Further Analysis

Kexing Ying

May 15, 2020

Contents

1 Euclidean Spaces

For $n \geq 1$, the n-dimensional *Euclidean space* denoted by $\mathbb{R}^n$, is the set of ordered n-tuples $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ for $x_i \in \mathbb{R}$. Recall that $\mathbb{R}^n$ is a vector space over $\mathbb{R}$, we can use the usual vector space operations, i.e. vector addition and scalar multiplication. Furthermore, $\mathbb{R}^n$ forms a inner product space with the operation,

$$\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} : (\mathbf{x}, \mathbf{y}) \mapsto \sum_{i=1}^n x_i y_i.$$

Thus, as a inner product space induces a normed vector space, we find a natural norm defined for $\mathbb{R}^n$ by,

$$\| \cdot \| : \mathbb{R}^n \to \mathbb{R} : \mathbf{x} \mapsto \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^n x_i^2}.$$

By manually checking, we find that this norm satisfy the norm axioms, i.e. it satisfy the *triangle inequality*, is *absolutely scalable*, and *positive definite* (In fact, we do not need the norm to be non-negative as it can deduced from the other axioms).

1.1 Preliminary Concepts in $\mathbb{R}^n$

Sequences in $\mathbb{R}^n$ can be defined similarly to that of $\mathbb{R}$, and we carry over all notations in all suitable places.

Definition 1.1 (Convergence in $\mathbb{R}^n$)**.** We say a sequence $(\mathbf{x}_i)_{i=1}^\infty \subseteq \mathbb{R}^n$ converges to $\mathbf{x} \in \mathbb{R}^n$ if and only if for all $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that for all $i \geq N$, $\|\mathbf{x}_i - \mathbf{x}\| < \epsilon$.

Proposition 1. *A sequence $(\mathbf{x}_i)_{i=1}^\infty \in \mathbb{R}^n$ converges to $\mathbf{x} \in \mathbb{R}^n$ if and only if each component of $\mathbf{x}_i$ converges to the corresponding component of $\mathbf{x}$.*

In the first dimension, we've considered the topology of $\mathbb{R}$ including the examination of open and closed sets. We extend this idea for higher dimensions. The most basic examples we have of an open set (or closed set for that matter) in $\mathbb{R}$ are the open and closed intervals respectively. This is extended in $\mathbb{R}^n$ to be sets of the form

$$\prod_{i=1}^{n}(a_i, b_i) := \{\mathbf{x} \mid a_i < \mathbf{x}_i < b_i, \forall 1 \leq i \leq n\},$$

and similarly for closed intervals. However, while this is nice to look at, it is not very useful. So for this reason, we again will extend the notion of open and closed sets for $\mathbb{R}^n$.

Definition 1.2 (Open Ball)**.** Let $\mathbf{x} \in \mathbb{R}^n$ and $r \in \mathbb{R}^+$, we define the open ball of radius r about $\mathbf{x}$ as the set

$$B_r(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{y}\| < r\}$$

.

Definition 1.3 (Open)**.** A set $U \subseteq \mathbb{R}^n$ is open in $\mathbb{R}^n$ if and only if for all $\mathbf{x} \in U$, there is some $r \in \mathbb{R}^+$ such that $B_r(\mathbf{x}) \subseteq U$.

Definition 1.4 (Closed)**.** A set $U \subseteq \mathbb{R}^n$ is closed if and only if its complement is open.

Straight away from the definition, we can see that every open ball is open (see here). Furthermore, we find the union and intersection of two open sets is open. In fact, the union and any collection of open sets is also open, however, this is not necessarily true for closed sets.

Definition 1.5 (Continuity at a Point)**.** Let $A \subseteq \mathbb{R}^n$ be an open set, and let $f : A \to \mathbb{R}^m$. We say f is continuous at $p \in A$ if and only if for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $x \in A \cap B_\delta(p)$, $\|f(x) - f(p)\| < \epsilon$.

If the function f is continuous at every point of A, then we say f is continuous on A.

Definition 1.6. Let $A \in \mathbb{R}^n$ be open in $\mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. For $p \in A$, we say that the limit of f as $\mathbf{x}$ tends to $\mathbf{p}$ in A is equal to $\mathbf{q} \in \mathbb{R}^m$ if and only if for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in A \cap B_\delta(p), x \neq p$, $\|f(x) - \mathbf{q}\| < \epsilon$.

This is the same notion we used for continuity in the first dimension to say that f is continuous at p if and only if $\lim_{x\to p} f(x) = q$.

Proposition 2. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function. Then f is continuous if and only if for all open subsets U of $\mathbb{R}^m$, $f^{-1}(U)$ is open in $\mathbb{R}^n$.*

Proof. See here for the proof. □

1.2 Derivative of Maps in Euclidean Spaces

Let Ω be a open in $\mathbb{R}^n$, and $f : \Omega \to \mathbb{R}^m$ be a "nice behaving map". We poses the question on how we should define the notion of derivatives for this mapping at some point $p \in \Omega$. We recall that the derivative at a point p in the first dimension is defined to be

$$f'(p) = \lim_{x\to p} \frac{f(x) - f(p)}{x - p}.$$

While we see that this equation makes no sense if we simply generalise this equation to higher dimensions, we see the following result.

Lemma 1.1. Let $f : S \to \mathbb{R}$ where $S \subseteq \mathbb{R}$, then f is differentiable at some $p \in S$ if and only if there exists some $\lambda \in \mathbb{R}$ such that

$$\lim_{x \to p} \left| \frac{f(x) - A_\lambda(x)}{x - p} \right| = 0,$$

where $A_\lambda(x) = \lambda(x - p) + f(p)$.

Proof. Follows from algebraic manipulation. □

With this, we can conclude that $f(x) - A_\lambda(x)$ tends to zero faster than $x - p$. We will generalise this result to higher dimensions.

We may rewrite $A_\lambda(x) = \lambda(x - p) + f(p) = \lambda x + (f(p) + \lambda p)$, so, we see that λ is the translation of a linear map λx, i.e. $A_\lambda = (x \mapsto x + (f(p) + \lambda p)) \circ (x \mapsto \lambda x)$. We can easily generalise such maps to higher dimensions and we call such maps *affine maps*.

Definition 1.7 (Differentiable Functions in $\mathbb{R}^n$)**.** Recall the definition of linear maps for general vector spaces which we will use in the context of Euclidean spaces. Let $L(\mathbb{R}^n; \mathbb{R}^m)$ denote the set of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$, $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}^m$ be a function. Then we say f is differentiable at some point $p \in \Omega$ if and only if there exists some $\Lambda \in L(\mathbb{R}^n; \mathbb{R}^m)$, such that

$$\lim_{x \to p} \frac{\|f(x) - (\Lambda(x - p) + f(p))\|}{\|x - p\|} = 0.$$

If this is true, we write $Df(p) = \Lambda$ and call Λ the derivative of f at p.

Remark. *Some book refers to $Df(p)$ as the total derivative of f at p.*

It is often useful to express the derivative criterion as

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

Proposition 3. *Let $f_i : (a, b) \to \mathbb{R}$ be differentiable for all i, then the function, $f : (a, b) \to \mathbb{R}^m : x \mapsto (f_i(x))_{i=1}^m$ is differentiable for all $p \in (a, b)$.*

Proof. Let the Jacobian $\Lambda = \mathrm{diag}(\lambda_i)$ where λ_i is the derivative of f_i at p. Then I claim, $Df(p) = \Lambda p$.

Consider

$$\lim_{h \to 0} \frac{\|f(p + h) - f(p) - \Lambda h\|}{\|h\|} = \lim_{h \to 0} \frac{\sqrt{\sum_{i=0}^m |f_i(p + h) - f_i(p) - \lambda_i h|}}{\|h\|}.$$

However, for all i, $\|h\| \geq |h_i|$, so

$$\lim_{h \to 0} \frac{\sqrt{\sum_{i=0}^{m} |f_i(p+h) - f_i(p) - \lambda_i h|}}{\|h\|} \leq \lim_{h \to 0} \sum_{i=0}^{m} \sqrt{\frac{|f_i(p+h) - f_i(p) - \lambda_i h|}{|h_i|}} = 0$$

□

A lot of results from the first dimension generalises easily to higher dimensions. Similar to that of the first dimension, the chain rule in general Euclidean spaces states,

Theorem 1. *Let $\Omega \subseteq \mathbb{R}^n$ and $\Omega' \subseteq \mathbb{R}^m$ be open sets with $g : \Omega \to \Omega'$ be differentiable at $p \in \Omega$ and $f : \Omega' \to \mathbb{R}^l$ be derivatives at $g(p)$. Then $h = f \circ g$ is differentiable at p with derivative*

$$Dh(p) = Df(g(p)) \circ Dg(p).$$

Proof. Similar to the proof of the Chain rule in the first dimension using algebraic manipulation. □

Omitted many examples here, check official lecture notes for these examples.

1.3 Directional Derivatives

Although the definition of derivative in dimension one and higher is similar, it is different in that we verify whether a linear map is the total derivative at a point instead of computing the limit of some equation. This is difficult as often times, it is not easy to guess what the derivative of a function is. Thus, it is useful to somehow identify candidate linear maps for the derivative.

Assume $\Omega \subseteq \mathbb{R}^n$ is open and $f : \Omega \to \mathbb{R}^m$ is a differentiable function at some $p \in \Omega$. Let $v \in \mathbb{R}^n$ be a unit vector. We would like to identify $Df(p)[v] \in \mathbb{R}^m$.

By the definition of derivatives in higher dimensions,

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

So, by letting $t \in \mathbb{R}$, we have

$$\begin{aligned} 0 &= \lim_{t \to 0} \frac{\|f(p+tv) - f(p) - \Lambda(tv)\|}{\|tv\|} \\ &= \lim_{t \to 0} \frac{\|f(p+tv) - f(p) - t\Lambda v\|}{|t|} \\ &= \lim_{t \to 0} \frac{\|f(p+tv) - f(p)\|}{|t|} - \Lambda v, \end{aligned}$$

So, $\lim_{t \to 0} \|f(p+tv) - f(p)\| / |t| = \Lambda v$. Thus, by finding the limits of the above equation for each basis vector $v \in B$, we find the Jacobian $[\Lambda]_B$.

Remark. *For notation, we denote the limit as* $\lim_{t\to 0} \|f(p+tv) - f(p)\|/|t|$ *as* $\partial f/\partial v \mid_p$ *and we call it the directional derivative of* f *in the direction of* v *at* p*. We will normally consider the directional derivatives in the direction of the standard basis and we call them the partial derivatives of* f *at* p*.*

Theorem 2. *Let* $\Omega \subseteq \mathbb{R}^n$ *be open, and* $f : \Omega \to \mathbb{R}^m : x \mapsto [f_i(x)]$ *for* $i \in \{1, \cdots m\}$ *is differentiable at some* $p \in \Omega$*. Then the Jacobian of* f *at* p *is* $[\partial f_i/\partial e_j \mid_p]_{i,j}$*.*

Proof. We recall that the Jacobian is simply the matrix form of the linear map that is the derivative. So, for all $v \in \mathbb{R}^n$, we have $Jv = Df(p)(v)$. As, $v \in \mathbb{R}^n$, it can be represented as a sum of the standard basis, that is there exists $v_i \in \mathbb{R}$, $v = \sum_{i=1}^n v_i e_i$, so, $Df(p)(v) = Df(p)(\sum_{i=1}^n v_i e_i) = \sum_{i=1}^n v_i Df(p)(e_i) = \sum_{i=1}^n v_i \partial f/\partial e_i \mid_p = [\sum_{i=1}^n v_i \partial f_j/\partial e_i \mid_p]_j = [\partial f_i/\partial e_j \mid_p]_{i,j} v$. (We used the fact that $[\partial f/\partial e_i]_j = \partial f_j/\partial e_i$.) $\square$

Remark. *We note that the reverse is not true, that is the existence of partial derivatives does not imply differentiability. A counter example of this is* $f : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto$ *if* $x = y = 0$*, then* 0*, else,* $\frac{xy}{\sqrt{x^2+y^2}}$*.*

Theorem 3. *Let* $\Omega \subseteq \mathbb{R}^n$ *is open, and* $f : \Omega \to \mathbb{R}$ *be a function. Suppose that the partial derivatives of* f*,* $D_i f(x)$ *exists for all* $i = 1 \cdots n$ *exists at all* $x \in \Omega$*. Furthermore, if the map* $x \mapsto D_i f(x)$ *is continuous for all* i *at some point* p*. Then* f *is differentiable at* p*.*

1.4 Higher Derivatives

Similar to the first dimension, we would like to think about how to differentiate more than once.

Let $\Omega \subseteq \mathbb{R}^n$ is open, and $f : \Omega \to \mathbb{R}^m$ be differentiable everywhere on Ω. Then we may consider the map

$$Df : \Omega \to \mathcal{L}(\mathbb{R}^n; \mathbb{R}^m) : p \mapsto Df(p).$$

As there is a trivial isomorphism between $\mathcal{L}(\mathbb{R}^n; \mathbb{R}^m)$ and the matrices of dimension $m \times n$, we can represent every linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ as an element of $\mathbb{R}^{m\times n}$. Thus, Df can be represented as a map from Ω to $\mathbb{R}^{m\times n}$. So, we may ask if Df is continuous at some p and furthermore, we can ask if Df is differentiable at some $p \in \Omega$. If this is the case, we have the second derivative

$$DDf : \Omega \to \mathcal{L}(\mathbb{R}^n; \mathbb{R}^{m\times n}).$$

Definition 1.8 (Second Derivative)**.** Let $\Omega \subseteq \mathbb{R}^n$ be open and $f : \Omega \to \mathbb{R}^m$ be differentiable everywhere on Ω with derivative $Df : \Omega \to \mathcal{L}(\mathbb{R}^n; \mathbb{R}^m)$. Then the second derivative of f at some $p \in \Omega$ is the linear map $\Lambda \in \mathcal{L}(\mathbb{R}^n; \mathbb{R}^{m\times n})$ such that

$$\lim_{x\to p} \frac{\|Df(x) - Df(p) - \Lambda(x-p)\|}{\|x-p\|} = 0.$$

Thus, with this definition, we can easily extend the notion of derivatives any number of times to get the k-th derivative of a function. However, this is formally difficult and requires the notion of multilinear maps. Luckily, instead of working with this difficult definition whenever we would like to work with higher derivatives, we can instead look at whether the k-th derivative exists and whether or not it is continuous by theorem 3.

Now that we have established the notion of higher derivatives we would like to ask how higher partial derivatives interacts. That is, when does $D_i D_j f(p) = D_j D_i f(p)$?

Theorem 4 (Schwartz' Theorem)**.** *Let $\Omega \subseteq \mathbb{R}^n$ be open and $f : \Omega \to \mathbb{R}$ be differentiable at every $p \in \Omega$. Suppose that for some $i, j \in \{1, \cdots, n\}$ the second partial derivatives $D_i D_j f$ and $D_j D_i f$ exists and is continuous for all $p \in \Omega$, then*

$$D_i D_j f(p) = D_j D_i f(p)$$

for all $p \in \Omega$.