\mathbf{b}) = \mathbf{a}$ and $\mathbf{F}\begin{pmatrix}\mathbf{g}(\mathbf{y})\\ \mathbf{y}\end{pmatrix} = \mathbf{0}$.

(ii) Prove that $\big[\mathbf{Dg}(\mathbf{b})\big] = -A^{-1}B$.

Proof

Part (a)

Since $\mathbf{F} : W \subset \mathbb{R}^n \to \mathbb{R}^{n-k}$, $\big[\mathbf{DF}\big]$ is a $(n-k)\times n$ matrix. And since $\big[\mathbf{DF}(\mathbf{c})\big]$ is onto ($\mathbb{R}^{n-k}$), all $n-k$ rows are independent, which means we have $n-k$ pivotal columns.

We reorder the variables so that the $n-k$ passive variables are first; then we have a square matrix A with $n-k$ independent columns. Therefore, A is invertible.

Part (b)

Let $\mathbf{c} = \begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix}$, such that $\mathbf{a}$ has $n-k$ components and $\mathbf{b}$ has k components. If $\mathbf{b}$ changes a small amount, to a general point $\mathbf{y}$, how does $\mathbf{a}$ need to change to $\mathbf{x} = \mathbf{g}(\mathbf{y})$ so that we still have $\mathbf{F}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \vec{0}$?

We start with $\mathbf{F}\begin{pmatrix}\mathbf{g}(\mathbf{y})\\ \mathbf{y}\end{pmatrix} = \vec{0}$ and take the derivative of both sides with respect to $\mathbf{y}$ (using the chain rule):

$$\begin{aligned}
\big[\mathbf{DF}(\mathbf{c})\big]\begin{bmatrix}\mathbf{Dg}(\mathbf{b})\\ I_k\end{bmatrix} &= \vec{0}\\
\Rightarrow \big[A \mid B\big]\begin{bmatrix}\mathbf{Dg}(\mathbf{b})\\ I_k\end{bmatrix} &= \vec{0}\\
\Rightarrow A\big[\mathbf{Dg}(\mathbf{b})\big] + BI_k &= \vec{0}\\
\Rightarrow \big[\mathbf{Dg}(\mathbf{b})\big] &= -A^{-1}B
\end{aligned}$$

■

Proof 11.2: Tangent space as a kernel

Suppose that $U \subset \mathbb{R}^n$ is an open subset, $\mathbf{F} : U \to \mathbb{R}^{n-k}$ is a C^1 mapping, and manifold M can be described as the set of points that satisfy $\mathbf{F}(\mathbf{z}) = 0$. Use the implicit function theorem to show that if $\left[\mathbf{DF}(\mathbf{c})\right]$ is onto for $\mathbf{c} \in M$, then the tangent space $T_{\mathbf{c}}M$ is the kernel of $\left[\mathbf{DF}(\mathbf{c})\right]$. You may assume that the variables have been numbered so that when you row-reduce $\left[\mathbf{DF}(\mathbf{c})\right]$, the first $n-k$ columns are pivotal.

Proof

($\subseteq$)

Since $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^{n-k}$, then $\left[\mathbf{DF}(\mathbf{c})\right]$ will be a $(n-k) \times n$ matrix. By assumption, this matrix is *onto*, so there will be $(n-k)$ independent columns. So the matrix A (from the first $n-k$ columns) will be invertible. And by the *Rank-Nullity Theorem*, there will be k nonpivotal columns, and dim ker $\left[\mathbf{DF}(\mathbf{c})\right] = k$.

We know that the manifold will also be the graph of some implicit function expressing the passive variables $\mathbf{x}$ as a function of the active variables $\mathbf{y}$: $\mathbf{x} = \mathbf{g}(\mathbf{y})$. Near the point $\mathbf{c}$, we can say that $\mathbf{a} = \mathbf{g}(\mathbf{b})$, and by the *Implicit Function Theorem*, $\left[\mathbf{Dg}(\mathbf{b})\right] = A^{-1}B$.

The tangent space of M at $\mathbf{c}$ is defined as the graph of $\left[\mathbf{Dg}(\mathbf{b})\right]$ so that vectors in $T_{\mathbf{c}}M$ are of the form

$$\begin{aligned}
\dot{\mathbf{z}} = \begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} \left[\mathbf{Dg}(\mathbf{b})\right]\dot{\mathbf{y}} \\ \dot{\mathbf{y}} \end{bmatrix} &= \begin{bmatrix} -A^{-1}B\dot{\mathbf{y}} \\ \dot{\mathbf{y}} \end{bmatrix} \\
\implies \left[\mathbf{DF}(\mathbf{c})\right]\dot{\mathbf{z}} &= \left[A \mid B\right] \begin{bmatrix} -A^{-1}B\dot{\mathbf{y}} \\ \dot{\mathbf{y}} \end{bmatrix} \\
&= -AA^{-1}B\dot{\mathbf{y}} + B\dot{\mathbf{y}} \\
&= -B\dot{\mathbf{y}} + B\dot{\mathbf{y}} \\
&= \vec{0}
\end{aligned}$$

This shows that $\dot{\mathbf{z}} \in T_{\mathbf{c}}M \implies \dot{\mathbf{z}} \in \ker\left[\mathbf{DF}(\mathbf{c})\right]$, so $\boxed{T_{\mathbf{c}}M \subseteq \ker\left[\mathbf{DF}(\mathbf{c})\right]}$.

($\supseteq$)

If $\dot{\mathbf{z}} = \begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{y}} \end{bmatrix} \in \ker\left[\mathbf{DF}(\mathbf{c})\right]$, then

$$\begin{aligned}
\left[\mathbf{DF}(\mathbf{c})\right]\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{y}} \end{bmatrix} &= \vec{0} \\
\implies \left[A \mid B\right]\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{y}} \end{bmatrix} &= \vec{0} \\
\implies A\dot{\mathbf{x}} + B\dot{\mathbf{y}} &= \vec{0} \\
\implies \dot{\mathbf{x}} &= -A^{-1}B\dot{\mathbf{y}} \\
&= \left[\mathbf{Dg}(\mathbf{b})\right]\dot{\mathbf{y}}
\end{aligned}$$

This shows that $\dot{\mathbf{z}} \in T_{\mathbf{c}}M$, so $\boxed{\ker\left[\mathbf{DF}(\mathbf{c})\right] \subseteq T_{\mathbf{c}}M}$.

Taken together, we've shown that $T_{\mathbf{c}}M = \ker\left[\mathbf{DF}(\mathbf{c})\right]$. ■

Proof 12.1: Constrained Critical Points and Tangent Space

Let $U \in \mathbb{R}^n$ be an open subset and let $f : U \to \mathbb{R}$ be a C^1 (continuously differentiable) function. Let $M \subset \mathbb{R}^n$ be a k-dimensional manifold.

Prove that $\mathbf{c} \in M \cap U$ is a critical point of f restricted to M if and only if

$$T_{\mathbf{c}}M \subset \ker \left[\mathbf{D}f(\mathbf{c})\right]$$

Proof

We have a manifold M that is given by a parametrization $\gamma : \mathbb{R}^k \to \mathbb{R}^n$; this expresses points on the manifold as a function of some parameters: $\mathbf{z} = \gamma(\mathbf{u})$.

We also have a function $f : \mathbb{R}^n \to \mathbb{R}$ (a "goal" function); we want to optimize f but only for points $\mathbf{z}$ on the manifold M. We take the derivative:

$$\left[\mathbf{D}\left(f \circ \gamma\right)(\mathbf{u})\right] = \left[\mathbf{D}f(\gamma(\mathbf{u}))\right]\left[\mathbf{D}\gamma(\mathbf{u})\right] = 0$$

(⇒)

If $\mathbf{c} = \gamma(\mathbf{u})$ is a critical point, then $\left[\mathbf{D}f(\mathbf{c})\right]\left[\mathbf{D}\gamma(\mathbf{u})\right] = 0$.
We also know that $T_{\mathbf{c}}M = \text{Img}\left[\mathbf{D}\gamma(\mathbf{u})\right]$, which means $\forall \vec{v} \in T_{\mathbf{c}}M, \exists \vec{w} : \vec{v} = \left[\mathbf{D}\gamma(\mathbf{u})\right]\vec{w}$. Then

$$\left[\mathbf{D}f(\mathbf{c})\right]\vec{v} = \underbrace{\left[\mathbf{D}f(\mathbf{c})\right]\left[\mathbf{D}\gamma(\mathbf{u})\right]}_{=0}\vec{w} = \vec{0}$$

So $\vec{v} \in T_{\mathbf{c}}M \implies \vec{v} \in \ker\left[\mathbf{D}f(\mathbf{c})\right]$, and $T_{\mathbf{c}}M \subset \ker\left[\mathbf{D}f(\mathbf{c})\right]$.

(⇐)

If $T_{\mathbf{c}}M \subset \ker\left[\mathbf{D}f(\mathbf{c})\right]$, then $\text{Img}\left[\mathbf{D}\gamma(\mathbf{u})\right] \subset \ker\left[\mathbf{D}f(\mathbf{c})\right]$. So, since the image is the space spanned by the columns of $\left[\mathbf{D}\gamma\right]$, any of its columns are in the kernel. Therefore,

$$\begin{aligned} &\left[\mathbf{D}f(\mathbf{c})\right]\left[\mathbf{D}\gamma(\mathbf{u})\right] = 0 \\ \Rightarrow &\left[\mathbf{D}(f \circ \gamma)(\mathbf{u})\right] = 0 \end{aligned}$$

So $f \circ \gamma$ has a critical point at $\mathbf{u}$, and in turn f has a critical point (constrained to being on M) at $\mathbf{c} = \gamma(\mathbf{u})$. ■

Proof 12.2: Lagrange Multipliers for Constrained Critical Points

Let M be a manifold known by a real-valued C^1 function $\mathbf{F}(\mathbf{z}) = \mathbf{0}$, where $\mathbf{F}$ goes from an open subset U of $\mathbb{R}^n$ to $\mathbb{R}^m$ and $\left[\mathbf{DF}(\mathbf{z})\right]$ is onto.

Let $f : U \to \mathbb{R}$ be a C^1 function.

Prove that $\mathbf{c} \in M$ is a critical point of f restricted to M if and only if there exist m Lagrange multipliers $\lambda_1, \ldots, \lambda_m$ such that

$$\left[\mathbf{D}f(\mathbf{c})\right] = \lambda_1 \left[\mathbf{D}F_1(\mathbf{c})\right] + \cdots + \lambda_m \left[\mathbf{D}F_m(\mathbf{c})\right]$$

Proof

$\left[\mathbf{DF}\right]$ is a $m \times n$ matrix (partial derivative of each constraint equation): $\left[\mathbf{DF}\right] = \begin{bmatrix} \mathbf{D}F_1 \\ \mathbf{D}F_2 \\ \vdots \\ \mathbf{D}F_m \end{bmatrix}$

($\Rightarrow$)

Assume that $\lambda_1, \lambda_2, \ldots, \lambda_m$ exist so that $\left[\mathbf{D}f\right] = \lambda_1 \left[\mathbf{D}F_1\right] + \lambda_2 \left[\mathbf{D}F_2\right] + \cdots + \lambda_m \left[\mathbf{D}F_m\right]$.
If $\vec{v} \in T_{\mathbf{c}}M = \ker \left[\mathbf{DF}\right]$, then $\left[\mathbf{DF}\right] \vec{v} = \vec{0}$. So all of the $\left[\mathbf{D}F_i\right] \vec{v} = 0$, and

$$\begin{aligned} \lambda_1 \left[\mathbf{D}F_1\right] \vec{v} + \lambda_2 \left[\mathbf{D}F_2\right] \vec{v} + \cdots + \lambda_m \left[\mathbf{D}F_m\right] \vec{v} &= \vec{0} \\ \Rightarrow \left[\mathbf{D}f\right] \vec{v} &= \vec{0} \end{aligned}$$

The last equation above means $\vec{v} \in \ker \left[\mathbf{D}f\right]$. So $T_{\mathbf{c}}M \subset \ker \left[\mathbf{D}f\right]$, which, from *Proof 12.1*, means $\mathbf{c}$ is a critical point of f constrained to the manifold M.

($\Leftarrow$)

Assume $\mathbf{c}$ is a critical point of f constrained to M. Then (by *Proof 12.1*)

$$T_{\mathbf{c}}M \subset \ker \left[\mathbf{D}f(\mathbf{c})\right] \implies \ker \left[\mathbf{DF}\right] \subset \ker \left[\mathbf{D}f\right].$$

Since $\left[\mathbf{DF}\right]$ is onto, the rows of $\left[\mathbf{DF}\right]$ are linearly independent.

Assume for contradiction that no such Lagrange multipliers exist where $\left[\mathbf{D}f\right] = \lambda_1 \left[\mathbf{D}F_1\right] + \lambda_2 \left[\mathbf{D}F_2\right] + \cdots + \lambda_m \left[\mathbf{D}F_m\right]$. Then $\left[\mathbf{D}f\right]$ is not in the span of the rows of $\left[\mathbf{DF}\right]$.

We can create a new matrix $A = \begin{bmatrix} \mathbf{D}F_1 & \mathbf{D}F_2 & \cdots & \mathbf{D}F_m & \mathbf{D}f \end{bmatrix}^T$ which will have linearly independent rows and thus is onto. Since A is onto, $\exists \vec{v} : A\vec{v} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}^T$. So

$$\underbrace{\left[\mathbf{D}F_1\right] \vec{v} = 0, \ \left[\mathbf{D}F_2\right] \vec{v} = 0, \ \ldots, \left[\mathbf{D}F_m\right] \vec{v} = 0}_{\vec{v} \in \ker \left[\mathbf{DF}\right]}, \ \underbrace{\left[\mathbf{D}f\right] \vec{v} = 1}_{\vec{v} \notin \ker \left[\mathbf{D}f\right]}$$

Together, these imply that $\ker\left[\mathbf{DF}\right] \not\subset \ker\left[\mathbf{D}f\right]$, or $T_{\mathbf{c}}M \not\subset \ker\left[\mathbf{D}f\right]$, which contradicts our assumption that $\mathbf{c}$ is a critical point of f on M.

Therefore, if there is a critical point of a function on a manifold, the Lagrange multipliers must exist. ■

Proof 12.3: The Root Test

Consider the infinite series $\sum a_n$ and let $\alpha = \limsup |a_n|^{\frac{1}{n}}$.

Using the Comparison Test, prove the following statements about $\sum a_n$:

(i) The series converges absolutely if $\alpha < 1$

(ii) The series diverges if $\alpha > 1$

(iii) If $\alpha = 1$, then nothing can be deduced conclusively about the behavior of the series

Proof

Part (a)

Let $\limsup |a_n|^{\frac{1}{n}} = \alpha < 1$. Then $\forall \varepsilon > 0, \exists N : \forall n > N, |a_n|^{\frac{1}{n}} < \alpha + \varepsilon$, which implies $|a_n| < (\alpha + \varepsilon)^n$.

We choose ε to be small enough so that $\alpha + \varepsilon < 1$, and find N for which the above is true. Then $\sum_{n=N+1}^{\infty} (\alpha + \varepsilon)^n = (\alpha + \varepsilon)^{N+1} \sum_{n=1}^{\infty} (\alpha + \varepsilon)^n$ is a geometric series with $r = \alpha + \varepsilon < 1$, which converges.

Therefore, $\sum_{n=N+1}^{\infty} |a_n|$ also converges by the *Comparison Test*. Since $\sum_{n=1}^{N} |a_n|$ is just a constant, the infinite series $\sum_{n=1}^{\infty} |a_n|$ converges, meaning $\sum a_n$ converges absolutely.

Part (b)

Let $\limsup |a_n|^{\frac{1}{n}} = \alpha > 1$. Then $\forall \varepsilon > 0,\ \forall N,\ \exists n > N : |a_n|^{\frac{1}{n}} > \alpha - \varepsilon$.

We choose ε small enough so that $\alpha - \varepsilon > 1$. then $\forall N, \exists n > N : |a_n|^{\frac{1}{n}} > \alpha - \varepsilon > 1$. This implies that $|a_n| > 1$, so $\lim a_n \neq 0$, and therefore $\sum a_n$ diverges.

Part (c)

We will choose two series.

First, let $a_n = \frac{1}{n}$. Then $\alpha = \limsup \left(\frac{1}{n}\right)^{\frac{1}{n}} = \lim \frac{1}{n^{1/n}} = 1$, but we already showed that $\sum \frac{1}{n}$ diverges.

Second, let $a_n = \frac{1}{n^2}$. Then $\alpha = \limsup \left(\frac{1}{n^2}\right)^{\frac{1}{n}} = \lim \left(\frac{1}{n^{1/n}}\right)^2 = 1$, but we already showed that $\sum \frac{1}{n^2}$ converges.

Since we have shown that $\alpha = 1$ for series that converge or diverge, nothing can be deduced conclusively by the Root Test. ■

Proof 13.1: Taylor's Theorem With Remainder

Let f be defined on (a, b) with $a < 0 < b$. Suppose that the nth derivative $f^{(n)}$ exists on (a, b).

Define the remainder

$$R_n(x) = f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(0)}{k!} x^k$$

Prove, by repeated use of Rolle's theorem, that for each $x \neq 0$ in (a, b), there is some y between 0 and x for which

$$R_n(x) = \frac{f^{(n)}(y)}{n!} x^n$$

Proof

Fix some $x \neq 0$, and write $R_n(x) = \frac{Mx^n}{n!}$. We want to show that for some $y \in (0, x)$, $f^{(n)}(y) = M$.

We'll create a "helper function", and evaluate it at 0 and x:

$$\begin{aligned} g(t) &= \sum_{k=0}^{n-1} \frac{f^{(k)}(0)}{k!} t^k - f(t) + \frac{Mt^n}{n!} \\ g(0) &= f(0) - f(0) + 0 = 0 \\ g(x) &= \sum_{k=0}^{n-1} \frac{f^{(k)}(0)}{k!} x^k - f(x) + \frac{Mx^n}{n!} = 0 \end{aligned}$$

g is continuous on $[0, x]$ and differentiable on $(0, x)$, so by *Rolle's Theorem* $\exists x_1 \in (0, x) : g'(x_1) = 0$. We will calculate $g'(t)$ and evaluate it at 0 and x_1:

$$\begin{aligned} g'(t) &= \sum_{k=1}^{n-1} \frac{f^{(k)}(0)}{k!} kt^{k-1} - f'(t) + \frac{Mnt^{n-1}}{n!} \\ g'(0) &= f'(0) - f'(0) + 0 = 0 \\ g'(x_1) &= 0 && \text{(Given by Rolle's Theorem above)} \end{aligned}$$

We use *Rolle's Theorem* again on $g'(t)$: $\exists x_2 \in (0, x_1) : g''(x_2) = 0$.

We keep doing this until $\exists x_n \in (0, x_{n-1}) : g^{(n)}(x_n) = 0$.

Let $y = x_n$. Then

$$g^{(n)}(y) = 0 = 0 - f^{(n)}(y) + M \implies M = f^{(n)}(y) \qquad y \in (0, x)$$

So $R_n(x) = \dfrac{f^{(n)}(y)}{n!} x^n$. ■

Proof 13.2: Infinite Triangle Inequality

Starting from the triangle inequality for two vectors, prove the triangle inequality for m vectors in $\mathbb{R}^n$, then prove the "infinite triangle inequality":

$$\left|\sum_{i=1}^{\infty} \vec{a}_i\right| \leq \sum_{i=1}^{\infty} |\vec{a}_i|$$

You may assume that the series $\sum_{i=1}^{\infty} \vec{a}_i$ is "absolutely summable" (the infinite series of lengths on the right is convergent) but you must prove that this series is "summable" (infinite sum of vectors on the left is convergent).

You may use Theorems 0.5.8 (if $\sum_{n=1}^{\infty} |a_n|$ converges, then so does $\sum_{n=1}^{\infty} a_n$) and 1.5.13 (a sequence of vectors in $\mathbb{R}^n$ converges if and only if each component converges).

Proof

We can write the ith vector in our sequence $\vec{a}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix} \in \mathbb{R}^n$. So a_{ij} is the jth component of the ith vector in the series. Furthermore, $|a_{ij}| \leq |\vec{a}_i| = \sqrt{a_{i1}^2 + a_{i2}^2 + \cdots + a_{in}^2}$.

By the *Comparison Test*, given that $\sum_{i=1}^{\infty} |\vec{a}_i|$ converges, then $\sum_{i=1}^{\infty} |a_{ij}|$ also converges for any component j. And since absolute convergence implies convergence, then $\sum_{i=1}^{\infty} a_{ij}$ converges. And given that if a series of vector components converges, then the series of vectors converges. So $\sum_{i=1}^{\infty} \vec{a}_i$ converges to some vector $\vec{s}$.

For some finite N, we want to show that $\left|\sum_{i=1}^{N} \vec{a}_i\right| \leq \sum_{i=1}^{N} |\vec{a}_i|$. We will prove by induction.

Base case (n = 2): $|\vec{a}_1 + \vec{a}_2| \leq |\vec{a}_1| + |\vec{a}_2|$ is true by the triangle inequality (*Proof 2.1b*).

Induction step:

$$\left|\sum_{i=1}^{N+1} \vec{a}_i\right| = \left|\sum_{i=1}^{N} \vec{a}_i + \vec{a}_{N+1}\right| \leq \left|\sum_{i=1}^{N} \vec{a}_i\right| + |\vec{a}_{N+1}| \leq \sum_{i=1}^{N} |\vec{a}_i| + |\vec{a}_{N+1}| = \sum_{i=1}^{N+1} |\vec{a}_i|$$

$s_N = \sum_{i=1}^{N} |\vec{a}_i|$ is a sequence of partial sums which converges to some number $\sum_{i=1}^{\infty} |\vec{a}_i|$ (which is given to exist). So we have

$$\begin{array}{ccc} \left|\sum_{i=1}^{N} \vec{a}_i\right| & \leq \sum_{i=1}^{N} |\vec{a}_i| & \leq \sum_{i=1}^{\infty} |\vec{a}_i| \\ t_N & \leq s_N & \leq s \end{array}$$

Since all of the above are nonnegative, $t_N \in [0, s]$ for all N. So $\lim_{N\to\infty} t_N \in [0, s]$, and

$$\lim_{N\to\infty} t_N = \left|\sum_{i=1}^{\infty} \vec{a}_i\right| \leq \sum_{i=1}^{\infty} |\vec{a}_i| = s$$

■