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:

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

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 $q = \{(n, m), (2n, 2m), (3n, 3m) \dots (-2n, -2m), (-3, -3m) \dots\}$

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:

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

  Where $s = \pm 1$, $N \in \mathbb{N}$, and $d_n \in 0, 1, 2, 3, \dots 8, 9$. It is also not the case that $d_n = d_{n+1} = \dots = 0$. 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.

<u>Claim:</u> $t \in [a, b]$. 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

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

Lemma: 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 $|x_{n_k} - x| < \epsilon$).

Proof for $d = 2$: 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. Proof: 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 <u>continuous</u> 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 $\vec{x} \in B_r(\vec{p})$.

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

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

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 <u>Minkowski Distance</u>

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

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

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

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

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: $|ac+bd| \leq |ac|+|bd| \leq \sqrt{a^2+b^2}\sqrt{c^2+d^2}$, or $|v\dot{u}| \leq ||u||||v||$ with equality iff $ab \geq 0$.

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 Linear Algebra

8.1 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).

8.2 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:

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

$\vec{f}$ 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.

8.3 Matrices

8.3.1 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 $a_{31} = 0$.

8.3.2 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.

8.3.3 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.

8.4 Unique Expression of Linear Functions

8.4.1 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.

8.4.2 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 $\vec{x} = \mathrm{A}\vec{x} = \mathrm{B}\vec{x}$.

9 Differentiability of Vector Valued Functions

9.1 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 $\vec{p}$, we write it as $D\vec{f}(\vec{p}) \in \mathbb{R}^{e \times d}$, and $D\vec{f}(\vec{p})_{ij}$ (Jacobian derivative) $= \partial_{x_j} f_i(\vec{p})$.

Claim: The above definition decomposes componentwise.

This is done by the basic norm bounds ($|v_j| \leq ||\vec{v}|| \leq$

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.

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) $= \partial_{x_j} f_i$.

9.2 Definition of Differentiability

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:

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

10 The Gradient Operator

10.1 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:

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