Contents

Chapter 1

Multivariable Calculus

1.1 Linear Algebra

1.1.1 Vector Spaces

Definition (Normed vector space): Let V be a vector space. A **norm** is a real valued function $\|\cdot\| : V \to \mathbb{R}$ which has the following properties

1. $\forall x \in V,\ \|x\| > 0.$
2. $\|x\| = 0 \implies x = 0.$
3. $\forall x \in V\ \forall \alpha \in \mathbb{F},\ \|\alpha x\| = |\alpha|\,\|x\|.$
4. $\forall x, y \in V\ \ \|x + y\| \leq \|x\| + \|y\|.$

Each normed vector space induces a metric space (V, d) where $d(x, y) = \|y - x\|$.

Theorem 1.1. *In every normed space* $(V, \|\cdot\|)$ *we have*

$$|\|v\| - \|w\|| \leq \|v - w\|$$

Hence the norm is Lipschitz continuous.

Definition: Assume V is a vector space and let $\|\cdot\|_1$, $\|\cdot\|_2$ be two norms for V. They are said to be equivalent when

$$\exists c_1, c_2 > 0\ \forall x : \qquad c_1 \|x\|_1 \leq \|x\|_2 \leq c_2 \|x\|$$

To check if the above definition is indeed an equivalence relation, we must show that following:

Reflexive $\|\cdot\|_1 \sim \|\cdot\|_1$.

Symmetric $\|\cdot\|_1 \sim \|\cdot\|_2 \implies \|\cdot\|_2 \sim \|\cdot\|_1$.

Transitive $\|\cdot\|_1 \sim \|\cdot\|_2$, $\|\cdot\|_2 \sim \|\cdot\|_3 \implies \|\cdot\|_1 \sim \|\cdot\|_3$.

Remark 1. Equivalent norms induce equivalent metrics, hence they induce the same topology.

Theorem 1.2. *All norms defined on a finite dimensional vector space V are equivalent.*

Proof. Let $\|\cdot\|$ be an arbitrary norm on V and $\{e_1, e_2, \ldots, e_n\}$ be a basis of V. Let $\|\cdot\|_2$ be L_2-norm (Euclidean norm). It will suffice to show $\|\cdot\| \sim \|\cdot\|_2$. Let

$$M = \max(\|e_1\|, \ldots, \|e_n\|)$$

Take $x \in V$, writing $x = \sum_{i=1}^{n} \xi_i e_i$ we have:

$$\|x\| = \left\| \sum_{i=i}^{n} \xi_i e_i \right\| \leq \sum_{i=1}^{n} |\xi_i| \|e_i\| \leq M\sqrt{n} \|x\|_2$$

Taking $c_2 = M\sqrt{n}$ proves the right inequality. For the left inequality we need the following lemma

Lemma 1.3. *If V is a normed vector space with $\|\cdot\|_2$, as defined above, is viewed as metric space $(V, \|\cdot\|_2)$ then $\|\cdot\| : V \to \mathbb{R}$ is continuous.*

Proof. Let $x_0 \in V$ and M be defined as above. For any $\epsilon > 0$ consider $\delta = \frac{\epsilon}{M\sqrt{n}}$ then if $\|x - x_0\|_2 < \delta$

$$|\|x\| - \|x_0\|| \leq \|x - x_0\| \leq M\sqrt{n} \|x - x_0\| \leq \epsilon$$

Now consider the sphere of radius $r = 1$ centered at 0, $S_1(0) = S_1 = \{x \in V : \|x\|_2 = 1\}$. One can show that S is compact (Theorem 1.4). Therefore, $\|x\|$ assumes its minimum on S. Let $a = \|x_0\|$ be the minimum. Since $0 \notin S$ then $a > 0$. By letting $y = x/\|x\|_2$, we have $y \in S$ and thus $a \leq \|y\|$ which is

$$a \|x\|_2 \leq \|x\|$$

Taking $c_1 = a$ proves the theorem. ■

Theorem 1.4. *Let $(V, \|\cdot\|)$ be a normed space over a normed complete field $\mathbb{F}$. The following are equivalent*

1. *V is finite dimensional.*
2. *every bounded closed set in V is compact.*
3. *the closed unit ball in V is compact.*

Proof. Item 1 $\implies$ Item 2: It is similar to proving a closed set $\mathbb{R}^n$ is compact using the fact a closed interval is compact in $\mathbb{R}$.

Item 2 $\implies$ Item 3: Trivial.

Item 3 $\implies$ Item 1: Requires the following lemma:

Lemma 1.5 (Riesz's lemma). *If V is a normed vector space and W is a closed proper subspace of V and $\alpha \in \mathbb{R}$ with $0 < \alpha < 1$, then there exists an $v \in V$ with $\|v\| = 1$ such that $\|v - w\| \geq \alpha$ for all $w \in W$*

Now suppose V were to be an infinite dimensional vector space. Then by the Lemma 1.5 there is sequence of unit vectors x_n such that $\forall m, n \in \mathbb{N}$, $\|x_n - x_m\| > \alpha$ for some $0 < \alpha < 1$. Which implies that no subsequence of $\{x_n\}$ is convergent and hence the closed unit ball can not be compact. ■

Example 1.1. The closed unit ball in the infinite dimensional vector space $C([0,1], \mathbb{R})$ with $\|f\| = \max f(x)$ is not compact. Take $f_n(x) = x^n$. Obviously $\|f_n\| = 1$, however f_n doesn't uniformly converge and hence f_n doesn't have a limit in $C([0,1], \mathbb{R})$ with the max norm. Consider the following norm

$$\|f\|_I = \int_0^1 |f(x)| \, \mathrm{d}x$$

Note that $\|\cdot\|_I$ and $\|\cdot\|_{\max}$ are not equivalent. Let $g(x) = 0$ for all $x \in [0,1]$. Then

$$\|f_n - g\|_I = \frac{1}{n+1} \to 0 \quad \text{as} n \to \infty.$$

Definition (Banach space): A normed vector space V that is complete is a **Banach space**. A **Hilbert Space** is a Banach space whose norm is induced by an inner product.

Proposition 1.6. *A normed finite dimensional vector space V over a normed complete field $\mathbb{F}$, is Banach space.*

Proof. Let $\{v_i\} \in V$ be a Cauchy sequence, and $\{e_1, \ldots, e_n\}$ be a basis for V with the norm L^1, that is if $v = (\xi^1, \ldots, \xi^n)$ then $\|v\| = \sum_{m=1}^n |\xi^m|$. Then if $v_i = (\xi_i^1, \ldots, \xi_i^n)$

$$\left|\xi_i^m - \xi_j^m\right| \leq \sum_{m=1}^n \left|\xi_i^m - \xi_j^m\right| \leq \|v_i - v_j\| < \epsilon$$

then $\{\xi_i^m\}_i$ are a Cauchy sequence in $\mathbb{F}$ and hence they converge $\xi_i^m \to \xi^m$. Then, clearly $v_i \to v = (\xi^1, \ldots, \xi^n)$ as each component converges. ■

Example 1.2. $\mathbb{Q}$ form a vector space itself over itself. It is finite dimensional as $\{\mathbb{1}_{\mathbb{Q}}\}$ is the basis, however the sequence

$$1, \; 1.4, \; 1.41, \; \ldots$$

does not converge even though it is Cauchy.

1.1.2 Linear Maps

Let V and W be a vector spaces over $\mathbb{F}$. A map $T : V \to W$ is **linear** if

$$T(x + \lambda y) = T(x) + \lambda T(y)$$

for all $x, y \in V$ and $\lambda \in \mathbb{F}$.

Definition: Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed spaces then, a linear transformation $T : V \to W$ is **bounded** if there exists a constant $C > 0$ such that

$$\|Tv\|_W \leq C \|v\|_V$$

for all $v \in V$. We denote the set of all linear map from $V \to W$ as $\mathcal{L}(V, W)$ and the set of all bounded linear maps as $\mathcal{B}(V, W)$. If $T \in \mathcal{L}(V, W)$ is bijective such that $T^{-1} \in \mathcal{L}(V, W)$, then T is called an **isomorphism** and V, W are **isomorphic**. An operator $T \in \mathcal{L}(V, W)$ is called **isometric** if $\|Tv\|_W = \|v\|_V$ for all $v \in V$.

Definition: If $(V, \|\cdot\|_V), (W, \|\cdot\|_W)$ are normed spaces then the **operator norm** of a linear transformation $T : V \to W$ is

$$\|T\| = \sup\left\{ \frac{\|Tv\|_W}{\|v\|_V} \middle| v \neq 0 \right\}$$

Proposition 1.7. *Let $T : U \to V$ and $T' : V \to W$ be two linear transformations.*

$$\|T' \circ T\| \leq \|T\| \, \|T'\|$$

Proof. for an arbitrary non-zero $x \in U$

$$\|T' \circ T(x)\|_W \leq \|T'\| \, \|Tx\|_V \leq \|T'\| \, \|T\| \, \|x\|_U$$

which implies

$$\|T' \circ T\| \leq \|T\| \, \|T'\|$$

Theorem 1.8. *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed spaces and $T : V \to W$ be a linear transformation. The following are equivalent*

1. *$\|T\|$ is finite.*
2. *T is bounded.*
3. *T is Lipschitz continuous.*
4. *T is continuous at a point.*
5. $\sup_{\|v\|_V = 1} \|Tv\|_W < \infty$.

Proof. item 1 $\Rightarrow$ item 2: Obviously

$$\frac{\|Tv\|_W}{\|v\|_V} \leq \|T\|$$
$$\implies \|Tv\|_W \leq \|T\| \, \|v\|_V$$

note that if $v = 0$ then $Tv = 0$ as well and thus the last inequality holds for all $v \in V$.

item 2 $\Rightarrow$ item 3:

$$\|Tv - Tu\|_W = \|T(u - v)\|_W \leq C \, \|u - v\|_V$$

item 3 $\Rightarrow$ item 4: Trivial.

item 4 $\Rightarrow$ item 5: Let T be continuous at $u \in V$. Then there is a $\delta > 0$ such that

$$\|v - u\| < \delta \implies \|Tv - Tu\|_W = \|T(v - u)\|_W < 1$$

Now for an arbitrary non-zero v we have

$$\left\| \left(\frac{\delta v}{2 \, \|v\|_V} + u \right) - u \right\|_V < \delta$$

Therefore

$$\left\| T\left(\frac{\delta v}{2\left\|v\right\|_V}\right)\right\|_W < 1$$
$$\left\| T\left(\frac{v}{\left\|v\right\|_V}\right)\right\|_W < \frac{2}{\delta}$$

item 5 $\Rightarrow$ item 1: Let $v \in V$ be an arbitrary vector. Then

$$\sup \left\| T\left(\frac{v}{\left\|v\right\|_V}\right)\right\|_W < \infty$$
$$\implies \sup \frac{\left\|Tv\right\|_W}{\left\|v\right\|_W} < \infty$$

Theorem 1.9. *If V is a finite dimensional normed vector space then any linear transformation $T : V \to W$ is continuous.*

Proof. Since V is finite dimensional, according to Theorem 1.2, any two norms are equivalent. Hence, take $\|\cdot\|_2$ to be Euclidean norm over a basis $\{e_1, \ldots, e_n\}$. Let x be such that $\|x\|_2 < \delta$ for some $\delta > 0$. Therefore, $|\xi_i| < \delta^2$

$$\|Tx\|_W = \left\| \sum_{i=1}^{n} \xi_i T(e_i) \right\|_W \leq \sum_{i=1}^{n} |\xi_i| \, \|T(e_i)\|_W \leq \delta^2 K$$

where $K = \max \|T(e_i)\|_W$. By letting $\delta = \sqrt{\frac{\epsilon}{K}}$ we proved continuity at 0 and hence the continuity by Theorem 1.8. ■

Another proof of Propostion 1.6

Proof. Let $\{e_1, \ldots, e_n\}$ be a basis for V and $\phi : V \to \mathbb{F}^n$ be the representation map for the basis. Since ϕ is a linear map and a bijection then ϕ is homeomorphism. Consider a Cauchy sequence $\{v_k\} \in V$ and let $x_k = \phi(v_k)$ then by continuity of ϕ and ϕ^{-1} we have

$$|x_i - x_j| = |\phi(v_i) - \phi(v_j)| \leq \|\phi\| \, \|v_i - v_j\| \leq \|\phi\| \left\|\phi^{-1}(x_i) - \phi^{-1}(x_j)\right\| \leq \|\phi\| \left\|\phi^{-1}\right\| |x_i - x_j|$$

hence $\{x_k\}$ are Cauchy in $\mathbb{F}^n$ which by completeness of $\mathbb{F}$ implies that they are convergent, $x_k \to x$. Let $v = \phi^{-1}(x)$ then by the right side of the inequality $v_k \to v$. ■

Remark 2. As seen in the last proof, for a bijective linear transformation T

$$1 \leq \|T\| \left\|T^{-1}\right\|$$

Definition (Dual space): Let V be a normed space over the normed field $\mathbb{F}$, then the **topological/continuous dual space** of the normed space V is

$$V^* = \mathcal{L}(V, \mathbb{F})$$

elements of V^* are called **bounded functionals** on V.

Remark 3. Dual space is defined for all vector spaces, however, in analysis we study the topological dual space which only in the finite dimensional case coincide with the algebraic dual space.

Proposition 1.10. *For a finite dimensional normed vector space V, $\dim V^* = \dim V$.*

Proof. Let $\{e_1, \dots, e_n\}$ be a basis for V then, consider the following linear functions

$$e_1^*, \dots, e_n^* \in V^*$$

where

$$e_i^*(e_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

we claim that $\{e_1^*, \dots, e_n^*\}$ is a basis for V^*. It is easy to see that they are as for each j

$$\left[\sum_{i=1}^{n} c_i e_i^*\right] e_j = c_j$$

and for each $\phi \in \mathcal{L}(V, \mathbb{F})$ we have

$$\phi(e_j) = \alpha_i = \sum_{i=1}^{n} \alpha_i e_o^*(e_i)$$

hence $\dim V^* = n = \dim V$. ■

Theorem 1.11. *For two normed vector spaces V, W, $(\mathcal{B}(V,W), \|T\|)$ is a normed vector space. Moreover, it is a Banach space when W is a Banach space.*

Proof. Clearly $\mathcal{B}(V,W)$ is a vector space. For its norm $\|T\|$ we have

1. $\|T\| \geq 0$ by definition.
2. if $\alpha \in \mathbb{F}_W$ then

$$\|\alpha T\| = \sup\left\{\frac{\|(\alpha T)v\|_W}{\|v\|_V} \middle| v \neq 0\right\} = |\alpha| \sup\left\{\frac{\|Tv\|_W}{\|v\|_V} \middle| v \neq 0\right\} = |\alpha| \, \|T\|$$

3. for the triangle inequality

$$\begin{aligned} \|T_1 + T_2\| &= \sup\left\{\frac{\|(T_1 + T_2)v\|_W}{\|v\|_V}\right\} \\ &\leq \sup\left\{\frac{\|T_1 v\|_W + \|T_2 v\|_W}{\|v\|_V}\right\} \\ &= \sup\left\{\frac{\|T_1 v\|_W}{\|v\|_V}\right\} + \sup\left\{\frac{\|T_2 v\|_W}{\|v\|_V}\right\} \\ &= \|T_1\| + \|T_2\| \end{aligned}$$

Suppose W is a Banach space and $\{T_i\} \in \mathcal{B}(V,W)$ is a Cauchy sequence. Then for all $v \in V$

$$\forall \epsilon\, \exists N \text{ s.t. } m, n > N \implies \|T_m v - T_n v\|_W \leq \|T_m - T_n\| \, \|v\|_V < \epsilon$$

$\{T_i v\}$ is a Cauchy sequence. Since W is complete then $T_i v \to Tv$ for some function T. We claim that T is a bounded linear map and is the limit of $T_i \to T$.

$$\begin{aligned} T(v + cu) &= \lim_{i\to\infty} T_i(v + cu) = \lim_{i\to\infty} T_i v + c T_i u \\ &= Tv + cTu \end{aligned}$$

Note that $|\|T_m\| - \|T_n\|| \leq \|T_m - T_n\|$ and hence $\|T_i\|$ is a Cauchy in sequence in $\mathbb{R}$ that has a limit t. There exists a N such that $|\|T_n\| - t| < 1$ for all $n \geq N$.

$$\frac{\|Tv\|_W}{\|v\|_V} = \lim_{i\to\infty} \frac{\|T_i v\|_W}{\|v\|_V} < t + 1$$

hence T is bounded and $T \in \mathcal{B}(V, W)$. Finally, we show that $T_i \to T$. For an arbitrary $v \neq 0$ and $\epsilon > 0$ there exist N such that

$$n \geq N \implies \|T_i v - Tv\|_W < \epsilon \|v\|_V$$

which means that

$$\|T_i - T\| = \sup \frac{\|T_i v - Tv\|_W}{\|v\|_V} < \epsilon$$

Therefore $T_i \to T$ as desired. ■

Theorem 1.12. *Let $(V, \|\cdot\|)$ be a normed space. Then any linear transformation $T : \mathbb{R}^n \to V$ is continuous. Furthermore, if T is a bijection, it is a homeomorphism.*

Proof. Since $\mathbb{R}^n$ is finite then by Theorem 1.9, T is continuous. Assuming T is bijective, we must show that its inverse T^{-1} is continuous as well. Since T is a bijection then T is a linear isomorphism and $\dim V = \dim \mathbb{R}^n = n$ hence $T^{-1} : V \to \mathbb{R}^n$ is a continuous map. ■

Theorem 1.13. *Let V, W be two finite dimensional normed vector spaces. $T : V \to W$ linear transformation is invertible if and only if there exists a c such that:*

$$c \|v\|_V \leq \|Tv\|_W$$

Proof. If T is invertible then $T^{-1} : W \to V$ is bounded and thus

$$\left\|T^{-1} w\right\|_V \leq c \|w\|_W$$

and since T is bijective then there exists v such that $w = Tv$ which implies

$$\|y\|_V \leq c \|Ty\|_W$$

If there exists such c then $\|Tx\| > 0$ for all non-zero x and hence $\ker T = 0$ which implies that T is a bijection and is invertible. ■

Remark 4. the supremum of such c is $\|T^{-1}\|^{-1}$ which is called the **conorm** of T.

Definition (General linear group): The **general linear group** of a vector space, written $\mathrm{GL}(V)$ is the set of all bijective linear transformation.

Proposition 1.14. *If V is a finite (also works for infinite) vector space then $\mathrm{GL}(V)$ is open in $\mathcal{L}(V, V)$, in fact, if $f \in \mathrm{GL}(V)$ then the open ball centered at f with radius $\|f^{-1}\|^{-1}$ remains in $\mathrm{GL}(V)$. Furthermore, the inverse operator $i : \mathrm{GL}(V) \to \mathrm{GL}(V)$, $i(T) = T^{-1}$ is continuous.*

Proof. First assume $f = \mathbb{1}_V$ then we prove that any linear g that $\|\mathbb{1}_V - g\| < 1$ is invertible which then implies bijectivity (true for linear maps). Let $\|v\| = 1$ then

$$\||v| - \|gv\|\| \leq \|v - gv\| \leq \|\mathbb{1}_V - g\| \|v\| < 1$$

Therefore

$$0 < \|gv\| < 2$$

which means $\ker g = \{0\}$ and since V is finite then then g is invertible. For a general f, we have that

$$\left\|1 - f^{-1} \circ g\right\| \leq \left\|f^{-1}\right\| \|f - g\| < 1$$

therefore $f^{-1} \circ g$ is invertible and as a consequence $g = f \circ f^{-1} \circ g$ is invertible. To prove inverse operator is continuous, fix $\epsilon > 0$ then for a $\delta > 0$ if $\|T - S\| < \delta$ then

$$\begin{aligned} &\left\|\mathbb{1}_V - T^{-1} \circ S\right\| = \left\|T^{-1} \circ T - T^{-1} \circ S\right\| \leq \left\|T^{-1}\right\| \|T - S\| < \delta \left\|T^{-1}\right\| \\ \implies &\left\|T^{-1} - S^{-1}\right\| \leq \left\|T^{-1} \circ S - \mathbb{1}_V\right\| \left\|S^{-1}\right\| < \delta \left\|T^{-1}\right\| \left\|S^{-1}\right\| \end{aligned}$$

note that by letting $\delta = \|T^{-1}\|^{-1}/2$ then

$$\|S\| > -\frac{\|T^{-1}\|^{-1}}{2} + \|T\| > \frac{\|T^{-1}\|^{-1}}{2}$$

also if for any invertible linear map R

$$\|R\| > a \implies \|Rx\| > a\|x\| \implies \frac{\|y\|}{a} = \frac{\|R \circ R^{-1}(y)\|}{a} > \left\|R^{-1}y\right\|$$

which means that $\|S^{-1}\| < 2\|T^{-1}\|$, hence by letting

$$\delta = \min\left\{\frac{\epsilon\|T^{-1}\|^2}{2}, \frac{\|T^{-1}\|^{-1}}{2}\right\}$$

we proved the continuity. ■

Definition: Let $V_1, V_2, \ldots, V_n$ be normed vector spaces. Then $\phi : V_1 \times \ldots \times V_n \to W$ is **n-linear** if by fixing any $n - 1$ component, ϕ is linear relative to the remaining component.

Proposition 1.15. *If $V_1, V_2, \ldots, V_n$ are normed vector spaces and $\phi : V_1 \times \ldots \times V_n \to W$ is a n-linear then the followings are equivalent*

1. *ϕ is continuous.*
2. *ϕ is continuous at 0.*
3. *ϕ is bounded, that is there exists a constant $C > 0$ such that*

$$\|\phi(v_1, \ldots, v_n)\|_W \leq C\|v_1\|_{V_1} \ldots \|v_n\|_{V_n}$$

Remark 5. As oppose to linear transformation, n-linear function's continuity does not imply uniform continuity.

Proof. Item 1 $\implies$ Item 2: Trivial.

Item 2 $\implies$ Item 3: For the sake of contradiction, suppose Item 3 is false. That is, for every $k \in \mathbb{N}$ there exists a point $v_k = (v_k^1, \ldots, v_k^n)$ such that

$$\left\|\phi\left(v_k^1, \ldots, v_k^n\right)\right\|_W > n^n \left\|v_k^1\right\|_{V_1} \ldots \left\|v_k^n\right\|_{V_k}$$

Note that v_k^m can not be zero for any k and m, otherwise $\phi(v_k) = 0$. Define

$$w_k^m = \frac{v_k^m}{n \left\|v_k^m\right\|_{V_k}} \to 0$$

which from the continuity at 0 implies that $w_k = (w_k^1, \ldots, w_k^n) \to 0$. However,

$$\left\|\phi(w_k) - \phi(0)\right\|_W > n^n \frac{1}{n} \ldots \frac{1}{n} = 1$$

which is a contradiction.

Item 3 $\implies$ Item 1. Let $v_n \to v$ and define the points

$$\bar{v}_k^m = (v^1, \ldots, v^m, v_k^{m+1}, \ldots, v_k^n), \qquad \bar{v}_k^0 = v_k$$

and $\bar{v}_k^n = v$. Note that v_k^m are bounded for sufficiently large $k \geq N_1$, therefore there exists M such that $\forall m$, $\left\|v_k^m\right\|_{V_m} \leq M$. Also, pick M such that $\forall m$, $\left\|v^m\right\|_{V_m} \leq M$ as well. Then

$$\begin{aligned}
\left\|\phi(v_k) - \phi(v)\right\|_W &\leq \sum_{i=1}^{n} \left\|\phi\left(\bar{v}_k^{i-1}\right) - \phi\left(\bar{v}_k^i\right)\right\|_W \\
&= \sum_{i=1}^{n} \left\|\phi\left(\bar{v}_k^{i-1} - \bar{v}_k^i\right)\right\|_W \\
&\leq \sum_{i=1}^{n} C \left\|v^1\right\|_{V_1} \ldots \left\|v^{i-1}\right\|_{V_{i-1}} \left\|v_k^i - v^i\right\|_{V_i} \left\|v_k^{i+1}\right\|_{V_{i+1}} \ldots \left\|v_k^n\right\|_{V_n} \\
&\leq CM^{n-1} \sum_{i=1}^{n} \left\|v_k^i - v^i\right\|_{V_i}
\end{aligned}$$

pick N_2 such that for all $k \geq N_2$, for each i, $\left\|v_k^i - v^i\right\|_{V_i} < \frac{\epsilon}{nCM^{n-1}}$ then

$$\left\|\phi(v_k) - \phi(v)\right\|_W < CM^{n-1} \sum_{i=1}^{n} \frac{\epsilon}{nCM^{n-1}} = \epsilon$$

We denote the set of all n-linear functions from $V_1 \times \ldots \times V_n \to W$ by $\mathcal{L}^n(V_1 \times \ldots \times V_n, W)$.

Proposition 1.16. *Let $V_1, \ldots, V_n, W$ be normed vector spaces. Then $\mathcal{L}^n(V_1 \times \ldots \times V_n, W)$ and $\mathcal{L}(V_1, \mathcal{L}(V_2, \ldots, \mathcal{L}(V_n, W)))$ are isomorphic.*

Proof. We want to prove

$$\mathcal{L}^n(V_1 \times \ldots \times V_n, W) \cong \mathcal{L}(V_1, \mathcal{L}(V_2, \ldots, \mathcal{L}(V_n, W)))$$

consider the mapping $T : \mathcal{L}(V_1, \mathcal{L}(V_2, \ldots, \mathcal{L}(V_n, W))) \to \mathcal{L}^n(V_1 \times \ldots \times V_n, W)$, such that for any $v_1 \in V_1$, $\ldots$, $v_n \in V_n$

$$\alpha((v_1)(v_2) \ldots (v_n)) = T(\alpha)(v_1, v_2, \ldots, v_n)$$

First note that T is linear. Then if $T(\alpha) = 0$ implies $\alpha = 0$, thus T is injective and hence bijective. ■

Definition (Positive definite): Matrix $A \in \mathcal{M}_n(\mathbb{R})$ is **positive definite** whenever A is symmetric and

$$\forall x \in \mathbb{R}^n \backslash \{0\}\,,\; x^T A x > 0$$

Theorem 1.17. *Every positive definite matrix A is diagonizable. In face, there exists an orthognal matrix P such that*

$$PAP^T = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \dots & 0 & \lambda_n \end{bmatrix}$$

where $\lambda_i > 0$ for each i.

Exercises

1. Show that for a linear transformation T, $\|T\| = \sup_{\|v\|_V \leq 1} \|Tv\|_W$.
2. Prove or disprove that if $x^T A x = x^T A^T x$ for all $x \in \mathbb{R}^n \backslash \{0\}$ then $A = A^T$.

1.2 Derivative

Let V, W be finite dimensional vector spaces and $f : U \subset V \to W$ where U is open. Then f is differentiable at x_0 when a linear transformation $T : V \to W$ such that

$$\lim_{\|h\|\to 0} \frac{\|f(x_0 + h) - f(x_0) - T(h)\|}{\|h\|} = 0$$

Or equivalently there exists a sublinear function $R(h)$ such that

$$f(x_0 + h) - f(x_0) - Th = R(h) \qquad \frac{R(h)}{\|h\|} \to 0$$

T if it exists is unique, represented by $f'(x_0)$, $\mathrm{D}f$, or $\mathrm{d}f(x)$ and called the **total derivative** or **Fréchet derivative**.

Example 1.3. Any linear function $f : V \to W$ with $f(v) = Tv + b$ where $b \in W$ is differentiable and $\mathrm{D}f(v) = T$. Since

$$\|h\|_V < \delta \implies \|f(v + h) - f(v) - (\mathrm{D}f(v))(h)\|_W = \|T(v + h) - Tv - Th\|_W = 0 < \epsilon \|h\|_V$$

Hence, the derivative of any linear function is constant. Consider $S : V \times V \to V$ with $S(v, u) = v + u$. S is differentiable because S is linear (why?). We claim that $\mathrm{D}S = S$ as

$$\|S((v + h), (u + k)) - S(v, u) - S(h, k)\| = 0$$

Example 1.4. Let $\mu : \mathbb{R} \times V \to V$ with $\mu(r, x) = rx$. Then μ is differentiable and $(\mathrm{D}\mu(r, x))(t, h) = rh + tx$ as

$$\begin{aligned} \|\mu((r + t), (x + h)) - \mu(r, x) - (\mathrm{D}\mu(r, x))(t, h)\| &= \|rx + rh + tx + th - rx - rh - tx\| \\ &= |t| \|h\| \le \epsilon \|(t, h)\| \end{aligned}$$

by letting $\|(t, h)\| = \sqrt{t^2 + \|h\|^2}$ and $\delta = \epsilon$.

Proposition 1.18. *Differentiability of f at x implies continuity at x.*

Proof.

$$\|f(x + h) - f(x)\| = \|(\mathrm{D}f(x))(h) + R(h)\| \le \|\mathrm{D}f(x)\| \|v\| + \|R(v)\| \to 0$$

as $v \to 0$. ■

Proposition 1.19. *Assume $f : U \subset V \to W$ is differentiable at x_0 and let $u \in V$ be a non-zero vector then*

$$f'(x_0)(u) = \lim_{t\to 0} \frac{f(x_0 + tu) - f(x_0)}{t}$$

Proof. Let $h = tu$ then

$$\begin{aligned} R(tu) &= f(x_0 + tu) - f(x_0) - T(tu) \\ &= f(x_0 + tu) - f(x_0) - tT(u) \\ \implies \frac{R(tu)}{t} &= \frac{f(x_0 + tu) - f(x_0)}{t} - T(u) \\ \implies \lim_{t\to 0} \frac{f(x_0 + tu) - f(x_0)}{t} &= T(u) \end{aligned}$$

Definition (Directional derivative): If we let $\|u\| = 1$ then the limit in Propostion 1.19 becomes the **directional derivative** of f in the direction of u and is denoted by $\mathrm{D}_u f$.

Remark 6. The existence of all directional derivatives of f doesnt imply its differentiability or even its continuity.

Remark 7. If $\mathrm{D}f : U \to \mathcal{L}(V, W)$ is continuous then each $\dfrac{\partial f_i}{\partial x_j}$ is continuous. Since

$$\mathrm{D}f(x) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(x) & \dots & \dfrac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \\ \dfrac{\partial f_m}{\partial x_1}(x) & \dots & \dfrac{\partial f_m}{\partial x_n}(x) \end{bmatrix}$$

and the reverse is true as well.

Theorem 1.20. *$f : V \to W$ has all of its partial derivative in a neighbourhood of $u \in U$ and they're continuous at u then f is differentiable at u. Especially, if $\dfrac{\partial f_i}{\partial x_j}$ exist and are continuous at every point of U then $f \in \mathcal{C}^1$.*

Proof. We prove that each f_i is differentiable. Let $\{e_1, \dots, e_n\}$ be a basis for V and take $\|x\| = \sum |\xi_j|$. Consider a convex neighbourhood E of a. Then, for a given $\epsilon > 0$ we will show there exists a $\delta > 0$ such that

$$\|h\| < \delta \implies \left\| f_i(a+h) - f_i(a) - \sum_{j=1}^{n} \big(\mathrm{D}_{e_j} f_i(a)\big)(h_j) \right\| \leq \epsilon \|h\|$$

Cosider the point sequence $a^k = \sum_{j<k} a_j e_j + \sum_{j \geq k} (a_j + h_j) e_j$ where $a^1 = a + h$ and $a^{n+1} = a$ then

$$\left\| f_i(a+h) - f_i(a) - \sum_{j=1}^{n} \big(\mathrm{D}_{e_j} f_i(a)\big)(h_j) \right\| \leq \sum_{k=1}^{n} \left\| f_i\big(a^k\big) - f_i\big(a^{k+1}\big) - (\mathrm{D}_{e_k} f_i(a))(h_k) \right\|$$

hence we are done if

$$\left\| f_i\big(a^k\big) - f_i\big(a^{k+1}\big) - (\mathrm{D}_{e_k} f_i(a))(h_k) \right\| \leq \epsilon |h_k|$$

for $k = n$

$$\| f_i(a^n) - f_i(a) - (\mathrm{D}_{e_n} f_i(a))(h_n) \|$$

which equivalent to the existence n_{th} partial derivative of a. and for $k < n$

$$\begin{aligned} &\left\| f_i\big(a^k\big) - f_i\big(a^{k+1}\big) - (\mathrm{D}_{e_k} f_i(a))(h_k) \right\| \\ &\leq \left\| f_i\big(a^k\big) - f_i\big(a^{k+1}\big) - \big(\mathrm{D}_{e_k} f_i\big(a^k\big)\big)(h_k) \right\| + \left\| \big(\mathrm{D}_{e_k} f_i\big(a^k\big)\big)(h_k) - (\mathrm{D}_{e_k} f_i(a))(h_k) \right\| \end{aligned}$$

which uses the existence of partial derivatives in neighbourhood and its continuity. ■

Proposition 1.21. *Let $f, g : V \to W$ be differentiable at x and $h : W \to U$ be differentiable at $y = f(x)$. Furthermore, let c be an scalar then*

1. $\mathrm{D}(f+cg) = \mathrm{D}f + c\,\mathrm{D}g$.
2. *$h \circ f$ is differentiable at x and*

$$\mathrm{D}(h \circ f) = ((\mathrm{D}h) \circ f) \circ \mathrm{D}f$$

Proof.

1. we have

$$\begin{aligned}&\|(f+cg)(x+k) - (f+cg)(x) - (\mathrm{D}f(x) + c\,\mathrm{D}g(x))(k)\| \\ &\leq \|f(x+k) - f(x) - (\mathrm{D}f(x))(h)\| + |c|\,\|g(x+k) - g(x) - (\mathrm{D}g(x))(h)\|\end{aligned}$$

2. we know that

$$\begin{cases} f(x+k) - f(x) - (\mathrm{D}f(x))(k) = R(k) \\ h(y+l) - h(y) - (\mathrm{D}h(y))(l) = S(l) \end{cases}$$

and we wish to prove that

$$h \circ f(x+k) - h \circ f(x) - (\mathrm{D}h(f(x)) \circ \mathrm{D}f(x))(k) = T(k)$$

where $\|T(k)\| \leq \epsilon\,\|k\|$ whenever $\|k\| < \delta$. Let $l = f(x+k) - f(x)$ and substituting into the second equation

$$\begin{aligned}&h(f(x+k)) - h(f(x)) - (\mathrm{D}h(y))(f(x+k) - f(x)) \\ &\qquad = h(f(x+k)) - h(f(x)) - (\mathrm{D}h(y))((\mathrm{D}f(x))(k) + R(k)) \\ &\qquad = h(f(x+k)) - h(f(x)) - (\mathrm{D}h(y) \circ \mathrm{D}f(x))(k) - (\mathrm{D}h(y))(R(k)) \\ &\qquad = T(k) - (\mathrm{D}h(y))(R(k)) = S(l) \\ &\implies T(k) = S(l) + (\mathrm{D}h(y))(R(k))\end{aligned}$$

Proposition 1.22. *$f : U \subset V \to W_1 \times \ldots \times W_n$ is differentiable at x_0 if and only if all its component is differentiable at x_0. Furthermore, $\mathrm{D}f = (\mathrm{D}f_1, \ldots, \mathrm{D}f_n)$.*

Proof. Define the following norm on $W_1 \times \ldots \times W_n$

$$\|(w_1, \ldots w_n)\| = \sum_{i=1}^{n} \|w_i\|_{W_i} \tag{1.1}$$

then

$$\|f(x_0+h) - f(x_0) - (\mathrm{D}f(a))(h)\| = \sum_{i=1}^{n} \|f_i(x_0+h) - f_i(x_0) - (\mathrm{D}f_i(a))(h)\|$$

and since every other norm is equivalent to the norm defined above, we are done. ■

Theorem 1.23 (Leibnitz rule). *Let $V_1, V_2, \ldots, V_n$ be finite dimensional vector spaces and $f : V_1 \times \ldots \times V_n \to W$ is a n-linear function. f is differentiable at $a = (a_1, \ldots, a_n)$ and*

$$(\mathrm{D}f(a))(h_1, \ldots h_n) = f(h_1, a_2, \ldots, a_n) + f(a_1, h_2, \ldots, a_n) + \cdots + f(a_1, a_2, \ldots, h_n)$$

Proof. we have that
$$f(a+h) = \sum_{\xi_i \in \{a_i, h_i\}} f(\xi_1, \dots, \xi_n)$$
therefore
$$f(a+h) - f(a) - \sum_{i=1}^{n} f(a_1, \dots, a_{i-1}, h_i, a_{i+1}, \dots, a_n) = \sum_{\substack{\xi_i \in \{a_i, h_i\} \\ \text{at least two } h_i}} f(\xi_1, \dots, \xi_n)$$
Let $\delta = 1$ then $\|h\| = \sum \|h_i\| < 1$ also $i, j,\ \|h_i\| \|h_j\| \leq \|h\|^2$. Hence if we define
$$A = \max \left\{ \prod_{i \in I} \|a_i\| \,\middle|\, I \subset \mathbb{N}_n \right\}$$
then
$$\sum_{\substack{\xi_i \in \{a_i, h_i\} \\ \text{at least two } h_i}} f(\xi_1, \dots, \xi_n) \leq (2^n - n - 1) A \|h\|^2$$
and letting $\delta = \min \left\{ 1, \dfrac{\epsilon}{(2^n - n - 1)(A+1)} \right\}$ we arrive at the conclusion. ∎

Example 1.5. Let $Z : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ with $Z(u, v) = u \times v$ be a bilinear function, $f, g : \mathbb{R} \to \mathbb{R}^3$ and $h(t) = f(t) \times g(t)$. $h = Z \circ \phi$ where $\phi(t) = (f(t), g(t))$. Then we have:
$$\begin{aligned} \mathrm{D}h(t) &= (\mathrm{D}Z)(\phi(t)) \circ \mathrm{D}\phi(t) \\ &= (\mathrm{D}Z)(\phi(t)) \circ (\mathrm{D}f(t), \mathrm{D}g(t)) \\ &= Z(\mathrm{D}f(t), g(t)) + Z(f(t), \mathrm{D}g(t)) \\ &= \mathrm{D}f(t) \times g(t) + f(t) \times \mathrm{D}g(t) \end{aligned}$$

Example 1.6. Consider $A = [f_{ij}(x_1, \dots, x_n)]$ where each f_{ij} is differentiable. Then
$$\mathrm{D}\det(A)$$
can be calculated using the Leibnitz rule, since determinant is n-linear function.

1.2.1 Mean value theorem

Mean value theroem of 1-dimensional does not generalize very well. For example, the continuous function $f(t) :\ [0, 1] \to \mathbb{R}^2$ with
$$t \mapsto (t^2, t^3)$$
is differentiable on $]0, 1[$, however
$$\begin{aligned} f(1) - f(0) = (1, 1) &= \mathrm{D}f(c)\,(1 - 0) \\ &= (2c, 3c^2) \end{aligned}$$
which has no solution for $c \in\ ]0, 1[$. Although it must be said that for $f : U \to \mathbb{R}$ where $U \subset V$ is convex, the mean value theorem holds.

Theorem 1.24. *Let V, W be normed finite dimensional vector spaces and $f : U \to W$ is differentiable and $A, B \in U$ are such that the line connecting in completely contained in U and for each p on that line*

$$\|\mathrm{D}f(p)\| \leq M$$

then

$$\|f(B) - f(A)\|_W \leq M \|B - A\|_V$$

First consider the following lemma:

Lemma 1.25. *If $\phi : [0,1] \to W$ is continuous, differentiable on $]0,1[$ and $\|\phi'(t)\| \leq M$ for all $t \in]0,1[$ then*

$$\|\phi(1) - \phi(0)\|_W \leq M$$

Proof. We provide three proofs for the lemma

1. Assuming the norm on W is induced by an inner product. Then, let $e = \frac{\phi(1)-\phi(0)}{\|\phi(1)-\phi(0)\|}$ be a unit vector in W then $\psi : [0,1] \to \mathbb{R}$, $\psi(t) = e \cdot \phi(t)$ is continuous and differentiable on $]0,1[$. By the mean the value theorem

$$\begin{aligned} |\psi(1) - \psi(0)| &= |\psi'(t_0)| \\ |e \cdot (\phi(1) - \phi(0))| &= |e \cdot \phi'(t)| \\ \|\phi(1) - \phi(0)\| &\leq M \end{aligned}$$

2. Using the Hahn-Banach theorem, that is for a finite dimensional vector space V and $e \in V$ with $\|e\| = 1$ there exists a linear function $\theta : V \to \mathbb{R}$ such that $\|\theta\| = 1$ and $\theta(e) = 1$. Now let $\psi(t) = \theta(\phi(t))$ and take e as defined above then

$$\begin{aligned} |\psi(1) - \psi(0)| &= |\psi'(t_0)| \\ |\theta(\phi(1) - \phi(0))| &= (\mathrm{D}\theta(\phi(t_0)))(\phi'(t_0)) \\ \|\phi(1) - \phi(0)\| &= \theta(\phi'(t_0)) \leq \|\theta\| \|\phi'(t_0)\| \leq M \end{aligned}$$

3. From Hoimander. For any ϵ consider the set T_ϵ.

$$T_\epsilon = \{t \in [0,1] \mid \forall s,\ 0 \leq s \leq t,\ \|\phi(s) - \phi(0)\| \leq (M+\epsilon)s + \epsilon\}$$

first note that $T_\epsilon = [0,c]$ for some $c > 0$ because for $s = 0$ the inequality is strict and both sides are continuous with respect to s. We claim that $c = 1$ because otherwise $c < 1$ and by differentiability of ϕ, there exists a $\delta < 1 - c$ such that if

$$\begin{aligned} \|h\| < \delta \implies \|\phi(c+h) - \phi(c) - (\mathrm{D}\phi(c))(h)\| &\leq \epsilon \|h\| \\ \implies \|\phi(c+h) - \phi(c)\| &\leq \|h\| (\epsilon + \|\phi'(c)\|) \\ &\leq \|h\| (\epsilon + M) \end{aligned}$$

also since $c \in T_\epsilon$

$$\begin{aligned} \|\phi(c) - \phi(0)\| &< (M+\epsilon)c + \epsilon \\ \implies \|\phi(c+h) - \phi(0)\| &< (M+\epsilon)(c+h) + \epsilon \qquad 0 < h < \delta \end{aligned}$$

hence $c + h \in T_\epsilon$ which is a contradiction and thus $c = 1$. ☐

Proof. Let $\sigma : [0,1] \to U$ be a parameterization of the line connecting the point A to point B, $\sigma(t) = (1-t)A + tB$. Let $\phi = f \circ \sigma$, then clearly ϕ is continuous on $[0,1]$ and differentiable on $]0,1[$ and we have

$$\begin{aligned} \phi'(t) &= (\mathrm{D}f(\sigma(t)))(\sigma'(t)) \\ &= (\mathrm{D}f(\sigma(t)))(B-A) \\ \implies \|\phi'(t)\| &\leq \|\mathrm{D}f(\sigma(t))\| \, \|B-A\|_V \leq M \, \|B-A\|_V \end{aligned}$$

therefore by the Lemma 1.25

$$\|f(B) - f(A)\|_W = \|\phi(1) - \phi(0)\|_W \leq M \, \|B-A\|_V$$

which concludes the proof. ■

Corollary 1.26. *Let $U \subset V$ is connected and open and $f : U \to W$ is differentiable and $\mathrm{D}f(u) = 0$ for all $u \in U$ then f is constant.*

Proof. Let $p \in U$ and $S = \{q \in U \mid f(q) = f(p)\}$. S is closed because f is continuous and hence the pre-image closed set $\{f(p)\}$ is closed. For each $q \in S$ there exists $r > 0$ such that $B_r(q) \subset U$ and since $B_r(q)$ is convex then for each $l \in B_r(q)$ we apply the Theorem 1.24

$$\|f(l) - f(q)\| \leq \sup \|\mathrm{D}f(t)\| \, \|l - q\| = 0$$

which implies that $f(l) = f(q) = f(p)$ hence S is open in U which by the connectedness of U means $S = U$. Therefore, f is constant on U. ■

Corollary 1.27. *Let V_1, V_2, W be finite dimensional normed vector space and $U \subset V_1 \times V_2$ is open such that for every $y \in V_2$ the intersection $(V_1 \times \{y\}) \cap U$ is connected. Assumne $f : U \to W$ is differentiable and $\mathrm{D}_{V_1} f(x,y) = 0$ for all $(x,y) \in U$ then for any two point $(x_1, y), (x_2, y) \in U$, $f(x_1, y) = f(x_2, y)$.*

Proof. Fix $y \in V_2$ and define the function $g : V_1 \to W$

$$g(x) = f(x,y)$$

therefore

$$\mathrm{D}g(x) = \mathrm{D}_{V_1} f(x,y) = 0$$

and since $(V_1 \times \{y\}) \cap U$, the domain of g is connected. Hence by applying the Corollary 1.26 we get that

$$g(x) = c \implies f(x_1, y) = f(x_2, y)$$

for all $y \in V_2$. ■

1.2.2 Fundamental theorem of calculus

Theorem 1.28. *Let U be an open set of V such that for every $A, B \in U$ the line segment connecting A and B remains in U and let $\sigma : [0,1] \to U$ be that line, $\sigma(t) = (1-t)A + tB$, and lastly let $f : U \to W$ is continuously differentiable. Then*

$$f(B) - f(A) = T(B - A)$$

where T is

$$T = \int_0^1 \mathrm{D}f(\sigma(t)) \, \mathrm{d}t$$

Proof. Let $g_i : [0,1] \to \mathbb{R}$ be

$$g_i(t) = \pi_i \circ f(\sigma(t))$$

is continuously differentiable then by the fundamental theorem of calculus for the real-valued functions we have

$$\begin{aligned}
g(1) - g(0) &= \int_0^1 g'(t)\,\mathrm{d}t \\
&= \int_0^1 \pi_i \circ \mathrm{D}f(\sigma(t))\,\mathrm{d}t \\
&= \pi_i \circ \int_0^1 \mathrm{D}f(\sigma(t))\,\mathrm{D}\sigma(t)\,\mathrm{d}t \\
&= \pi_i \circ \int_0^1 \mathrm{D}f(\sigma(t))\,(B-A)\,\mathrm{d}t \\
\implies \pi_i \circ (f(B) - f(A)) &= \pi_i \circ T(B-A) \\
\implies f(B) - f(A) &= T(B-A)
\end{aligned}$$

which was what was wanted. ■

Theorem 1.29. *Consider the continuous function $T : U \times U \to \mathcal{L}(V, W)$ which is such that*

$$f(B) - f(A) = (T(A,B))(B-A)$$

then $f \in \mathcal{C}^1$ and $\mathrm{D}f(A) = T(A,A)$

Proof. We have

$$f(A+h) - f(A) = (T(A+h, A))(h)$$

hence

$$\begin{aligned}
\|f(A+h) - f(A) - (T(A,A))(h)\| &= \|(T(A+h,A))(h) - (T(A,A))(h)\| \\
&\le \|T(A+h,A) - T(A,A)\|\,\|h\|
\end{aligned}$$

now by continuity of T, there exists a $\delta > 0$ such that

$$\|(h,k)\| < \delta \implies \|T(A+h, A+k) - T(A,A)\| < \epsilon$$

By letting $k = 0$ we get $\mathrm{D}f(A) = T(A,A)$. Since T is continuous then $f \in \mathcal{C}^1$ as well. ■

Corollary 1.30. *Let V be a normed finite dimensional vector space and U is open subset of V. If*

$$f : [a,b] \times U \to \mathbb{R}$$

is continuous then

$$F(y) = \int_a^b f(x,y)\,\mathrm{d}x$$

is continuous. Furthermore, if $\dfrac{\partial f}{\partial y_i}$ exists and is continuous then $\dfrac{\partial F}{\partial y_i}$ exists and is continuous as well.

$$\frac{\partial F}{\partial y_i} = \int_a^b \frac{\partial f}{\partial y_i}(x,y)\,\mathrm{d}x$$

Proof. Firstly, we want to show that there exists a $\delta > 0$ such that for each $y \in U$

$$\|h\| < \delta \implies \|F(y+h) - F(y)\| < \epsilon$$

we have that

$$\begin{aligned}\|F(y+h) - F(y)\| &= \left\| \int_a^b f(x, y+h)\, f(x,y)\,\mathrm{d}x \right\| \\ &\leq (b-a) \sup_{x \in [a,b]} \{f(x, y+h)\, f(x,y)\}\end{aligned}$$

note that from the continuity of f for each $x \in [a,b]$ and $y \in U$ there are open balls $I_{x,y}$ around x and $J_{x,y}$ around y such that

$$x' \in I_{x,y},\ y' \in J_{x,y} \implies \|f(x',y') - f(x,y)\| < \frac{\epsilon}{b-a}$$

Fix y_0, then $\cup I_{x,y_0} \supset [a,b]$ which by the compactness of the interval implies that there is a finite family of there open set the covers $[a,b]$. Setting δ to the minimum radius of J_{x,y_0} yields the result. Secondly, we show that there exists a $\delta > 0$ such that

$$|h| < \delta \implies \left\| \frac{F(y+he_i) - F(y)}{h} \right\| < \epsilon$$

and we have that

$$\begin{aligned}\frac{F(y+he_i) - F(y)}{h} &= \frac{1}{h}\int_a^b f(x, y+he_i) - f(x,y)\,\mathrm{d}x \\ &= \frac{1}{h}\int_a^b \frac{\partial f}{\partial y_i}(x, y+the_i)\, h\,\mathrm{d}x \\ &= \int_a^b \frac{\partial f}{\partial y_i}(x, y+the_i)\,\mathrm{d}x\end{aligned}$$

from the previous part we know that we can make

$$\left\| \frac{\partial f}{\partial y_i}(x,y') - \frac{\partial f}{\partial y_i}(x,y) \right\|$$

as small as we want by making $\|y - y'\| < \delta$ small independently of x. Therefore, there exist a $\delta > 0$ such that if $|th| < |h| < \delta$ then

$$\left\| \frac{\partial f}{\partial y_i}(x,y') - \frac{\partial f}{\partial y_i}(x,y) \right\| < \frac{\epsilon}{b-a}$$

hence

$$\left\| \frac{F(x, y+he_i) - F(x,y)}{h} - \int_a^b \frac{\partial f}{\partial y_i}(x,y)\,\mathrm{d}x \right\| = \left\| \int_a^b \frac{\partial f}{\partial y_i}(x, y+the_i) - \frac{\partial f}{\partial y_i}(x,y)\,\mathrm{d}x \right\| < \frac{\epsilon}{b-a}$$

and the continuity of $\dfrac{\partial F}{\partial y_i}$ comes as a result of applying the first part to $\dfrac{\partial f}{\partial y_i}$. ■

1.2.3 Higher derivative

Let V, W be finite dimensional normed vector spaces with $(e_1, \ldots, e_n)$ is an ordered basis for V. Consider $U \subset V$ is an open set and $f : U \to W$. If f is differentiable then its partial derivatives

$$\mathrm{D}_i f : U \to E \quad \text{with} \quad (\mathrm{D}_i f)(x) = (\mathrm{D}f(x))(e_i)$$

Then, clearly if $\mathrm{D}_i f$ is differentiable one can define its partial derivatives $(\mathrm{D}_j)(\mathrm{D}_i f)$ also denoted by

$$(\mathrm{D}_j)(\mathrm{D}_i f) = \frac{\partial^2 f}{\partial x_j \partial x_i} = \mathrm{D}_{ji} f$$

For Fréchet derivative, if $\mathrm{D}f : U \to \mathcal{L}(V, W)$ is differentiable at x, then f is twice differentiable and

$$\mathrm{D}^2 f(x) = (\mathrm{D}(\mathrm{D}f))(x) : U \xrightarrow{\text{linear map}} \mathcal{L}(V, W)$$

is a linear map. Therefore,

$$\mathrm{D}^2 f : U \to \mathcal{L}(V, \mathcal{L}(V, W))$$

which by the Propostion 1.16 is equivalent to $\mathcal{L}^2(V \times V, W)$ and one can define

$$\mathrm{d}^2 f : U \to \mathcal{L}^2(V \times V, W)$$

where $\mathrm{d}^2 = T(\mathrm{D}^2)$ as defined in Propostion 1.16. With this definition, for the higher order derivatives $n \geq 2$

$$\mathrm{d}^n : U \to \mathcal{L}^n(V^n, W)$$

Example 1.7. Let $A : V \to W$ be a affine function $A(x) = Lx + b$ where L is linear. Then, $\mathrm{D}A(x) = L$ and hence $\mathrm{D}^2 A = 0$.

Example 1.8. Let $\beta : V \times V \to W$ be a bilinear function. By the Leibnitz rule

$$(\mathrm{D}\beta(x_1, x_2))(h_1, h_2) = \beta(x_1, h_2) + \beta(h_1, x_2)$$

therefore $\mathrm{D}\beta : V \times V \to \mathcal{L}(V \times V, W)$ is a linear a function itself, since

$$(\mathrm{D}\beta(x_1 + x_1', x_2 + x_2'))(h_1, h_2) = \beta(x_1, h_2) + \beta(x_1', h_2) + \beta(x_1, h_2) + \beta(b_1, x_2')$$

which means $(\mathrm{D}(\mathrm{D}\beta))(x) = \mathrm{D}\beta$ independent of x.

Theorem 1.31. *If f is twice differentiable at p then its second partial derivatives exist at p. Conversely, if its second partial derivatives exist at a neighbourhood of p and they are continuous, then f is differentiable.*

Proof. Assume that $\mathrm{D}^2 f(p)$ exists. Then

$$\begin{aligned}
\mathrm{D}_j \left(\mathrm{D}_i f(p)\right) &= \lim_{h \to 0} \frac{(\mathrm{D}f(p + he_j))(e_i) - (\mathrm{D}f(p))(e_i)}{h} \\
&= \left(\lim_{h \to 0} \frac{\mathrm{D}f(p + he_j) - \mathrm{D}f(p)}{h}\right)(e_i)
\end{aligned}$$

which exists sincce $\mathrm{D}f$ is differentiable at p. Conversely, assume that the second partials exist and are continuous at p. Then

$$(\mathrm{D}(\mathrm{D}f))(p) = \left[\frac{\partial\, \mathrm{D}f}{\partial x_1}(p) \quad \ldots \quad \frac{\partial\, \mathrm{D}f}{\partial x_n}(p)\right]$$

note that each $\dfrac{\partial\,\mathrm{D}f}{\partial x_i}(p)$ is in $\mathcal{L}(V,W)$. In fact, since

$$\mathrm{D}f = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \dots & \dfrac{\partial f}{\partial x_n} \end{bmatrix}$$

then

$$\frac{\partial\,\mathrm{D}f}{\partial x_i} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_i \partial x_1} & \dots & \dfrac{\partial^2 f}{\partial x_i \partial x_n} \end{bmatrix}$$

which is continuous at p and hence $\dfrac{\partial\,\mathrm{D}f}{\partial x_i}(p)$ is continuous and by Theorem 1.20, $\mathrm{D}f$ is differentiable at p. ■

Remark 8. In general, one can show that $f \in \mathcal{C}^r$ is equivalent to its partial being in $\mathcal{C}^r$.

Let $f : U \to \mathbb{F}$ then $\mathrm{D}f : U \to \mathcal{L}(V, \mathbb{F})$ which is the topological dual space V^* therefore

$$\mathrm{D}f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} e_i^*$$

then for the second derivative of f, $\mathrm{D}^2 f(x) : U \to V^*$

$$\begin{aligned}
\mathrm{D}^2 f(x) &= \sum_{i=1}^{n} \frac{\partial\,\mathrm{D}f}{\partial x_i}(x)\, e_i^* \\
&= \sum_{i=1}^{n} \frac{\partial \sum_{j=1}^{n} \dfrac{\partial f}{\partial x_j}}{\partial x_i}(x)\, e_j^* e_i^* \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} e_j^* e_i^* \\
\implies \big(\big(\mathrm{D}^2 f(x)\big)(e_i)\big)(e_j) &= \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \\
\implies \mathrm{d}^2 f(x)(e_i, e_j) &= \frac{\partial^2 f}{\partial x_i \partial x_j}(x)
\end{aligned}$$

Definition (Hessian matrix): If for a function $f : U \to \mathbb{F}$ all of its second partial derivatives exist then **hessian matrix** is

$$\begin{bmatrix}
\dfrac{\partial^2 f}{\partial x_1^2}(x) & \dots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\
\vdots & \ddots & \vdots \\
\dfrac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \dfrac{\partial^2 f}{\partial x_n^2}(x)
\end{bmatrix}$$

Theorem 1.32. *If f is twice differentiable at x, $\mathrm{d}^2 f(x)$ is symmetric. That is,*

$$\mathrm{d}^2 f(x)(h,k) = \mathrm{d}^2 f(x)(k,h)$$

Proof. Let $\|h\|$ and $\|k\|$ be sufficiently small such that $a+th, a+tk, a+th+tk$ and the lines connecting them stays in U for some $t \in \mathbb{R}$. Consider

$$\Delta(t,h,k) = f(a+th+tk) - f(a+th) - f(a+tk) + f(a)$$

Assuming f is a real-valued twice differentiable function then if we prove

$$\left(\mathrm{d}^2 f(x)\right)(h,k) = \lim_{t\to 0} \frac{\Delta(t,h,k)}{t^2}$$

we are done since, Δ is symmetric with respect to h and k. Now consider

$$g(s) = f(a+th+tsk) - f(a+tsk)$$

then by the the Mean value theorem

$$\begin{aligned}\Delta(t,h,k) = g(1) - g(0) = g'(\xi)& \\ &= (\mathrm{D}f(a+th+t\xi k))(tk) - (\mathrm{D}f(a+t\xi k))(tk)\end{aligned}$$

and since $\mathrm{D}f$ is differentiable then by definition

$$\implies \mathrm{D}f(a+x) = \mathrm{D}f(a) + \left(\mathrm{D}^2 f(a)\right)(x) - R(x)$$

therefore

$$\begin{aligned}\Delta(t,h,k) = t\left(\mathrm{D}f(a) + \left(\mathrm{D}^2 f(a)\right)(th+t\xi k) - R(th+t\xi k)\right)(k)& \\ - t\left(\mathrm{D}f(a) + \left(\mathrm{D}^2 f(a)\right)(t\xi k) - R(t\xi k)\right)(k)&\end{aligned}$$

then

$$\begin{aligned}\Delta(t,h,k) &= t\left(\left(\mathrm{D}^2 f(a)\right)(th+t\xi k) - \left(\mathrm{D}^2 f(a)\right)(t\xi k)\right)(k) - t\left(R(t\xi k) - R(th+t\xi k)\right)(k) \\ &= t^2\left(\left(\mathrm{D}^2 f(a)\right)(h)\right)(k) - t\left(R(t\xi k) - R(th+t\xi k)\right)(k) \\ \implies \frac{\Delta(t,h,k)}{t^2} &= \left(\left(\mathrm{D}^2 f(a)\right)(h)\right)(k) - \frac{(R(t\xi k) - R(th+t\xi k))(k)}{t} \to \left(\left(\mathrm{D}^2 f(a)\right)(h)\right)(k)\end{aligned}$$

which is what we wanted. ■

Theorem 1.33. *The k_{th} derivative of a k-times differentiable function is a symmetric k-linear function.*

Proof. it's generalization of above. ■

Proposition 1.34. *If $f, g \in \mathcal{C}^r$ are two functions then $f \circ g \in \mathcal{C}^r$.*

Proof. Let $f : V' \to V''$ and $g : V \to V'$ be two $\mathcal{C}^r$ functions and $\beta : \mathcal{L}(V', V'') \times \mathcal{L}(V, V') \to \mathcal{L}(V, V'')$ is a bilinear function such that

$$\beta(\phi, \psi) = \phi \circ \psi$$

Now note that

$$\begin{aligned}(\mathrm{D}(f\circ g))(a) &= (\mathrm{D}f \circ g)(a) \circ \mathrm{D}g(a) \\ &= \beta((\mathrm{D}f\circ g)(a)\,, \mathrm{D}g(a))\end{aligned}$$

Consider the following functions

$$a \xmapsto[\mathcal{C}^\infty]{\Delta} (a,a) \xmapsto[\mathcal{C}^{r-1}]{(\mathrm{D}f\circ g, \mathrm{D}g)} ((\mathrm{D}f\circ g)(a)\,, \mathrm{D}g(a)) \xmapsto[\mathcal{C}^\infty]{\beta} (\mathrm{D}(f\circ g))(a)$$

therefore $\mathrm{D}(f\circ g) \in \mathcal{C}^{r-1}$ and hence $f \circ g \in \mathcal{C}^r$. ■

Example 1.9. The inverse operator $i : \mathrm{GL}(V) \to \mathcal{L}(V,V)$ is in C^∞. Remember that

$$((\mathrm{D}i)(A))(M) = -A^{-1}MA^{-1}$$

Let $\gamma : \mathcal{L}(V,V) \times \mathcal{L}(V,V) \to \mathcal{L}(\mathcal{L}(V,V), \mathcal{L}(V,V))$ with

$$(\gamma(A,B))(M) = -AMB$$

is a bilinear function. Therefore

$$((\mathrm{D}i)(A))(M) = \big(\gamma\big(A^{-1}, A^{-1}\big)\big)(M)$$

now

$$A \overset{i}{\mapsto} A^{-1} \xmapsto[\mathcal{C}^\infty]{\Delta} (A^{-1}, A^{-1}) \xmapsto[\mathcal{C}^\infty]{\gamma} (\mathrm{D}i)(A)$$

Since we have proved that i is differentiable then $\mathrm{D}i$ is differentiable which means i is twice differentiable and so on. Hence $i \in C^\infty$.

As a matter of notation if $\phi : V_1 \times \ldots \times V_n$ be an n-linear then

$$\phi \cdot h_1 \ldots h_n := \phi(h_1, \ldots, h_n)$$

particularly if $V_1 = \cdots = V_n$

$$\phi \cdot h^n := \phi(h, \ldots, h)$$

Now one can describe a homogeneous polynomial of degree k with a symmetric k-linear function

$$p(x) = \phi \cdot x^k$$

Then, $p(x)$ is differentiable since

$$x \xmapsto[\mathcal{C}^\infty]{\Delta} (x, \ldots, x) \xmapsto[\mathcal{C}^\infty]{\phi} p$$

and

$$\begin{aligned}(\mathrm{D}p(x))(h) &= (\mathrm{D}\phi(\Delta(x)) \circ \mathrm{D}\Delta(x))(h) \\ &= \mathrm{D}\phi \cdot x^n \circ \Delta(h) \\ &= k\phi \cdot x^{k-1}h \\ \implies \mathrm{D}p(x) &= k\phi \cdot x^{k-1}\end{aligned}$$

Theorem 1.35 (Taylor approximation). *Let $f : U \to W$ be k-times differentiable at a, then*

$$p_k(x) = f(a) + \mathrm{d}f(a) \cdot (x-a) + \frac{1}{2!}\mathrm{d}^2 f(a) \cdot (x-a)^2 + \cdots + \frac{1}{k!}\mathrm{d}^k f(a)\,(x-a)^k$$

*is k_{th} degree **Taylor** polynomial. Then the followings hold*

1.

$$\lim_{x \to a} \frac{f(x) - p_k(x)}{\|x-a\|^k} = 0$$

2. *$p_k(x)$ is the only k_{th} degree polynomial with such property.*
3. *Additionally, if f is $(k+1)$-times differentiable in a neighbourhood of a then the remainder*

$$R(x) = f(x) - p_k(x)$$

can be estimated with

$$\|R(b)\| \leq \frac{1}{(k+1)!} \sup \left\{ \left\| \mathrm{D}^{k+1} f(\xi) \right\| \right\} \|b-a\|^{k+1}$$

where ξ is on line connecting a to b.

Proof.

1. for $k = 1$ it is equivalent to differentiability of f. By induction, assume it is true for $k = n-1$ and let $g(x) = f(x) - p_k(x)$ then [1]

$$\begin{aligned} \mathrm{D}g(x) &= \mathrm{D}f(x) - \mathrm{D}p_k(x) \\ &= \mathrm{d}f(x) - \mathrm{D}\left[f(a) + \mathrm{d}f(a) \cdot (x-a) + \cdots + \frac{1}{n!}\, \mathrm{d}^n f(a)\, (x-a)^n \right] \\ &= \mathrm{d}f(x) - \left[\mathrm{d}f(a) + \frac{1}{1!}\, \mathrm{d}^2 f(a) \cdot (x-a) + \cdots + \frac{1}{(n-1)!}\, \mathrm{d}^n f(a)\, (x-a)^{n-1} \right] \end{aligned}$$

which is equivalent to the proposition at $n-1$ for $\mathrm{d}f(a)$ and hence there exists a $\delta > 0$ such that if $\|x-a\| < \delta$

$$\|\mathrm{D}g(x)\| \leq \epsilon \|x-a\|^{n-1}$$

by the Theorem 1.24 we have

$$\begin{aligned} \|g(x)\| = \|g(x) - g(a)\| &\leq \|x-a\| \sup \|\mathrm{D}g(\xi)\| \\ &\leq \epsilon \|x-a\| \, \|\xi - a\|^{k-1} \\ &\leq \|x-a\|^k \end{aligned}$$

2. If there were two such polynomial p_1, p_2 then for $q = p_1 - p_2$ we have that

$$\lim_{x \to a} \frac{q(x)}{\|x-a\|^k} = 0$$

then one can show that $q(x) \equiv 0$.

3. Define $g : [0,1] \to W$ as such

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

therefore

$$g^{(n)}(t) = \mathrm{d}^k f(a + t(b-a)) \cdot (b-a)^k$$

[1] Differentiablity of order k implies differentiability of order $k-1$ in a neighbourhood.

For each component of g we apply the single variable Taylor's approximation

$$g_i(1) - \sum_{n=0}^{k} \frac{g_i^{(n)}(0)}{n!} = \frac{g_i^{(k+1)}(\xi_i)}{(k+1)!}$$

or equivalently

$$\begin{aligned}\|R(b)\| &= \left\| f(b) - \sum_{n=0}^{k} \frac{\mathrm{d}^n f(a)\cdot(b-a)^n}{n!} \right\| \\ &= \frac{1}{(k+1)!} \left\| \left[\mathrm{d}^{k+1} f_1(a+\xi_1(b-a))\cdot(b-a)^k \quad \dots \quad \mathrm{d}^{k+1} f(a+\xi_m(b-a))\cdot(b-a)^k \right] \right\|\end{aligned}$$

Theorem 1.36. *Let $f : U \to \mathbb{R}$ and p is an extremum of the function then*

$$\forall h,\ (\mathrm{D}f(p))(h) = 0$$

Proof. For all h define $g_h : \,]-\epsilon, \epsilon[\, \to \mathbb{R}$

$$g_h(t) = f(p+th)$$

then $g_h'(0) = 0$. ■

Theorem 1.37. *Let $f : U \to \mathbb{R}$ be of $\mathcal{C}^2$, p be a critical point of f, and $\mathrm{D}^2 f(p)$ be positive definite. Then, p is a local minimum of f. (If $\mathrm{D}^2 f(p)$ is negative definite then p is local maxima.)*

Assuming the following lemma

Lemma 1.38. *If $\mathrm{D}^2 f$ is continuous and positive definite at point p then it is positive definite in a neighbourhood of p.*

Proof. We wish to prove that there exists a $\delta > 0$ for all unit vectors in V, e, $0 < t < \delta$

$$f(p) \le f(p+te)$$

To do so, define $g_e : \,]0, \delta[\, \to \mathbb{R}$

$$g_e(t) = f(p+te)$$

then by the Taylor's theorem

$$g_e(t) = g(0) + g'(0)\, t + \frac{g''(\xi)}{2!} t^2$$

where $\xi \in \,]0, t[\,$. Equivalently

$$\begin{aligned} f(p+te) &= f(p) + (\mathrm{D}f(p))(e) + \frac{\mathrm{d}^2 f(p+t\xi)\cdot e^2}{2} t^2 \\ &= f(p) + \frac{\mathrm{d}^2 f(p+t\xi)\cdot e^2}{2} t^2 \end{aligned}$$

Using the Lemma 1.38 there exists a neighbourhood of p such that

$$\mathrm{d}^2 f(p+t\xi)\cdot h^2 > 0$$

for all h in the neighbourhood. Therefore,

$$f(p+te) > f(p)$$

which is what we wanted. ■

1.2.4 Smoothness Classes

Let $f \in \mathcal{C}^r$ then one can define the norm

$$\|f\|_r = \max\left\{\sup_{x\in U}\|f(x)\|, \ldots, \sup_{x\in U}\|\mathrm{D}^r f(x)\|\right\}$$

and let the set of all such f with $\|f\|_r < \infty$ be denoted as $\mathcal{C}^r(U, W)$.

Theorem 1.39. *Uniform convergence in $\mathcal{C}^r$ is equivalent to Cauchy.*

Proof.

Theorem 1.40. *$\mathcal{C}^r(U, W)$ under $\|\cdot\|_r$ is a Banach space.*

Definition (Local convergence): A functional sequence f_n is **locally convergent** if for each $x \in U$ there exists a open set $x \in V \subset U$ such that $f_n|_V$ is uniformly convergent.

Theorem 1.41. *Let V, W be normed finite dimensional spaces, $U \subset V$ is open and connected, $x_0 \in U$ and $f_n : U \to W$ is a sequence of differentiable function that*

1. *$f_n(x_0)$ is convergent.*
2. *$\mathrm{D}f_n : U \to \mathcal{L}(V, W)$ is locally convergent to some function $g : U \to \mathcal{L}(V, W)$*

then the sequence f_n is locally convergent to $f : U \to W$ and $\mathrm{D}f = g$. Furthermore, because of connectedness of U for each $x \in U$, $f_n(x)$ is convergent.

Proof. take open ball W around x_0 such that $\mathrm{D}f_n|_W$ is uniformly convergent. then prove the first statement.

$$\|f_m(x) - f_n(x)\| \leq \|(f_m - f_n)(x) - (f_m - f_n)(x_0)\| + \|f_m(x_0) - f_n(x_0)\|$$

apply MVT here and make the bounds smaller using (2). Then prove the differentiability with e/3. To prove (3) use open/close argument. ■

1.2.5 Inverse function theorem

Consider a function f, we wish to find all the solutions to the equation

$$f(x) = y_0$$

To do so, we can define another function F_{y_0} such that

$$F_{y_0}(x) = x - f(x) + y_0$$

then if x is a solution to the equation, it is a fixed point of F_{y_0}.

Theorem 1.42 (Banach fixed point). *Let (X, d) be a complete metric space and $f : X \to X$ is such that for some $0 \leq \lambda < 1$*

$$\forall x, y \in X,\ d(f(x), f(y)) \leq \lambda d(x, y)$$

Then for each $x \in X$ the sequence $\{f^n(x)\}$ is convergent to $p \in X$ such that $f(p) = p$.

Proof. Let $x_n = f^n(x)$ for $n \geq 0$ then

$$d(x_n, x_{n+1}) \leq \lambda d(x_{n-1}, x_n) \leq \lambda^n d(x_0, x_1)$$

thereofore

$$d(x_n, x_m) \leq \sum_{i=n}^{m-1} \lambda^i d(x_0, x_1) \leq d(x_0, x_1) \frac{\lambda^n}{1-\lambda}$$

hence $\{f^n(x)\}$ is Cauchy and it is convergent to a point p. Lastly,

$$\begin{aligned} d(f(p), p) &\leq d(f(p), f(x_n)) + d(f(x_n), p) \\ &\leq \lambda d(p, x_n) + d(x_{n+1}, p) < \epsilon \end{aligned}$$

Theorem 1.43 (Inverse function theorem). *Let V, W be finite dimensional normed vector space such that $\dim V = \dim W$ and $U \subset V$ is open. If $f : U \to W$ is continuously differentiable and for some $a \in U$, $\mathrm{D}f(a)$ is invertible. Then, there are open set $S \subset V$ and $T \subset W$such that $a \in S \subset U$ and $f(a) \in T$ such that $f|_S$ is bijective and $(f|_S)^{-1} = g$ where $g \in \mathcal{C}^1$ and*

$$\mathrm{D}g(f(x)) = (\mathrm{D}f(x))^{-1}$$

Proof. Let S be an open convex set around a such that for all $x \in S$

$$\|\mathrm{D}f(x) - \mathrm{D}f(a)\| < \frac{1}{2} \left\|\mathrm{D}f^{-1}(a)\right\|^{-1}$$

hence $\mathrm{D}f(x)$ is invertible. Let $T = f(S)$ then we shall prove the following

1. $f|_S$ is bijective.

   Let $\psi : S \to V$ with

$$\begin{aligned} \psi_y(x) &= x - (\mathrm{D}f(a))^{-1} (f(x) - y) \\ \implies \mathrm{D}\psi_y(x) &= \mathbb{1}_V - (\mathrm{D}f(a))^{-1} \mathrm{D}f(x) \\ &= (\mathrm{D}f(a))^{-1} \circ (\mathrm{D}f(a) - \mathrm{D}f(x)) \\ \implies \|\mathrm{D}\psi_y(x)\| &\leq \left\|(\mathrm{D}f(a))^{-1}\right\| \|\mathrm{D}f(a) - \mathrm{D}f(x)\| \\ &< \frac{1}{2}[(\mathrm{D}f(a))^{-1}][(\mathrm{D}f(a))^{-1}]^{-1} = \frac{1}{2} \end{aligned}$$

   therefore by mean value theorem

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

   which follows that ψ_y has at most one fixed point because

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

   is a contradiction, and for that fixed point

$$\psi_y(x) = x - (\mathrm{D}f(a))^{-1} (f(x) - y) = x \implies y = f(x)$$

   which means f is injective. By the definition of T, f is surjective as well.

2. T is open.

   We wish to prove that for each $f(x_0) = y_0 \in T$ we wish to prove there exist a $\sigma > 0$ such that $B_\sigma(y_0)$ is contained in T. In other words, $\forall y \in B_\sigma(y_0)$

$$\exists x \in S,\ f(x) = y \iff \psi_y(x) = x$$

   To apply the contraction fixed point we must find complete metric space X such that $\psi_y(X) = X$. Choose ρ as small as needed that $\overline{B_\rho(x_0)} \subset S$, which makes a complete metric space. Let $\sigma = \dfrac{r\rho}{2}$ where $r = \left\|(\mathrm{D}f(a))^{-1}\right\|^{-1}$. Lastly, we show that for each $y \in \overline{B_\sigma(y_0)}$, $\psi_y\Big(\overline{B_\rho(x_0)}\Big) = \overline{B_\rho(x_0)}$. That is, $x \in \overline{B_\rho(x_0)}$ implies that $\psi_y(x) \in \overline{B_\rho(x_0)}$.

$$\begin{aligned}
&\|\psi_y(x) - x_0\| \leq \|\psi_y(x) - \psi_y(x_0)\| + \|\psi_y(x_0) - x_0\| \\
&\leq \frac{1}{2}\|x - x_0\| + \left\|(\mathrm{D}f(a))^{-1}(y - y_0)\right\| \leq \frac{\rho}{2} + \frac{\rho}{2} = \rho
\end{aligned}$$

3. $g = (f|_S)^{-1} : T \to S$ is continuously differentiable. Writting the differentiability criteria

$$\|g(y+h) - g(y) - (\mathrm{D}g(y))(h)\| \leq \epsilon\|h\|$$

   Let $y = f(x)$ and $y + h = f(x+k)$ then $h = f(x+k) - f(x)$ and note

$$\|\psi_y(x+k) - \psi_y(x)\| = \left\|k - \big((\mathrm{D}f(a))^{-1}\big)(h)\right\| \leq \frac{1}{2}\|k\|$$

   which implies

$$\frac{1}{2}\|k\| \leq \left\|\big((\mathrm{D}f(a))^{-1}\big)(h)\right\| \leq \frac{3}{2}\|k\|$$

$$\begin{aligned}
\left\|k - \big((\mathrm{D}f(x))^{-1}\big)(f(x+k) - f(x))\right\| &= \left\|\big((\mathrm{D}f(x))^{-1}\big)((\mathrm{D}f(x))(k) - f(x+k) - f(x))\right\| \\
&\leq \left\|(\mathrm{D}f(x))^{-1}\right\|\|f(x+k) - f(x) - (\mathrm{D}f(x))(k)\| \\
&\leq \left\|(\mathrm{D}f(x))^{-1}\right\|\epsilon\|k\| \\
&\leq 2\left\|(\mathrm{D}f(x))^{-1}\right\|\left\|(\mathrm{D}f(a))^{-1}\right\|\epsilon
\end{aligned}$$

   which proves the differentiability of g as $\left\|(\mathrm{D}f(x))^{-1}\right\|$ is bounded in S. As shown, the inverse operator is i is continuous and therefore if $\mathrm{D}f$ is continuous, then $i(\mathrm{D}f)$ is continuous. In fact, if $f \in C^k$ then $g \in C^k$ as well. ■

1.2.6 Implicit function

Theorem 1.44. *Let V, W be finite dimensional normed vector spaces and $U \subset V \times W$ is open. If $f : U \to W$, $f \in \mathcal{C}^1$ where $f(a,b) = 0$ and $\big(\mathrm{D}f|_{\{a\}\times W}\big)(a,b)$ is invertible then there exist open set S are a and T around b such that $S \times T \subset U$, and a continuously differentiable function $\phi : S \to T$ such that $\phi(a) = b$ and $f^{-1}(0) \cap (S \times T)$ is the graph of ϕ.*

Proof. To apply the inverse function theorem, we need a function whose domain and range have the same dimension. So define, $F : U \to V \times W$

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

Then

$$\mathrm{D}F(a,b) = \left[\begin{array}{c|c} I_n & \mathbb{0}_n \\ \hline \big(\mathrm{D}f|_{\{b\}\times U}\big)(a,b) & \big(\mathrm{D}f|_{\{a\}\times W}\big)(a,b) \end{array}\right]$$

Since I_n and $\big(\mathrm{D}f|_{\{a\}\times W}\big)(a,b)$ are both invertible then $\mathrm{D}F(a,b)$ is invertible as well. By inverse function theorem there are open set Ω_1 around (a,b) and Ω_2 around $(a,0)$ such that $F|_{\Omega_1}$ is $\mathcal{C}^1$ diffeomorphism from Ω_1 to Ω_2. Let $G : \Omega_2 \to \Omega_1$ be the local inverse of F andv S, T be such that $V \times W \subset \Omega_1$. Let $\phi : S \to T$

$$f(x) = (\pi_2 \circ G)(x,0)$$

Corollary 1.45. *$U \subset \mathbb{R}^n$ is open and $f : U \to \mathbb{R}^k$ is continuously differentiable, $n \geq k$, $f(a) = 0$, are the rank of $\mathrm{D}f(a)$ is equal to k. Then, there exists an open set V around a such that $f^{-1}(0) \cap V$ is the graph of $\mathcal{C}^1$ function from a coordinate subspace $n-k$ of $\mathbb{R}^n$ to its complement.*

Proof.

1.2.7 Rank theorem

A generalization of $PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$

Theorem 1.46. *Let $f : U \to W$ be of class $\mathcal{C}^1$ and*

$$\forall x \in U,\ \operatorname{rank} \mathrm{D}f(x) = k$$

then for each $p \in U$ there exist open subsets $p \in U_0$ and $f(p) \in W_0$ and diffeomorphisms

$$\alpha : U_0 \to U_0'$$
$$\beta : V_0 \to V_0'$$

such that

$$\beta \circ f \circ a^{-1}(x_1, \ldots, x_n) = (x_1, \ldots, x_k, 0, \ldots, 0)$$

Proof. Let $A = \mathrm{D}f(p)$ then there are invertible matrices P, Q such that

$$QAP^{-1} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

Exercises

1. Using l'Hopital's rule show that

$$\lim_{t\to 0} \frac{\Delta(t,h,k)}{t^2} = \frac{(\mathrm{d}f(a))(h,k) + (\mathrm{d}f(a))(k,h)}{2}$$