Multivariable Calculus

Avery Karlin

Fall 2015

Contents

Teacher: Stern

1 Multivariable Calculus Basics

Multivariable calculus is the study of the analogs, if any, of the fudemental theorem of calculus to higher dimensions, and any restrictions that may exist on higher dimensions

Let R be a simple closed curve, where simple means that each point is crossed by the curve at most once and closed means that the curve does not have a unqiue starting point and end point.

When integrating in 2 dimensions, we also need to pick an "interval." In this case, R, the bounded region, would be an interval (analogous to (a, b) in single variable calculus) and C would be the boundary (analogous to a, b).

If we were to integrate a function $f(x, y)$ over R, it is denoted by:

$$\iint_R f(x,y)dA_{xy}$$

The dA_{xy} is the area element, an infinitesimally small piece of area, analogous to dx, the length element in single variable calculus.

Note that in single variable calculus, there is an implied orientation, going left to right is the positive "direction," In multivariable calculus, it is accepted that the positive direction for the curve to go in is the counterclockwise direction, such that the opposite direction adds a negative.

1.1 Green's Theorem

This is one of the FTC's generalization to higher dimensions. The Green's Theorem works with functions that take in 2 variables.

Suppose there exists $f(x, y)$ and $g(x, y)$, and a region R bounded by a positively oriented, simple closed curve C. Then Green's theorem states that:

$$\iint_R (\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y})dA_{xy} = \int_C f dx + g dy = \int_C (fx'(t) + gy'(t))dt$$

1.2 Generalized FTC

The goal of this course in multivariable calculus is to reach the following conclusion:

For some function ω evaluated over the region M, where ∂M is the boundry of M:

$$\int_M d\omega = \int_{\partial M} \omega$$

2 Real Number Set

2.1 Definitions of Number Structures

2.1.1 $\mathbb{N}$

We can define the natural number system by sets, like the following:

$$0 = \emptyset$$

And from there we introduce a succession operation:

$$n + 1 = n \cup \{n\}$$

So for example, $1 = \{0\} = \{\emptyset\}$, $1 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}$, etc.

Set theory is the deepest concept in mathematics, such that it must be assumed as postulates.

2.1.2 $\mathbb{Z}$

Positive integers are defined as an ordered pair of natural numbers, for example, $2 = (2, 0)$, and $-2 = (0, 2)$

2.1.3 $\mathbb{Q}$

Rationals are defined as an infinite set of ordered pairs of integers. A rational number $q = \frac{n}{m}$, then $q = \{(n, m), (2n, 2m), (3n, 3m) \ldots (-2n, -2m), (-3, -3m) \ldots\}$

2.1.4 $\mathbb{R}$

There are several definitions of the real number system

- Each real number can be thouhgt of as an infinite sequence in the following format:

  $$(s, N, d_1, d_2, d_3 \ldots)$$

  Where $s = \pm 1$, $N \in \mathbb{N}$, and $d_n \in 0, 1, 2, 3, \ldots 8, 9$. It is also not the case that $d_n = d_{n+1} = \cdots = 0$. If there is a terminal decimal, we express it as 9 repeated.

- Each real number forms a subdivision of $\mathbb{Q}$ into two disjoint sets that cover the entirety of $\mathbb{Q}$, one of which lies entirely to the left of the other.

2.2 Basic Structure of $\mathbb{R}^1$

$\mathbb{R}$ is an instance of many kinds of mathematical structures, such as:

- Field
  - A set closed under addition and multiplication (results are within set)
  - Obeys all associated laws of addition and multiplication
- Ordered Set
  - The set has an ordering that reflect the operations in the field (such that x, y ¿ 0, then x + y and xy ¿ 0)
  - This structure is what allows for comparisons, like the $<$ function
- Metric Space
  - There exists a standard distance operation between numbers
  - In $\mathbb{R}$, $dist(x, y) := |x - y|$ (:= means "is defined as")
- Vector Space
  - Elements can be thought of as vectors from the origin
- Geometric Space
  - This structure means that you can measure angle in a meaningful way
  - In $\mathbb{R}^1$, the angle measure can be either 0 or π
  - In higher dimensions there are more angles possible

2.3 Properties of $\mathbb{R}^n$

In higher dimensions, several of the properties of $\mathbb{R}$ no longer hold. $\mathbb{R}^n, n > 1$ is not a field, and not an ordered set, but it is a vector, metric, and geometric space.

2.4 Basic Axioms for $\mathbb{R}^1$

$\mathbb{R}$ is a field under + and $\cdot$ $(x, y \in \mathbb{R})$

1. Additive closure: $x + y \in \mathbb{R}$
2. Associative Property of Addition: $x + (y + z) = (x + y) + z$
3. Communicative Property of Addition: $x + y = y + x$
4. 0 is the identity element of addition: $x + 0 = x$
5. Every element has an additive inverse: $x + (-x) = 0$
6. Multiplicative closure: $xy \in \mathbb{R}$
7. Associative Property of Multiplication: $x(yz) = (xy)z$

8. Communicative Property of Multiplication: $xy = yx$
9. 1 is the identity element of multiplication: $x(1) = x$
10. Every element (except 0) has a multiplicative inverse: $x \cdot \frac{1}{x} = 1$
    - Theorem: $x(0) = 0$
    - Proof: $x(0) = x(0+0)$, then we apply the distributive law, and get $x(0) = x(0) + x(0)$. Now we add $-x(0)$ to both side, and we get: $0 = x(0)$
11. Distributive Law: $x(y+z) = xy + xz$

$\mathbb{R}$ is an ordered field and has a proper subset (aka not the entire set) $\mathbb{R}^+$ (the positives) such that:

1. $\mathbb{R}^+$ is closed under $+$ and $\cdot$
2. $1 \in \mathbb{R}^+, 0 \notin \mathbb{R}^+$
3. **Trichotomy Property**: for any $x \in \mathbb{R}$, x is either 0, $\in \mathbb{R}^+$ or $\notin \mathbb{R}^+$
4. Definition of $<$ and $>$:
    - $x < y$ means $y - x \in \mathbb{R}^+$
    - $x > y$ means $y < x$

2.5 The Seperation Axiom

The main difference between the $\mathbb{Q}$ and the $\mathbb{R}$ is the separation axiom which the rationals do not have. If $\mathcal{A} \subseteq \mathbb{R}$ and $\mathcal{B} \subseteq \mathbb{R}$, and:

1. $\mathcal{A} \cap \mathcal{B} = \emptyset$
2. $\mathcal{A} \neq \emptyset$, $\mathcal{B} \neq \emptyset$
3. $\mathcal{A} < \mathcal{B}$ "$\mathcal{A}$ is to the left of $\mathcal{B}$"
    - $\forall$ (for all) $a \in \mathcal{A}, \forall b \in \mathcal{B}, a < b$

2.5.1 Proof of Irrationals

$$\mathcal{A} = \mathbb{Q}^- \cup \{0\} \cup \{q \in \mathbb{Q}^+ | q^2 < 2\}$$
$$\mathcal{B} = \{q \in \mathbb{Q}^+ | q^2 \geq 2\}$$

Then, $\exists$ (there exists at least 1) $c \in \mathbb{R}$ such that $\mathcal{A} \leq c \leq \mathcal{B}$

We know that $\mathcal{A} \neq \emptyset$ and $\mathcal{B} \neq \emptyset$ because 0 is in $\mathcal{A}$ and 2 is in $\mathcal{B}$. We also know that $\mathcal{A} \cup \mathcal{B} = \mathbb{Q}$ and $\mathcal{A} \cap \mathcal{B} = \emptyset$. As well as the fact that $\mathcal{A} < \mathcal{B}$ (all elements of A are less than all elements of B.

Now, if we want to find the boundary element, q_0, which separates $\mathcal{A}$ and $\mathcal{B}$. We know that $\mathcal{A} \leq q_0 \leq \mathcal{B}$. So that must mean $q_0 = \sqrt{2}$. However, $\sqrt{2} \notin \mathbb{Q}$. Therefore, we know that the set of rational numbers do not follow the separation axiom.

3 Sequence Theorems

3.1 The Least Upper Bound Theorem

If $\mathcal{A} \subseteq \mathbb{R}$ is non-empty, and is bounded above (So $\exists b_1 \in \mathbb{R}$ such that $\mathcal{A} < b_1$), then $\mathcal{A}$ has a least upper bound, i.e. a number $b_0 \in \mathbb{R}$ such that $\mathcal{A} \leq b_0$ and for any b with $\mathcal{A} \leq b$, $b_0 \leq b$

$$\mathcal{A} \subseteq \mathbb{R}, \mathcal{A} \neq \emptyset, (\exists\ b_1 \in \mathbb{R} : \mathcal{A} \leq b_1) \rightarrow [\exists\ b_0 \in \mathbb{R} : \mathcal{A} < b_0, \forall b \in \mathbb{R}(\mathcal{A} \leq b \rightarrow b_0 \leq b)]$$

b_0 is known as the least possible upper bound, or the *supremum* of $\mathcal{A}$, we write $b_0 = \sup \mathcal{A}$.

Similarly, for any non-empty set $\mathcal{A}$ bounded below, it has a greatest lower bound, $\inf \mathcal{A}$, called the *infimum* of $\mathcal{A}$

3.1.1 Proof

Define $\mathcal{B}$ to be the set of all upper bounds of $\mathcal{A}$. Let $\mathcal{C} = \mathbb{R} \setminus \mathcal{B}$. Clearly $\mathcal{B}$ is nonempty; also $\mathcal{C}$ is non-empty because it contains $x_0 - 1$, where $x_0 \in \mathcal{A}$. By the way in which we defined $\mathcal{C}$, $\mathcal{B} \cap \mathcal{C} = \emptyset$. Pick any $c \in \mathcal{C}$ and $b \in \mathcal{B}$. By the definition of $\mathcal{C}$, $\exists\ x_1 \in \mathcal{A} : c < x_1$. But $x_1 \leq b$ by the definition of $\mathcal{B}$. Therefore $\mathcal{C} < \mathcal{B}$. By the separation postulate, $\exists\ b_0 \in \mathbb{R} : \mathcal{C} \leq b_0 \leq \mathcal{B}$. Note that $\mathcal{A} \setminus \{b_0\} \subseteq \mathcal{C}$. Thus, b_0 is an upper bound for $\mathcal{A}$. Morever, it is the least upper bound because $b_0 \leq \mathcal{B}$.

3.2 Bounded Monotone Sequence Theorem

For any sequence $\{a_n\}$

1. If $a_n \leq a_{n+1}$ for all $n \geq 1$, and $\exists\ b \in \mathbb{R}$ such that $a_n \leq b$ for all $n \geq 1$, then $\lim_{n\to\infty} a_n$ exists and is less than or equal to b
2. If $a_n \geq a_{n+1}$ for all $n \geq 1$, and $\exists\ b \in \mathbb{R}$ such that $a_n \geq b$ for all $n \geq 1$, then $\lim_{n\to\infty} a_n$ exists and is greater than or equal to b

3.2.1 Proof

We first convert the sequence $\{a_n\}$, which is bounded by b into the set $\mathcal{A} = \{a_n | n \geq 1\}$. We know that $\mathcal{A} \neq \emptyset$ because the sequence has some terms. We also know that $\mathcal{A}$ is bounded above by b; $\mathcal{A} < b$ By the Least Upper Bound Theorem, $\exists\ b_0 = sup\mathcal{A}$. We now show that $b_0 = \lim_{n\to\infty} a_n$. By the definition of limits, to say $b_0 = \lim_{n\to\infty} a_n$ means to say $\forall \epsilon > 0, \exists\ N > 0, \forall n \geq N$, $|a_n - b_0| < \epsilon$

If we look at the number $b_0 - \epsilon$, it is not an upper bound on $\mathcal{A}$ because b_0 is the least upper bound and $\epsilon > 0$. Therefore, $\exists\ a_N > b_0 - \epsilon$. Since $\{a_n\}$ is increasing, $\forall n > N$, $a_n > b_0 - \epsilon$. If we rearrange the terms, we get $b_0 - a_n < \epsilon$. Therefore, b_0 (which exists by the least upper bound theorem) is the limit of a_n as $n \to \infty$.

3.3 Archimedean Property

For any positive numbers x and y, it is possible to find some $n \in \mathbb{N}$ such that $nx > y$.

$$\forall x, y > 0, \exists\, n \in \mathbb{N} : nx > y$$

3.3.1 Proof

Assume that $\neg\exists\, n : nx > y$, this is logically equivalent to $\forall n : nx \leq y$. Let $\mathcal{C} = \{nx | n \in \mathbb{N}\}$. Then $\mathcal{C} \leq y$, let $c = \sup \mathcal{C}$. We claim that $\exists N : c - \frac{1}{2}x < Nx \leq c$. This is true because if such N does not exist, then $c - \frac{1}{2}x$ would be an upper bound, but c is the least upper bound, so such N must exist. Now we've established the existence of N, let us consider $(N+1)x$. $(N+1)x = Nx + x > (c - \frac{1}{2}x) + x = c + \frac{1}{2}x > c$. But $(N+1)x \in \mathcal{C}$, so it should be $< c$. We have a contradiction. This shows that the original assumption is false, so $\forall x, y > 0, \exists n \in \mathbb{N} : nx > y$

3.3.2 Consequences

This property can be used to show that $\lim_{n\to\infty} \frac{1}{n} = 0$. If we consider the definition of limits, the statement is equivalent to saying that $\forall \epsilon \in \mathbb{R} > 0, \exists N \in \mathbb{N}, \forall n > N, n \in \mathbb{R}, \frac{1}{n} < \epsilon$. If we rearrange the term, we get that we need to show $1 < \epsilon N$ for any ϵ. This is true because of the Archimedean Property. $\forall n > N$, since $\epsilon > 0$, $1 < \epsilon N < \epsilon n$. Therefore we know the limit is truely 0.

3.4 Sunrise Lemma

The Sunrise Lemma states for any sequence $(a_n)_n^\infty = 1$ in $\mathbb{R}$, $\exists$ monotone subsequence $(a_{n_k})_{k=1}^\infty$, where $(n_k)_{k=1}^\infty$ is a strictly increasing sequence in $\mathbb{N}$ and $n_k \geq k$ for all $k \in \mathbb{N}$

Vistas are points in a sequence, a_n, where $N \in \mathbb{N}$, such that $a_N > a_n$ for all $n > N$

This means that for any sequence, there exists a subset of points within, such that within that sequence, the sequence is monotone

3.4.1 Well-Ordering Property

For any set $A \subseteq \mathbb{N}, A \neq \emptyset, min(A)$ exists

3.4.2 Proof

<u>Case I:</u> The set V of vistas, is infinite, such that $n_1 = min(v)$ and $n_k = min(V \cap n_{k-1}^\infty$, where $k \geq 2$, then a_{n_k} is strictly decreasing
<u>Case II:</u>

$$n_1 = \begin{cases} 1 & if v = \emptyset \\ 1 + max(v) & if v \neq \emptyset \end{cases}$$

$n_k = choice\{n > n_{k-1} | a_n \geq a_{n_{k-1}}$, thus $n_k \neq \emptyset$ because V is finite, thus a_{n_k} is increasing

3.5 Bolzano-Weierstrass Theorem

Every bounded sequence in $\mathbb{R}$ has at least one convergent subsequence

3.5.1 Proof

Let $(a_n)_{n=1}^{\infty}$ be a bounded sequence. For any monotone sequence $(a_{n_k})_{k=1}^{\infty}$, then $(a_{n_k})_{k=1}^{\infty}$ is both bounded and monotone, so it converges.

4 Extreme Value Theorem

For some function $f : [a,b] \to \mathbb{R}$ (for the codomain, not the range) that is continuous ($f(x_0) = \lim_{x \to x_o} f(x)$ for any x $\in (a,b)$, $f(a) = \lim_{x \to a^+} f(x)$, $f(b) = \lim_{x \to b^-} f(x)$), then $\exists c, d \in [a,b]$ such that $f(c) \leq f(x) \leq f(d)$.

4.1 Proof

4.1.1 $\exists M > 0$ such that $f(x) \leq < M \forall x \in [a,b]$ (f(x) is bounded above)

Lets assume that f is not bounded above, such that for any $n \in \mathbb{N}, \exists x_n \in [a,b]$ such that $f(x_n) > n$. The sequence $(x_n)_{n=1}^{\infty}$ is bounded between a and b, such that it has a convergent subsequence $(x_{n_k})_{k=1}^{\infty}$ converging to some point t, when $\lim_{k \to \infty}(x_{n_k})$, by the below claim.

Claim: $t \in [a,b]$. Let $t > b$, then $\epsilon > 0$ such that $[a,b] \cap (t-\epsilon, t+\epsilon) = \emptyset$. But $\exists N$ such that $x_N \in (t-\epsilon, t+\epsilon)$ and $x_N \in [a,b]$, which is a contradiction.

By the previous claim, $lim_{k \to \infty} f(x_{n_k}) = f(t)$ by the assumed continuity of f. Thus, $\exists$ K such that for all $k \geq K, f(x_{n_k}) < f(t) + 1 \in \mathbb{R}$. On the other hand, $f(x_{n_k}) > n_k > f(t) + 1$ when k is sufficiently large. Thus, there is a contradiction, and f(x) is bounded from above. This can be reversed to show it is bounded from below, as well.

4.1.2 The function reaches the supremum and infimum

We now know $R := f([a,b]) = \{f(x) | x \in [a,b]\}$ is bounded, such that S := sup(R) and I := inf(R). By the definition of infimum and supremum, $\exists (y_n)_{n=1}^{\infty}$ such that $y_n \in R \forall n$, and $\lim_{n \to \infty} y_n = S$. Since $y_n \in R, \exists x_n [a,b]$ such that $f(x_n) = y_n$. Now $(x_n)_{n=1}^{\infty}$ is bounded between a and b so it has a convergent subsequence, $(x_{n_k})_{k=1}^{\infty}$, converging to $t \in [a,b]$. Also, by continuity of f, $\lim_{k \to \infty} f(x_{n_k}) = f(t)$. Thus, f(t) = S. This can be reversed to apply to the infimum.

5 Higher Dimensional Mathematics

5.1 Higher Dimensions

5.1.1 Cartesian Product

The Cartesian Plane represents the set $\mathbb{R}^2 := \{(x,y) + | + x, y \in (R)\}$. This is known as the **Cartesian Product** of $\mathbb{R}$ with itself. The Cartesian Product of two sets $\mathcal{S}$ and $\mathcal{T}$, $\mathcal{S} \times \mathcal{T} := \{(s,t) + | + s \in \mathcal{S}, t \in \mathcal{T}\}$.

5.1.2 Shapes

- Open-ball: $B_r(P) := \{X + | + dist(X,P) < r\}$
- Closed-ball: $\bar{B}_r := \{X + | + dist(X,P) \leq r\}$
- Sphere: $S_r := \{X + | + dist(X,P) = r\}$

5.1.3 Boundary

Given $D \subseteq \mathbb{R}^2, D \neq \emptyset$. We say $(a,b) \in \partial D$, i.e. (a,b) is on a boundry point of D if $\forall \epsilon > 0$, there are points $(x,y) \in D$ and $(u,v) \in D^c$ ($D^c := \mathbb{R}^2 \setminus D$) such that $dist(x,y+;+a,b) < \epsilon$ and $dist(u,v+;+a,b) < \epsilon$. Since the definition is symmetrical, $\partial D^c = \partial D$.

5.1.4 Interior

Let $D \subseteq \mathbb{R}^2$. We say (a,b) is an interior point of D if $\exists + r > 0 : B_r(a,b) \subseteq D$. The set of all interior points is called the interior of D and is written as int D.

5.1.5 Exterior

Let $D \subseteq \mathbb{R}^2$. We say (a,b) is an exterior point for D if it is an interior point of D^c. $\exists r > 0 : B_r(a,b) \subseteq D^c$. The set of all exterior points for D is the exterior of D, written as ext D.

Thm: For any $D \subseteq \mathbb{R}^2$, $\mathbb{R}^2 = \text{int } D \cup \partial D \cup \text{ext } D$. And $\text{int } D \cap \partial D = \emptyset$, $\text{int } D \cap \text{ext } D = \emptyset$, $\partial D \cap \text{ext } D = \emptyset$.

Thm: $\text{int } D = \text{ext } D^c$ and $\text{ext } D = \text{int } D^c$

Thm: $\text{ext } D \subseteq D^c$

5.1.6 Closure

The **closure** of $D \subseteq \mathbb{R}^2$:

$$\bar{D} := D \cup \partial D$$

Thm: $\partial \bar{D} = \partial D$, $\text{int } \bar{D} = \text{int } D$, and $\text{ext } \bar{D} = \text{ext } D$

Thm: int $D \subseteq D \subseteq \bar{D}$.

5.1.7 Ordered Pair

The ordered pair (a, b) can be thought of as a set, but a set is inheritly unordered. To express the order, we can do the following: $(a, b) = \{\{a\}, \{a, b\}\}$. Now we know that a is the first element because it appears in both subsets. We can then expand this into higher dimensions like the following: $(a, b, c) = ((a, b), c)$. Note that this means that $((a, b), c) \neq (a, (b, c))$. But this does not matter to us.

Fundamental Postulate of Ordered Pairs: $(a_1, a_2, a_3, \ldots, a_n) = (b_1, b_2, b_3, \ldots, b_n)$ if and only if $a_1 = b_1 \wedge a_2 = b_2 \wedge \cdots \wedge a_n = b_n$.

5.1.8 Vector and Points

Vectors are quantities of directionality and length, its location does not matter. Points are just positions in space. In higher dimensions with no ambiance space (flat space surrounding the surface, i.e. the shortest distance in the ambiant space is the straight line), we define a vector as all the lines with the same direction at a certain point.

However, the nice thing about $\mathbb{R}^d$ is that there is always ambiance space, so we will not make any notational distinction between a point and a vector.

The length of a vector in d space is defined as:

$$||\vec{a}|| := dist(\vec{0}, \vec{a}) = \sqrt{\sum_{i=1}^{d} a_i^2}$$

5.1.9 Space and Lines

$\mathbb{R} = \{(x_1, x_2, ..., x_d)|x_1, x_2, ...x_d \in \mathbb{R}\}$ A line, l, can be defined such that l $= \{(a_1 + tb_1, a_2 + tb_2, ..., a_d, tb_d)|t \in \mathbb{R}\}$

5.2 Dot Product

5.2.1 Definition and Perpendicularity

The dot product arises naturally through the idea of geometric distance, such that if $a \neq \emptyset, b \neq \emptyset$, then $a \perp bi$ iff $dist(a;b)^2 = dist(\emptyset, a)^2 + dist(\emptyset, b)^2$, where $a = (a_1, a_2), b = (b_1, b_2)$. Thus, by expanding out, $a \perp b$ iff $a_1b_1 + a_2b_2 = 0$ where $a \neq \emptyset, b \neq \emptyset$. In addition, orthoganal refers to both perpendicular vectors and where $a = \emptyset$ and/or $b = \emptyset$, so that no vector can be perpendicular to itself.

By extension, in $\mathbb{R}$, $a\dot{b} = a_1b_1 + a_2b_2 + a_3b_3 + ... + a_db_d$, such that it forms a scalar, rather than a vector.

5.2.2 Properties

The dot product is:

- Commutative
- Distributive over Vector Sums

5.3 Functions in Higher Dimensions

5.3.1 Domain, Range, and One-to-One

The domain is a subset of $\mathbb{R}^{\ltimes}$. Let the function of f(x, y) be an ordered pair within some curve, such that $(x, y) \in D$. Thus, the range of f, G = $x, y, f(x, y)|(x, y) \in D \subseteq R^{n+1}$. Functions are defined as one-to-one if for f(x, y), $(x, y), (u, v) \in D, f(x, y) = f(u, v)$ iff (if and only if) x = u $\wedge$ (and) y = v (such that for every z value, there is only one point that will create it.

5.3.2 Bolzano-Weirstrauss in Higher Dimensions

<u>Theorem:</u> A bounded sequence $\in \mathbb{R}$ has a convergent subsequence.

<u>Lemma:</u> If $(x_n)_{n=1}^{\infty}$ converges to $x \in \mathbb{R}$, then every subsequence $(x_{n_k})_{k=1}^{\infty}$ also converges to x, following from the definition of convergence and limits ($\forall \epsilon > 0, \exists N, \forall n \geq N : |x_n - x| < \epsilon$, thus $\exists K, \forall k \geq K : n_k \geq N$, since $n_k \to \infty$ as $k \to \infty$, such that $|x_{n_k} - x| < \epsilon$).

<u>Proof for $d = 2$:</u> Let $P_n = (x_n, y_n)$. Consider $(x_n)_{n=1}^{\infty} \in \mathbb{R}$. Since $\forall x_n, -M \leq x_n \leq M, (x_n)_{n=1}^{\infty}$ is bounded. For some $(x_{n_k})_{k=1}^{\infty}$, converging to some $x \in \mathbb{R}$. Consider $(y_{n_k})_{k=1}^{\infty}$ is bounded by the same rationale, thus $(y_{n_{k_j}})_{j=1}^{\infty}$ converges to some $y \in \mathbb{R}$. Since $(x_{n_{k_j}})_{j=1}^{\infty}$ is a subsequence of a converging sequence, it converges to the same value, x. Thus, $P_{n_{k_j}} = (x_{n_{k_j}}, y_{n_{k_j}}) \to (x, y) = P$.

5.3.3 Cauchy Sequence

A cauchy sequence in $\mathbb{R}$ is a sequence $(x_n)_{n=1}^{\infty}$ such that: $\forall \epsilon > 0, \exists N, \forall n, m \geq N : |x_n - x_m| < \epsilon$. This defines a sequence where as $n \to \infty$, the distance between values of points on the sequence decreases.

5.3.4 Convergence as Cauchy

Note that any convergent sequence is cauchy, because as terms get together to a limit, they also go very closely together. <u>Proof:</u> Let $(x_n)_{n=1}^{\infty}$ be convergent, with limit $x \in \mathbb{R}$. Then, by definition, $\forall \epsilon > 0, \exists N_\epsilon, \forall n \geq N_\epsilon : |x_n - x| < \epsilon$. Note that we can replace ϵ with $\frac{\epsilon}{2}$, all we have to change is the cutoff point from N_ϵ to $N_{\frac{\epsilon}{2}}$. Now if we take two subscripts $n, m \geq N_{\frac{\epsilon}{2}} \longrightarrow |x_n - x_m| = |(x_n - x) + (x - x_m)| \leq |x_n - x| + |x_m - x|$ because of the Triangle Inequality for Absolute Values. However, note that $|x_n - x| \leq \frac{\epsilon}{2}$ and $|x_m - x| \leq \frac{\epsilon}{2}$. Therefore, $|x_n - x_m| \leq |x_n - x| + |x_m - x| \leq \epsilon$.

5.3.5 Cauchy's Convergence Theorem

In $\mathbb{R}$, every cauchy sequence converges to a limit in $\mathbb{R}$.

Lemma #1: Every cauchy sequence is bounded.

Let us take $\epsilon = 1$, then the definition of "cauchiness" becomes:

$$\exists + N_1, \forall n, m \geq N_1 : |x_n - x_m| < 1$$

Let $M := \max\{|x_1|, |x_2|, \ldots, |x_{N_1-1}|, |x_{N_1}|+1\}$. We claim that $|x_n| \leq M$, for all $n \geq 1$. This is true because when $n \in \{1, 2, \ldots, N_1 - 1\}$, the statement is true by definition of M. When $n \geq N_1$, we know that $|x_n| \leq |x_{N_1}|+1 \leq M$ because we can let $m = N_1$, then by the definition of "cauchiness," we know that $|x_n - x_{N_1}| < 1$.

Now we see that M is a bound on the sequence for all $n \geq 1$. Therefore the sequence is bounded.

Lemma #2: If a subsequence of a cauchy sequence converges to $x \in \mathbb{R}$, the whole sequence must converge to x.

Say $(x_n)_{n=1}^{\infty}$ is cauchy, and $(x_{n_k})_{k=1}^{\infty}$ converges to x. For any arbitrary $\epsilon > 0$, we try to find N such that $\forall n \geq N : |x_n - x| < \epsilon$. If we prove the existence of N for all ϵ, we will have proven that the original sequence converges.

We know that $\forall \epsilon > 0, \exists K_\epsilon, \forall k \geq K_\epsilon : |x_{n_k} - x| < \epsilon$. We add and subtract x_{n_k} and group terms, and use the Triangle Inequality: $|x_n - x| = |(x_n - x_{n_k}) + (x_{n_k} - x)| \leq |x_n - x_{n_k}| + |x_{n_k} - x|$. Note that $|x_{n_k} - x| < \epsilon$ provided $k \geq K_\epsilon$ from the convergent subsequence condition. We also know that $|x_n - x_{n_k}| < \epsilon$ provided that $n, n_k \geq N_\epsilon$, which we call the "cauchy cutoff." This is true from the "cauchiness" condition.

We know that $k \to \infty$ implies $n_k \to \infty$. This means eventually $n_k > N_\epsilon$ provided that $k > L_{N_\epsilon}$. Now let $k = \max\{L_{N_\epsilon}, K_\epsilon\}$ and $n \geq N_\epsilon$, which implies $|x_n - x_{n_k}| < \epsilon$ and $|x_{n_k} - x| < \epsilon$. Now we know: $|x_n - x| \leq 2\epsilon$ provided $n \geq N_\epsilon$. Therefore the cauchy sequence converges to x.

With these two lemmas, the theorem becomes very easy to prove:

Because of Lemma #1 and the Bolzano-Weierstrass Theorem, we know that for all cauchy sequences, there is a bounded subsequence that converges to some value x. Then by Lemma #2, we know that the entire cauchy sequence converges to x as well, therefore the sequence converges.

5.3.6 Cauchy Sequences in Higher Dimensions

$(P_n)_{n=1}^{\infty}$ is cauchy if $\forall \epsilon > 0, \exists N_\epsilon, \forall n, m \geq N_\epsilon : dist(P_n, P_m) < \epsilon$. This is easy to prove due to the coordinate nature of $\mathbb{R}^d$.

5.4 Metric Topology in $\mathbb{R}^{\ltimes}$

5.4.1 Continuity

Let $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}^2$, $D \neq \emptyset$. Let $(a, b) \in D$. We say that f is continuous at (a, b) if:

$$f(a,b) = \lim_{\substack{(x,y)\to(a,b)\\(x,y)\in D}} f(x,y)$$

Or in other terms:

$$\forall\epsilon > 0, \exists + \delta > 0, \forall(x, y) \in D : dist(x, y+; +a, b) < \delta \to |f(x, y) - f(a, b)| < \epsilon$$

$f : D \to \mathbb{R}, D \subseteq \mathbb{R}^{\nvDash}, D \neq \emptyset$ is continuous if f is continuous at $(a, b)\forall(a, b) \in D$.

5.4.2 Directional Limits

Let $D \subseteq \mathbb{R} and a \in D \cup \partial D$ (the boundry, both already included in D, and not), $\lim_{x\to a^+} f(x) = L$ means $\forall\epsilon > 0, \exists\delta(\epsilon) > 0 : \forall x \in D \cap (a,\infty), |x - a| < \delta(\epsilon) \Rightarrow |f(x) - L| < \epsilon$. The limit only exists if both directions equal the same value.

5.4.3 Limits in Higher Dimensions

The same theory can be applied to higher dimensions, such that if two arbitrary approaches are not the same, it doesn't exist, but if several approaches yield the same result, the definition of a limit is used. Polar coordinate substitutions can be used to give format to directions of approach.

Due to difficulty defining approaching through lines, it is said that $(x, y) \to (a, b) iff dist(x, y : a, b) \to 0$.

Let $f : D \to \mathbb{R}, D \subseteq \mathbb{R}^{\nvDash}, D \neq \emptyset, and(a, b) \in D \cup \delta D$. Then, $L = \lim_{(x,y)\in D\to(a,b)} f(x) iff \forall\epsilon > 0, \exists\delta > 0, \forall(x, y) \in D : dist(x, y : a, b) < \delta \to |f(x, y) - L| < \epsilon$. As a corollary, when (a, b) is on the boundry, the approach can only be from the domain.

5.5 Properties of Domain

For the extreme value theorem to apply to a domain, the set must be compact, such that it must be bounded and closed over limits. On the $\mathbb{R}$ dimension, this applies to all closed intervals, as well as the empty set, though functions except the empty set cannot accept it as a domain.

5.5.1 Bounded

$If D \subseteq (R^2)$ is bounded if $\exists M > 0 : D \subseteq [-M, M]x[-M, M]$. Thus, a sequence is considered bounded if the set of all values within the sequence is bounded.

5.5.2 Closed

The term closed is used to apply to sets which are closed under limits. On $mathbbR$, if $x_n \in [a,b] for \forall n \in mathbbN, and x_n \to x \in \mathbb{R}, then x \in [a,b]$.

$D \subseteq \mathbb{R}^{\not\models} is closed iff for any points (x_n, y_n) \in D (for all n \in \mathbb{N}, if (x_n, y_n) \to (x,y) \in \mathbb{R}^{\not\models}, then (x,y) \in D. (x_n, y_n) \to (a,b) as n \to \infty$ means $d_n = \sqrt[2]{(x_n - a)^2 + (y_n - b)^2} \to 0 as n \to \infty$.

This applies the definition of limits to sequences, such that $(x_n, y_n) \to (a,b) if dist(x_n, y_n : a, b) \to 0 as n \to \infty$.

Thus, $D \subseteq \mathbb{R}^{\not\models}$ is closed if for any sequence $((x_n, y_n))_{n=1}^{\infty}$ in D that converges, the limit poiint (a, b) of the sequence also lies in D.

5.5.3 Open Set Theorem

$D \subseteq \mathbb{R}^{\not\models}$ is open if $\forall (a,b) \in D, \exists r > 0 :$ the disk of radius r, $B_r(a,b) \subseteq D$, such that D = Interior of D

It follows that for any open set, the complement set within the space is a closed set.

Proof: Assume D is open, we prove D^c is closed. Choose any convergent sequence $((x_n, y_n))_{n=1}^{\infty}$, converging to (a,b), where $(x_n, y_n) \in D^c$ for all $n \geq 1$. We prove this by contradiction. Assume that $(a,b) \in D$. Since D is open, $\exists + r > 0 : B_r(a,b) \subseteq D$. $\exists + N, \forall n \geq \mathbb{N}, (x_n, y_n) \in B_r(a,b)$ since $(x_n, y_n) \to (a,b)$. But we assumed that $(x_n, y_n) \in D^c$, and $(x_n, y_n) \in D$. But $D \cap D^c = \emptyset$. Therefore D^c is closed under taking limits.

For the other direction, we can pick some point within D, then assume there is no ball, such that any ball contains some ball not in D, even as radius $\to 0$, creating a sequence of points converging on the point, P, a contradiction.

Theorem: An open-ball $B_r(\vec{p}) = \{\vec{x} \in \mathbb{R} + | + dist(\vec{x}, \vec{p}) < r\}$, $r > 0$, is an open set.

Proof: $\forall \vec{q} \in B_r(\vec{p})$, $B_\epsilon(\vec{q}) \subseteq B_r(\vec{p})$, $\epsilon > 0$.

$d = dist(\vec{p}, \vec{q}) < r$. THerefore, $r - d > 0$. Take $\epsilon = \frac{1}{2}(r - d) > 0$. Let $\vec{x} = B_\epsilon(\vec{q})$, show $\vec{x} \in B_r(\vec{p})$.

$dist(\vec{x}, \vec{q}) < \epsilon$

$dist(\vec{x}, \vec{p}) \leq dist(\vec{x}, \vec{q}) + dist(\vec{q}, \vec{p}) < \epsilon + d = d + \frac{1}{2}r - \frac{1}{2}d = \frac{1}{2}r + \frac{1}{2}d < \frac{1}{2} \cdot 2 \cdot r = r$

Theorem: The union of any number of open sets is an open set.

Proof: Let U = $u_1 \cup u_2 \cup \cdots \cup u_n$, where u_n is an open set. We know that $\forall \vec{p} \in U, \vec{p} \in u_i$, which means $\exists r > 0 : B_r(\vec{p}) \in u_i \in U.$. Therefore U is an open set.

5.6 Extreme-Value Theorem

Let $f : D \to \mathbb{R}$ be continuous, where D $\subseteq \mathbb{R}$ is compact. Then $\exists P, Q \in D$, which do not need to be unique, such that $\forall X \in D : f(P) \leq f(x) \leq f(Q)$.

Proof: Assume that f is not bounded above, such that $\forall n \geq 1, f(P_n) > n$, where $P_n \in D$. For some subsequence $P_{n_k} \in D$ converging to P by the Balzano-Weirstrauss, by closure of D, $P \in D$. This is a contradiction since $f(P_{n_k}) \to \infty$ and $\to f(P)$, such that it must be bounded from above.

Thus, $\exists M = sup_{x \in D} f(x)$. We can find $P_n \in D$, with $f(P_n) \to M$. For some convergent subsequence $P_{n_k} \in D, P_{n_k} \to Q$.

6 Distance Functions

In axiomatic geometry, certain axioms including the definition of euclidean distance are taken as assumed. In actuality, standard distance functions must qualify under several non-geometric requirements, of which only the Euclidean distance qualifies.

Distance functions must be translation-iniant, or for any translation of two points, the distance must remain the same, such that $T_{h,k} : (x, y) \mapsto (mapsto)(x+h, y+k), then dist(x+h, y+k; x'+h, y'+k) = dist(x, y; x', y')$.

Thus, $dist(x, y; \tilde(x), \tilde(y)) = f(|x-\tilde(x)|, |y-\tilde(y)|)$, where f the distance function defined on $[0, \infty)x[0, \infty)$. As a result, it must be symmetrical, such that $dist(x, y; \tilde(x), \tilde(y)) = dist(\tilde(x), \tilde(y); x, y)$.

In addition, it must have basic reflection symmetry (isotropy), such that $dist(x, y; 0, 0) = dist(y, x; 0, 0)$. Thus, f(u, v) = f(v, u) for any $u \geq 0, v \geq 0$. It must also have the self-distance of (0, 0), such that dist(0,0; 0,0) = 0.

It must recreate the standard distance function on each axis, such that $dist(x, 0; \tilde(x), 0) = |x - \tilde(x)|, dist(0, y; 0, \tilde(y)) = |y - \tilde(y)|. Therefore, f(u, 0) = u, f(0, v) = v \forall u \geq 0, v \geq 0$.

As a result, it must have asymptotic flatness, where if a line is drawn to (x, y), where y is fixed, such that $dist(0, 0; 0, y) = v_0, while dist(0, 0; x, 0) = u. Then, \lim_{u \to \infty} f(u, v_0)/u = 1$. This also applies in the opposite direction, where x is fixed.

It must be continuous in its variables, such that with a minute movement of a point, the distance changes minutely as well.

The set of isometries (distance preserving one-to-one functions) that fix the origin onto itself (f(0) = 0) is an infinite set.

Based on these requirements, an ansatz (educated guess, verified by later results) is made, such that $f(u, v) = F(G(u) + G(v)), where F : [0, \infty) \to \mathbb{R} and G : [0, \infty) \to \mathbb{R}$. The use of G(u) and G(v) is needed to assure symmetry. The use of addition is mandated by symmetry, using addition rather than another symmetrical operation simply due to ease of calculations.

Theorem: $\exists only one suitable pair F, G; G(x) = x^2, F(x) = \sqrt{x}, that fits all requirements. If G(x) = x^n, F(x) = \sqrt[n]{x}$, it would have all required properties except infinite set of isometries.

Property: Iff $dist(p; q) = 0$, then $p = q$, where p and q are asome vector $\in \mathbb{R}$

6.1 Euclidean Distance

The distance function in one space between two points a and b is simply $|a - b|$. However, we can also write it in the following way: $\sqrt{(a - b)^2}$

In $\mathbb{R}^2$, the distance function is:

$$dist(x, y++;+a, b) := \sqrt{(x-a)^2 + (y-b)^2}$$

And in $\mathbb{R}^3$, the distance function is:

$$dist(x, y, z+;+a, b, c) := \sqrt{(x-a)^2 + (y-b)^2 + (c-z)^2}$$

The generalized form of Euclidean Distance in N space is:

$$dist(\vec{p}, \vec{q}) = \sqrt{\sum_{j=1}^{N} (p_j - q_j)^2}$$

This is known as the Euclidean Distance. We use this specific definition of distance because this is preserved under an infinite set of rigid or isometric motions, such as rotation, reflection, translation, etc.

6.2 Geometric Distance

Basic Transformations:

- $T_h : x \mapsto x + h$
- $R : x \mapsto -x$

Properties:

1. $dist(\vec{p}, \vec{q}) = dist(\vec{q}, \vec{p})$
2. $dist(\vec{p}, \vec{q}) \geq 0$
3. $dist(\vec{p}, \vec{q}) = 0 \leftrightarrow \vec{p} = \vec{q}$

6.3 Basic Distance Bounds Lemma

$\forall \vec{p}, \vec{q} \in \mathbb{R}^d$, and $\forall j \in \{1, 2, 3, \ldots, d\}$:

$$|p_j - q_j| \leq dist(\vec{p}, \vec{q}) \leq \sqrt{d} \max_{1 \leq k \leq d} |p_k - q_k|$$

Proof Note that $(p_j - q_j)^2 \leq \sum_{k=1}^{d} (p_k - q_k)^2$ is trivial, because you can only add positive nunder when you add squares. Now let's take the square root, and we get

$$\sqrt{(p_j - q_j)^2} = |p_j - q_j| \leq \sqrt{\sum_{k=1}^{d} (p_k - q_k)^2} = dist(\vec{p}, \vec{q})$$

To prove the other inequality, it is trivial as well. We can just factor out the length of the vector d and multiply that with the maximum value of the distance vector. Then we get:

$$dist(\vec{p}, \vec{q}) = \sqrt{\sum_{k=1}^{d}(p_k - q_k)^2} \leq \sqrt{d \max_{1\leq k\leq d}(p_k - q_k)^2} = \sqrt{d} \max_{1\leq k\leq d}|p_k - q_k|$$

Cor: Componentwise Nature of Convergence

Let $(\vec{p}_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{R}^d$, and let $\vec{p} \in \mathbb{R}^d$. Then $\vec{p}_n \to \vec{p}$ if and only if $p_{n|j} \to p_j$ ($\vec{p} = (p_1, p_2, p_3, \ldots, p_d)$ and $\vec{p}_n = (p_{n|1}, p_{n|2}, \ldots, p_{n|d})$). Otherwise known as convergence of points can be reduced to conversion of dimensions.

This follows directly from the inequality, because if the total distance goes to 0, then $|p_j - q_j|$ goes to 0. Therefore if the points converge, the corresponding coordinates must converge.

To prove the converse, we prove using the other side of the distance bounds. If all d coordinates are going to 0, then if we take the maximum, that would be going to 0. (the maximum of a sequence is less than the sum of the sequence, but if every term of the sum is going to 0, then the sum is going to 0, then the maximum is going to 0). Therefore the distance must also be going to 0. Thus the two points converges.

6.4 Other Distance

Of course, there are other distance formulas, like the Minkowski Distance

$$((x-a)^p + (y-b)^p)^{\frac{1}{p}} + + + + (p > 1)$$

This is another distance formula, but under this, only reflection preserves distance.

6.5 Distances Between Sets

We define the distance between a point $\vec{p}$ and a set D as:

$$dist(\vec{p}; D) = \inf_{\vec{d}\in D} dist(\vec{d}; \vec{p})$$

We also define the distance between two sets D_1 and D_2 as:

$$dist(D_1; D_2) = \inf_{\substack{\vec{p}\in D_1 \\ \vec{q}\in D_2}} dist(\vec{p}; \vec{q})$$

7 Inequalities

7.1 Level of Operations

Powers/root $\to$ Multiplation/division $\to$ addition/subtraction $\to$ succession/pretrition

7.2 AM-GM

$$\mu = [x_1, x_2, \dots, x_n] \text{ and } x_1, x_2, x_3, \dots, x_n \geq 0$$

"Multiset" $\mu = \{(x, n), (y, m), \dots\}$

We define the arithmetic mean of a multiset as:

$$A(\mu) = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

And the geometric mean as:

$$G(\mu) = \sqrt[n]{x_1 x_2 x_3 \dots x_n}$$

7.2.1 AM-GM Inequality

$A(\mu) \geq G(\mu)$, with equality iff all elements of μ are the same.

7.2.2 Proof

This is done by mathematical induction. Base case is $n = 2$, then $\mu = [x, y]$. Then $A(\mu) = \frac{x+y}{2}, G(\mu) = \sqrt{xy}$

We know by the trivial inequality that $(\sqrt{x} + \sqrt{y})^2 \geq 0$, with equality case happening iff $x = y$. Then we get:

$$\begin{aligned} x - 2\sqrt{xy} + y &\geq 0 \\ \frac{x+y}{2} &\geq \sqrt{xy} \\ A(\mu) &\geq G(\mu) \end{aligned}$$

Now we induce on n, we seek to prove that case n implies case $2n$.

$\mu = [x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n]$

Then we know that

$$\begin{aligned} A(\mu) &= \frac{A(\mu_x) + A(\mu_y)}{2} \\ &\geq \frac{G(\mu_x) + G(\mu_y)}{2} \\ &\geq \sqrt{G(\mu_x)G(\mu_y)} \\ &= G(\mu) \end{aligned}$$

Note that in all inequalities used, the equality case is always when all x_n and y_n are the same element, therefore the equality case holds in all cases where the length of the list is 2^n.

Now we prove that case n implies $n - 1$

$\mu = [x_1, x_2, x_3, \dots, x_{n-1}]$

Note that we can construct $\mu' = [x_1, x_2, x_3, \dots, x_{n-1}, A(\mu)]$

Note that $A(\mu') = A(\mu)$, and since the AM-GM inequality is true for μ' by the assumption, we know

$$\begin{aligned} A(\mu') = A(\mu) &\geq \sqrt[n]{x_1 x_2 x_3 \dots x_{n-1} A(\mu)} \\ &\geq \sqrt[n]{G(\mu)^{n-1} A(\mu)} \\ &\geq \sqrt[n]{G(\mu)^{n-1} G(\mu)} \\ &\geq G(\mu) \end{aligned}$$

7.3 Young's Inequality

7.3.1 Hölder Conjugate

q is said to be the Hölder Conjugate of p:

$$q := p^* := \frac{p}{p-1}$$

Note that $q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$

7.3.2 Young's Inequality

$a, b \geq 0$; $p > 1$; $q = p^*$, then Young's Inequality states that:

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

With equality case iff $a^p = b^q$

7.3.3 Proof

We first proof Young's Inequality assuming that $p, q \in \mathbb{Q}$. We can rewrite $p = \frac{n+m}{n}$ and $q = \frac{n+m}{m}$ for some $n, m \in \mathbb{N}$. Now Young's Inequality turns into:

$$ab \leq \frac{na^{\frac{n+m}{n}}}{n+m} + \frac{mb^{\frac{m+n}{m}}}{n+m}$$

If we let $x = a^{\frac{1}{n}}$ and $y = b^{\frac{1}{m}}$. Then the inequality turns into:

$$ab = x^n y^m =\leq \frac{nx^{n+m} + my^{n+m}}{n+m}$$

And that is true by weighted AM-GM, with equality iff $x = y$, which equals to $a^{\frac{1}{n}} = b^{\frac{1}{m}}$, which equals to

We can prove the inequality for irrational by taking limits, because $\forall n \geq n_0 : f(n) \leq g(n)$, and the limits of both $f(x)$ and $g(x)$ as $n \to \infty$ exists and are finite, then we know that $\lim_{n\to\infty} f(n) \leq \lim_{n\to\infty} g(n)$. When p and q are irrational, we construct $\{p_n\}$ and $\{q_n\}$ as two sequences of rationals that approaches p and q, the left hand side of Young's Inequality is unaffected by the limit, and by what we've just said about limits, we know that:

$$ab \leq \lim_{n\to\infty} \frac{a^{p_n}}{p_n} + \frac{b^{q_n}}{q_n} = \frac{a^p}{p} + \frac{b^q}{q}$$

7.4 Hölder's Inequality

7.4.1 p-norm

In $\mathbb{R}^2$: let $||(a,b)||_p = (|a^p| + |b^p|)^{1/p}$ for any $p \in \mathbb{R} > 1$, this is known as the p-norm of a vector. Note that $||(a,b)||_2 = ||(a,b)|| = \sqrt{a^2+b^2}$

7.4.2 Hölder's Inequality

$\forall (a,b),(c,d) \in \mathbb{R}^{\nvDash}, p > 1, q = p^* = \frac{p}{p-1}$:

$$0 \leq |ac| + |bd| \leq ||(a,b)||_p ||(c,d)||_q$$

Equality happens iff $(\frac{a}{s})^p = (\frac{c}{t})^q$ and $(\frac{b}{s})^p = (\frac{d}{t})^q$ where $s = ||(a,b)||_p$ and $t = ||(c,d)||_q$, such that $||(a,b)||_p ||(c,d)||_q = (|a|^p + |b|^p)^{1/p}(|c|^q + |d|^q)^{1/q}$

7.4.3 Proof

We know that the absolute value of the product is equal to the product of the absolute value. If we apply Young's Inequality to $|\frac{a}{s}||\frac{c}{t}|$ and $|\frac{b}{s}||\frac{d}{t}|$, we get:

$$|\frac{a}{s}||\frac{c}{t}| \leq \frac{|a|^p}{p|s|^p} + \frac{|c|^q}{q|t|^q}$$
$$|\frac{b}{s}||\frac{d}{t}| \leq \frac{|b|^p}{p|s|^p} + \frac{|d|^q}{q|t|^q}$$

Now we add:

$$\frac{1}{st}(|ac| + |bd|) \leq \frac{|a|^p + |b|^p}{p|s|^p} + \frac{|c|^q + |d|^q}{q|t|^q}$$

Note that $|a|^p + |b|^p = |s|^p$ and $|c|^q + |d|^q$, so everything cancels

$$\frac{1}{st}(|ac| + |bd|) \leq \frac{1}{p} + \frac{1}{q}$$

But we know that $q = p^*$, therefore, $\frac{1}{p} + \frac{1}{q} = 1$, and we get:

$$|ac| + |bd| \leq st = ||(a, b)||_p \cdot ||(c, d)||_q$$

The equality case occurs at basically the same way as Young's Inequality's equality case.

7.5 Cauchy-Schwarz Inequality

This is a special case of Hölder's Inequality, where $p = q = 2$. (This is very important, 2 is the *only* value that is its own conjugate, this is why Euclidean distance is so special)

If we plug in 2 for p and q and use the Triangle Inequality: $|ac+bd| \leq |ac|+|bd| \leq \sqrt{a^2 + b^2}\sqrt{c^2 + d^2}$. We can set $\vec{u} = (a, b), \vec{v} = (c, d)$, such that $|v\dot{u}| \leq ||u||\dot{|}|v||$ with equality iff $ab \geq 0$, which can be written as iff $\vec{v}||\vec{u}$.

Then, by Young's inequality, $|\frac{ac}{st}| = |\frac{a}{s}||\frac{c}{t}| \leq \frac{|a|^p}{p|s|^p} + \frac{|c|^q}{q|t|^q}$, and the same is true for bd.

It follows that $\frac{1}{st}(|ac| + |bd|) \leq \frac{|a|^p+|b|^p}{p|s|^p} + \frac{|c|^q+|d|^q}{q|t|^q} = \frac{1}{p} + \frac{1}{q} = 1$, or $|ac| + |bd| \leq st$.

7.6 Triangle Inequality

$$dist(\vec{p}, \vec{q}) + dist(\vec{q}, \vec{r}) \geq dist(\vec{p}, \vec{r})$$

This can be generalized by mathematical induction to $dist(\vec{p_0}, \vec{q_n}) \leq \sum_{j=1}^{n} dist(\vec{p}_{j-1}, \vec{p}_j)$ (Otherwise known that the shortest distance between two points is the straight line, or the **Generalized Triangle Inequality** or the "Broken Line Inequality")

This can be thought of algebraically, such that $|a + b| \leq |a| + |b|$ with equality iff $ab \geq 0$

7.7 Reverse Triangle Inequality

From the Triangle Inequality we know:

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

Let $\vec{z} = \vec{x} + \vec{y}$, then we subtract, we get:

$$|| + \vec{z} + || \leq || + \vec{x} + || + || + \vec{z} - \vec{x} + ||$$
$$|| + \vec{z} + || - || + \vec{x} + || \leq || + \vec{z} - \vec{x} + ||$$

Because $\vec{x}$ and $\vec{z}$ are just variables, we can switch them, and we get:

$$|| + \vec{x} + || - || + \vec{z} + || \leq || + \vec{x} - \vec{z} + || = || + \vec{z} - \vec{x} + ||$$

Since the right hand side is greater than both of the above qualities, we can just say it's greater than the absolute value of the difference. Hence we get the Reverse Triangle Inequality:

$$||+\vec{z}-\vec{x}+|| \geq |||+\vec{z}+||-||+\vec{x}+|||$$

7.8 Minkowski's Inequality

Mikowski's states that $||u+v||_p \leq ||u||_p + ||v||_p$, with equality if v = tu or u = tv for some $t \geq 0$.

This can be thought of as the triangle inequality for the p-norm, rather than the ordinary norm.

7.8.1 Rational Power Proof

The rational power of some nunder, m, exists if there is some sequence, qn, where $n \to \infty, qn \to$ the rational nunder, only true if for any sequence which does this, the limit is equal. This is proven by for any two sequences, qn and rn, $m^{qn}/m^{rn} = m^{qn-rn}$, such that as $n \to \infty$, it equals 1.

7.8.2 Proof

The calculation works in any dimension, for simplicity's sake, let's work in $\mathbb{R}^2$, let $\vec{u} = (a, b)$ and $\vec{v} = (c, d)$

$$||+\vec{u}+\vec{v}+||_p^p = |+a+c+|^p + |+b+d+|^p = |+a+c+||+a+c+|^{p-1} + |+b+d+||+b+d+|^{p-1}$$

Now we factor and use the Triangle Inequality for Absolute Value:

$$\leq (|a|+|c|)|+a+c+|^{p-1} + (|b|+|d|)|+b+d+|^{p-1}$$

Now we rearrange the terms:

$$= (|a||a+c|^{p-1} + |b||b+d|^{p-1}) + (|c||a+c|^{p-1} + |d||b+d|^{p-1})$$

Now we apply Hölder's Inequality, we get:

$$\leq (|a|^p + |b|^p)^{\frac{1}{p}} (|a+c|^{(p-1)q} + |b+d|^{(p-1)q})^{\frac{1}{q}} + (|c|^p + |d|^p)^{\frac{1}{p}} (\dots)$$

Note that $q = p^*$, therefore $(p-1)q = p$:

$$= (||+\vec{u}+||_p + ||+\vec{v}+||_p)||+\vec{u}+\vec{v}+||_p^{\frac{p}{q}}$$

Since $q = p^*$, we know that $\frac{p}{q} = p - 1$, and if we bring the inequality to the original left hand side:

$$\leq (||+\vec{u}+||_p + ||+\vec{v}+||_p) + ||+\vec{u}+\vec{v}+||_p^{p-1}$$

Now we divide:

$$||+\vec{u}+\vec{v}+||_p \leq ||+\vec{u}+||_p + ||+\vec{v}+||_p$$

Now let's consider the equality cases. If one of the vectors is 0, then the inequality is trivially true.

If neither vectors are the 0 vector, we see the equality cases of all the inequalities used to prove Minkowski's. First we used the triangle inequality, which only has equality when $ac \geq 0$ and $bd \geq 0$. Next we applied Hölder's, which has equality case when both coordinates are proportional. Therefore, the two vectors must be positive multiples of one another.

8 Extreme Value Theorem in $\mathbb{R}^n$

8.1 Theorem

Let $f : D \to \mathbb{R}$ be a continuous function mapping from the compact set D to the reals. Then $\exists P, Q \in D$, not necessarily distinct, such that $\exists x \in D : f(P) \leq f(x) \leq f(Q)$.

8.2 Proof

First we prove that f must be bounded from above. Assume that it is not, take a sequence $(\vec{P_n})_{n=1}^{\infty} \in D$, the assumption implies that implies for every positive integer n, $f(\vec{P_n}) > n$. Because it is bounded by D, we can pick a convergent subsequence $\vec{P_{n_k}}$, which converges to $\vec{P}$. However, since D is closed, we know that $\vec{P} \in D$. However, now we have a contradiction. Because f is continuous, $\lim_{n\to\infty} f(\vec{P_n}) \to \vec{P}$. This is a contradiction, because the right hand side $\vec{P}$ is finite (it's within D), but the left hand side goes to infinity by the assumption. Therefore the assumption is false, thus the function f is bounded from above.

Now we know that f is bounded from above, we know $M := \sup(f(D))$ exists where $0 \leq M < \infty$. We can chose a sequence $(P_n)_{n=1}^{\infty} \in D$ such that $f(P_n) \to M$. The sequence is bounded, so it has a convergent subsequence, $\vec{P_{n_k}} \to \vec{P}$. By the closure of D, $\vec{P} \in D$. Note that $f(\vec{P}) = \lim_{k\to\infty} f(\vec{P_{n_k}}) = M$ by the continuity of f. Therefore, the function actually takes on its maximum value at that point.

9 Heine-Borel Theorem and Domain Properties

9.1 Heine-Borel Theorem

Let K be a compact set in $\mathbb{R}$, and let $U_\lambda | \lambda \in \Lambda$ be a family of open sets in $\mathbb{R}$, which covers K, such that $K \subseteq \cup_{x\in\Lambda} U_\lambda$. Then $\exists \lambda_1, \lambda_2, \lambda_3...$ such that$K \subseteq U_{\lambda_1} \cup U_{\lambda_2} \cup ... \cup U_{\lambda_n}$.

9.1.1 Proof

Since K is bounded, it can be fully contained within some rectangle (R), which can then be split into 4 congruent parts, $R_{i_1}(R_1, R_2, R_3, R_4)$. Suppose one quadrant cannot be covered by any finite collection of U_λ. By extension, if that quadrant is divided further, one of the subquadrants $(R_{i_1 i_2})$ cannot be covered. This can continue for countable infinity divisions (able to be counted with an infinite amount of integer). Since $R_{i_1 i_2 \ldots i_n} \neq \emptyset$, let P_n be any point in $K \cap R_{i_1 i_2 \ldots i_n}$, such that there is a bounded sequence of points in K. Thus it must have a convergent subsequence such that $P_{n_k} \to P, P \in K$. Thus, $P \in U_{\lambda^*}$ for some $\lambda^* \in \Lambda$. Since U_{λ^*} is open, $\exists r > 0,$ such that$B_r(P) \subseteq U_{\lambda^*}$. Diameter/diagonal length (diam) of $R_{i_1 i_2 \ldots i_n} = \frac{diam(R)}{2^n} < r$ as $n \to \infty$. Thus, $R_{i_1 i_2 \ldots i_n} \subseteq B_r(P)$ as $n_k \to \infty$. $dist(P, P_{n_k}) < \frac{r}{2}$ and $dist(P_{n_k}, x) \leq diam(R_{i_1 i_2 \ldots i_n})$ and by the triangle inequality, $diam(R_{i_1 i_2 \ldots i_n}) < \frac{r}{2}$. Thus, there is a contradiction, and it must be covered by a finite number of sets.

9.2 Uniform Continuity

Let $f : D \to \mathbb{R}$, where $D \subseteq \mathbb{R}^d$. We say f is uniformally continuous on D if:

$$\forall \epsilon > 0, \exists + \delta > 0 : \forall \vec{x}, \vec{y} \in D, || + \vec{x} - \vec{y} + || < \delta \to |f(\vec{x}) - f(\vec{y})| < \epsilon$$

The difference between this and regular continuity is that the δ in regular continuity is defined by both ϵ and the specific point we are considering. Uniform continuity, however, the value δ is independent to the point you chose within the domain and is just dependent on ϵ.

For example, consider $y = \tan x$ where $D = (-\frac{\pi}{2}, \frac{\pi}{2})$. the value required for δ for a fixed ϵ gets smaller and smaller as x approaches both endpoints. This function is continuous but not uniformally so. If it were uniformally continuous, that δ value would NOT change.

Theorem: If f is uniformally continuous on D, then f is continuous for every point in D.

9.3 Uniform Continuity Theorem

Theorem: If $f : K \to \mathbb{R}$, where $K \subseteq \mathbb{R}^d$ is compact, and f is continuous at each $\vec{x} \in K$. Then f is uniformally continuous on K.

Proof: Fix $\epsilon > 0$. For each $\vec{x} \in K$, let $u_{\vec{x}}$ be an open ball, centered at $\vec{x}$ such that for any $\vec{y} \in K \cap 2u_{\vec{x}}$ [$2u_{\vec{x}}$ is an open ball centered at $\vec{x}$ with twice the radius of $u_{\vec{x}}$], $|f(\vec{x}) - f(\vec{y})| < \frac{\epsilon}{2}$. Because the function is continuous at $\vec{x}$, there is a radius 2δ around $\vec{x}$ such that $|f(\vec{x}) - f(\vec{y})| < \frac{\epsilon}{2}$. Now we see that every $\vec{x} \in K$ is covered by at least one such open ball, namely $u_{\vec{x}}$. The collection $\{u_{\vec{x}} + | + \vec{x} \in K\}$ is an open covering of K. By Heine-Borel, we can select a finite set of points $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ such that K is covered by $u_{\vec{x}_1} \cup u_{\vec{x}_2} \cup \cdots \cup u_{\vec{x}_n}$. Take $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_n\} > 0$ where δ_j is the radius of $u_{\vec{x}_j}$ for $j = 1, 2, 3, \ldots, n$.

Let $\vec{p}, \vec{q} \in K$ such that $|| + \vec{p} - \vec{q} + || < \delta$. We need to prove that $|f(\vec{p}) - f(\vec{q})| < \epsilon$. $\vec{p} \in K \to \exists + j : \vec{p} \in u_{\vec{x}_j} \to || + \vec{p} - \vec{x}_j + || < \delta_j$. But we know that $|| + \vec{p} - \vec{q} + || < \delta_j$. Now by the Triangle Inequality, we get:

$$|| + \vec{q} - \vec{x}_j + || \leq || + \vec{q} - \vec{p} + || + || + \vec{p} - \vec{x}_j + || \leq 2\delta_{\vec{x}_j}$$

This means that $\vec{q} \in 2u_{\vec{x}_j}$. By the way we picked our δ, we know that $|f(\vec{q}) - f(\vec{x}_j)| < \frac{\epsilon}{2}$. Similarly, we also have $|f(\vec{p}) - f(\vec{x}_j)| < \frac{\epsilon}{2}$. Now if we apply the triangle inequality again, we get:

$$|f(\vec{p}) - f(\vec{q})| = |f(\vec{p}) - f(\vec{x}_j) + f(\vec{x}_j) - f(\vec{q})| \leq |f(\vec{p}) - f(\vec{x}_j)| + |f(\vec{q}) - f(\vec{x}_j)| < \epsilon$$

Therefore, f is uniformally continuous over K.

9.4 Connected

9.4.1 Definition

$D \subseteq \mathbb{R}^d$ is said to be connected if $\forall \vec{p}, \vec{q} \in D$, $\exists$ a continuous function $\vec{\gamma} : [a, b] \to \mathbb{R}^d$ that $\vec{\gamma}(a) = \vec{p}$, $\vec{\gamma}(b) = \vec{q}$, and $\forall t \in [a, b], \vec{\gamma}(t) \in D$

9.4.2 Theorem

Let $D \subseteq \mathbb{R}^d$ be connected, $D \neq \emptyset$. Suppose $D = A \cup B$, where A and B are open. Such that $A \cap B = \emptyset$. Then $A = \emptyset$ or $B = \emptyset$.

Proof:

Assume that D can be broken up into two open, non-empty sets A and B such that $D = A \cup B$. Because $A \neq \emptyset$, we can pick $\vec{p} \in A$ and similarly we can pick $\vec{q} \in B$. Clearly, $\vec{p}, \vec{q} \in D$. Since D is connected, $\exists$ continuous $\vec{\gamma} : [a, b] \to D$ such that $\vec{\gamma}(a) = \vec{p}$, $\vec{\gamma}(b) = \vec{q}$. Let $S = \{t \in [a, b] + | + \vec{\gamma}(t) \in A\}$. We know that $a \in S$, so $S \neq \emptyset$. Note that $S \leq b$, and since it has a bound, it has a supremum. Let $t_0 = \sup S$, note that $a \leq t_0 \leq b$. Since $\vec{\gamma}(t_0) \in D$, it is either in A or in B (since $A \cap B = \emptyset$). Suppose $\vec{\gamma}(t_0) \in A$. Since A is open, we can draw an open ball (B_ϵ) around $\vec{\gamma}(t_0)$. Now consider $\vec{\gamma}(t_0 + \delta)$ for some small positive δ. If δ is small enough, $\delta \in B_\epsilon \in A$. This is a contradiction, as then it means that $t_0 + \delta \in S$. This is a contradiction, as then $t_0 \neq \sup S$. A similar contradiction can be made for the case that $\vec{\gamma}(t_0) \in B$. Therefore the original assumption was false.

10 Intermediate Value Theorem

10.1 Theorem in $\mathbb{R}$

If f: [a, b] $\to \mathbb{R}$ is continuous, and $f(a) \neq f(b)$, $\forall y \in (f(a), f(b))$, then $\exists c \in (a, b) : f(c) = y$.

10.1.1 Proof

Let there exist y, without loss of generality, such that $f(a) < y < f(b)$. Take the set $S = \{x \in [a, b] | f(x) \leq y\}$. Since $a \in S$, $S \neq \emptyset$, we also know that $S \leq b$. Now take $c = \sup S$, $a \leq c \leq b$ ($c \in [a, b]$). There are three possible cases, $f(c)$ is either $> y, < y$, or $= y$.

Consider the case in which $f(c) < y$. If this is the case, then we can chose an arbitrarily small $\delta > 0$ such that $f(c + delta) < y$. However, then $c + \delta \in S$. But this causes a contradiction because $c = \sup S$ and there should not be any element of S that's larger than c. Therefore, this case is impossible.

Now consider the case in which $f(c) > y$. Take some arbitrarily small $u \in [0, \delta], \delta > 0$ such that $f(c - u) > y$. However, then $c - \delta$ is therefore an upper bound of the set S, we once again reach the same contradiction.

Therefore, since c exists, $f(c) = y$.

10.2 Theorem in $\mathbb{R}^{\ltimes}$

If $f : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}^d$ is connected and continuous. $\forall \vec{p}, \vec{q} \in D$ such that $f(\vec{p}) \neq f(\vec{q})$ and if $y \in \mathbb{R}$ is between $f(\vec{p})$ and $\vec{q}$, then $\exists \vec{r} \in D$ such that $f(\vec{r}) = y$.

10.2.1 Proof

Without loss of generality, let us assume $f(\vec{p}) < f(\vec{q})$.

Since D is connected, there is some path $\vec{\gamma}(t) = (x(t), y(t)) \in D$, $a \leq t \leq b$ such that $\vec{\gamma}(a) = \vec{p}$ and $\vec{\gamma}(b) = \vec{q}$ and is continuous over $[a, b]$. Now we construct $g(t) = f(\vec{r}(t)) = (f \cdot \vec{r})(t) \in mathbbR$. Because $g(t)$ is a composition of continuous functions, $g(t)$ is also continuous. $g(a) = f(\vec{\gamma}(a)) = f(\vec{p})$ and $g(b) = f(\vec{\gamma}(b)) = f(\vec{q})$. Now we see that $g(a) < y < g(b)$. Therefore, by the IVT for single-variable functions, $\exists + t_0$ such that $g(t_0) = y$. Now we plug t_0 into $\vec{\gamma}$, $\vec{r} = \vec{\gamma}(t_0) \in D$. Then $f(\vec{r}) = f(\vec{\gamma}(t_0)) = g(t_0) = y$.

11 Vector Valued Functions

11.1 Definition

$f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}^d$ is known as real-valued functions.

$\vec{f} : D \to \mathbb{R}^e$, $D \subseteq \mathbb{R}^d$ is known as vector-valued/point-valued functions.

$\vec{f}(\vec{p}) = (f_1\vec{p}, f_2\vec{p}, \ldots, f_e(\vec{p}))$ where $f_1, f_2, \ldots, f_e$ are real-valued and are called the component functions of f.

11.2 Continuity

$f : \mathbb{R}^d \to \mathbb{R}^e$ is continuous at $\vec{a} = (a_1, a_2, a_3, \ldots, a_d) \in D$ iff $\forall \epsilon, \exists + \delta > 0 : \forall \vec{x} \in D$:

$$|| + \vec{x} - \vec{a} + ||_d < \delta \to || + f(\vec{x}) - f(\vec{a}) + ||_e < \epsilon$$

11.2.1 Component-wise Nature of Continuity

$f : D \to \mathbb{R}^e$ is continuous at a point $\vec{a} \in D$ iff $f_1, f_2, \ldots, f_e$ are all continuous at $\vec{a}$.

Proof:

First fix an ϵ, then by the basic distances bound lemma, we get a bunch of inequalities:

$$|f_1(\vec{x}) - f_1(\vec{p})| < \frac{\epsilon}{\sqrt{e}}, \forall \vec{x} \in D \cap B_{\delta_1}(\vec{p})$$
$$|f_2(\vec{x}) - f_2(\vec{p})| < \frac{\epsilon}{\sqrt{e}}, \forall \vec{x} \in D \cap B_{\delta_2}(\vec{p})$$
$$\dots\dots \quad \dots\dots$$
$$|f_e(\vec{x}) - f_e(\vec{p})| < \frac{\epsilon}{\sqrt{e}}, \forall \vec{x} \in D \cap B_{\delta_e}(\vec{p})$$

Then let $\delta = \min\{\delta_1, \delta_2, \dots, \delta_e\} > 0$, which must exist and satisfy all the distance inequalities, specifically $\max_{1 \le j \le e}(|f_j(\vec{x}) - f_j(\vec{p})|)$. Then again, by the basic distance bounds lemma, we know that $|| + \vec{f}(\vec{x}) - \vec{f}(\vec{p}) + ||_e \le \max_{1 \le j \le e}(|f_j(\vec{x}) - f_j(\vec{p})|)$. Therefore we get that for any fixed ϵ, we can find a δ such that $|| + \vec{x} - \vec{p} + ||_d < \delta \to || + f(\vec{x}) - f(\vec{p}) + ||_e < \epsilon$

11.2.2 Composition of Continuous Functions

If $\vec{f} : D \to E$, where $D \subseteq \mathbb{R}^d$, $E \subseteq \mathbb{R}^e$ and $\vec{g} : E \to \mathbb{R}^k$ are both continuous on their respective domains. Then $\vec{h} = \vec{g} \circ \vec{f}$ is continuous on D

Proof:

To prove this, fix $\epsilon > 0$. There's a $\eta > 0$ such that $\forall \vec{y} \in E \cap B_\eta(\vec{f}(\vec{p}))$, because we know that $\vec{g}$ is continuous at $\vec{f}(\vec{p})$, we get:

$$|| + \vec{g}(\vec{y}) - \vec{g}(\vec{f}(\vec{p})) + || < \epsilon$$

To guarantee that $\vec{y} = \vec{f}(\vec{x})$ lies within η units of $\vec{f}(\vec{p})$ i.e. $|| + \vec{f}(\vec{x}) - \vec{f}(\vec{p}) + || < \eta$, we can take $\vec{x} \in D \cap B_\delta(\vec{p})$ where $\delta > 0$ corresponding to η [using the continuity of $\vec{f}$ at $\vec{p} \in D$].

Now, as long as $\vec{x} \in D \cap B_\delta(\vec{p})$, we have $\vec{f}(\vec{x}) \in E \cap B_\eta(\vec{f}(\vec{p}))$. Thus, if we take $\vec{y} = \vec{f}(\vec{p})$, we get: $|| + \vec{g}(\vec{f}(\vec{x})) - \vec{g}(\vec{f}(\vec{p})) + || < \epsilon$, or $|| + \vec{h}(\vec{x}) - \vec{h}(\vec{p}) + || < \epsilon$. So $\vec{h}$ is continuous at $\vec{p}$.

11.3 Compactness Theorem

11.3.1 Theorem

Let $\vec{f} : D \to \mathbb{R}^e$ be a contnuous function, where $D \subseteq \mathbb{R}^d$ is compact. Then its *range* $\vec{f}(D) := \{f(\vec{p}) + | + \vec{p} \in D\}$ is also compact. In other words: compactness is preserved under continuous mappings. Note that this is the generalization of the Extreme Value Theorem.

11.3.2 Proof

To prove this, write $R := f(D)$. We need to show that R is closed and bounded in $\mathbb{R}^e$. Boundedness is easy. Since each component function f_j of f is real valued, by EVT each component function f_j has an absolute bound M_j, so that $| + f_j(\vec{p}) + | \le M_j$ for all $\vec{p} \in D$. Take $M := \max\{M_1, M_2, \dots, M_e\}$. Then for all $\vec{p} \in D$, we have

$$||+\vec{f}(\vec{p})+|| = \sqrt{f_1(\vec{p})^2+\cdots+f_e(\vec{p})^2} \leq \sqrt{e\times M^2} = M\sqrt{e}$$

This says that the range R lies within the closed ball of radius $M\sqrt{e}$ centered at $\vec{0}$ in $\mathbb{R}^e$. It therefore certainly lies within some closed cube centered at $\vec{0}$, and hence is bounded.

To prove closedness, let $(\vec{y_n})_{n=1}^{\infty}$ be a convergent sequence in $\mathbb{R}^e$ with limit $\vec{y}$, such that $\vec{y_n} \in R$ for each $n \geq 1$. We need to prove that $\vec{y} \in R$. Since R is the range of f, we must have $\vec{y} = \vec{f}(\vec{x_n})$, where $\vec{x_n} \in D$. Because D is bounded, we can pick a convergent subsequence $\vec{x}_{n_k} \to \vec{x}$. But because D is closed, we know that $\vec{x} \in D$. Because $\vec{f}$ is continuous, we get that $\vec{y}_{n_k} = \vec{f}(\vec{x}_{n_k}) \to \vec{f}(\vec{x})$. However, we know that $\vec{y}_{n_k} \to \vec{y}$. But since each sequence converges to one point, we know that $\vec{y} = \vec{f}(\vec{x})$, where $\vec{x} \in D$. Therefore, $\vec{y} \in R$.

Therefore, R is closed and bounded.

11.4 Connectedness Theorem

11.4.1 Theorem

$\vec{f}: D \to \mathbb{R}^e$, $D \in \mathbb{R}^d$ is continuous on D. If D is connected, $E = \vec{f}(D)$ (the range of the domain), is also connected. Note that this is the generalization of the Intermediate Value Theorem.

11.4.2 Proof

$\forall \vec{u}, \vec{v} \in E$, we can find two points $\vec{p}, \vec{q}$ such that $\vec{f}(\vec{p}) = \vec{u}$ and $\vec{f}(\vec{q}) = \vec{v}$. Now because D is connected, $\exists + \vec{\gamma}: [a,b] \to \mathbb{R}^d$, $\vec{\gamma}([a,b]) \subseteq D$. Now we consider $\vec{\delta} = \vec{f}(\vec{\gamma}(t))$, $\vec{\delta}: [a,b] \to \mathbb{R}^e$. Since it's a composition of continuous functions, $\vec{\delta}$ is also continuous. And $\forall t \in [a,b]$, $\vec{\delta}(t) = \vec{f}(\vec{\gamma}(t)) \in E$. We also know that $\vec{\delta}(a) = \vec{f}(\vec{\gamma}(a)) = \vec{f}(\vec{p}) = \vec{u}$, and $\vec{\delta}(b) = \vec{f}(\vec{\gamma}(b)) = \vec{f}(\vec{q}) = \vec{v}$. Therefore, $\forall \vec{u}, \vec{v} \in E$, there is a continuous path that connects $\vec{u}$ to $\vec{v}$ and stays within E. Therefore, E is connected.

12 Sequences of Functions

12.1 Infinity Norm

For $f: [a,b] \to \mathbb{R}$ that is bounded, we say:

$$||f||_\infty := ||f||_D := \sup_{x\in[a,b]} |f(x)|$$

12.2 Pointwise Convergence:

Definition: For $f_n: [a,b] \to \mathbb{R}$ for $n \geq 1$, and assume that they are all bounded on $[a,b]$. ($\exists + M_n > 0: |f_n(x)| \leq M_n$ for all $x \in [a,b]$), and $f: [a,b] \to \mathbb{R}$, f is bounded ($\exists + M > 0: |f(x)| \leq M$ for all $x \in [a,b]$).

If $\forall x \in [a,b] : \lim_{n\to\infty} f_n(x) = f(x)$, then we say that $f_n \xrightarrow{p} f(x)$ (f_n converges "pointwise" to f).

Theorem: If $\vec{f_n} \overset{u(\text{uniform})}{\rightarrow} \vec{f} : \mathbb{R} \to \mathbb{R}$, then $\vec{f_n} \overset{p(\text{pointwise})}{\rightarrow} \vec{f}$ on D: $\vec{f_n}(\vec{x}) \to \vec{f}(\vec{x})$ for every point $\vec{x} \in D$ (for any converging sequences, any point will also converge similarly, such that uniform convergence implies pointwise convergence, though the converse is untrue)

Proof: $\vec{f_n} \xrightarrow{u} \vec{f}$ on D (uniform convergence): $||\vec{f_n} - \vec{f}||_D = sup_{\vec{x}\in D}||\vec{f_n}(\vec{x}) - \vec{f}(\vec{x})|| \to 0$ as $n \to \infty$. $Then 0 \leq ||\vec{f_n}(\vec{x}) - \vec{f}(\vec{x})|| \leq ||\vec{f_n} - \vec{f}|| \to 0$.

Then, by the squeeze theorem, for a uniform sequences, $f_n(x) \to f(x)$ on D, giving pointwise convergence.

12.3 Uniform Convergence

12.3.1 In $\mathbb{R}$

$f_n : [a,b] \to \mathbb{R}$ for $n \geq 1$, and assume that they are all bounded on $[a,b]$. ($\exists + M_n > 0 : |f_n(x)| \leq M_n$ for all $x \in [a,b]$).

$f : [a,b] \to \mathbb{R}$, f is bounded ($\exists + M > 0 : |f(x)| \leq M$ for all $x \in [a,b]$).

We then claim that $f_n \xrightarrow{u} f$ as $n \to \infty$ if $\forall\epsilon$, $\exists + N_\epsilon$ such that $\forall n \geq N_\epsilon : || + f_n - f + ||_\infty < \epsilon$. In other words, this forces that the greatest vertical difference between the two functions will be arbitrarily small after N_ϵ. This forces the two functions to be "close" as a whole.

12.3.2 In $\mathbb{R}^\ltimes$

Let $D \in \mathbb{R}^d$ be a non-empty set, and let $\vec{f} : D \to \mathbb{R}^e$ and $\vec{f_n} : D \to \mathbb{R}^e$ for $n \geq 1$. We say that $\vec{f_n} - \vec{f}$ on D if:

$$|| + f_n - f + ||_D \to 0 \text{ as } n \to \infty$$

12.4 Uniform Convergence Theorem

12.4.1 Theorem

If $f_n \xrightarrow{u} f$ on $[a,b]$, where each f_n is continuous on $[a,b]$. Then f is also continuous on $[a,b]$

12.4.2 Proof

Pick any $\vec{p} \in D$, and fix $\epsilon > 0$. We need to find δ such that $\forall \vec{x} \in D$ with $||\vec{x} - \vec{p}|| < \delta$, then $||\vec{f}(\vec{x}) - \vec{f}(\vec{p})|| < \epsilon$. We can apply the triangle inequality and we get:

$$||\vec{f}(\vec{p}) - \vec{f}(\vec{p})|| \leq ||\vec{f}(\vec{x})\vec{f_n}(\vec{x})|| + ||\vec{f_n}(\vec{x}) - \vec{f_n}(\vec{p})|| + ||\vec{f_n}(\vec{p}) - \vec{f}(\vec{p})||$$

For large enough n, we know that $||\vec{f_n} - \vec{f}||_D < \frac{\epsilon}{3}$ because $\vec{f_n} - \vec{f}$. Now we know that the first and third ter are bounded by $\frac{\epsilon}{3}$. The second term is bounded b $\frac{\epsilon}{3}$ because $\vec{f_n}$ is uniformally continuous.

Therefore, we know that $||\vec{f}(\vec{p}) - \vec{f}(\vec{p})|| \leq \epsilon$ for any δ we pick. $\therefore \vec{f}$ is continuous on D

13 Differentiation

13.1 Differentiable in $\mathbb{R}$

Let $f : [a, b] \to \mathbb{R}$, let $p \in (a, b) = [a, b]^o(int[a, b])$,, then f is <u>differentiable</u> at p if $\exists a \in \mathbb{R}$ such that $a = \lim h \to 0 \frac{f(p+h)-f(p)}{h}$, called the <u>derivative</u> at point p ($a = f'(p) = \frac{df}{dx}(p)$).

13.2 Differentiable in $\mathbb{R}^\Bbbk$

This is generalized to high dimensions, such that for $f : D \subseteq \mathbb{R} \to \mathbb{R}$, and $\vec{p} \in D^o(\exists r > 0 : B_r(\vec{p} \subseteq D)$. Thus, it can be approached from any given direction, due to the ball existing in all directions.

$\vec{h} \to \vec{0}$ iff $h_d \to 0$, such that it can be substituted for the limit. Due to the lack of vector division though, the definition of the derivative has to be changed.

By the previous definition, $\lim h \to 0 |\frac{f(p+h)-f(p)}{h} - a| = 0$, rewritten $\lim h \to 0 \gamma(p, h) = 0$.

$$|h|\gamma(p, h) = |f(p + h) - f(p) - ah| < \epsilon|h| \text{ as } h \to 0(|h| < \delta).$$

$\exists a \in \mathbb{R}$, such that $\forall \epsilon > 0, \exists \delta > 0 : |h| < \delta$, then $|f(p + h) - f(p) - ah| < \epsilon|h|$, defining a as the derivative at p.

This is due to the idea that for some function, g(x), with linear approximation l(x), $\frac{|g(x)-l(x)|}{|x|} \to 0$ as $x \to 0$, such that the y distance decreases far faster than the x decay, called super-linear (suplinear) decay. Then, there must at most be 1 line that can superlinearly approximate g at x = 0. The superlinear decay curve is then the derivative function, where f(p + h) - f(p) = g(x) and ah = l(x).

<u>Proof:</u> Assume there are two function, l(x) = ax and m(x) = bx, then $\frac{|g(x)-ax|}{|x|} \to 0$ as $x \to 0$ and $\frac{|g(x)-bx|}{|x|} \to 0$ as $x \to 0$. Then $0 \leq |a - b| \frac{|ax-bx|}{|x|} = \frac{|(g(x)-bx)-(g(x)-ax|)}{|x|} \leq \frac{|g(x)-bx|}{|x|} + \frac{|g(x)-ax|}{|x|} \to 0$ as $x \to 0$, such that a = b. This proof can be done in each component for higher dimensions, using the dot product to remove all other components, by the properties of the dot product.

The definition can now be easily moved to higher dimensions, by superlinear decay, such that there is only one object in $\mathbb{R}^{\Bbbk - \Bbbk\!\!\!/}$, a hyperplane, or the set of all vectors orthoganal to the non-zero normal vector, all anchored to a specific point, $\vec{p_0}$, such that superlinear decay takes place.

<u>Definition:</u> A hyperplane in $\mathbb{R}^{+\Bbbk\!\!\!/}$ is a set of the form, Q $= \{\vec{x} \in \mathbb{R}^{+\Bbbk\!\!\!/} | (\vec{x} - \vec{p_0} \cdot \vec{n} = 0)$, where $\vec{p_0}, \vec{n} \in \mathbb{R}^{+\Bbbk\!\!\!/}$ with $\vec{n} \neq \vec{0}$, where n is a normal of Q ($\vec{n} \perp Q$).

Thus to summerize differentiability for real valued functions, $\exists!$ (exists exactly one) or $\exists\vec{a} \in \mathbb{R}$ such that $\forall \epsilon > 0, \exists \delta > 0 : ||\vec{h}|| < \delta$, then $|f(\vec{p} + \vec{h}) - f(\vec{p}) - \vec{a}\vec{h}| < \epsilon||\vec{h}||$. If the latter is true, f(x) is said to be <u>differentiable</u> at $\vec{p}$, since it is a unique value, denoted f'(p).

13.3 Gradient

This is then defined specifically such that $\vec{a} = (\vec{\nabla} f)(p)$, or the gradient of f at p.

Claim: The graph of $y = f(\vec{p}) + \vec{a} \cdot (\vec{x} - \vec{p})$ is a hyperplane in $\mathbb{R}^{+\mathbb{K}}$, such that the gradient is a hyperplane.

Proof: Let $\vec{N}$ (the normal vector to the hyperplane$\in \mathbb{R}^{+\mathbb{K}} = (-a_1, -a_2, ..., -a_d, 1), and \vec{P_0} \in \mathbb{R}^{+\mathbb{K}} = (p_1, p_2, ..., p_d, f(\vec{p}))$and $\vec{X} \in \mathbb{R}^{+\mathbb{K}} = (x_1, x_2, ..., x_d, y)$. Thus, $\vec{(X)} - \vec{P_0} = (x_1 - p_1, x_2 - p_2, ...x_d - p_d, y - f(\vec{p}))$, giving the equation of a hyperplane $(\vec{X} - \vec{P_0}) \cdot \vec{N} = 0$.

If $\vec{p} + \vec{h} = \vec{X}$, then $|f(\vec{X}) - [f(\vec{p}) + \vec{a} \cdot (\vec{X} - \vec{p})|| < \epsilon||\vec{h}||$. This is equal to the difference between the point on the graph and the function approximation. This can also be seen to be easily equal to the equation of the hyperplane from above, such that the gradient is the hyperplane.

13.3.1 Gradient Representation

Since $\vec{N}$ is the normal vector to the plane at that point, equal to $(-\vec{\nabla} f(\vec{p}), 1)$, or the projection of the negation of the normal vector onto the domain, such that it is the vector of the direction and magnitude of the fastest increase of the function at $\vec{p}$.

Theorem: If $f : D \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable at $\vec{p} \in D^o$, then $\vec{u_0} = \frac{\vec{\nabla} f(\vec{p})}{||\vec{\nabla} f(\vec{p})||}$ is the direction of steepest ascent for f at $\vec{p}$, and $||\vec{\nabla} f(\vec{p})|| = max_{||\vec{u}||=1} \partial_{\vec{u} f(\vec{p})}$, where $\vec{u}$ is any unit vector, such that as a result, $\partial_{\vec{u}} f(\vec{p}) = \vec{\nabla} f(\vec{p}) \cdot \vec{u}$.

This is by the Cauchy-Schwartz equality case $(-||\vec{u}||||\vec{v}|| \leq \vec{u} \cdot \vec{v} \leq ||\vec{u}||||\vec{v}||$, with equality with the lower bound if $\vec{u}$ and $\vec{v}$ are in opposite directions, the higher bound if in the same direction), such that $-\vec{u_0}$ is the direction of steepest descent for f at $\vec{p}$ and $-||\vec{\nabla} f(\vec{p})|| = min_{||\vec{u}||=1} \partial_{\vec{u}} f(\vec{p})$.

13.3.2 Gradient Calculation

Take $\vec{h} = h\vec{e_j}$, where $h \to 0$ and the set of $\vec{e_j}$ is known as the *standard basis vectors* in $\mathbb{R}^d$:

$$\vec{e_j} = \begin{cases} \vec{e_1} & = (1, 0, 0, \dots, 0) \\ \vec{e_2} & = (0, 1, 0, \dots, 0) \\ \vdots & \vdots \quad \vdots \quad \vdots \quad \vdots \\ \vec{e_d} & = (0, 0, 0, \dots, 1) \end{cases}$$

Note that because $||\vec{e_j}|| = 1$, $||\vec{h}|| = |h|||\vec{e_j}|| = |h|$. Now if we fix a $j \in \{1, 2, 3, \dots, d\}$ and apply the definition of differentiability, the unique gradient $(\vec{a})$ must satisfy:

$$\forall \epsilon > 0, \exists \delta > 0 : \forall h \text{ with } |h| < \delta \implies |f(\vec{p} + h\vec{e_j}) - f(\vec{p}) - \vec{a} \cdot h\vec{e_j}| < \epsilon|h|$$

However, note that when we dot $\vec{a}$ with $\vec{e_j}$, the result is the j^{th} component of $\vec{a}$, or a_j. Now if we divide through by $|h|$, we get:

$$\left|\frac{f(\vec{p}+h\vec{e}_j)-f(\vec{p})}{h}-a_j\right|<\epsilon$$

This says is that $\left|\frac{f(\vec{p}+h\vec{e}_j)-f(\vec{p})}{h}\right|$ approaches a_j indefinitely, therefore, we can rewrite the relationship as a limit statement:

$$\boxed{a_j=\lim_{h\to 0}\left|\frac{f(\vec{p}+h\vec{e}_j)-f(\vec{p})}{h}\right|}$$

We call this a_j as a **partial derivative** of $f(\vec{x})$ at j^{th} component, which can be written as $\partial_{x_j}f(\vec{p})$ or $\frac{\partial f}{\partial x_j}(\vec{p})$.

Note that:

$$\partial_{x_j}f(\vec{p})=\frac{d}{dx_j}\left|_{x_j=p_j}f(p_1,p_2,\ldots,p_{j-1},x_j,p_{j+1},\ldots,p_d)\right.$$

In other words, we can hold all other components of f constant and differentiate based on only one component, and plug in the value p_j after the differentiation. Now we know how to compute the gradient of f, it is simply the vector of all the partial derivatives:

$$\vec{\nabla}f=(\partial_{x_1}f(\vec{p}),\partial_{x_2}f(\vec{p}),\ldots,\partial_{x_d}f(\vec{p}))$$

13.4 Directional Derivative

Take $\vec{u}\in\mathbb{R}^d$, $\vec{u}\neq\vec{0}$ and take a point on the function f, $\vec{p}$, we define the directional derivative $\partial_{\vec{u}}f(\vec{p})$ as:

$$\begin{aligned}\partial_{\vec{u}}f(\vec{p})&=\lim_{h\to 0}\frac{f(\vec{p}+h\vec{u})-f(\vec{p})}{h}\\&=\frac{d}{dt}\Big|_{t=0}f(\vec{p}+t\vec{u})\end{aligned}$$

This quality describes how fast the function f is changing at $\vec{p}$ in the direction of $\vec{u}$.

The latter definition gives the single variable function, such that g(t) = $\vec{p}+t\vec{u}$, where $\vec{u}\neq\vec{0}$, is called the uniform rectilinear motion curve, due to being a line with constant curve speed.

However, the magnitude of $\vec{u}$ also has meaning, its magnitude is the rate of change of the function in the direction of $\vec{u}$.

Thm: Let $\vec{u}$ be any unit vector, then:

$$\partial_{\vec{u}}f(\vec{p})=\vec{\nabla}f(\vec{p})\cdot\vec{u}$$

13.5 Geometric Interpretation of the Gradient

If $f : D \subseteq \mathbb{R}^d \to \mathbb{R}$ is differentiable at $\vec{p} \in D^\circ$, then:

$$\vec{u}_0 = \frac{\vec{\nabla} f(\vec{p})}{||\vec{\nabla} f(\vec{p})||}$$

i.e. the unit vector of the gradient is the direction of steepest ascent for f at $\vec{p}$.

$$||\vec{\nabla} f(\vec{p})|| = \max_{||\vec{u}||=1} \partial_{\vec{u}} f(\vec{p})$$

And the magnitude of the gradient is the rate at which the function is ascending at the point.

This is a combination of the Cauchy-Schwarz Inequality and the theorem which states that $\partial_{\vec{u}} f(\vec{p}) = \vec{\nabla} f(\vec{p}) \cdot \vec{u}$. By the Cauchy-Schwarz Inequality, we know that the length of the gradient is an upper bound of the directional derivative. This maximum is also obtained, because in Cauchy-Schwarz inequality, the equality case happens when the two vector have the same direction. Therefore, we know that the gradient is the direction of fastest ascent of a function at any given point $\vec{p}$

However, the Cauchy-Schwarz inequality also states that the dot product is bounded from below by negative of the product of the lengths, which is achieved when the two vector are anti-directional. Therefore we also know that the negative of the gradient points in the direction of steepest descent and the magnitude can be written as:

$$-||\vec{\nabla} f(\vec{p})|| = \min_{||\vec{u}||=1} \partial_{\vec{u}} f(\vec{p})$$

13.6 Vector Valued Directional Derivative

Given a vectored valued function $\vec{\gamma}(t)$:

$$\vec{\gamma}(t) = \begin{bmatrix} \gamma_1(t) \\ \gamma_2(t) \\ \vdots \\ \gamma_d(t) \end{bmatrix}$$

We define the "speed" vector of $\vec{\gamma}$ as its derivative, which is defined as:

$$\frac{d\vec{\gamma}}{dt}(t) := \lim_{h \to 0} \frac{\vec{\gamma}(t+h) - \vec{\gamma}(t)}{h}$$

But because of the componentwise nature of limits, we can distribute the limit into each component of $\vec{\gamma}$, so we can rewrite the speed vector as:

$$\frac{d\vec{\gamma}}{dt}(t) = \begin{bmatrix} \gamma_1'(t) \\ \gamma_2'(t) \\ \vdots \\ \gamma_d'(t) \end{bmatrix}$$

13.7 Differentiability Determination

The mere existance of the set of partial derivatives at $\vec{p}$ with respect to each variable is not enough to guarantee differentiability or continuity at $\vec{p}$, since there must exist some tangential hyperplane, but rather must have a partial derivative in all directions.

13.7.1 Sufficient Condition

Theorem: There is a sufficient condition (true implies, but false does not imply the opposite) for differeniability, where if $f : D \subseteq \mathbb{R} \to \mathbb{R}$ and $\vec{p} \in D^o$, and if $\vec{\nabla} f(\vec{x})$ exists for all points $\vec{x} \in B_\delta(\vec{p})(\exists \delta > 0)$ and is continuous at $\vec{p}$, then f is differentiable at $\vec{p}$.

Proof: Let $\vec{h} = (h_1, h_2, \ldots, h_d)$ and consider the point $\vec{p} + \vec{h}$. We will construct a path from $\vec{p}$ to $\vec{p} + \vec{h}$ such that in each "step" we move h_d in the d^{th} axial direction, and we call each intermediate point $\vec{p}_d$ in the following manner where $\vec{e}_d$ is the d^{th} standard basis vector.

$$\begin{aligned} \vec{p}_0 &= \vec{p} \\ \vec{p}_1 &= \vec{p}_0 + h_1 \vec{e}_1 \\ \vec{p}_2 &= \vec{p}_1 + h_2 \vec{e}_2 \\ \vdots &= \vdots \vdots \vdots \\ \vec{p}_d &= \vec{p}_{d-1} + h_d \vec{e}_d \end{aligned}$$

Now consider the function $\Delta f = f(\vec{p} - \vec{h}) - f(\vec{p})$, we can write this as a telescoping sum:

$$\Delta f = \sum_{j=1}^{d} \{ f(\vec{p}_j) - f(\vec{p}_{j-1}) \}$$

Within each term, we can use the mean value theorem from one dimensional calculus. We can do so because $\vec{p}_j$ and $\vec{p}_{j-1}$ only differ in 1 coordinate, like the following:

$$\begin{aligned} \vec{p}_{j-1} &= (p_1 + h_1, \ldots p_{j-1} + h_{j-1}, \boxed{p_j}, \ldots p_d) \\ \vec{p}_j &= (p_1 + h_1, \ldots p_{j-1} + h_{j-1}, \boxed{p_j + h_j}, \ldots p_d) \end{aligned}$$

We also know that the partial derivatives of f exists for all points within $B_r(\vec{p})$, therefore we can indeed apply the Mean Value Theorem.

Now we apply the MVT:

$$\Delta f = \sum_{j=1}^{d} \partial_{x_j} f(\vec{q_j}) h_j$$

Where $\vec{q_j} = (p_1 + h_1, \ldots, p_{j-1} + h_{j-1}, p_j + \theta h_j, \ldots, p_d)$ where $0 < \theta < 1$. In other words, $\vec{q_j}$ is somewhere between $\vec{p}_{j-1}$ and $\vec{p_j}$ in the j^{th} coordinate.

Now let us consider the definition of differentiability, we need to prove that

$$|\Delta f - \vec{\nabla} f(\vec{p}) \cdot \vec{h}| = |\sum_{j=1}^{d} \partial_{x_j} f(\vec{q}) h_j - \sum_{j=1}^{d} \partial_{x_j} f(\vec{p}) h_j|$$

Now we can apply the triangle inequality on the summations and get:

$$|\Delta f - \vec{\nabla} f(\vec{p}) \cdot \vec{h}| < \sum_{j=1}^{d} |\partial_{x_j} f(\vec{q_j}) - \partial_{x_j} f(\vec{p})||h_j|$$

Now we divide both sides by the length of $\vec{h}$:

$$\begin{aligned} \frac{|\Delta f - f(\vec{p}) \cdot \vec{h}|}{||\vec{h}||} &\leq \sum_{j=1}^{d} |\partial_{x_j} f(\vec{q_j}) - \partial_{x_j} f(\vec{p})| \frac{|h_j|}{||\vec{h}||} \\ &\leq \sum_{j=1}^{d} |\partial_{x_j} (\vec{q_j}) - \partial_{x_j} f(\vec{p})| \end{aligned}$$

Now let $\vec{h} \to \vec{0}$. Then each $\vec{q_j}$ approaches $\vec{p}$. Then we know:

$$\sum_{j=1}^{d} |\partial_{x_j} (\vec{q_j}) - \partial_{x_j} f(\vec{p})| \to 0$$

Since the number of terms in a sum is fixed and each term within the sum is going to 0, the entire sum is going to 0, which means $||\vec{h}|| < \delta(\epsilon)$, we can say that the sum is less than ϵ, which means that

$$\frac{|\Delta f - f(\vec{p}) \cdot \vec{h}|}{||\vec{h}||} \leq \sum_{j=1}^{d} |\partial_{x_j} f(\vec{q_j}) - \partial_{x_j} f(\vec{p})| \frac{|h_j|}{||\vec{h}||} \leq \epsilon$$

Which means that a superlinear decay is possible, which means that f is differentiable.

13.8 Higher-Order Partial Derivatives

13.8.1 Notations

$f \in C^2$ if $\partial_{x_i}\partial_{x_j} f$ is continuous for any $i, j \in \{1, 2, \ldots, d\}$.

Classically, we also define second order partial derivatives as follows:

$$\partial_{x_i}\partial_{x_i} = \partial^2_{x_i} f = \frac{\partial^2 f}{\partial x_i^2}$$
$$\partial_{x_i}\partial_{x_j} = \partial_{x_i}\partial_{x_j} f = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

13.8.2 Partial Equality Theorem

If $f : D \subseteq \mathbb{R} \to \mathbb{R}$ is C^2 (On $B_r(\vec{p})\partial_{x_j}\partial x_j$ is continuous for any $i, j \in 1, 2, 3, ..., d) near \in D^o$, then $\partial_{x_j}\partial_{x_i} f(\vec{p}) = \partial_{x_i}\partial_{x_j} f(\vec{p})$.

Proof: WLOG, assume $i \leq j$. Let $x = x_i$ and $y = x_j$.

Define $g(x, y) = f(p_1, \ldots, p_{i-1}, x, p_{i+1}, \ldots, p_{j-1}, y, p_{j+1}, \ldots, p_d)$.

In terms of g, we can rewrite the partials:

$$\partial_{x_j}\partial_{x_i} f(p_1, \ldots, x, \ldots, y, \ldots p_d) = \partial_y \partial_x g(x, y)$$
$$\partial_{x_i}\partial_{x_j} f(p_1, \ldots, x, \ldots, y, \ldots p_d) = \partial_x \partial_y g(x, y)$$

We know that both partials exist in some neighborhood of p_i, p_j and both are continuous at p_i, p_j. Conside a new function $\Delta(h) = g(x + h, y + h) + g(x, y) - g(x + h, y) - g(x, y + h)$. Let us rearrange some terms, $\Delta(h) = \{g(x+h, y+h) - g(x+h, y)\} - \{g(x, y+h) - g(x, y)\}$. If we define $G(x) = g(x, y + h) - g(x, y)$, we can rewrite $\Delta(h) = G(x + h) - G(x)$. Now apply the MVT, and we get:

$$\Delta(h) = G'(x + \theta_h h) \times h \text{ where } 0 < \theta_h < 1$$

Note that $G'(x) = \partial_x g(x, y + h) - \partial_x g(x, y)$. The derivates also must exist because $C^2 \implies C^1$. Now if we expand $\Delta(h)$, we get:

$$\Delta(h) = [\partial_x g(x + \theta_h h, y + h) - \partial_x g(x + \theta_h h, y)]h$$

Note that here only y is changing, so let us apply the MVT again, and we get:

$$\Delta(h) = \partial_y(\partial_x g)(x + \theta_h h, y + \phi h)h^2 \text{ where } 0 < \phi < 1$$

Now take the limit of $\frac{\Delta(h)}{h^2}$ as $h \to 0$, we get:

$$\lim_{h\to 0} \frac{\Delta(h)}{h^2} = \partial_y(\partial_x g)(x, y)$$

But note that we could have first applied MVT on the y and then on the x due to the symmetry of the function $\Delta(h)$. Therefore, $\partial_y(\partial_x g)(x, y) = \partial_x(\partial_y g)(x, y)$

Corollary: If f is C^k on $B_r(\vec{p})$ for $k \geq 2[\partial x_{i_1} \partial x_{i_2} ... \partial x_{i_k}$ is continuous on $B_r(\vec{p})$ for all $i_1, i_2, ..., i_k \in 1, 2, ..., d$, and if $(i_1, i_2, ..., i_k)$ and $(j_1, j_2, ..., j_k)$ are permutations of one another, then $\partial_{x_{i_1}} ... \partial x_{i_k} f(\vec{p}) = \partial x_{j_1} \partial x_{j_k} f(\vec{p})$.

This is proven fairly trivially by grouping the initially done partials (since function composition is associative), such that there are only 2 partials either grouped, or not grouped, which can then be switched by the theorem.

13.8.3 Partial Continuity-Differentiability Theorems

Theorem: If f is $C^k(k \geq 1)$, then f is also C^{k-1}. Thus, a function can be said to be C^0 if it is continuous itself on that region.

Theorem: If f is differentiable at $\vec{p} \in D^o$, then f is continuous at $\vec{p}$.

Proof: $\forall \epsilon > 0, \exists \delta > 0, \forall \vec{h}$ with $||\vec{h}|| < \delta, |f(\vec{p} + \vec{h}) - f(\vec{p}) - \vec{\nabla} f(\vec{p}) \cdot \vec{h}| < \epsilon ||\vec{h}||$.

Next, by the triangle and Cauchy-Schwartz inequalities, $|f(\vec{p} + \vec{h}) - f(\vec{p})| = |f(\vec{p} + \vec{h}) - f(\vec{p}) - \vec{a} \cdot \vec{h} + \vec{a} \cdot \vec{h}| \leq |f(\vec{p} + \vec{h}) - f(\vec{p}) - \vec{a} \cdot \vec{h}| + |\vec{a} \cdot \vec{h}| < \epsilon ||\vec{h}|| + ||\vec{a}|| ||\vec{h}|| = (1 + ||\vec{a}||)||\vec{h}|| = M||\vec{h}||$ (where M is some constant), such that it can be made less than any constant, showinng continuity.

This is due to the fact that if the partial derivatives exist and are continuous around the point, then it must be differentiable, and as a result, continuous.

Thus, the base case of C^1, then C^0 is true, and it can be done by induction, using the induction step, stating $\partial_{x_{i_1}} \partial_{x_{i_2}} ... \partial x_{i_k} f = \partial x_{i_1} g$, which must be continuous for any $i_1 \in 1, ..., d$, since the original function was continuous. Thus, g is C^1, such that g must be continuous.

14 Linear Algebra

14.1 Linear Mappings/Functions

Consider a function $\vec{l} : \mathbb{R}^d \to \mathbb{R}^e$. $\vec{l}$ is called a **linear mapping** if the image of any k-flat in $\mathbb{R}^d$ $(0 \leq k \leq d)$ is a $\tilde{\text{k}}$-flat in $\mathbb{R}^e$, where $\tilde{k} \leq k$. Basically these definitions “preserve flatness.”

Thus the rigorous definition of a linear function is if $\vec{l} : \mathbb{R} \to \mathbb{R}$ is linear, then $\exists A \in \mathbb{R}^{\times}$ such that $\vec{l}(\vec{x}) = A\vec{x}$, where $\vec{x}$ is viewed as a column matrix (d $\times$1).

14.2 Sufficient Conditions for Linear Mapping

For $\vec{l}$ to be a linear function, it must have the following properties:

1. Additivity: $\vec{l}(\vec{x}+\vec{y}) = \vec{l}(\vec{x}) + \vec{l}(\vec{y})$
2. Homogeneity: $\vec{l}(c\vec{x}) = c\vec{l}(\vec{x})$

Proof: Take k-flat in $\mathbb{R}^d$ $F = \{t_1\vec{a}_1 + t_2\vec{a}_2 + \cdots + t_k\vec{a}_k | t_1, t_2, \ldots, t_k \in \mathbb{R}\}$. Here, $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k \in \mathbb{R}^d$ such that all of them are independent of each other, i.e. no $\vec{a}_i$ is a linear combination of $\{\vec{a}_j | j \neq i\}$. Because otherwise $\vec{a}_i$ can be broken up and the dimension of F will be reduced. This can also be phrased as the set of $\vec{a}_i$ must satisfy the following condition: $F = \vec{0} \implies t_1 = t_2 = \cdots = t_k = 0$. Because if there exists a non-trivial solution, we can subtract all the terms with their coefficient being 0, and divide through the non-zero coefficient, then we would express $\vec{a}_i$ as a linear combination of others. Therefore, if all the vectors are independent, the equation $F = \vec{0}$ only has the trivial solution.

So now consider the linear mapping $\vec{l}$. We can distribute because the mapping is additive, and we can factor out the coefficients because it is homogeneous.

$$\begin{aligned}\vec{l}(F) &= \{\vec{l}(t_1\vec{a}_1 + \cdots + t_k\vec{a}_k) | t_1, \ldots, t_k \in \mathbb{R}\} \\ &= \{t_1\vec{l}(\vec{a}_1) + \cdots + t_k\vec{l}(\vec{a}_k)\}\end{aligned}$$

If we denote $\vec{l}(\vec{a}_i) := b_i$, we can rewrite $\vec{l}(F) = \{t_1 b_1 + \cdots + t_k b_k\}$. This is a $\tilde{\text{k}}$-flat for some $\tilde{\text{k}} \leq$ k, because we can always pick a subset of $\vec{b}_i$ such that they are all independent of each other (unless they are all $\vec{0}$, but in that case $\vec{l}(F)$ is just the 0-flat), and then we reduce, and the final result would be a $\tilde{\text{k}}$-flat.

Example: Homogeneity but not additive and therefore non-linear.

Note that if we are dealing with one dimension, all functions that have an unlimit range are linear mappings. However, in $\mathbb{R}^2$, homogeneity is not enough.

Consider the function:

$$\vec{f}(x,y) = \begin{cases} (\frac{x^3+y^3}{x^2+y^2}, \frac{xy^2}{x^2+y^2}) & \text{if } (x,y) \neq (0,0) \\ (0,0) & \text{if } (x,y) = (0,0) \end{cases}$$

$\vec{f}$ is homogeneous as it is a composition of homogeneous functions, but it does not preserve flatness for obvious reasons, and that is because this function is not additive.

14.3 Matrices

14.3.1 Definition

Matrices are list of vectors, with each column being a single vector. For example, $((1,2,0),(-1,3,4))$ can be rewritten as

$$\begin{vmatrix} 1 & -1 \\ 2 & 3 \\ 0 & 4 \end{vmatrix}$$

This is known as a matrix, and an element of a matrix can be denoted with two subscripts with the lower case of the matrix' name, with the first subscript denoting the row number, and the second denoting the column number. If the above matrix is A, then $a_{11} = 1$ and $a_{31} = 0$.

14.3.2 Determinant

Geometrically, if you treat the matrix $A \in \mathbb{R}^{d \times n}$ as an d-dimensional strucutre composed of n d-dimensional vectors, then the determinant is the "volume" of said structure.

Algebraically

$$\det A = \sum_{\sigma \in S_n} (sgn\sigma) a_{1\sigma_1} a_{2\sigma_2} \dots a_{n\sigma_n}$$
$$\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$$

where

$$sgn\sigma = \frac{\Pi_{i<j}(x_{\sigma_i} - x_{\sigma_j})}{\Pi_{i<j}(x_i - x_j)}$$

where $x_1, x_2, \dots, x_n$ are distant values. This function, in a more explicit/less percise form is:

$$sgn = (-1)^N$$

where N is the number of ordered pairs (i, j) where $i < j$ but $\sigma_j < \sigma_i$.

And S_n is the set of permutations of σ

Properties of the Determinant Function:

1. $\det I = 1$ (unit property)
2. $\det[\vec{a}_1, \dots, c\vec{a}_j, \dots, \vec{a}_n] = c \det A$
3. $\det A = \det A^T$

14.3.3 Operations

The Transpose operation: $B = A^T$ (B is A transposed), then $b_{ij} = a_{ji}$

The dot product: If we have $A = [\vec{a}_1 \vec{a}_2 \dots \vec{a}_n] \in \mathbb{R}^{d \times n}$ and $B = [\vec{b}_1 \vec{b}_2 \dots \vec{b}_m] \in \mathbb{R}^{d \times m}$, then we say that the dot product of the two is:

$$A \cdot B = \begin{vmatrix} \vec{a}_1 \cdot \vec{b}_1 & \vec{a}_1 \cdot \vec{b}_2 & \dots & \vec{a}_1 \cdot \vec{b}_m \\ \vec{a}_2 \cdot \vec{b}_1 & \vec{a}_2 \cdot \vec{b}_2 & \dots & \vec{a}_2 \cdot \vec{b}_m \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_n \cdot \vec{b}_1 & \vec{a}_n \cdot \vec{b}_2 & \dots & \vec{a}_n \cdot \vec{b}_m \end{vmatrix}$$

This operation is distributive, which means that $A \cdot (B + C) = A \cdot B + A \cdot C$. However, this product is not associative, i.e. $A \cdot (B \cdot C) \neq (A \cdot B) \cdot C$.

To solve this problem, we have the matrix multiplication operator.

Matrix Multiplation: For $A \in \mathbb{R}^{n \times d}$, $B \in \mathbb{R}^{d \times m}$, we define the matrix multiplication product to be:

$$\boxed{AB := A^T \cdot B} \in \mathbb{R}^{n \times m}$$

If we call $C = AB$, then we can say that

$$c_{ij} = \sum_{k=1}^{d} a_{ik} b_{kj} \ (1 \leq i \leq n, 1 \leq j \leq m)$$

As a whole, the C column would look like this (here we denote $\vec{\alpha}_i$ as the i^{th} row of A and suppose A has p rows). Then:

$$C = \begin{vmatrix} \vec{\alpha}_1 \cdot \vec{b}_1 & \dots & \vec{\alpha}_1 \cdot \vec{b}_m \\ \vdots & \ddots & \vdots \\ \vec{\alpha}_p \cdot \vec{b}_1 & \dots & \vec{\alpha}_m \cdot \vec{b}_m \end{vmatrix}$$

Note this operation is both distribut and associative, quick proof:

$$A(BC) \stackrel{?}{=} (AB)C$$

$$[A(BC)]_{ij} = \sum_k a_{ik}[BC]_{kj} = \sum_k a_{ik}(\sum_l b_{kl}c_{lj}) = \sum_{(k,l)} a_{ik}b_{kl}c_{lj}$$

$$[(AB)C]_{ij} = \sum_l [AB]_{il}c_{li} = \sum_l (\sum_k a_{ik}b_{kl})c_{lj} = \sum_{(l,k)} a_{ik}b_{kl}c_{lj}$$

Therefore the values of the two matrices are the same. We also have to prove that they are the same size. If $A \in \mathbb{R}^{n \times d}$, $B \in \mathbb{R}^{d \times e}$ and $C \in \mathbb{R}^{e \times m}$. $BC \in \mathbb{R}^{d \times m}$ and $A(BC) \in \mathbb{R}^{n \times m}$. $AB \in \mathbb{R}^{n \times e}$ and $(AB)C \in \mathbb{R}^{n \times m}$. Therefore matrix multiplication is associative.

14.3.4 Examples

:

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = O_{2\times 2} = O$$

Note that in the world of matrices, the product of two non-zero matrices can result in the zero matrix.

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Note that the communicative property does not hold for matrix multiplication.

14.3.5 Norm

Given $A \in \mathbb{R}^{e\times d} = [\vec{a}_1 \vec{a}_2 \dots \vec{a}_d]$, we define the norm of A, $||A||$ as:

$$||A|| := \sqrt{\sum_{j=1}^{d} ||\vec{a}_j||^2} = \sqrt{\sum_{j=1}^{d}\sum_{i=1}^{e} a_{ij}^2}$$

In other words, the norm of a matrix is the sqareroot of the sum of the squares of every element in the matrix.

Properties of the Norm:

1. $||A|| > 0, ||A|| = 0 \leftrightarrow A = O$
2. $||cA|| = |C|||A||$
3. $||A + B|| \leq ||A|| + ||B||$ (Triangle Inequality)
4. $||AB|| \leq ||A||||B||$ (Generalized Cauchy-Schwarz Inequality)

To prove these, we first establish a correspondence between any matrix $A \in \mathbb{R}^{e\times d}$ and a vector in $\mathbb{R}^{ed}$, we define the mapping Ψ from $\mathbb{R}^{e\times d}$ to $\mathbb{R}^{ed}$:

$$A \overset{\Psi}{\longleftrightarrow} (a_{11}, \dots, a_{1d}, a_{21}, \dots, a_{2d}, \dots, a_{e1}, \dots, a_{ed})$$

Note that under this operation, all vector space properties of the matrix are preserved. Note that $||A|| = ||\Psi(A)||$, $\Psi(cA) = cA$, and $\Psi(A + B) = A + B$.

However, note that the product is not exactly preserved with this transformation to vector, since the size of the matrix plays an integral part in matrix multiplication. To prove property 4 of the norm, we need to do a bit of work.

Let $C = AB$, then wwe know that

$$\begin{aligned}||AB||^2 = \sum_i \sum_j \left(\sum_k a_{ik} b_{kj} \right)^2 &= \sum_i \sum_j (\vec{\alpha}_i \cdot \vec{b}_j)^2 \\ &\leq \sum_i \sum_j ||\vec{\alpha}_i||^2 ||\vec{b}_j||^2 \\ &= (\sum_i ||\vec{\alpha}_i||^2)(\sum_j ||\vec{b}_j||^2) \\ &= ||A^T||^2 ||B||^2 \\ &= ||A||^2 ||B||^2\end{aligned}$$

Therefore, $||AB|| \leq ||A|| ||B||$.

14.3.6 Inverses

Suppose there exists $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$, we say that B is a **two sided inverse** of A if $AB = BA = I (= I_n)$, where I is the identity matrix, which functions like the number one in real number multiplication.

Thm: If A has a two-sided inverse, then it has exactly one, namely A^{-1}.

Suppose that B and C are both two-sided inverses for A, i.e. $AB = BA = I$ and $AC = CA = I$. We know we can represent $B = BI = B(AC) = BA(C) = IC = C$. Therefore $B = C$.

Thm: A is invertible iff $\det A \neq 0$.

Let $B = A^{-1}$, $AB = I$. Now we take the determinant of both sides, we get:

$$\det(AB) = \det I$$

Since the determinant is distributive, we get:

$$(\det A)(\det B) = 1$$

Now it's clear that $\det A \neq 0$

14.3.7 Components of Linear Functions

Theorem

Given a linear mapping $\vec{l}(\vec{x})\mathbb{R}^d \to \mathbb{R}^e = (l_1(\vec{x}), l_2(\vec{x}), \ldots, l_e(\vec{x}))$, then we know that each l_i is linear $(\mathbb{R}^d \to \mathbb{R})$.

Proof

Let us first check additivity:

$$\vec{l}(\vec{x}+\vec{y}) = (l_1(\vec{x}+\vec{y}), \ldots, l_e(\vec{x}+\vec{y}))$$
$$\vec{l}(\vec{x}) + \vec{l}(\vec{y}) = (l_1(\vec{x}) + l_1(\vec{y}), \ldots, l_e(\vec{x}) + l_e(\vec{y}))$$

Now if we inspect each element, each l_i is additive. A similar argument can be made to prove homogeneity. Therefore, each l_i is a linear mapping.

14.3.8 Cancellation Law of the Dot Product

Theorem

If $\vec{a} \cdot \vec{x} = \vec{b} \cdot \vec{x}$ for all $\vec{x} \in \mathbb{R}^d$, where $\vec{a}, \vec{b} \in \mathbb{R}^d$, then $\vec{a} = \vec{b}$.

Proof

Take $\vec{x} = \vec{e}_1 = (1, 0, 0, \ldots, 0)$, this yields that $a_1 = b_1$. Then we can take $\vec{x}$ to any of the basic component vectors, we get that $a_2 = b_2, a_3 = b_3, \ldots, a_d = b_d$. Therefore, $\vec{a} = \vec{b}$.

14.3.9 Matrix Valued Functions

It is entirely possible for functions to give out matrices as its output. Suppose $A : D \subseteq \mathbb{R}^d \to \mathbb{R}^{e \times k}$. Similar to how vector valued functions have component functions, matrix valued functions have entry functions. A would look something like this:

$$A(\vec{x}) = \begin{bmatrix} a_{11}(\vec{x}) & \ldots & a_{1k}(\vec{x}) \\ \vdots & \vdots & \vdots \\ a_{e1}(\vec{x}) & & a_{ek}(\vec{x}) \end{bmatrix}$$

Continuity for such functions is an entrywise property: A is continuous at $\vec{p}$ iff a_{ij} is continuous at $\vec{p}\ \forall i, j$

14.4 Unique Expression of Linear Functions

14.4.1 Theorem

For any linear map $\vec{l} : \mathbb{R}^d \to \mathbb{R}^e$. There is an unique matrix $A \in \mathbb{R}^{e \times d}$ (d columns and e rows) such that $\vec{l}(\vec{x}) = A\vec{x}$ where the right hand size is a matrix product, where $\vec{x}$ is regarded as a $d \times 1$ column.

14.4.2 Proof

$\vec{l}(\vec{X}) = (l_1(\vec{x}), l_2(\vec{x}), ..., l_e(\vec{x}))$, such that each l_j is a linear real-valued function.

Claim: A real valued linear function, $\Gamma : \mathbb{R} \to \mathbb{R}$ has the form $\Gamma(\vec{x}) = \vec{a} \cdot \vec{x}$ for some $\vec{a} \in \mathbb{R}$.

$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + ... + x_d\vec{e}_d = (x_1, x_2, ..., x_d)$, and $\Gamma(\vec{x}) = \Gamma(x_1\vec{e}_1 + ... + x_d\vec{e}_d) = \Gamma(\vec{e}_1)x_1 + \Gamma(\vec{e}_2)x_2 + ... + \Gamma(\vec{e}_d)x_d$. If each term, $\Gamma(\vec{e}_n)x_n = a_n$, then $\Gamma(\vec{x}) = \vec{a} \cdot \vec{x}$.

Claim: If $\cdot\vec{x} = \vec{b} \cdot \vec{x}$ for all $\vec{x} \in \mathbb{R}$, where $\vec{a}$, $\vec{b} \in \mathbb{R}$, then $\vec{a} = \vec{b}$.

For some $\vec{x} = \vec{e_1} = (1, 0, 0, ..., 0)$, then $a_1 = \vec{a} \cdot \vec{e_1} = \vec{b} \cdot \vec{e_1} = b_1$, and so forth.

Thus, $\vec{l}(\vec{x}) = (\vec{\alpha}_1 \cdot \vec{x}, \vec{\alpha}_2 \cdot \vec{x}, ..., \vec{\alpha}_d \cdot \vec{x})$, able to be written as a column matrix of the vector components, which by the definition of the dot product, can be written as a column matrix of α^T, multiplied by $\vec{x} = \mathrm{A}\vec{x} = \mathrm{B}\vec{x}$.

15 Differentiability of Vector Valued Functions

15.1 Definition of Differentiability

Let $\vec{f} : D \subseteq \mathbb{R}^d \to \mathbb{R}^e$. Let $\vec{p} \in D^\circ$. Then $\vec{f}$ is differentiable at $\vec{p}$ if $\exists A \in \mathbb{R}^{e \times d}$ such that

$$\boxed{\forall \epsilon > 0, \exists \delta > 0 : 0 < ||\vec{h}|| < \delta \implies ||\vec{f}(\vec{p} + \vec{h}) - \vec{f}(\vec{p}) - A\vec{h}|| < \epsilon ||\vec{h}||}$$

If A exists, then we call it the derivative of $\vec{f}$ at point $\vec{p}$, say that $\vec{f}$ is differentiable at $\vec{p}$. We write the derivative as $D\vec{f}(\vec{p}) \in \mathbb{R}^{e \times d}$, and $D\vec{f}(\vec{p})_{ij}$ (Jacobian derivative) $= \partial_{x_j} f_i(\vec{p})$.

15.2 The Jacobian Derivative

Claim: The above definition decomposes componentwise.

This is done by the basic norm bounds ($|v_j| \leq ||\vec{v}|| \leq \sqrt{d} max_{1 \leq k \leq d} |v_k|$).

It can then be shown that $|f_i(\vec{p} + \vec{h}) - f_i(\vec{p}) - \vec{\alpha}_i \cdot \vec{h}| < \epsilon ||\vec{h}||$, where $\vec{\alpha}_i^T$ is the i-th row of A. Thus, $\vec{\alpha}_i = \vec{\nabla} f_i(\vec{p})$, such that $\vec{\alpha}_i$ is unique, such that A is unique as well.

Then D$\vec{f}$ is the column matrix of the transposed gradient of each component function, or the column matrix of the Jacobian derivative of each component function, or the matrix of the partial derivatives, such that $[D\vec{f}]_i j$ (the ij-th entry) $= \partial_{x_j} f_i = \frac{\partial f_i}{\partial x_j}$. The final interpretation is the row matrix of $\partial_{x_i} \vec{f}$.

16 The Gradient Operator

16.1 Basic Rules

Since all rules of single-variable derivative operators apply to partial derivative operators, due to functioning similarly, just holding all but one variable constant:

- $\vec{\nabla}(f + g) = \vec{\nabla} f + \vec{\nabla} g$
- $\vec{\nabla}(cf) = c\vec{\nabla} f$
- $\vec{\nabla}(fg) = f\vec{\nabla} g + g\vec{\nabla} f$
- $\vec{\nabla}(\frac{f}{g}) = \frac{g\vec{\nabla} f - f\vec{\nabla} g}{g^2}$

16.2 Chain Rule

If $\vec{f} : D \subseteq \mathbb{R} \to \mathbb{R}$, $\vec{g} : E \subseteq \mathbb{R} \to \mathbb{R}^{\urcorner}$, $\vec{f}(D) \cap E \neq \emptyset$, and $\vec{p} \in D^o$ such that $\vec{f}(\vec{p}) \in E^o$. Then, if $\vec{f}$ is differentiable at $\vec{p}$, and $\vec{g}$ is differentiable at $\vec{q} = \vec{f}(\vec{p})$, then $\vec{g} \circ \vec{f}$ is also differentiable at $\vec{p}$, and $D(\vec{g} \circ \vec{f})(\vec{p}) = D\vec{g}(\vec{f}(\vec{p}))D\vec{f}(\vec{p})$.

Thus, it is the same rule as in single-variable calculus, though matrix multiplication is used rather than regular multiplication.

In other words, $[D(\vec{g} \circ \vec{f})]_{ij} = \partial_{x_j}(g_i \circ \vec{f})(\vec{p})$. Classically, this has also been written as $\frac{\partial z_i}{\partial x_j}(\vec{p})$. Note that $[D\vec{g}(\vec{q})D\vec{f}(\vec{p})]_{ij} = \sum_{l=1}^{e} \partial_{y_l} g_i(\vec{q}) \partial_{x_j} f_l(\vec{p})$. This has been classically been written as:

$$\frac{\partial z_i}{\partial x_j} = \sum_{l=1}^{e} \frac{\partial z_i}{\partial y_l}(\vec{f}(\vec{p}))\frac{\partial y_l}{\partial x_j}(\vec{p})$$

The “legit” definition of the chain rule is:

$$\partial_{x_j}(g_i \circ \vec{f})(\vec{p}) = \sum_{l=1}^{e} \partial_{y_l} g_i(\vec{q}) \partial_{x_j f_l(\vec{p})}$$

This is also clasically written as:

$$\frac{\partial z_i}{\partial x_j} = \sum_{l=1}^{e} \frac{\partial z_i}{\partial y_l}\frac{\partial y_l}{\partial x_j}$$

16.2.1 Proof

For simplicity's sake, let $B := D\vec{g}(\vec{f}(\vec{p}))$ and $A := D\vec{f}(\vec{p})$. We know that A satisfies $\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0, \forall \vec{h}$ with $||\vec{h}|| < \delta(\varepsilon)$:

$$||\Delta \vec{f}(\vec{p}, \vec{h}) - A\vec{h}|| < \varepsilon ||\vec{h}||$$

Let us call this condition $*_{\vec{f}}$

We also know that B satisfies $\forall \varepsilon' > 0, \exists \delta'(\varepsilon') > 0, \forall \vec{k}$ with $||\vec{k}|| < \delta'(\varepsilon')$:

$$||\Delta \vec{g}(\vec{q}, \vec{k}) - B\vec{k}|| < \varepsilon' ||\vec{k}||$$

Let us call this condition $*_{\vec{g}}$

We seek to prove the following:

$\forall \varepsilon_2 > 0, \exists \delta_2(\varepsilon_2) > 0, \forall \vec{h}$ with $||\vec{h}|| < \delta_2(\varepsilon_2)$:

$$||(\vec{g} \circ \vec{f})(\vec{p} + \vec{h}) - (\vec{g} \circ \vec{f})(\vec{p}) - (BA)\vec{h}|| \overset{?}{<} \varepsilon_2 ||\vec{h}||$$

Let's define a vector $\vec{k} := \Delta\vec{f} = \vec{f}(\vec{p}+\vec{h}) - \vec{f}(\vec{p})$. Note that since $\vec{q} = \vec{f}(\vec{p})$, $\vec{q}+\vec{k} = \vec{f}(\vec{p}+\vec{h})$. We can therefore rewrite the equation as:

$$||\vec{g}(\vec{q}+\vec{k}) - \vec{g}(\vec{q}) - B\vec{k} + B\vec{k} - (BA)\vec{h}|| \overset{?}{<} \varepsilon_2||\vec{h}||$$

Now we left factor B out of the left hand side:

$$||\vec{g}(\vec{q}+\vec{k}) - \vec{g}(\vec{q}) - B\vec{k} + B(\vec{k} - A\vec{h})|| \overset{?}{<} \varepsilon_2||\vec{h}||$$

Now we can apply the triangle inequality and the generalized Cauchy-Schwarz Inequality to get the following:

$$||\vec{g}(\vec{q}+\vec{k}) - \vec{g}(\vec{q}) - B\vec{k} + B(\vec{k} - A\vec{h})|| \leq ||\vec{g}(\vec{q}+\vec{k}) - \vec{g}(\vec{q}) - B\vec{k}|| + ||B||||\vec{k} - A\vec{h}||$$

But note that the first part of the right hand side is condition $*_{\vec{g}}$, so we get:

$$||\vec{g}(\vec{q}+\vec{k}) - \vec{g}(\vec{q}) - B\vec{k}|| + ||B||||\vec{k} - A\vec{h}|| < \varepsilon'||\vec{k}|| + ||B||||\vec{k} - A\vec{h}|| = \varepsilon'||\vec{k} - A\vec{h} + A\vec{h}|| + ||B||||\vec{k} - A\vec{h}||$$

Now we can apply the same trick, except WLOG we assume that $||\vec{k}|| < \delta'(\varepsilon')$ and assume $\varepsilon' \leq 1$. we get:

$$\varepsilon'||\vec{k} - A\vec{h} + A\vec{h}|| + ||B||||\vec{k} - A\vec{h}|| \leq \varepsilon'||\vec{k} - A\vec{h}|| + \varepsilon'||A||||\vec{h}|| + ||B||||\vec{k} - A\vec{h}||$$

Now we factor and apply the bounds on ε':

$$\varepsilon'||\vec{k} - A\vec{h}|| + \varepsilon'||A||||\vec{h}|| + ||B||||\vec{k} - A\vec{h}|| \leq (||B|| + 1)||\vec{k} - A\vec{h}|| + ||A||||\vec{h}||\varepsilon'$$

Note that $||\vec{k} - A\vec{h}|| = ||\vec{f}(\vec{p}+\vec{h}) - \vec{f}(\vec{p}) - A\vec{h}|| < \varepsilon||\vec{h}||$. From this we get:

$$(||B|| + 1)||\vec{k} - A\vec{h}|| + ||A||||\vec{h}||\varepsilon' \leq (||B|| + 1)\varepsilon||\vec{h}|| + ||A||||\vec{h}||\varepsilon'$$

If we take $\varepsilon = \varepsilon'$, we get:

$$(||B|| + 1)\varepsilon||\vec{h}|| + ||A||||\vec{h}||\varepsilon' \leq (||A|| + ||B|| + 1)\varepsilon||\vec{h}||$$

Note that $(||A|| + ||B|| + 1)\varepsilon$ can be arbitarily small because it is a constant times an arbitarily small value. If we call this value ε_2, we get that

$$||(\vec{g} \circ \vec{f})(\vec{p}+\vec{h}) - (\vec{g} \circ \vec{f})(\vec{p}) - (BA)\vec{h}|| < \varepsilon_2||\vec{h}||$$

16.2.2 Special Cases

Consider the following "chain": $\mathbb{R}^d_{\vec{x}} \supseteq D \xrightarrow{\vec{f}} \mathbb{R}^e_{\vec{y}} \supseteq E \cap \vec{f}(D) \xrightarrow{\vec{g}} \mathbb{R}^k_{\vec{z}}$. The special case we're interested in is if $d = 1$ and $k = 1$. In this case, the original $\vec{x}$-axis can be thought of as the time axis, and $\vec{f}$ can be thought of as a path $\vec{\gamma} : I \to \mathbb{R}^e$ where $D = I$, an interval, $I \subseteq \mathbb{R}$. The difference between this ans the general chain rule is that the domain is connected (because the domain is time). This way, the "chain" reduces to:

$$\mathbb{R} \supseteq D \xrightarrow{(g \circ \vec{\gamma})} \mathbb{R}$$

Therefore the derivative is an ordinary single-calculus derivative. However, to calculate the derivative with the chain rule, we have to do the following;

$$(g \circ \vec{\gamma})'(t) = Dg(\vec{\gamma}(t))D\vec{\gamma}(t)$$

This can be written as:

$$(g \circ \vec{\gamma})'(t) = g(\vec{\gamma}(t))^T \vec{\gamma}'(t) = \boxed{g(\vec{\gamma}(t)) \cdot \vec{\gamma}'(t)}$$

Here, $\vec{\gamma}'(t)$ is known as the **velocity vector** of the curve, which is the column matrix (vector) of the derivatives of the component functions of $\vec{\gamma}$ evaluated at t. It turns out that this value is also equal to:

$$\vec{\gamma}'(t) = \lim_{h \to 0} \frac{\vec{\gamma}(t+h) - \vec{\gamma}(t)}{h}$$

This is because all the operations used here distributes across components. This is the physics-y interpretation of the derivative. The magnitude of this vector is the instataneous speed of the curve.

$$||\vec{\gamma}'(t)|| = \left|\left| \lim_{h \to 0} \frac{\vec{\gamma}(t+h) - \vec{\gamma}(t)}{h} \right|\right| = \lim_{h \to 0^+} \frac{||\vec{\gamma}(t+h) - \vec{\gamma}(t)||}{h}$$

However, this physical interpretation is also the same as the Jacobian Derivative of $\vec{\gamma}$. But since there is only one input variable, there is only one partial to compute, $\partial_t \gamma_1$, otherwise known as γ_1'. Therefore, the Jacobian Derivative becomes a column vector of the derivatives of the component functions.

16.3 Rolle's Theorem

16.3.1 Theorem

If f is differentiable on $(a - r, b + r)$ where $r > 0$ and $a < b$, and $f(a) = f(b)$. Then $\exists c \in (a, b)$ such that $f'(c) = 0$.

16.3.2 Proof

f is either constant or it is not on $[a,b]$. If it is constant, then the theorem is trivially true. If not, then there exists $p \in [a,b]$ such that $f(p) \neq f(a) = f(b)$. If $f(p) > f(a)$, then take c to be a global maximum point for the function f on the interval $[a,b]$ (EVT). Note that $c \neq a$, $c \neq b$, so $c \in (a,b)$. Let us then consider $f'(c)$. We are given that

$$\lim_{h \to 0} \frac{f(c+h) - f(c)}{h}$$

But if we consider the limits from two directions, we get:

$$\lim_{h \to 0^+} \frac{f(c+h) - f(c)}{h} \leq 0$$

$$\lim_{h \to 0^-} \frac{f(c+h) - f(c)}{h} \geq 0$$

Therefore, $f'(c) = 0$. A similar argument can be made if $f(p) < f(a$

16.4 Mean Value Theorem

16.4.1 Review of the Single Variable Theorem

If $f : [a,b] \to \mathbb{R}$ is differentiable on $(a-r, b+r)$ where $r > 0$ and continuous on $[a,b]$, then $\Delta f = f(b) - f(a) = f'(c)(b-a)$ where $c \in (a,b)$.

Proof

For the given function f, let us create a function $l : y = f(a) + \frac{f(b)-f(a)}{b-a}(x-a)$, or a secant line connecting a to b. Let $g(x) = f(x) - l(x)$. Then, $g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$. Note also $g(a) = 0 = g(b)$. Then by Rolle's Theorem, $\exists c \in (a,b)$ such that $g'(c) = 0$. This implies that at c, $f(c)$ has the same derivative as $l(c)$, or $\frac{f(b)-f(a)}{b-a}$.

16.4.2 Theorem

Exists $f : D \subseteq \mathbb{R}^d \to \mathbb{R}$, $\vec{p}, \vec{q} \in D^\circ$, $[\vec{p}, \vec{q}] \subseteq D^\circ$ [1]. Also assume that f is differentiable on D. Then $\exists \vec{r} \in (\vec{p}, \vec{q})$ such that:

$$\boxed{f(\vec{q}) - f(\vec{p}) = f(\vec{r}) \cdot (\vec{q} - \vec{p})}$$

[1] This represents a **closed segment** with the end points of $\vec{p}$ and $\vec{q}$, which can be expressed as $[\vec{p}, \vec{q}] = \{(1-t)\vec{p} + t\vec{q} | 0 \leq t \leq 1\}$. An **open segment** is written as $(\vec{p}, \vec{q}) = \{(1-t)\vec{p} + t\vec{q} | 0 < t < 1\}$

16.4.3 Proof

Let $g(t) = f((1-t)\vec{p} - t\vec{q})$, where $-\varepsilon \leq t \leq 1+\varepsilon$ for some $\varepsilon > 0$. We can do this because $\vec{p}, \vec{q} \in D^{\circ}$. Now we can try to take the derivative of $g(t)$. To do this we use the chain rule: $g'(t) = f((1-t)\vec{p} + t\vec{q}) \cdot (\vec{q} - \vec{p})$. We now apply the Mean Value Theorem to g, we know that $\exists c \in (0,1)$ such that $g'(c) =$

1 - 0 = f() - f().$Nowwehave:$

$$f(\vec{q}) - f(\vec{p}) = f((1-c)\vec{p} + c\vec{q}) \cdot (\vec{q} - vecp)$$

Now we simply call $(1-c)\vec{p} + c\vec{q}\ \vec{r}$, and we're done.

16.4.4 For Vector Valued Functions

Sadly there is no generalization of the MVT to vector valued functions. If we apply the MVT to each of its component functions, we see that for each $\vec{r}$ may be different. Therefore there is no easy, consistent generalization.

In order to actually generalize this, we will need to make the statement weaker.

We can at least provide a bound for $|f(\vec{p}) - f(\vec{q})|$. By Cauchy-Schwarz inequality, we get:

$$|f(\vec{q}) - f(\vec{p})| \leq ||f(\vec{r})||||\vec{q} - \vec{p}||$$

for some $\vec{r} \in (\vec{p}, \vec{q})$. However, we can write the most general case as:

$$\boxed{|f(\vec{q}) - f(\vec{p})| \leq \left(\sup_{\vec{x} \in [\vec{p}, \vec{q}]} ||f(\vec{x})||\right) ||\vec{q} - \vec{p}||}$$

This is known as the **Mean Value Inequality**. And as a corollary, we can take any B such that if $||f|| \leq B$ on $[\vec{p}, \vec{q}]$ then:

$$|f(\vec{q}) - f(\vec{p})| \leq B||\vec{q} - \vec{p}||$$

To generalize the Mean Value Inequality to vector valued function, let $\vec{f} : D \subseteq \mathbb{R}^d \to \mathbb{R}^e$, and $[\vec{p}, \vec{q}] \subseteq D^{\circ}$. If $\vec{f}$ is C^1 [2] in some open set U such that $[\vec{p}, \vec{q}] \subseteq U \subseteq D^{\circ}$.

$$\boxed{\therefore ||\vec{f}(\vec{q}) - \vec{f}(\vec{p})|| \leq \left(\max_{[\vec{p}, \vec{q}]} ||D\vec{f}||\right) ||\vec{q} - \vec{p}||}$$

Let $M := \max_{[\vec{p}, \vec{q}]} ||D\vec{f}||$. Fix $\varepsilon > 0$. Let $D = \{t \in [0,1] || ||\vec{f}(\vec{q}_t) - \vec{f}(\vec{p})|| \leq (M+\varepsilon)||\vec{q}_t = \vec{p}||\}$ where $\vec{q}_{:} = (1-t)\vec{p} + t\vec{q}$. Note that $\vec{q}_0 = \vec{p}$ and $\vec{q}_1 = \vec{q}$.

[2] For vector valued function, C^n over U means that every $\frac{\partial^n f_i}{(\partial x_j)^n}(\vec{x})$ is continuous for all $\vec{x} \in U$, $1 \leq i \leq e$ and $q \leq j \leq d$

Note that $0 \in S$, which means that $S \neq \emptyset$. Also note that $S \leq 1$ because $t \in [0,1]$. Therefore $\sup S$ exists and finite (bounded by $[0,1]$), let us call this value T. This means that $\exists t_n \in S$ such that $t_n \to T^-$. Therefore, $\forall n, ||\vec{f}(\vec{q}_{t_n}) - || \leq (M+\varepsilon)||\vec{q}_{t_n} - \vec{p}||$. Now take the limit as $n \to \infty$ and we get:

$$||\vec{f}(\vec{q}_T) - || \leq (M+\varepsilon)||\vec{q}_T - \vec{p}||$$

This means that $T \in S$ as well. Assume for contradiction that $T < 1$. It will follow that $T = 1$. Because $\vec{q}_1 = \vec{q}$, it follows that:

$$||\vec{f}(\vec{q}) - || < (M+\varepsilon)||\vec{q} - \vec{p}||$$

Now take the limit as $\varepsilon \to 0^+$. We can then conclude:

$$||\vec{f}(\vec{q}) - || \leq M||\vec{q} - \vec{p}||$$

Say $T < 1$. Choose $\delta > 0$ such that $\tau = T + \delta \leq 1$. We now seek to show:

$$||\vec{f}(\vec{q}_\tau) - \vec{f}(\vec{p})|| \overset{?}{\leq} (M+\varepsilon)||\vec{q}_\tau - \vec{p}||$$

We first add a bunch of terms to the left hand side and use the triangle inequality.

$$||\vec{f}(\vec{q}_\tau) - \vec{f}(\vec{p})|| = ||\vec{f}(\vec{q}_\tau) - \vec{f}(\vec{q}_T) - A_T(\vec{q}_\tau - \vec{q}_T)||$$

16.5 Implicit Function Theorem

Let $F : D \subseteq \mathbb{R}^{d+1} \to \mathbb{R}$ be C^1 on D^o. Suppose $F(\vec{A}) = 0$, where $\vec{A} = (\vec{a}, \alpha), \vec{a} \in \mathbb{R}^d, \alpha \in \mathbb{R}$. Also suppose $\partial_y F(\vec{A}) \neq 0$. Here, we write $F(\vec{x}, y) = F(x_1, x_2, ..., x_d, y) = F(\vec{X})$. Then $\exists$ an open set $U \subseteq \mathbb{R}^d$ and an open interval $I \subseteq \mathbb{R}$, such that $\vec{a} \in U$ and $\alpha \in I$ and $UxI \subseteq D^o$, and $\exists$ a function $f : U \to I$, such that $\{\vec{X} \in UxI | F(\vec{X}) = 0\} = G_f := \{(\vec{x}, f(\vec{x})) | \vec{x} \in U\}$. Also f is also C^1, and

$$\vec{\nabla} f(\vec{x}) = -\frac{1}{\partial_y F(\vec{x}, f(\vec{x}))} \vec{\nabla}_{\vec{x}F(\vec{x}, f(\vec{x}))}$$

, where $\vec{\nabla}_{\vec{x}F(\vec{x})=(\partial_{x_1}F, \partial_{x_2}F, ..., \partial_{x_d}F)}$, called the partial gradient.

16.5.1 Meaning

If F is C^1 on its domain $D \subseteq \mathbb{R}^{d+1}, \vec{A} \in D^o, F(\vec{A}) = 0$, and $\partial_y F(\vec{A}) \neq 0, F(\vec{x}, y) = 0$ can be solved inplicticly for y as a function of $vecx$ locally in some neighborhood of a solution point, $\vec{A} = (\vec{a}, \alpha)$.

16.5.2 Proof

The implicit function theorem is done in three parts, first by proving the existance of f, then proving the continuity of f, and finally that f is C^1, allowing the equation to be proved.

For the first part, a cube (or any convex set) is drawn with an axis from the center of the top and bottom faces, where the midpoint of the axes is said to be $\vec{A}$. $\partial_y F(\vec{A}) \neq 0$ and it can be assumed to be either positive or negative, not changing the proof. Let $g_{\vec{a}(y)=F(\vec{a},y)}$. When $y = \alpha, g'_{\vec{a}(y)>0}$, meaning $\partial_y F(\vec{x}, y)$ is continuous since F is C^1. By continuity, we can assume that the partial derivative is positive for all points within the cube, such that there exists a cube where that is true by the continuity. Thus, the derivative is defined on the axis, such that $\vec{a}$ is fixed. By the theorem assumptions, $F(\vec{A}) = 0, g_{\vec{a}(\alpha)=0}$, such that $\exists$ some point $p > \alpha$ and $-p < \alpha$, such that the function value is greater and less than 0 respectively.

By the continuity of F, $g_{\vec{a}(y)}$ is continuous, such that a square can be drawn around $g_{\vec{a}(p)}$, such that F ¿ 0 for all points within by continuity. We can do the same for p, gaining edge lengths of r^+ and r^- respectively, such that the minimum edge length forms the open boundary of U, and I := (-p, p). For each $\vec{a}$ by the Intermediate Value Theorem and the derivative, there must be exactly one point such that $F(\vec{a}, \alpha) = 0$, such that the function of those α is f, proving the existance of the implicit function.

For the second part, since for all $\vec{x}, F(\vec{x}, f(\vec{x})) = 0$ for all $\vec{x} \in U$, we can add $\vec{h}$ to $\vec{x}$, such that it is still in U, such that $F(\vec{x} + \vec{h}, f(\vec{x} + \vec{h})) - F(\vec{x}, f(\vec{x})) = 0$, or $\vec{\nabla} F(\vec{X}^*) \cdot \vec{H} = 0$, by the mean value theorem where $\vec{X}^* \in [(\vec{x}, f(\vec{x})), (\vec{x} + \vec{h}, f(\vec{x} + \vec{h}))]$. Further, we can seperate the gradient, such that $\vec{\nabla}_{\vec{x}F(\vec{X}^*)\cdot\vec{h}+\partial_y F(\vec{X}^*)\delta f} = 0$. Since if f is continuious, $\delta f = f(\vec{x} + \vec{h}) - f(\vec{x}) \to 0$ as $\vec{h} \to 0$ is true, we use that to prove the continuity of f.

We take the absolute value of the two sides of the equation, such that $|-\partial_y F(\vec{X}^*)\delta f| = |\vec{\nabla}_{\vec{x}F(\vec{x}^*)\cdot\vec{h}| \leq ||\vec{\nabla}_{\vec{x}F(\vec{X}^*)}|| ||\vec{h}|| \leq ||\vec{\nabla} F(}$ by Cauchy-Schwartz Inequality. Since F is C^1, then $\vec{\nabla} F$ is contnuous on U x I, and $||\vec{\nabla} F||$ is continuous on U x I, let M := $max_{UxI}||\vec{\nabla} F||$. We then use the fact that the box is convex, such that any two points within the convex region, the line segment joining them is contained within the region. Thus, $|\partial_y F(\vec{X}^*)||\delta f| \leq M||\vec{h}||$, isolating δf. We can also say that $\partial_y F(\vec{X}^*) \geq \frac{1}{2}\partial_y F(\vec{A})$, by continuity saying we can create some point $\vec{X}^*$, such that it is true, such that we can create a cube subset of the original cube, such that it is true for all points, $\vec{X}$. Since $\frac{2M||\vec{H}||}{\partial_y F(\vec{A})} \to 0$ as $\vec{h} \to \vec{0}$, we can state the same thing of δf, since it must be greater than 0, proving continuity.

For the third part, let $\vec{h} = h\vec{e}_j = (0, ..., 1, ..., 0)$, where the 1 is in the j^{th} place. By the previous equation, $-h(\vec{\nabla}_{\vec{x}F(\vec{X}^*)\cdot\vec{e}_j)=\partial_y F(\vec{X}^*)\delta f}$, after which we divide by h on both sides. As $h \to 0, \frac{\delta f}{h} = -\frac{\vec{\nabla}_{\vec{x}F(\vec{X}^*)\cdot\vec{e}_j}}{\partial_y F(\vec{X}^*)} \to -\frac{\vec{\nabla}_{\vec{x}F(\vec{x},f(\vec{x}))\cdot\vec{e}_j}}{\partial_y F(\vec{x},f(\vec{x}))} = -\frac{\partial_{x_j} F(\vec{x},f(\vec{x}))}{\partial_y F(\vec{x},f(\vec{x})}$. Since this is equal to $\partial_{x_j} f(\vec{x})$, both can be shown easily to be differentiable.

16.5.3 Theorem Summary

$F(\vec{x}, y) = 0$ can be written as $y = f(x)$ locally in a neighborhood (for all $(\vec{x}, y) \in UxI$) of some solution point, $(\vec{a}, \alpha) \in D_F^o \subseteq \mathbb{R}^{d+1}$, when $\partial_y F(\vec{a}, \alpha) \neq 0$ and F is C^1 on D^o.

In addition, it follows that f is also C^1, and $\vec{\nabla} f = \frac{-1}{\partial_y Fo(Idxf)} \vec{\nabla}_{\vec{x}Fo(Idxf)}$, where $(Idxf)(\vec{x}) =$

$(\vec{x}, f(\vec{x}))$, where Id is the identity function which makes that valid.

16.5.4 Generalization

$F(\vec{x}, y) = 0$ is equivalent to $y = f(\vec{x})$ locally in a neighborhood of some solution point, $(\vec{a}, \alpha)$ if F is C^1, $F(\vec{a}, \alpha) = 0$, $(\vec{a}, \alpha) \in D_F^\circ \subseteq \mathbb{R}^{d+1}$ and $\partial_y F(\vec{a}, \alpha)$

However, if $\vec{F}$ is a vector-valued function, we also have this type of equivalency ($\vec{F}(\vec{x}, \vec{y}) = \vec{0}$ is equivalent to $\vec{y} = \vec{f}(\vec{x})$ for $\vec{x} \in U \subseteq \mathbb{R}^d$ and $\vec{y} \in V \subseteq \mathbb{R}^e$) locally around some solution point $\vec{F}(\vec{a}, \vec{b}) = \vec{0} \in \mathbb{R}^e$ if $\vec{F}$ is C^1 on $D_{\vec{F}}^\circ$[3] and $(\vec{a}, \vec{b}) \in D_{\vec{F}}^\circ \subseteq \mathbb{R}^{d+e}$ and $\partial_{\vec{y}} \vec{F}(\vec{a}, \vec{b}) \neq 0$ where $\partial_{\vec{y}} \vec{F} = \det D_{\vec{y}} \vec{F}$ where $D_{\vec{y}}$ (known as the **partial Jacobian**) can be described as:

$$D_{\vec{y}} F = \left[\frac{\partial F_i}{\partial y_j} \right]_{\substack{1 \le i \le e \\ 1 \le j \le e}}$$

Similarly, we know that if all the conditions are met, we know that $\vec{f}$ is C^1. We also have a semi-explicit formula for the Jacobian of f:

$$\boxed{D\vec{f} = -(D_{\vec{y}} \vec{F})^{-1} D_{\vec{x}} \vec{F}}$$

By definition we know $\vec{F}(\vec{x}, \vec{f}(\vec{x})) \equiv \vec{0}$ for all $\vec{x} \in U$. Now we can differentiate:

$$D_{\vec{x}} \vec{F}(\vec{x}, \vec{f}(\vec{x})) \equiv D_{\vec{x}} \vec{0} = O \in \mathbb{R}^{e \times d}$$

But we can rewrite the left hand side as $D_{\vec{x}}[\vec{F} \circ (Id \times \vec{f})]$, here we can apply the chain rule:

$$(D_{\vec{x}} \vec{F})(\vec{x}, \vec{f}(\vec{x}))(D_{\vec{x}}(Id \times \vec{f}))(\vec{x}) = O$$

We can express $(Id \times \vec{f})(\vec{x})$ as $(\vec{x}, \vec{f}(\vec{x}))$, which can be written as

$$\begin{bmatrix} x_1 \\ \vdots \\ x_d \\ f_1(\vec{x}) \\ \vdots \\ f_e(\vec{x}) \end{bmatrix} = \begin{bmatrix} \vec{x} \\ \vec{f}(\vec{x}) \end{bmatrix}$$

Since the Jacobian Operator works row by row, we can distribute it

$$D_{\vec{x}} \begin{bmatrix} \vec{x} \\ \vec{f}(\vec{x}) \end{bmatrix} = \begin{bmatrix} D_{\vec{x}} \vec{x} \\ D_{\vec{x}} \vec{f}(\vec{x}) \end{bmatrix} = \begin{bmatrix} I_d \\ D\vec{f}(\vec{x}) \end{bmatrix}$$

[3] A vector valued function is C^1 if all of its component functions are C^1

16.5.5 Example

$$(*)\begin{cases} F(u,v,x,y,z)=0 \\ G(u,v,x,y,z)=0 \\ H(u,v,x,y,z)=0 \end{cases}$$

Suppose (a,b,p,q,r) is a specific solution for (u,v,x,y,z) in $(*)$ where $(a,b,p,q,r)\in D^{\circ}$

We can rewrite this as:

$$\begin{aligned} F&: px+qy+rz=-(au+bv) \\ G&: p'x+q'y+r'z=-(a'u+b'v) \\ H&: p''x+q''y+r''z=-(a''u+b''v) \end{aligned}$$

Here we can rewrite this as the matrix equation $A\vec{x}=\vec{b}$ where

$$A=\begin{bmatrix} p & q & r \\ p' & q' & r' \\ p'' & q'' & r'' \end{bmatrix} \vec{x}=\begin{bmatrix} x \\ y \\ z \end{bmatrix} \vec{b}=\begin{bmatrix} -(au+bv) \\ -(a'u+b'v) \\ -(a''u+b''v) \end{bmatrix}$$

To solve this matrix equation we need that $\det A\neq 0$. We can also write A as:

$$\begin{bmatrix} F \\ G \\ H \end{bmatrix}$$

We know that

$$\partial_{(x,y,z)}(F,G,H)=\begin{vmatrix} \partial_x F & \partial_y F & \partial_z F \\ \partial_x G & \partial_y G & \partial_z G \\ \partial_x H & \partial_y H & \partial_z H \end{vmatrix}$$

The Implicit Function Theorem says that if $\partial_{(x,y,z)}(F,G,H)|_{(a,b,p,q,r)}\neq 0$ then $(*)$ is locally equivalent near (a,b,p,q,r) to:

$$\begin{cases} x=f(u,v) \\ y=g(u,v) \\ z=h(u,v) \end{cases}$$

And we call the vector function (f,g,h) $\vec{f}$, we also know that f,g,h are C^1 on their domain. We also have the semi-explicit representation of $D\vec{f}$:

$$D\begin{bmatrix} f \\ g \\ h \end{bmatrix}=\begin{bmatrix} \partial_u f & \partial_v f \\ \partial_u g & \partial_v g \\ \partial_u h & \partial_v h \end{bmatrix}=-\{D_{(x,y,z)}(F,G,H)\}^{-1}D_{(u,v)}(F,G,H)$$

To prove this we use an inductive argument, the base case is 1 equation one variable ($F(\vec{x}, y) = 0$), which is done. Suppose that we can solve 2 equations in 2 unknowns locally and with C^1 solutions. We seek to prove that we can solve 3 equations in 3 unknows.

To be definite, suppose $\partial_z H \neq 0$ in a neighborhood of (a, b, p, q, r)

We have the following:

$$H(u, v, x, y, z) = 0 \longleftrightarrow z = \phi(u, v, x, y)$$

where ϕ is some C^1 explicit function in the neighborhood of (a, b, p, q, r).

Let us define $\tilde{F}(u, v, x, y) := F(u, v, x, y, \phi(u, v, x, y))$ and $\tilde{G}(u, v, x, y) := G(u, v, x, y, \phi(u, v, x, y))$. We know that both of these equal to 0, as well as the fact that $\tilde{F}$ and $\tilde{G}$ are C^1.

Then we only need to show the following:

$$\partial_{(x,y)}(\tilde{F}, \tilde{G}) = \begin{vmatrix} \partial_x \tilde{F} & \partial_y \tilde{F} \\ \partial_x \tilde{G} & \partial_y \tilde{G} \end{vmatrix} \neq O$$

To do this we use the chain rule:

$$\begin{aligned} \partial_x \tilde{F} &= \partial_x F + (\partial_z F)(\partial_x \phi) = F_x + F_z \phi_x \\ \partial_y \tilde{F} &= \partial_y F + (\partial_z F)(\partial_y \phi) = G_x + G_z \phi_x \end{aligned}$$

The same holds true for $\partial_x \tilde{G}$ and $\partial_y \tilde{G}$. Note also that $\partial_x H = H_x + H_z \phi_x = 0$ and $\partial_y H = 0$ as well. Therefore:

$$\partial_{(x,y)}(\tilde{F}, \tilde{G}) = \begin{vmatrix} F_x + F_z \phi_x & F_y + F_z \phi_y \\ G_x + G_z \phi_x & G_y + G_z \phi_y \end{vmatrix} = \begin{vmatrix} F_x + F_z \phi_x & F_y + F_z \phi_y & F_z / H_z \\ G_x + G_z \phi_x & G_y + G_z \phi_y & G_z / H_z \\ H_x + H_z \phi_x & H_y + H_z \phi_y & 1 \end{vmatrix}$$

We are given that

$$\begin{vmatrix} F_x & F_y & F_z \\ G_x & G_y & G_z \\ H_x & H_y & H_z \end{vmatrix} = H_z \begin{vmatrix} F_x & F_y & F_z / H_z \\ G_x & G_y & G_z / H_z \\ H_x & H_y & 1 \end{vmatrix} \neq 0$$

However, note that we can get from here to the original 3×3 determinant by column operations, therefore we know that $\partial_{(x,y)}(\tilde{F}, \tilde{G}) \neq 0$. Therefore we have the degeneracy condition.

Then we can write

$$\begin{cases} x = f(u, v) \\ y = g(u, v) \end{cases}$$

by the assumption. Now we plug back into the formula for z and we get:

$$z = h(u,v) = \phi(u,v,f(u,v),g(u,v))$$

Now we have successfully written all three components as C^1 functions of u and v.

16.5.6 Alternate Form

$$D\vec{f} = -\{D_{\vec{x}}\vec{F}\}^{-1}(D_{\vec{u}}\vec{F})$$

Let $\vec{G}(\vec{u}) = \vec{F}(\vec{u},\vec{f}(\vec{u})) \equiv \vec{0}$

Now we can Jacobian both sides

$$D\vec{G}(\vec{u}) \equiv D\vec{0} = O$$

Note that $\vec{G} = \vec{F} \circ (Id \times \vec{f})$. Now we can apply the chain rule:

$$\begin{aligned} D\vec{G} &= D(\vec{F} \circ (Id \times \vec{f})) \\ &= \{(D\vec{F}) \circ (Id \times \vec{f})\}D(Id \times \vec{f}) \end{aligned}$$

Now we try to write these out in a type of "block decomposition." Let us start with $D(Id\times\vec{f})$

$$D(Id \times \vec{f})(\vec{u}) = \begin{bmatrix} D_{\vec{u}}\vec{u} \\ (D\vec{f})(\vec{u}) \end{bmatrix} = \begin{bmatrix} D(Id) \\ D\vec{f} \end{bmatrix}(\vec{u}) = \begin{bmatrix} I \\ D\vec{f} \end{bmatrix}(\vec{u}) = \begin{bmatrix} I \\ D\vec{f}(\vec{u}) \end{bmatrix}$$

We also have a block decomposition of $D\vec{F}$. We know:

$$\vec{F}(\vec{u},\vec{x}) = \begin{bmatrix} F_1(\vec{u},\vec{x}) \\ F_2(\vec{u},\vec{x}) \\ \vdots \\ F_e(\vec{u},\vec{x}) \end{bmatrix}$$

Therefore we know:

$$\begin{aligned} D\vec{F} &= \left[\begin{array}{ccc|ccc} \partial_{u_1}F_1 & \dots & \partial_{u_d}F_1 & \partial_{x_1}F_1 & \dots & \partial_{x_e}F_1 \\ \vdots & & \vdots & \vdots & & \vdots \\ \partial_{u_1}F_2 & \dots & \partial_{u_d}F_e & \partial_{x_1}F_e & \dots & \partial_{x_e}F_e \end{array}\right] \\ &= [D_{\vec{u}}\vec{F}|D_{\vec{x}}\vec{F}] \end{aligned}$$

Now we can bring everything back together and we have

$$
\begin{aligned}
D\vec{G}(\vec{u}) &= [D_{\vec{u}}\vec{F} \circ (Id \times \vec{f}) | D_{\vec{x}}\vec{F} \circ (Id \times \vec{f})] \begin{bmatrix} I \\ D\vec{f}(\vec{u}) \end{bmatrix} \\
&= D_{\vec{u}}\vec{F} \circ (Id \times \vec{f}) + (D_{\vec{x}}\vec{F} \circ (Id \times \vec{f}))(D\vec{f}) \equiv O
\end{aligned}
$$

Now if we isolate $D\vec{f}$, we get

$$\boxed{D\vec{f} = -(D\vec{F})^{-1}(D_{\vec{u}}F)}$$

We know that the inverse matrix exists because we assumed that

$$\det D_{\vec{x}}F \neq 0$$

16.6 Inverse Function Theorem

16.6.1 Theorem

Suppose we have two sets of variables $\vec{x}$ and $\vec{u}$ where

$$
\begin{cases}
x_1 &= f_1(u_1, u_2, \ldots, u_d) \\
x_2 &= f_2(u_1, u_2, \ldots, u_d) \\
\vdots & \\
x_d &= f_d(u_1, u_2, \ldots, u_d)
\end{cases}
$$

or that $\vec{x} = \vec{f}(\vec{u})$. Assume that we have a particular solution $\vec{b} = \vec{f}(\vec{a})$ where $\vec{f}$ is C^1 and $\vec{a} \in (\vec{f})^\circ$ and $\partial_{\vec{x}}\vec{f}(\vec{a}) \neq 0$. Then there exists an open set U around $\vec{a} \in \mathbb{R}^d_{\vec{u}}$ and there exists an open set V around $\vec{b} \in \mathbb{R}^d_{\vec{x}}$. There also exists a C^1 function $\vec{g} : V \to U$ such that $\vec{f} \circ \vec{g} = Id_V$ and $\vec{g} \circ \vec{f}|_{\vec{g}(V)} = Id_{\vec{g}(V)}$. $\vec{g}$ is known as the **local inverse** of $\vec{f}$ near $\vec{b}$.

16.6.2 Random Fact

If $\vec{f}$ is linear, it can be written as $\vec{x} = \vec{f}(\vec{u}) = A\vec{u}$, and its inverse function would be $\vec{u} = A^{-1}\vec{x}$ (if we assume that $\det A \neq 0$). However, note that $\vec{f}(\vec{u}) = A\vec{u}$ means that $D\vec{f}(\vec{u}) \equiv A$.

If we look at the component functions of $\vec{f}$, we can see a list of partials

$$
\begin{aligned}
f_1(\vec{u}) = A\vec{u} \cdot \vec{e}_1 &= a_{11}u_1 + a_{12}u_2 + \cdots + a_{1d}u_d \\
\partial_{u_1} f_1(\vec{u}) &= a_{11} \\
\partial_{u_2} f_1(\vec{u}) &= a_{12}
\end{aligned}
$$

$$\boxed{\partial_{u_j} f_i(\vec{u}) = a_{ij}}$$

Therefore, $\det A = \partial\vec{f}(\vec{u}) \neq 0$ at $\vec{a}$.

16.6.3 Proof

Let us create a function $\vec{F}(\vec{u},\vec{x}) := \vec{x} - \vec{f}(\vec{u})$. By the Implicit Function Theorem, we know that this is equivalent to $\vec{x} = \vec{f}(\vec{u})$. However, there is nothing in the Implicit Function Theorem that says we cannot solve for $\vec{u}$. A couple of conditions must be met beforehand, however, including that $\vec{F}$ has to be C^1, that $(\vec{a},\vec{b}) \in (\vec{F})^\circ$, and that $\partial_{\vec{u}}\vec{F}(\vec{a},\vec{b}) \neq 0$.

Let us start with the last condition first. $D_{\vec{u}}\vec{F} = D_{\vec{u}}Id(\vec{x}) - (D_{\vec{u}}\vec{f})(\vec{u}) = -D\vec{f}(\vec{u})$

$$\begin{aligned}\partial_{\vec{u}}\vec{F} = \det D_{\vec{u}}\vec{F} &= \det(-D\vec{f}(u))\\ &= (-1)^d \det D\vec{f}(\vec{u})\\ &= (-1)^d \partial\vec{f}(\vec{u}) \neq 0 \text{ at } (\vec{a}, vecb)\end{aligned}$$

We also know that $\vec{F}$ is C^1 by assumption. Now all we have to check is that $(\vec{a},\vec{b}) \in (\vec{F})^\circ$. $\vec{F} = (\vec{f} \times \mathbb{R}^d_{\vec{x}}) \subseteq \mathbb{R}^{2d}_{(\vec{u},\vec{x})}$. So $\vec{b}$ is not a problem. We also know that $\vec{a} \in \vec{f}$ by assumption.

Therefore we conclude, using the Implicit Function Theorem that $\exists$ open sets U, V around $\vec{a},\vec{b}$ respectively, and $\exists \vec{g} : V$ such that $\vec{F}(\vec{g}(\vec{x}),\vec{x}) \equiv \vec{0}$ for all $\vec{x} \in V$. This implies that $\vec{x} - \vec{f}(\vec{g}(\vec{x})) = \vec{0}$. Therefore $\boxed{\vec{f}(\vec{g}(\vec{x})) \equiv \vec{x} \text{ for all } \vec{x} \in V}$.

To prove the converse, take any point $\vec{u} \in \vec{g}(\vec{V})$, plug in $\vec{f}(\vec{u})$ for $\vec{x}$, then $\vec{f}(\vec{u}) = \vec{f}(\vec{g}(\vec{x}_0)) = \vec{x}_0 \in V$. Now we plug the thing back again and we get $\vec{f}(\vec{g}(\vec{f}(\vec{u}))) = \vec{f}(\vec{u})$. Now since $\vec{f}$ is one-to-one locally, we can cancel out the outmost function, so we get $\vec{g}(\vec{f}(\vec{u})) = \vec{u}$. Therefore, we can say $\vec{f}$ and $\vec{g}$ are local inverses of each other.

To prove that $\vec{f}$ is one-to-one, suppose $\vec{f}(\vec{u}_1) = \vec{f}(\vec{u}_2)$ but $\vec{u}_1 \neq \vec{u}_2$, and $\vec{u}_1, \vec{u}_2 \in \vec{g}(V)$. Then we know, by the continuity of $\vec{g}$ that there exists $\vec{x}_1, \vec{x}_2 \in V$ such that $\vec{u}_1 = \vec{g}(\vec{x}_1)$ and $\vec{u}_2 = \vec{g}(\vec{x}_2)$. However, then we know that $\vec{f}(\vec{g}(\vec{x}_1)) = \vec{f}(\vec{g}(\vec{x}_2))$, which implies that $x_1 = x_2$, which implies that $\vec{g}(\vec{x}_1) = \vec{g}(\vec{x}_2)$, which means $\vec{u}_1 = \vec{u}_2$. Therefore $\vec{f}$ is one-to-one.

16.6.4 Implications

Let $\vec{f} : D \subseteq \mathbb{R}^d_{\vec{u}} \to \mathbb{R}^d_{\vec{x}}$ be a C^1 transformation such that $\partial\vec{f}(\vec{u}) \neq 0$ for all $\vec{u} \in D^\circ$. Let K be a compact subset of D°, Then $\vec{f}(K^\circ) = [\vec{f}(K)]^\circ$. In other words, the interior of the pre-image maps onto the interior of the image.

Moreover, if $\vec{f}$ is also one-to-one on some open superset U of K, then $\vec{f}(\partial K) = \partial\{\vec{f}(K)\}$.

17 Lagrange Multiplier

17.1 Level-Sets as Manifolds

17.1.1 Definitions

Level Sets

Let $g : D \subseteq \mathbb{R}^d \to \mathbb{R}$ be C^1 on D°. The Level-set at level $c \in g(D)^\circ$ of the function g, written as $L_c g := \{\vec{x} \in D | g(\vec{x}) = c\}$. Note that $L_c g \neq \emptyset$ and $L_c g \neq \{\vec{x}_0\}$.

K-Patch

A K-patch in d-space is a set $P \subseteq \mathbb{R}^d$ which is the C^1 and one-to-one image of some open connected set in $\mathbb{R}^k (k \leq d)$ by a mapping of "full rank", i.e. $P = \vec{\phi}(U)$ where $U \subseteq \mathbb{R}^k$ is open and connected, $\vec{\phi}$ is C^1 and one-to-one and $\{\partial_{u_1}\vec{\phi}, \partial_{u_2}\vec{\phi}, \ldots, \partial_{u_k}\vec{\phi}\}$ are linearly independent for all $\vec{u} \in U$.

In layman terms, the geometric transformation is one that bends the surface in $\mathbb{R}^k$ "smoothly," it is a "bending" of the original set without any sharp edges or elimination of dimensions. In other way, the dimensionality is reserved. Which is why $\vec{\phi}$ must be C^1 and one-to-one.

The function $\vec{\phi}$ is known as a parameterization of the set P because as the argument transverses U, $\vec{\phi}$ traces out P. Note that our "full rank" condition, which is that $\{\partial_{u_1}\vec{\phi}, \partial_{u_2}\vec{\phi}, \ldots, \partial_{u_k}\vec{\phi}\}$ are linearly independent. This is because in the domain, pick an arbitrary point $\vec{p}$ and move in the u_1 direction, since $\vec{\phi}$ is C^1 and one-to-one, this movement in the domain will trace out $\vec{\phi}(\vec{p})$ in the co-domain. The tangent vector to $\vec{\phi}(\vec{p})$ is $\partial_{u_1}\vec{\phi}(\vec{p})$. If we move along u_2 direction in the domain of $\vec{p}$, the co-domain tangent velocity would be $\partial_{u_2}\vec{\phi}(\vec{p})$, which should be linearly independent from $\partial_{u_2}\vec{\phi}(\vec{p})$. In other words, independent movements from the domain maps onto independent movements in the co-domain. This prevents foldings and crimping.

It so happens that if P has a parameterization, it has an infinitely many of them.

K-Manifold

A K-manifold in $\mathbb{R}^d$ is a set $M \subseteq \mathbb{R}^d$ that can be completely covered by countably many (possibly overlapping) K-patches.

17.1.2 Level-Set Theorem

Let $g : D \subseteq \mathbb{R}^d \to \mathbb{R}$ be C^1, and choose $c \in g(D)^\circ$. Then $L_c g = \{\vec{x} \in D | g(\vec{x}) = c\}$ is a $(d-1)$-manifold in $\mathbb{R}^d$ provided that $g \neq \vec{0}$ for any point $\vec{x} \in L_c g$.

Proof:

We know that $g(x_1, x_2, \ldots, x_d) = c$, now let's pick $\vec{p} \in L_c g$, we know that $g(p_1, p_2, \ldots, p_d) = c$ and that $\partial_{x_1} g(\vec{p}), \partial_{x_2} g(\vec{p}), , \partial_{x_d} g(\vec{p})$ cannot all be 0 (as $g(\vec{p}) \neq 0$), therefore $\exists j \in \{1, 2, \ldots, d\} : \partial_{x_j} g(\vec{p}) \neq 0$. To be definite, let $j = d$ (wlog). By the Implicit Function Theorem, we get the equivalency between $g(\vec{x}) = c$ and $x_d = \phi(x_1, x_2, \ldots, x_{d-1})$ in a neighborhood $U \times I \subseteq \mathbb{R}^{d-1} \times \mathbb{R}$ of $\vec{p}$. In this neighborhood, we can define a parameterization function $\vec{\phi}(\vec{u}) : U \subseteq \mathbb{R}^{d-1} \to \mathbb{R}^d$ as:

$$\vec{\phi}(u_1, u_2, \ldots, u_{d-1}) = (u_1, u_2, \ldots, u_{d-1}, \phi(u_1, u_2, \ldots, u_{d-1}))$$

Therefore the range, $\vec{\phi}(U) = L_c g \cap N_{\vec{p}}$ where $N_{\vec{p}}$ is some open neighborhood of $\vec{p}$ in $\mathbb{R}^d$.

We now seek to show that $\vec{\phi}(U)$ is a $(d-1)$-patch in $\mathbb{R}^d$. We have to prove:

1. $\vec{\phi}$ is one-to-one
2. $\vec{\phi}$ is C^1 on U
3. $\{\partial_{u_1}\vec{\phi}, \ldots, \partial_{u_{d-1}}\vec{\phi}\}$ are linearly independent

Condition 2 is trivial, because $\vec{\phi} = Id \times \phi$, and since both the identity function and ϕ are C^1, $\vec{\phi}$ must be C^1.

Condition 1 is also not hard to see. suppose $\vec{\phi}(\vec{u}_1) = \vec{\phi}(\vec{u}_2)$, we then know that the first $d-1$ coordinates for $\vec{u}_1$ and $\vec{u_2}$ are the same, but $\vec{u}_1$ and $\vec{u}_2$ are both of dimension $d-1$. Therefore $\vec{u}_1 = \vec{u}_2$.

Condition 3, or the "full rank" condition is a bit harder to prove, first we need the Jacobian Matrix:

$$D\vec{\phi}(\vec{u}) = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ \frac{\partial\phi}{\partial u_1} & \frac{\partial\phi}{\partial u_2} & \frac{\partial\phi}{\partial u_3} & \ldots & \frac{\partial\phi}{\partial u_{d-1}} \end{bmatrix}$$

From here is is kind of obvious that the columns are linearly independent. Therefore the "full rank" condition is fulfilled.

Therefore this is the correct parameterization. Therefore the level set is indeed a manifold.

17.1.3 Applications

Find a local extremum of $f(x_1, \ldots, x_d)$ subject to the constraint $g(x_1, \ldots, x_d) = c$ where f, g are C^1.

This has tremendous application in engineering, economics, etc.

17.2 Optimization

17.2.1 Critical Points

In single variable calculus, we optimize by taking the derivative of the function and setting it to zero, which gives us the set of **critical points**, which is made up of local extremas and "saddle points," which are points in the function that just happen to have 0 derivative but the second order derivative changes as well.

In multivariable calculus we do much of the same, except now we set the entire gradient to zero. Given $f : D \subseteq \mathbb{R}^d \to \mathbb{R}$ where f is C^1, any point $\vec{p} \in D$ where $f(\vec{p}) = \vec{0}$ is called a critical point.

Thm:If $\vec{p} \in D^\circ$ and $\vec{p}$ is a local extremum for $\vec{p}$, then $f(\vec{p}) = \vec{0}$.

Proof: Note that:

$$\partial_{x_j} f(\vec{p}) = \lim_{h \to 0^+} \frac{f(\vec{p} + h\vec{e_j}) - f(\vec{p})}{h}$$

because f is C^1. WLOG assume $\vec{p}$ is a local maximum, then $f(\vec{p})$ should be greater than all surrounding values. Therefore $f(\vec{p}+h\vec{e_j}) \leq f(\vec{p})$, thus the numerator is negative, but the denominator is always greater than zero. Therefore the partial as a whole is negative or zero.

Of course since the derivative is a two sided limit,

$$\partial_{x_j} f(\vec{p}) = \lim_{h \to 0^-} \frac{f(\vec{p} + h\vec{e_j}) - f(\vec{p})}{h}$$

by similar logic the partial must therefore be positive or zero.

Combining the two statements it is obvious that the partial is in fact, 0.

17.2.2 Second Partial Derivative Test

Finding the set of the

17.2.3 Example

Given

$$f(x, y) = 3x^2 - 2xy + y^2 - x + 1$$

We can construct the following system of equations:

$$\begin{cases} \partial_x f & = 6x - 2y - 1 = 0 \\ \partial_y f & = -2x + 2y = 0 \end{cases}$$

Now we can solve and we get the solution point $(\frac{1}{4}, -\frac{1}{4})$. Now we use the Second Partial Test, and we find out that this point is a local minimum.

18 Integral Calculus

18.1 Riemann Integral

18.1.1 Definition

The Riemann Integral is defined in two ways: the lower and upper Riemann Integrals.

The **lower Riemann Integral** is defined as:

$$\underline{\int_a^b} f(x)dx = \sup\{\sum_{i=1}^{n}[\inf\{f(x)|x_{i-1} \leq x \leq x_i\}]\Delta x_i | a = x_0 < x_2 < \cdots < x_n = b\}$$

This is known as the **down-and-up** procedure, for that we first take the lower bound estimate with the infimum, but then we take the maximum possible lower approximation. Note that the supremum is necessary because the maximum doesn't actually exist in the set of lower estimations.

The **upper Riemann Integral** is defined as:

$$\overline{\int_a^b} f(x)dx = \inf\{\sum_{i=1}^{n}[\sup\{f(x)|x_{i-1} \leq x \leq x_i\}]\Delta x_i | a = x_0 < x_2 < \cdots < x_n = b\}$$

This is also known as the **up-and-down** procedure.

Note that both of these rely on the existence of the infimum and supremum, so they only exist over bounded regions.

It should also come as no suprise that

$$\underline{\int_a^b} f \leq \overline{\int_a^b} f$$

If $\underline{\int_a^b} f = \overline{\int_a^b} f$, we say f is **Riemann Integrable**, and the value is the shared value.

18.1.2 Sufficient Condition for Riemann Integration

A function f is Riemann Integrable is all the discontinuity points can be covered by a set of arbitrarily small open intervals.