# Advanced Calculus

Georges Khater
American University of Beirut, Math 223

February 25, 2026

## Contents



## 1 Metric Spaces

**Definition 1.1.** Let $X$ be a set, a *metric* $d$ is a map

$$d\colon X \times X \to \mathbb{R}$$

such that:

(i) $d(x,y) \geq 0$ with equality iff $x = y$.

(ii) $d(x,y) = d(y,x)$

(iii) $d(x,z) \leq d(x,y) + d(y,z)$

**Example 1.2.** Given a vector space $V$ and norm $\|.\|$ we can define the metric

$$d(x,y) = \|x - y\|$$

**Question.** Given a vector space $V$ and a metric $d$ on $V$, does $d$ "come from" a norm? i.e does there exist some norm $\|\|$ s.t

$$d(x,y) = \|x - y\|$$

The answer in general is no (consider the discrete metric, or even any bounded metric: this contradicts homogeneity).

**Notation.** Take $(X, d)$ metric space $\delta > 0$ and $x_0 \in X$, the *$\delta$-neighborhood* of $x_0$

$$N_\delta(x_0) := \{x \in X : d(x, x_0) < \delta\}$$

**Definition 1.3.** $(X, d)$ metric space and $E \subset X$

(i) $x_0 \in X$ is an *interior point* of $E$ iff $\exists \delta > 0$ s.t

$$N_\delta(x_0) \subset E$$

(ii) $x_0 \in X$ is a *limit point* iff $\forall \delta > 0$

$$N^*_\delta(x_0) \cap E \neq \varnothing$$

(We denote by $E'$ the set of limit points of $E$)

(iii) $E$ is *open* iff every $x_0 \in E$ is an interior point of $E$.

(iv) $E$ is *closed* iff every $E' \subset E$.

(v) The *closure* of $E$ denoted $\overline{E}$ is

$$\overline{E} := E \cup E'$$

**Review.** Fix $(X, d)$ a metric space.

1. $N_\delta(x_0)$ is open $\forall \delta > 0,\ \forall x_0 \in X$.
2. $E$ is open iff $E^c$ is closed.
3. $\overline{E}$ is closed, in fact $\overline{E}$ is the smallest closed set containing $E$.
4. The union of any collection of open sets is open.
5. A finite intersection of a collection of open sets is open.
6. Intersection of any collection of closed sets is closed.
7. Finite union of closed is closed.
8. If $x \in E'$ then $N_\delta(x) \cap E'$ is an infinite set.

**Definition 1.4.** Let $x_n$ be a sequence in $X$, we say that $x_n$ converges to $x_0 \in X$ iff $\forall \varepsilon > 0\ \exists N \in \mathbb{N}$ s.t

$$d(x_n, x_0) < \varepsilon \quad \forall n \geq N$$

1. The limit of a sequence (if exists) is unique.
2. $x_n$ converges to $x_0$ in $(X, d)$ iff $d(x_n, x_0)$ converges 0 in $(\mathbb{R}, \|\|)$.

**Example 1.5.** $x_n = (1/n + 1, e^{-n})$ in $\mathbb{R}^2$. Find the limit of $x_n$ in $(\mathbb{R}^2, \|.\|_\infty)$, $(\mathbb{R}^2, \|.\|_1)$, $(\mathbb{R}^2, \|.\|_2)$.

- $$\|(1/n + 1, e^{-n}, (1, 0))\|_\infty = \max(1/n, e^{-n}) \leq 1/n + e^{-n} \to 0$$
- $x_0 = (1, 0)$
$$\|(1/n + 1, e^{-n}, -(1, 0))\|_1 = \|(1/n, e^{-n})\|_1 = |1/n| + |e^{-n}| \to 0$$
- $$\|(1/n + 1, e^{-n}) - (1, 0)\|_2 = \sqrt{1/n^2 + e^{-2n}} \to 0$$

**Remark 1.6.** $x_n \to x_0$ in $(\mathbb{R}^k, \|.\|_\infty)$

$$(x_{1,n}, x_{2,n}, \cdots, x_{k,n}) \to (x1, 0, \cdots, x_{k,0})$$

Iff

$$\begin{gathered} x_{1,n} \to x_{1,0} \\ x_{2,n} \to x_{2,0} \\ \vdots \\ x_{k,n} \to x_{k,0} \end{gathered}$$

in $(\mathbb{R}, \|.\|)$.

*Proof.* $|x_{i,n} - x_{i,0} \leq \|x_n - x_0\|_\infty \leq |x_{1,n} - x_{1,0}| + \cdots + |x_{k,n} - x_{k,0}|$ Therefore $x_n \to x_0$ in $(\mathbb{R}^k, \|.\|_\infty)$ $\iff$ $\|x_n - x_0\|_\infty \to 0$ and $\iff$ $|x_{i,n} - x_{i,0}| \to 0\ \forall i$ and hence $\iff$ $x_{i,n} \to x_{i,0}\ \forall i$. $\square$

**Remark 1.7.** Take $x \in \mathbb{R}^n$, then

$$\|x\|_\infty = \max_i(x_i) \leq |x_1| + \cdots + |x_n| = \|x\|_1 \leq n\|x\|_\infty$$

Therefore

$$\|x\|_\infty \leq \|x\|_1 \leq n \cdot \|x\|_\infty$$

Similarly, show something for $\|x\|_2$ and $\|x\|_\infty$.

**Definition 1.8.** Given a vector space $V$ and norms $\|.\|_a$ and $\|.\|_b$ we say that $\|.\|_a$ and $\|.\|_b$ are equivalent iff $\exists c_1, c_2 > 0 \in \mathbb{R}$ s.t $\forall x \in V$

$$c_1\|x\|_b \leq \|.\|_a \leq c_2\|x\|_b$$

**Lemma 1.9.** *$(X, d)$ metric space and $E \subseteq X$, then $E$ is closed iff for sequence $x_n \in E$ that converges $\lim_{n\to\infty} x_n \in E$.*

*Proof.* $\implies$ Assume that there exists a sequence $x_n$ in $E$ that converges to $x_0$ s.t $x_0 \notin E$, take $\varepsilon > 0$, then since $x_n \to x_0$ then $\exists N \in \mathbb{N}$ s.t

$$x_n \in N_\varepsilon(x_0) \quad \forall n \geq N$$

But $x_n \in E$,

$$N^*_\varepsilon(x_0) \cap E \neq \varnothing$$

Therefore

$$x_0 \in E'$$

Contradiction to the fact that $E$ is closed.

- Take $x_0 \in E'$, let $\varepsilon_n = 1/n$ then

$$\exists x_n \in N^*_n(x_0) \cap E$$

Consider $\{x_n\}_{n\in\mathbb{N}}$ $x_n$ is a sequence in $E$ s.t

$$d(x_n, x_0) < 1/n \implies x_n \to x_0$$

and hence $x_0 \in E$.

□

**Theorem 1.10.** *Given a vector space $V$ and norms $\|.\|_a$ and $\|.\|_b$ then the following are all equivalent:*

1. *$\|.\|_a$ and $\|.\|_b$ are equivalent ($c_1\|x\|_b \leq \|x\|_a \leq c_2\|x\|_b$).*
2. *$x_n \to x_0$ in $\|.\|_a \iff x_n \to x_0$ in $\|.\|_b$.*
3. *$E$ is closed in $(V, \|.\|_a) \iff E$ is closed in $(V, \|.\|_b)$.*
4. *$U$ is open in $(V, \|.\|_a) \iff U$ is open in $(V, \|.\|_b)$.*

*Proof.* $1 \implies 2$

$$c_1\|x_n - x_0\|_b \leq \|x_n - x_0\|_a \leq c_2\|x_n - x_0\| \leq \frac{c_2}{c_1}\|x_n - x_0\|_a$$

Apply Squeeze.

$2 \implies 3$ By the Lemma.

$3 \implies 4$ Trivial.

$4 \implies 1$ $B_{\|.\|_a}(0,1)$ open in $(V, \|.\|_a)$ therefore it is open in $(V, \|.\|_b)$. Hence $0$ is an interior point in $(V, \|.\|_b)$. Hence $\exists\delta > 0$

$$B_{\|.\|_b}(0,\delta) \subseteq B_{\|.\|_a}(0,1)$$

Let $x \in V$, $x \neq 0$.

$$\frac{\delta}{2} \cdot \frac{x}{\|x\|_b} \in B_{\|.\|_b}(0,\delta) \subseteq B_{\|.\|_a}(0,1)$$

Therefore

$$\|\frac{\delta}{a}\frac{x}{\|x\|_b}\|_a < 1$$

We get

$$\frac{\delta}{2\|x\|_b}\|x_a\| < 1 \implies \|x\|_a \leq \frac{2}{\delta}\|x\|_b \; \forall x \in V$$

□

## 1.1 Compact Sets

**Definition 1.11.** $(X, d)$ metric space, $K \subseteq X$. We say that $K$ is *compact* iff for every open cover

$$\mathcal{G} = \{G_\alpha\}_{\alpha \in I}$$

of $K$ has a finite subcover.

**Example 1.12.** $(0, 1]$ is not compact in $(R, |.|)$, take

$$\mathcal{G} = \{(1/n, 10)\}_{n \in \mathbb{N}}$$

which has no finite subcover.

**Theorem 1.13** (Heine-Borel). *In* $\mathbb{R}$ $[a, b]$ *is compact.*

**Example 1.14.** If $X$ is finite, then $X$ is compact.

*Proof.* Let $X = \{x_1, \cdots, x_n\}$ and let

$$\mathcal{G} = \{G_\alpha\}_{\alpha \in I}$$

be an open cover $X$. Then for every $x_i$, fix some $G_{\alpha_i}$ s.t

$$x = \{x_1, \cdots, x_n\} \subseteq \bigcup_{i=1}^{n} G_{\alpha_i}$$

□

**Proposition 1.15.** $(X, d)$ *metric space and* $K \subseteq X$ *then every closed subset* $E$ *of* $K$ *is compact.*

*Proof.* Let $\mathcal{G} = \{G_\alpha\}_{\alpha \in I}$ be an open cover of $E$, then

$$E \subseteq \bigcup_{\alpha \in I} G_\alpha$$

Therefore

$$K \subseteq \bigcup G_\alpha \cup E^c, \quad E^c \text{ open}$$

Therefore

$$\mathcal{G}' = \mathcal{G} \cup \{E^c\} \quad \text{is an open cover of } K$$

Therefore it has a finite subcover and

$$K \subseteq G_{\alpha_1} \cup \cdots \cup G_{\alpha_k} \cup E^c$$

Hence

$$E \subseteq G_{\alpha_1} \cup \cdots \cup G_{\alpha_k}$$

Which is a finite subcover of $\mathcal{G}$. □

**Definition 1.16.** $(X, d)$ metric space and $x_n$ sequence in $X$ and

$$\varphi \colon \mathbb{N} \to \mathbb{N} \quad \text{strictly increasing}$$

We say that $b_n = a_{\varphi(n)}$ is a *subsequence* of $a_n$.

**Proposition 1.17.** *$x_n$ converges to $x_0 \iff$ every subsequence $x_{n_k}$ converges to $x_0$.*

**Definition 1.18.** $(X, d)$ metric space and $K \subseteq X$. We say that $K$ is sequentially compact iff every sequence $x_n \in K$ has a convergent subsequence in $K$.

**Proposition 1.19.** *$(X, d)$ metric space, $K$ sequentially compact iff every infinite $E \subseteq K$ has a limit point in $K$.*

*Proof.* $\implies$ Let $K$ be sequentially compact, take $E \subseteq K$ an infinite set. We can extract from $E$ a sequence of distinct elements $x_n \in E \subseteq K$. Therefore $x_n$ has a convergence subsequence $x_{n_k}$ s.t the limit of $x_{n_k} = x_0 \in K$. Moreover, $x_{n_k}$ are distinct in $E$ therefore $x_0 \in E' \cap K$.

$\impliedby$ Let $x_n$ be a sequence in $K$, take

$$E = \{x_n\}_{n \in \mathbb{N}}$$

We have two cases:

- $E$ **is finite.** Then $\exists x_0$ and $\exists n_1 < n_2 < \cdots$ (infinitely many) s.t

$$x_{n_k} = x_0, \quad \forall k \in \mathbb{N}$$

  Therefore

$$\lim_{k \to \infty} x_{n_k} = x_0 \in E$$

  We found a converging subsequence.

- $E$ **is infinite.** Then by the assumption, we have that $E$ has a limit point $x_0$ in $K$.

  $\varepsilon = 1$ $\exists n_1$ s.t $x_{n_1} \in N_1^*(x_0) \cap E$.

  $\varepsilon = 1/2$ $\exists n_2 > n_1$ s.t $x_{n_2} \in N_{1/2}^*(x_0) \cap E$ (since $N_{1/2}^*(x_0) \cap E$ is infinite).

  we can construct $n_k > n_{k-1}$ s.t $x_{n_k} \in N_{1/k}^*(x_0) \cap E$, therefore

$$\lim_{n \to \infty} x_{n_k} = x_0 \in K$$

□

**Example 1.20.** In $(\mathbb{Q}, |.|)$, $K = \mathbb{Q} \cap [0, 1]$ is closed and bounded. $P_n$ increasing sequence in $\mathbb{Q} \cap [0, 1]$ s.t

$$P_n \to 1/\sqrt{2} \in \mathbb{R}$$

Consider

$$\mathcal{G} = \Big\{(-1, P_1) \cap \mathbb{Q}, (-1, P_2) \cap \mathbb{Q}, \cdots, (-1, P_n) \cap \mathbb{Q}, \cdots, (1/\sqrt{2}, 10) \cap \mathbb{Q}\Big\}$$

**Proposition 1.21.** *$K \subseteq X$ is compact $K$ is sequentially compact.*

*Proof.*

- Let $E \subseteq K$ infinite, suppose that $E' \cap K = \varnothing$. Take $x \in K \implies x \notin E'$, therefore $\exists \delta_x > 0$ s.t

$$N_{\delta_x}^*(x) \cap E = \varnothing$$

Let $\mathcal{G} = \{N_{\delta_x}(x)\}_{x \in K}$, is an open cover of $K$ compact, hence it has a finite subcover

$$K \subseteq N_{\delta_{x_1}}(x_1) \cup \cdots \cup N_{\delta_{x_1}}(x_k)$$

But $E \subset K \implies E = (E \cap K)$ hence

$$\begin{aligned} E &\subseteq (N_{\delta_{x_1}}(x_1) \cap E) \cup \cdots \cup (N_{\delta_{x_1}}(x_k) \cap E) \\ &\subseteq \{x_1, \cdots, x_k\} \end{aligned}$$

therefore $E$ is finite, contradiction.

□

## 1.2 Complete sets

**Definition 1.22.** Given a sequence $x_n$, we say that $x_n$ is Cauchy iff for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ s.t

$$d(x_n, x_m) < \varepsilon \quad \text{for } n, m \geq N$$

**Proposition 1.23.** *If $x_n$ converges, then $x_n$ is Cauchy.*

*Proof.* Triangle inequality. □

**Remark 1.24.** The converse is false. In $(\mathbb{Q}, |.|)$, in $(\mathbb{R}, |.|)$ there exists some sequence $x_n \in \mathbb{Q}$ s.t

$$x_n \to \sqrt{2} \quad \text{as } n \to \infty$$

But $x_n \to \sqrt{2}$ implies that $x_n$ is Cauchy in $(\mathbb{R}, |.|)$, therefore $x_n$ is Cauchy in $\mathbb{Q}$. Suppose $x_n$ converges to $P$ in $(\mathbb{Q}, |.|)$, therefore $x_n \to p \in (\mathbb{R}, |.|)$ by uniqueness of limit, $p = \sqrt{2}$, contradiction.

**Definition 1.25.** $(X, d)$ metric space and $K \subseteq X$, we say that $K$ is complete iff every Cauchy Sequence in $K$ converges in $K$.

**Example 1.26.** $(\mathbb{R}, |.|)$ is complete (Math 210)

**Proposition 1.27.** *If $K$ is complete then $K$ is closed.*

*Proof.* Let $x_n$ be a sequence in $K$ s.t $x_n \to x_0 \in X$. Since $x_n$ converges, then $x_n$ is Cauchy and $K$ is complete then $x_n$ converges in $K$, therefore $x_0 \in K$. □

**Proposition 1.28.** *$K$ complete and $E \subseteq K$ is closed, then $E$ is complete.*

*Proof.* Let $x_n$ be a Cauchy sequence in $E$, then $x_n$ is Cauchy in $K$ which is complete, therefore $x_n \to x_0 \in K$. But since $E$ is closed, we get that $x_0 \in E$. □

**Proposition 1.29.** *$(X, d)$ metric space and $x_n$ Cauchy sequence in $K \subseteq X$. If $x_n$ has a convergent subsequence $x_{n_k}$ then $x_n$ converges to the subsequential limit.*

*Proof.* Triangle inequality. □

**Proposition 1.30.** *$(X, d)$ metric space and $K \subseteq X$ is sequentially compact then $K$ is complete (and therefore closed).*

*Proof.* Let $\{x_n\}$ be a Cauchy sequence in $K$, $K$ is sequentially compact therefore $x_n$ has a convergent subsequence $x_{n_k}$ that converges in $K$. Therefore by the previous proposition we get that $x_n$ converges in $K$. □

**Theorem 1.31.** *Let $V$ be a finite dimensional inner product space over $\mathbb{R}$, then $V$ is complete w.r.t norm induced by the inner product. Therefore $K \subseteq V$ is complete $\iff$ $K$ is closed.*

*(For the converse $K \subseteq V$ which is closed, then $K$ is complete).*

*Proof.* Let $S = \{v_1, \cdots, v_l\}$ be an orthonormal basis of $V$, take $x_n$ be a cauchy sequence in $V$. Therefore

$$x_n = r_{1,l}v_1 + \cdots + r_{l,n}v_l$$

Take $r_{1,n} = \langle x_n, v_1 \rangle$ therefore

$$\begin{aligned} |r_{1,m} - r_{1,n}| &= |\langle x_m, v_1 \rangle - \langle x_n, v_1 \rangle| \\ &= |\langle x_m - x_n, v_1 \rangle| \\ &\leq \|x_m - x_n\| \|v_1\| \quad \text{By C.S} \end{aligned}$$

Let $\varepsilon > 0$ then $\exists N \in \mathbb{N}$ s.t

$$\|x_m - x_n\| < \varepsilon \implies r_{1,n} \text{ is Cauchy in } \mathbb{R}$$

THerefore $r_{1,n}$ converges to $r_1 \in \mathbb{R}$. Do the same fore very coordinate, and define

$$x = r_1 v_1 + \cdots + r_n v_n$$

Therefore

$$\begin{aligned} \|x_n - x\| &= \|(r_{1,n} - r_1)v_1 + \cdots + (r_{1,l} - r_l)v_l\| \\ &\leq \|r_{1,n} - r_1\| \|v_1\| + \cdots + \|r_{l,n} - r_l\| \|v_l\| \end{aligned}$$

Therefore

$$x_n \to x$$

□

## 1.3 Total Boundedness

**Definition 1.32.** $(X, d)$ metric space and $K \subseteq X$ we say that $K$ is totally bounded iff for every $\varepsilon > 0$ $\exists x_1, \cdots, x_t \in K$ s.t

$$K \subseteq N_\varepsilon(x_1) \cup \cdots \cup N_\varepsilon(x_t)$$

**Definition 1.33.** $K \subseteq X$ is bounded iff $\exists R > 0$ s.t

$$K \subseteq N_R(x_0) \quad \text{for some } x_0 \in K$$

We say that a sequence is bounded iff

$$E = \{x_n\}_{n \in \mathbb{N}} \quad \text{is bounded}$$

**Proposition 1.34.** *Cauchy sequences are bounded.*

**Proposition 1.35.** *$K \subseteq X$ is totally bounded it is bounded.*

Note that the converse is false.

**Example 1.36.** $(\mathbb{R}, d_{disc})$, $\mathbb{R}$ is bounded since

$$\mathbb{R} \subseteq N_2(0)$$

but it isn't totally bounded, take $\varepsilon = 1/2$ then if it is totally bounded, then $\exists x_1 \cdots x_t \in \mathbb{R}$ s.t

$$\mathbb{R} \subseteq N_{1/2}(x_1) \cup \cdots \cup N_{1/2}(x_t) = \{x_1, \cdots, x_t\}$$

*Proof.* Take $\varepsilon = 1$ then $K \subseteq N_1(x_1) \cup N_1(x_t)$ for $x_1, \cdots, x_t \in K$. Take $x_1$ and let $d = \max_{k \in \{1, \cdots, t\}}(x_1, x_t)$, therefore for any $x \in K$ we have that

$$d(x, x_1) \leq d(x, x_i) + d(x_i, x_1)$$

for some $x_i$ s.t $x \in N_1(x_i)$ (Exists by the above), therefore

$$d(x, x_1) \leq 1 + d \implies x \in N_{1+d}(x_1)$$

□

**Proposition 1.37.** *$K$ is sequentially compact then $K$ is totally bounded.*

*Proof.* Suppose not, then $\exists \varepsilon_0$ s.t $K$ cannot be covered by finitely many $\varepsilon_0$-neighborhoods. Fix some $x_1$ then $\exists x_2 \in K \setminus N_{\varepsilon_0}(x_1)$, Similarly there exists some $x_3 \in K \setminus N_{\varepsilon_0}(x_1) \cup N_{\varepsilon_0}(x_2)$ and in general

$$\exists x_n \in K \setminus (N_{\varepsilon_0}(x_1) \cup \cdots \cup N_{\varepsilon_0}(x_{n-1}))$$

Therefore we have that for $n \neq m$

$$d(x_n, x_m) \geq \varepsilon_0$$

Therefore $x_n$ doesn't have a convergent subsequence; contradiction. □

**Lemma 1.38.** *Every subset of a totally bounded set is totally bounded.*

*Proof.* $K \subseteq X$ totally bounded, and $E \subseteq K$; take $\varepsilon > 0$, therefore $\exists x_1, x_2, \cdots, x_t \in K$ s.t

$$K \subseteq N_{\varepsilon/2}(x_1) \cup \cdots \cup N_{\varepsilon/2}(x_t)$$

Ignore the neighborhood that don't intersect $E$, renumbering we have $x_1, \cdots, x_l$ s.t

$$N_{\varepsilon/2}(x_i) \cap E \neq \varnothing$$

We get

$$E \subseteq N_{\varepsilon/2}(x_1) \cup \cdots \cup N_{\varepsilon/2}(x_l)$$

Take $y_i \in N_{\varepsilon/2}(x_i) \cap E$, then

$$N_{\varepsilon}(y_i) \supseteq N_{\varepsilon/2}(x_i)$$

Therefore $E \subseteq N_{\varepsilon}(y_1) \cup \cdots \cup N_{\varepsilon}(y_l)$ and hence $E$ is totally bounded. □

**Example 1.39.** $[0,1]$ is totally bounded (since it is compact), therefore $(0,1)$ is totally bounded.

**Proposition 1.40.** *$(X,d)$ Metric space, $K \subseteq X$ is complete and totally bounded then $K$ is compact.*

*Proof.* Assume $K$ is not compact, fix $\mathcal{G} = \{G_\alpha\}_{\alpha \in I}$ an open cover of $K$ with no finite subcover. Let $\varepsilon = 1$, $K$ is totally bounded then $K$ can be covered by finitely many 1-balls with center in $K$. Hence $\exists x_1 \in K$ s.t $K_1 = N_1(x_1) \cap K$ cannot be covered by finitely many open sets in $\mathcal{G}$. But $K_1$ is totally bounded (subset of $K$), now let $\varepsilon = 1/2$, $\exists x_2 \in K_1$ s.t $K_2 = N_{1/2}(x_2) \cap K_1$ cannot be covered by finitely many open sets in $\mathcal{G}$. Similarly, we construct a sequence $x_n$ and

$$K_n = N_{1/n}(x_n) \cap K_{n-1} \quad x_n \in K_{n-1}$$

such that $K_n$ cannot be covered by finitely many $G_\alpha$'s. Notice that

$$K_0 = K \supseteq K_1 \supseteq K_2 \supseteq \cdots$$

**Claim.** $x_n$ is Cauchy. Indeed, let $\varepsilon > 0$ then $\exists N$ s.t

$$1/N < \varepsilon/2$$

Now take $m, n > N$, then

$$x_n \in K_{n-1} \subseteq K_N, \quad x_m \in K_{m-1} \subseteq K_N$$

therefore $x_n, x_m \in K_N = N_{1/N}(x_N) \cap K_{n-1}$, and hence

$$d(x_n, x_m) \leq 2/N < \varepsilon$$

By completeness of $K$, we get that $x_n$ converges to some $x_0 \in K$; since $\mathcal{G} = \{G_\alpha\}$ is an open cover of $K$, there exists $\alpha_0 \in I$ s.t $x_0 \in G_{\alpha_0}$ (open set). So $\exists \varepsilon_0$ s.t

$$N_{\varepsilon_0}(x_0) \subseteq G_{\alpha_0}$$

But $x_n \to x_0$ then $\exists N_1$ s.t

$$x_N \in N_{\varepsilon_0/2}(x_0) \text{ and } 1/N < \varepsilon_0/2$$

Therefore

$$K_N \subseteq N_{1/N}(x_n) \subseteq N_{\varepsilon_0}(x_0) \subseteq G_{\alpha_0}$$

contradiction. □

**Theorem 1.41.** *$V$ finite dimensional product normed space (over $\mathbb{R}$). Let $S = \{v_1, \cdots, v_n\}$ a basis of $V$. Then the cube*

$$Q_S = \{r_1 v_1 + r_2 v_2 + \cdots + r_n v_n : 0 \leq r_i \leq 1\}$$

*is compact.*

*Proof.* Let $x^k = r_{1,k} v_1 + r_{2,k} v_2 + \cdots + r_{n,k} v_n$ a sequence in $Q_S$, now clearly

$$r_{1,k} \text{ is a sequence in } [0,1]\text{, compact set}$$

Therefore $r_{1,k}$ has a convergent subsequence $r_{1,k_n}$. Similarly $r_{2,k_l} \in [0,1]$ therefore $r_{2,k_l}$ has a convergent subsequence $r_{2,k_{l_t}}$. Doing this $n$ times we get that $\exists$ a subsequence $\{k_j\}$ s.t

$$r_{1,k_j},\ r_{2,k_j}, \cdots,\ r_{n,k_j} \to r_1, r_2, \cdots, r_n$$

take

$$x_0 = r_1 v_1 + \cdots + r_n v_n$$

We can show that

$$\|x_0 - x^{k_j}\| \leq |r_{1,k_j} - r_1| \|v_1\| + \cdots + |r_{n,k_j} - r_n| \|v_n\| \to 0$$

then

$$x^{k_j} \to 0$$

Hence $Q_S$ is (sequentially) compact. □

**Theorem 1.42.** *Let $V$ be a finite inner product space then*

$$K \subseteq V \text{ is compact} \iff K \text{ closed and bounded}$$

*Proof.*
- $K$ compact $\implies$ complete $\implies$ closed and $\implies$ totally bounded $\implies$ bounded.
- $K$ is bounded, therefore $\exists R > 0$ s.t $K \subseteq B_{\|.\|}(0, R)$. Let $S = \{v_1, \cdots, v_n\}$ be a basis of $V$ and Take $x \in K$ therefore

  $$x = r_1 v_1 + \cdots + r_n v_n$$

  then $r_i = \|\langle x, v_i \rangle\| \leq \|x\| \cdot \|v_i\|$

  $$B_{\|.\|}(0, R) \subseteq \{r_1 v_1 + \cdots + r_n v_n : r_i \leq R\}$$

  Therefore $K \subseteq$ some cube and hence $K$ is compact.

□

**Proposition 1.43.** *In a finite dimensional inner product space we get that bounded $\iff$ totally bounded.*

*Proof.* $\implies$ Bounded $\implies$ can be put in a cube (compact) $\implies$ Totally bounded.
$\impliedby$ True in general. ∎

## 1.4 Continuity

**Definition 1.44.** $X, Y$ metric spaces and $E \subseteq X$ and $f \colon E \to Y$ map; $x_0 \in E'$ we say that

$$\lim_{x \to x_0} f(x) = y \in Y$$

iff for every $\varepsilon > 0$ $\exists \delta > 0$ s.t

$$f(x) \in N_\varepsilon(y_0) \text{ for } x \in N^*_\delta(x_0) \cap E$$

**Definition 1.45.** $f \colon E \to Y$ continuous at $x_0$ iff for every $\varepsilon > 0$, $\exists \delta > 0$ s.t $f(x) \in N_\varepsilon(x_0)$ for $x \in N_\delta(x_0) \cap E$; i.e

$$d_Y(f(x), f(x_0)) < \varepsilon \quad \text{for } 0 < d_X(x, x_0) < \delta$$

**Remark 1.46.** $f$ is continuous at $x_0 \iff x_0$ is an isolated point or $x_0 \in E'$ and $\lim x \to x_0 f(x) = f(x_0)$.

**Proposition 1.47.** $\lim_{x \to x_0} f(x) = y_0$ *iff for every sequence* $x_n \to x_0$ *we have that* $f(x_n) \to y_0$. *Therefore $f$ is continuous at $x_0$ iff for every $x_n \to x_0$ we have that $f(x_n) \to f(x_0)$.*

**Corollary 1.48.** *if $f, g$ are continuous then $f + g,\ cf,\ f \cdot g,\ f/g$ are all continuous at $x_0$.*

**Example 1.49.** $f(x, y) = \frac{xy}{x^2+y^2}$ we want to find

$$\lim_{(x,y) \to (0,0)} f(x, y)$$

Take the sequence $(x_n, y_n) = (1/n, 0)$ then

$$\lim_{n \to \infty} (x_n, y_n) = (0, 0)$$

and

$$f(x_n, y_n) = 0 \to 0 \text{ as } n \to \infty$$

Now take $(x_n, y_n) = (1/n, 1/n)$, then

$$\lim_{n \to \infty} (x_n, y_n) = (0, 0)$$

but

$$f(x_n, x_n) = \frac{1/n \cdot 1/n}{1/n^2 + 1/n^2} = 1/2 \to 1/2 \text{ as } n \to \infty$$

Therefore

$$\lim_{n \to \infty} f(x, y) \text{ does not exist}$$

Now we do the same for

$$f(x,y) = \frac{x^2y}{x^2+y^2}$$

Notice that

$$0 \le |f(x,y)| = |\frac{x^2}{x^2+y^2}y| \le |y|$$

but

$$\lim_{(x,y)\to(0,0)} |y| = 0$$

therefore by squeeze we get that

$$\lim_{(x,y)\to 0} |f(x,y)| = 0$$

Hence

$$\lim_{(x,y)\to 0} f(x,y) = 0$$

**Some useful inequalities**

$$x^2+y^2 \ge x^2$$
$$x^2+y^2 \ge y^2$$
$$x^2+y^2 \ge 2(xy)$$

**Example 1.50.** Let $(V, \|.\|)$ a normed space and

$$\varphi\colon V \to \mathbb{R}, \quad \varphi(x) = \|x\|$$

Then $\varphi$ is continuous.

Let $x_n$ a sequence in $V$ that converges to $x_0$ in $(V, \|.\|)$ therefore

$$\|x_n - x_0\| \to 0$$

We want to show that $\varphi(x_n) \to \varphi(x_0)$ i.e that

$$|\|x_n\| - \|x_0\|| = 0$$

But as shown in PS1, we have that

$$|\|x_n\| - \|x_0\|| \le \|x_n - x_0\|$$

**Proposition 1.51.** *$f\colon X \to Y$ is continuos iff $f^{-1}(G)$ open in $X$ for every $G$ open in $Y$ iff $f^{-1}(H)$ is closed in $X$ for every $H$ closed in $Y$.*

*Proof.* $\Longrightarrow$ let $x_n$ be a sequence in $f^{-1}(H)$ that converges to $x_0$ in $X$. Hence $f(x_n)$ is a sequence in $H$ by continuity of $f$ we have that

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

Moreover since $H$ is closed, we get that $f(x_0) \in H$. Hence

$$x_0 \in f^{-1}(H)$$

Therefore if $f$ is continuous $f^{-1}(G)$ is open in $X$ for every $G$ open in $Y$.

$\Longleftarrow$ Now we show that if $f^{-1}(G)$ is open in $X$ for all $G$ open in $Y$ then $f$ is continuous. Fix $x_0 \in X$, let $\varepsilon > 0$ then consider the open set in $Y$ $N_\varepsilon(f(x_0))$ hence $f^{-1}\left(N_\varepsilon(f(x_0))\right)$ is open in $X$. But $x_0 \in f^{-1}\left(N_\varepsilon(f(x_0))\right)$ hence $\exists \delta > 0$ s.t

$$f\left(N_\delta(x_0)\right) \subseteq N_\varepsilon(f(x_0))$$

□

**Example 1.52.** $E = \{x, y \in \mathbb{R}^2 \colon y \geq x\}$ is open in $\mathbb{R}^2$, define the continuous function

$$f \colon \mathbb{R}^2 \to \mathbb{R}, f(x, y) = y - x$$

Then

$$E = f^{-1}\left((0, \infty)\right)$$

which is an open set.

**Example 1.53.** $f \colon X \to Y$ continuous and consider the $c$-level set with $c \in Y$

$$E = \{x \in X \colon f(x) = c\} = f^{-1}\left(\{c\}\right)$$

is closed in $X$.

Therefore $E = \left\{(x, y) \colon x^2 + 3y^2 = 1\right\}$ is closed in $\mathbb{R}^2$, $S^n$ is closed.

**Remark 1.54.** $V, W$ normed spaces and

$$\|.\|_a, \ \|.\|_b \text{ equivalent norms on } V$$

so $f \colon (V, \|.\|_a) \to W$ is continuous iff $f \colon (V, \|.\|_b) \to W$ is continuous.

Equivalently: $V, W$ normed spaces and

$$\|.\|_a, \ \|.\|_b \text{ equivalent norms on } W$$

so $f \colon V \to (W, \|.\|_a)$ is continuous iff $f \colon V \to (W, \|.\|_b)$ is continuous.

**Example 1.55.** Take $f \colon \mathbb{R}^m \to \mathbb{R}^n$ defined as follows

$$f(x) = (f_1(x), \cdots, f_n(x)) \quad x = (x_1, \cdots, x_m)$$

Then $f$ is continuous iff $f_i \colon \mathbb{R}^m \to R^n$ is continuous for $1 \leq i \leq n$.

**Theorem 1.56** (Extreme Value Theorem)**.** *$X, Y$ metric spaces and $K \subseteq X$ compact and $f \colon K \to Y$ continuous. Then $f(K)$ is compact.*

**Corollary 1.57.** *If $Y = \mathbb{R}$, then the above implies that $f$ attains its maximum and minimum in $K$.*

*Proof.* Let $y_n$ be a sequence in $f(K)$, therefore $\exists x_n \in K$ s.t $f(x_n) = y_n$. Since $K$ is compact, $x_n$ has a convergent subsequence $x_{n_k}$ that converges to $x_0 \in K$, using continuity we get that

$$y_{n_k} \to f(x_0) \in f(K)$$

□

**Theorem 1.58.** *All norms in a finite dimensional vector space are equivalent.*

*Proof.* $V$ is a finite dimensional vector space, then by PS1 we can define an inner product on $V$. Let $\|.\|$ be the norm induced by our choice of inner product. Let $S = \{v_1, \cdots, v_n\}$ an orthonormal basis of $V$. Let $\|.\|_a$ be an arbitrary norm on $V$, we will show that $\|.\|$ and $\|.\|_a$ are equivalent. Let $x \in V$, then

$$x = r_1 v_1 + \cdots + r_n v_n, \quad |r_i| = \langle x, v_i \rangle \leq \|x\|$$

Therefore $\forall x \in V$

$$\begin{aligned}
\|x\|_a &= \|r_1 v_1 + \cdots + r_n v_n\| \\
&\leq |r_1| \|v_1\|_a + \cdots + |r_n| \|v_n\|_a \\
&\leq (\|v_1\|_a + \cdots + \|v_n\|_a) \cdot \|x\| \\
&\leq c_1 \cdot \|x\|
\end{aligned}$$

Now define the function $\varphi \colon (V, \|.\|) \to \mathbb{R}$ s.t

$$\varphi(x) = \|x\|_a$$

We claim that $\varphi$ is continuous, indeed let $x_n \in V$ s.t $x_n \to x_0$ in $\|.\|$. Therefore

$$\begin{aligned}
|\varphi(x_n) - \varphi(x_0)| &= |\|x_n\|_a - \|x_0\|_a| \\
&\leq \|x_n - x_0\| \\
&\leq c_1 \|x_n - x_0\| \to 0
\end{aligned}$$

Therefore $\lim_{n\to\infty} \varphi(x_n) = \varphi(x_0)$, so $\varphi$ is continuous. Let $K = \{x \in V \colon \|x\| = 1\}$ closed and bounded in the inner product space $(V, \|.\|)$. Then it is compact in $(V, \|.\|)$ therefore $\varphi(K)$ is compact in $\mathbb{R}$ so it attains its minimum in $K$. Notice that $\varphi(x) = 0 \iff x = 0$, hence $\forall x \in K, \ \varphi(x) > 0$. Then

$$\min_{x \in K} \varphi(x) > 0$$

Let $x \in V$ and $x \neq 0$ then $\frac{x}{\|x\|} \in K$, therefore

$$\varphi(\frac{x}{|x|}) = \|\frac{x}{\|x\|}\|_a \geq \min \varphi(K)$$

Therefore

$$\|x\|_a \geq \min(\varphi(K))\|x\|$$

□

**Corollary 1.59.** *If $V$ is finite dimensional then compact iff closed and bounded (independently of the choice of norm).*

## 1.5 Connected sets

**Definition 1.60.** $(X, d)$ metric space and $\Omega \subseteq X$, we say that $\Omega$ is disconnected iff $\exists U, V \subseteq X$ open sets s.t

$$\Omega = (U \cup V) \cap \Omega$$

With $U \cap \Omega,\ V \cap \Omega \neq \varnothing$ and $U \cap \Omega \cap V = \varnothing$.

Otherwise $\Omega$ is connected.

**Example 1.61.** Let $\Omega = (1, 2] \cup (3, 4)$ disconnected, since

$$\Omega \subseteq (0, 2.5) \cup (2.5, 17)$$

$\mathbb{Q}$ is disconnected in $\mathbb{R}$ since

$$\mathbb{Q} \subseteq (-\infty, \sqrt{2}) \cup (\sqrt{2}, \infty)$$

**Proposition 1.62.** *$X$ is a connected metric space iff the only sets that are both open and closed are $\varnothing, X$.*

*Proof.* $\Longrightarrow$ Assume that $U$ is open and closed an $U \notin \{\varnothing, X\}$ then $U^c$ is also not open and closed and not in $\{\varnothing, X\}$ but

$$X = U \cup U^c$$

Therefore $X$ is disconnected.

$\Longleftarrow$ Assume that $X$ is disconnected, then $X = U \cup V$ with $U, V$ open disjoint and nonempty. Therefore $V = U^c$ and $U$ is open then $V$ is closed. Hence $V$ is a clopen set of $X$ but $V \notin \{\varnothing, X\}$ contradiction.

□

**Definition 1.63.** $I \subseteq \mathbb{R}$ is an interval for $x, y \in I$ s.t $x < y$ then $z \in I,\ \forall x < z < y$.

**Theorem 1.64.** *$\Omega \subseteq \mathbb{R}$ is connected iff $\Omega$ is an interval.*

*Proof.* $\Longrightarrow$ Assume that $\Omega$ is not an interval, therefore $\exists x, y \in \Omega$ s.t $x < z < y$ and $z \notin \Omega$, therefore

$$\Omega \subseteq (-\infty, z) \cup (z, \infty)$$

Therefore $\Omega$ is disconnected.

$\Longleftarrow$ Let $\Omega$ be an interval. Suppose $\Omega$ is not connected, then $\exists U, V$ open and disjoint in $\Omega$ in $\mathbb{R}$ s.t

$$\Omega \subseteq U \cup V$$

Then $\exists x \in U \cap \Omega$ and $y \in V \cap \Omega$ and assume wlog that $x < y$; therefore $[x, y] \subseteq \Omega$ consider

$$z = \sup \{[x, y] \cup U\}$$

Note that $z \in [x, y]$ therefore $z \in \Omega$ therefore either $z \in U$ or $z \in V$. Notice that $z > x$ since $x \in U$ an open set. Also notice that $z < y$ since if $z = y \in V$ we have that $\exists \varepsilon_1$ s.t

$$(y - \varepsilon, y) = (z - \varepsilon, z) \subseteq [x, y] \cap V$$

since $z$ is the sup we get that

$$(z - \varepsilon, z) \cap [x, y] \cap U \neq \varnothing$$

contradiction.
Therefore $z \in (x, y)$, suppose that $z \in U$ then $(z, z + \varepsilon_0) \subseteq (x, y) \cap U$ contradiction because $z + \varepsilon_0/2 > z = \sup[x, y] \cap y$. Suppose that $z \in V$ then $\exists \varepsilon_1$ s.t $(z - \varepsilon_1, z) \subseteq (x, y) \cap V$. Since $z$ is the sup then $\exists z^*$ an element s.t

$$z - \varepsilon_1 < z^* < z \text{ and } z^* \in U$$

$z^* \in U \cap V$ contradiction.

□

**Theorem 1.65** (Intermediate Value Theorem). *$f\colon X \to Y$ continuous, if $X$ is connected then $f(X)$ is connected. In particular if $Y = \mathbb{R}$ we have the following:*

- *If $y_0 < y_1 \in f(X)$ then $\forall \alpha$ s.t $y_0 < \alpha < y_1$ $f^{-1}(\{c\}) \neq \varnothing$.*

*Proof.* Suppose that $f(X)$ is disconnected, then $\exists U, V \subseteq Y$ open sets that separate $f(X)$. Therefore

$$\begin{aligned} X &= f^{-1}(f(X)) \\ &= f^{-1}(U \cap V) \cap f(X) \\ &= \left[f^{-1}(U) \cup f^{-1}(V)\right] \cap X \end{aligned}$$

But $f$ is continuous therefore $f^{-1}(U),\ f^{-1}(V)$ are open and disjoint. Now $U \cap f(X) \neq \varnothing$ therefore $f^{-1}(U) \neq \varnothing$, same for $V$; then $X$ is disconnected. □

**Definition 1.66.** $X$ is path connected iff for every $a, b \in X$ there exists a continuous function

$$\alpha\colon [0, 1] \to X$$

s.t $\alpha(0) = a,\ \alpha(1) = b$.

**Remark 1.67.** Let $X$ be a vector space, the case where for every $a, b \in X$ the line segment $(1 - t)a + tb \in X\ \forall t \in [0, 1]$ we say that $X$ is convex.

**Example 1.68.** $B_R(x_0)$ is always convex and hence path connected in a vector space. Indeed, take $a, b \in B_R(x_0)$ consider

$$\begin{aligned}\|(1-t)a + tb - x_0\| &< (1-t)\|a - x_0\| + t\|b - x_0\| \\ &< (1-t)R + R = R\end{aligned}$$

**Proposition 1.69.** *$(X, d)$ metric space then $X$ is path connected $\implies$ $X$ is connected.*

*Proof.* Let $X$ be disconnected then we have $U, V$ nonempty disjoint and open

$$X = U \cup V$$

Then $\exists a \in U$ and $b \in V$, since $X$ is path connected $\exists \alpha \colon [0,1] \to X$ s.t

$$\alpha(0) = a, \quad \alpha(1) = b$$

$[0,1]$ is connected, therefore $\alpha([0,1])$ is also connected but

$$[0,1] = \alpha^{-1}(U) \cup \alpha^{-1}(V)$$

But $\alpha$ is continuous, therefore $\alpha^{-1}(U) \ni 0$ and $\alpha^{-1}(V) \ni V$ which is a contradiction since $[0,1]$ is connected. □

**Theorem 1.70.** *$(V, \|.\|)$ vector space and $\Omega \subseteq V$ open then $\Omega$ is connected iff $\Omega$ is path connected.*

*Proof.* $\impliedby$ True in general.

$\implies$ We will show that $A$ is closed by showing that $A$ is open. Take $z \in A^c$ so there is no path joining $z$ to $x_0$. But $\Omega$ is open, so $\exists \delta > 0$ s.t

$$B_\delta(z) \subseteq \Omega$$

Now every point in $B_\delta(z)$ cannot be joined to $x_0$; Hence

$$B_\delta(z) \subseteq A^c$$

so $A^c$ is open.
Therefore $A$ open, closed and nonempty in $\Omega$ a connected space, hence $A = \Omega$.

□

**Example 1.71** (Topological sine)**.** $\Omega = \{(x, \sin(1/x)) \mid \ x > 0\} \cup \{(0,0)\}$; let

$$A = \{(x, \sin 1/x) \colon x > 0\}$$

then $A$ is connected (image of a continuous function), with

$$\overline{A} = A \cup (\{0\} \times [-1,1])$$

Therefore $A \subseteq \Omega \subseteq \overline{A}$.

Suppose that $\Omega$ was connected, take $x_0 > 0$, suppose that $\exists \alpha\colon [0,1] \to \Omega$ from $(0,0)$ to $(x_0, \sin 1/x_0)$. let $t^* = \sup\{t \in [0,1] \mid x(t)\} = 0$; therefore

$$x(t^*) = 0$$

with $x(t^*) < 1$, hence

$$\alpha(t^*) = (0,0)$$

Take $\varepsilon = 1$, then by the IVT $\exists t^* < t_1 < t^* + 1$ and $t_1 \in [0,1]$ s.t

$$x(t_1) = \frac{1}{(4n_1+1)\pi/2}$$

Take $\varepsilon = 1/2$, then $\exists t_2 \in [0,1]$ s.t $t^* < t_2 < t^* + 1/2$ s.t

$$x(t_2) = \frac{1}{(4n_2+1)\pi/2}$$

We take a sequence $t_k \in [0,1]$ s.t $t^* < t_k < t^* + 1/k$

$$x(t_k) = \frac{1}{(4n_k+1)\pi/2}$$

hence

$$\alpha(t_k) = (\frac{1}{(4n_k+1)\pi/2}, 1)$$

With $t_k \to t^*$, therefore

$$\alpha(t^*) = (0,1)$$

contradicting the continuity of $\alpha$.

## 2 Differentiability

Review linear transformations.

**Proposition 2.1.** *Let $V$ and $W$ normed spaces $T\colon V \to W$ linear transformation, then the following are all equivalent:*

- *$T$ is continuous on $V$.*
- *$T$ is continuous on $0_V$.*
- *$\exists M > 0$ s.t*

$$\|Tx\|_W \le M\|x\|_V$$

*Proof.* 1 $\implies$ 2 Trivial.

2 $\implies$ 3 $T$ is continuous at $0$, hence for $\varepsilon = 1$ $\exists \delta > 0$ s.t

$$\|Tx - T0\|_W < 1 \quad \forall \|x - 0\|_V < \delta$$

Therefore

$$\|Tx\|_W < 1 \quad \forall \|x\|_V < \delta$$

Take any $x \neq 0$ then

$$\|\frac{x}{\|x\|_V} \cdot \delta/2\|_V = \delta/2$$

Therefore

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

therefore

$$\|Tx\|_W < \frac{2}{\delta}\|x\|$$

$3 \implies 1$

$$\|Tx - Ty\|_W = \|T(x-y)\|_W \leq M\|x-y\|_V$$

□

**Example 2.2** (The derivative as a linear transformation). Take $C^1([0,1])$ the space of continuously differentiable functions, with the sup norm

$$\|f\|_\infty = \sup |f(x)| \colon x \in [0,1]$$

and take $C([0,1])$ to be the space of continuous functions with the sup norm. Take

$$\frac{d}{dx} \colon C^1([0,1]) \to C([0,1])$$

Note that this is a linear transformation.

**Is it continuous?** If it was then $\exists M$ s.t

$$\|f\|_\infty \leq M\|f\|_\infty \quad \forall f$$

take $f_n(x) = x^n$, then

$$\|f_n\|_\infty = 1$$

but $f_n'(x) = nx^{n-1}$ with

$$\|f_n'\|_\infty = n$$

Contradiction to the Archimedean property.

**Theorem 2.3.** *Let $V$ be finite dimensional, and $T \colon V \to W$ linear transformation then $T$ is continuous.*

*Proof.* Let $\alpha = \{v_1, \cdots, v_n\}$ be a basis of $V$. Let $x \in V$, then $\exists r_1, \cdots, r_n$ s.t

$$x = r_1 v_1 + \cdots + r_n v_n$$

Since all norms are equivalent in a finite dimensional vector space, wlog work with the inner product norm and make $\alpha$ an orthonormal basis.

$$\begin{aligned}\|Tx\|_W &= \|r_1 Tv_1 + \cdots + r_n Tv_n\|_W \\ &\leq |r_1|\|Tv_1\|_W + \cdots + |r_n|\|Tv_n\|\end{aligned}$$

With

$$|r_i| = |\langle x, v_i \rangle| \leq \|x\|$$

Therefore

$$\|Tx\|_W \leq \|x\|_V(\|Tv_1\| + \cdots_{\|} TV_n\|) = M\|x\|$$

□

**Motivation.** In one dimension, $f\colon \mathbb{R} \to \mathbb{R}$ and $x_0 \in \mathbb{R}$

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

if it exists we say that $f$ is differentiable at $x_0$ and we define

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

In other words, we can say that

$$\begin{aligned}
L = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} &\iff \lim_{x \to x_0} \left( \frac{f(x) - f(x_0)}{x - x_0} - L \right) = 0 \\
&\iff \lim_{x \to x_0} \left( \frac{f(x) - f(x_0) - L(x - x_0)}{x - x_0} \right) = 0 \\
&\iff \exists L \in \mathbb{R} \text{ s.t } f(x) = f(x_0) + L(x - x_0) + o(|x - x_0|) \text{ with } \frac{o(t)}{|t|} \to 0 \\
&\iff \exists L \in \mathbb{R} \text{ s.t } f(x) = f(x_0) + L(x - x_0) + o(|x - x_0|) \\
&\quad \text{with } \frac{o(|x - x_0|)}{|x - x_0|} \to 0 \text{ as } x \to x_0
\end{aligned}$$

**Definition 2.4.** $V, W$ normed spaces and $U \subseteq V$ open set with $x_0 \in U$ we say that $f$ is differentiable at $x_0$ iff there exists a continuous linear transformation $T \in \mathcal{L}(V, W)$ s.t

$$f(x) = f(x_0) + T(x - x_0) + o(\|x - x_0\|_V)$$

or

$$f(x_0 + h) = f(x_0) + T(h) + o(h)$$

with $\frac{o(\|h\|)}{\|h\|_V} \to 0$ as $h \to 0$.

Iff $\exists T \in \mathcal{L}(V, W)$ s.t

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

**Remark 2.5.** If such a $T$ exists, then it is unique.

*Proof.* Suppose that $\exists T_1, T_2$ continuous linear transformations in $\mathcal{L}(V, W)$ s.t

$$f(x_0 + h) = f(x_0) + T_1(h) + o_1(h), \quad f(x_0 + h) = f(x_0) + T_2(h) + o_2(h)$$

Therefore

$$(T_1 - T_2)(h) = (o_1 - o_2)(h) = o_3(h)$$

take $v$ s.t $\|v\| = 1$, therefore

$$\begin{aligned}(T_2 - T_1)(\lambda v) &= \lambda o_3(\|\lambda v\|) \\ \lambda(T_2 - T_1)(v) = &= \lambda o_3(|\lambda|) \\ (T_2 - T_1)(v) &= \frac{o_3(|\lambda|)}{\lambda}\end{aligned}$$

Therefore as $\lambda \to 0$ we get that

$$(T_2 - T_1)(v) = 0 \quad \forall v \text{ s.t } \|v\| = 1$$

Therefore $T_2 = T_1$ $\square$

**Definition 2.6.** If $f$ is differentiable at $x_0$, then we let $Df(x_0)$ to be the linear transformation in $\mathcal{L}(V, W)$ s.t

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + o(\|x - x_0\|) \iff \lim_{x \to x_0} \frac{\|f(x) - f(x_0) - Df(x - x_0)\|_W}{\|x - x_0\|_V}$$

**Notation.** $\mathbb{R}^n = \{x = (x_1, \cdots, x_n) \colon x_i \in \mathbb{R}\}$ as rows vectors, observe $f \colon \mathbb{R}^{1\times n} \to \mathbb{R}^{m\times 1}$ differentiable at $x_0$ then we have that

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + o(|x - x_0|)$$

Then

$$Df(x_0) \in \mathcal{L}(\mathbb{R}^m, R^n)$$

Then $Df(x_0)$ can be represented by a matrix called the *the Jacobian matrix* $[Df(x_0)]$ (an $m \times n$ matrix). Therefore

$$Df(x_0)(x - x_0) = [Df(x_0)](x - x_0)^t$$

Note that we take a row vector and output a column vector.

**Example 2.7.** In 1 variable we have that

$$[Df(x_0)] = [f'(x_0)]$$

Then

$$Df(x_0) \colon t \to [f'(x_0)] \cdot t = f'(x_0) \cdot t$$

Find the derivative

$$f(x, y) = x^2 + y^2$$

we get that

$$\begin{aligned} f((x,y)+(h_1,h_2)) &= f(x_0+h_1, y_0+h_2) \\ &= (x_0+h_1)^2+(y_0+h_2)^2 \\ &= x_0^2+y_0^2+2y_0h_1+2y_0h2+h_1^2+h_2^2 \\ &= f(x_0,y_0)+(2x_0,2y_0)(h_1,h_2)+o(|h|) \end{aligned}$$

since $\frac{h_1^2+h_2^2}{\sqrt{h_1^2+h_2^2}} \to 0$ so we get

$$Df(x_0,y_0)\colon (v_1,v_2) \to (2x_0,2y_0)\cdot(v_1,v_2)$$

therefore we get

$$[Df(x_0)] = \begin{bmatrix} 2x_0 & 2y_0 \end{bmatrix}$$

**Example 2.8.** Take $T\colon V \to W$ be a continuous linear transformation we want to find $DT(x)$; but

$$T(x_0+h) = T(x_0)+T(h)+0$$

therefore

$$DT(x_0) = T$$

**Proposition 2.9.** *$f\colon V \to W$ if $f$ is differentiable at $x_0$ therefore $f$ is continuous at $x_0$.*

*Proof.* Let $f(x) = f(x_0)+Df(x_0)(h)+o(\|x-x_0\|)$ but notice that

$$o(\|x-x_0\|) \to 0$$

but $Df(x_0)$ is continuous hence $\exists M$ s.t

$$\|Df(x_0)(x-x_0)\| \le M\|x-x_0\| \to 0 \implies \lim_{x\to x_0} f(x) = f(x_0)$$

therefore $f$ is continuous. □

## 2.1 Partial derivatives

**Definition 2.10.** $U \subseteq \mathbb{R}^n$ open and $f\colon U \to \mathbb{R}$, then

$$\frac{\partial f}{\partial x_i}(x) = \lim_{\substack{h\to 0\\ h\in\mathbb{R}}} \frac{f(x+he_i)-f(x)}{h}$$

We call this the partial derivative w.r.t $x_i$.

**Example 2.11.**

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

We have shown that this is not continuous at $(0,0)$ hence $f$ is not differentiable at $(0,0)$. But notice that

$$\frac{f(h,0)-f(0,0)}{h} = 0 \to 0 \text{ as } h \to 0$$

therefore $\frac{\partial f}{\partial x}(0,0)$ exists and equal to $0$. Similarly we get that $\frac{\partial f}{\partial y}(0,0)$ exists and equal to $0$.

What if $f\colon U \to \mathbb{R}$ is differentiable at $x_0$ then $Df(x_0) \in \mathcal{L}(R^n, \mathbb{R})$ hence $[Df(x_0)]$ is a $1 \times n$ matrix.

$$f(x_0 + he_1) = f(x_0) + Df(x_0)(he_1) + o(|h|)$$

therefore

$$\frac{f(x_0+he_1)-f(x_0)}{h} = Df(x_0)e_1 + \frac{o(h)}{h} \to Df(x_0)e_1$$

therefore $\frac{\partial f}{\partial x_1}$ exists and is the first entry in $[Df(x_0)]$. Similarly we get $\frac{\partial f}{\partial x_i}$ exists and is the $i$'th entry in $[Df(x_0)]$. We get that

$$[Df(x_0)] = (\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n})$$

this is called the gradient of $f$ denoted $\nabla f$.

If $f$ is differentiable then

$$\begin{aligned} f(x) &= f(x_0) + \nabla f(x_0)(x_0 - x)^t + o(|x - x_0|) \\ &= f(x_0) + \nabla f(x_0) \cdot (x - x_0) + o(|x - x_0|) \end{aligned}$$

**Theorem 2.12.** *$U \subseteq \mathbb{R}^n$ open, and $f\colon U \to \mathbb{R}$ is differentiable iff all partial derivatives exist and*

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

*Proof.* $\implies$ Suppose that $f$ is differentiable, then

$$f(x_0 + h) = f(x_0) + Df(x_0)h + o(h)$$

Therefore

$$f(x_0 + he_i) = f(x_0) + Df(x_0)he_i + o(h)$$

Hence

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

Therefore all partial derivatives exist with

$$\nabla f(x_0) \cdot e_i = Df(x_0)e_i$$

Hence

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

$\Longleftarrow$ Suppose that all partial derivatives exist and

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

We claim that

$$[Df(x_0)] = \nabla(x_0)$$

Indeed, notice that by definition

$$f(x_0+h) - f(x_0) - \nabla(x_0)\cdot h = o(h)$$

Hence

$$f(x_0+h) = f(x_0) + \nabla(x_0)\cdot h + o(h)$$

$\square$

**Example 2.13.** $f(x,y) = \begin{cases} \frac{x^2y}{x^2+y^2} & \text{if } (x,y)\neq(0,0) \\ 0 & \text{o.w} \end{cases}$ then

$$f_x(0,0) = 0, \quad f_y(0,0) = 0$$

and

$$\frac{f(h_1,h_2) - f(0,0) - \nabla f(0,0)\cdot(h_1,h_2)}{\sqrt{h_1^2+h_2^2}} = \frac{h_1^2h_2}{(h_1^2+h_2^2)\sqrt{h_1^2+h_2^2}}$$

take $h_1 = 1/n, h_2 = 1/n$ we get that

$$\frac{1}{1\sqrt{2}} \not\to 0$$

hence this is not differentiable.

**Example 2.14.** $f(x,y) = x\cdot\sqrt{y}$ for $x\in\mathbb{R},\ y\geq 0$ then

$$\frac{\partial f}{\partial x} = \sqrt{y} \quad x\in\mathbb{R},\ y\geq 0$$

$$\frac{\partial f}{\partial y} = \frac{x}{2\sqrt{y}} \quad x\in\mathbb{R}, y>0$$

But what happens to $\frac{\partial f}{\partial y}$ at $(x,0)$?

$$\begin{aligned}\frac{\partial f}{\partial y}(x,0) &= \lim_{h\to 0}\frac{f(x,h)-f(x,0)}{h} \\ &= \lim_{h\to 0}\frac{x\sqrt{h}}{h} \\ &= \lim_{h\to 0}\frac{x}{\sqrt{h}}\end{aligned}$$

If $x\neq 0$ we have that

$$\lim_{h\to 0}\frac{x}{\sqrt{h}} \text{ doesn't exist}$$

hence $\frac{\partial f}{\partial y}$ doesn't exist at $(x,0)$ for $x \neq 0$. But for $x=0$ we have that

$$\frac{\partial f}{\partial y}(0,0) = 0$$

Now we ask, is $f(x,y) = x\sqrt{y}$ differentiable at $(0,0)$. Indeed

$$\nabla f(0,0) = (f_x(0,0), f_y(0,0)) = (0,0)$$

and

$$\begin{aligned}
\frac{|f\left((0,0)+(h_1,h_2)\right) - f(0,0) - (0,0)\cdot(h_1,h_2)|}{\|h\|} &= \frac{|h_1\sqrt{h_2}|}{\sqrt{h_1^2+h_2^2}} \\
&= \sqrt{\frac{h_1^2|h_2|}{h_1^2+h_2^2}} \\
&\leq \sqrt{|h_2|} \to 0
\end{aligned}$$

therefore $f$ is differentiable at $(0,0)$.

**Theorem 2.15.** *$U \subseteq \mathbb{R}^n$ and $f\colon U \to \mathbb{R}^n$ with $f = (f_1, \cdots, f_m)$ then $f$ is differentiable iff $f_1, \cdots, f_n\colon U \to \mathbb{R}$ are all differentiable. Then*

$$[Df(x_0)]_{m\times n} = \begin{pmatrix} \nabla f_1(x_0) \\ \nabla f_2(x_0) \\ \vdots \\ \nabla f_m(x_0) \end{pmatrix}$$

*Therefore*

$$Df(x_0)\begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix} = \begin{pmatrix} \nabla f_1(x_0)\cdot h \\ \nabla f_2(x_0)\cdot h \\ \vdots \\ \nabla f_m(x_0)\cdot h \end{pmatrix}$$

**Example 2.16.** $f(x,y) = (x^2y^2, x^2+y^2, x^2-y^2)$ then

$$[Df(x,y)] = \begin{pmatrix} 2xy^2 & 2x^2y \\ 2x & 2y \\ 2x & -2y \end{pmatrix}$$

Therefore

$$[Df(1,2)] = \begin{pmatrix} 2 & 4 \\ 2 & 4 \\ 2 & -4 \end{pmatrix}$$

Hence

$$Df(1,2)\colon \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \to \begin{pmatrix} 8h_1 + 4h_2 \\ 2h_1 + 4h_2 \\ 2h_1 - 4h_2 \end{pmatrix}$$

*Proof.*

- Suppose that $f$ is differentiable, then

  $$f(x_0 + h) = f(x_0) + Df(x_0)h + o(\|h\|)$$

  Taking the dot with $e_i \in \mathbb{R}^m$ we get

  $$f_i(x_0 + h) = f_i(x_0) + Df(x_0)(h) \cdot e_i + o(\|h\|)$$

  Therefore $f_i$ is differentiable with

  $$Df_i(x_0)\colon h \to (Df(x_0)(h)) \cdot e_i$$

- Suppose that each $f_i$ is differentiable then

  $$f_i(x_0 + h) = f_i(x_0) + Df_i(x_0)h + o_i(\|h\|)$$

  Therefore

  $$f(x_0 + h) = f(x_0) + (Df_1(x_0)h, \cdots, Df_m(x_0)h) + (o_1(\|h\|), \cdots, o_m(\|h\|))$$

  Hence $f$ is differentiable with

  $$Df(x_0) = \begin{pmatrix} Df_1(x_0) \\ Df_2(x_0) \\ \vdots \\ Df_m(x_0) \end{pmatrix}$$

  but $Df_i(x_0) = \nabla f_i(x_0)$. i.e

  $$Df(x_0)\colon h \to (\nabla f_i(x_0) \cdot h)_{1 \le i \le m}$$

$\square$

**Remark 2.17.** $f, g$ differentiable, then $f + \alpha g$ is differentiable, and

$$D(f + \alpha g) = Df + \alpha Dg$$

## 2.2 Chain Rule.

**Theorem 2.18.** *Let $U \subseteq \mathbb{R}^n$ and $f\colon U \to \mathbb{R}^m$ and $V \subset \mathbb{R}^m$ open and $g\colon V \to \mathbb{R}$, take $x_0 \in U$ s.t $f(x_0) \in V$.*

*If $f$ is differentiable at $x_0$, and $g$ is differentiable at $f(x_0)$. Then $h := g \circ f$ is differentiable at $x_0$ with*

$$Dh(x_0) = D_g(f(x_0)) \circ D_f(x_0)$$

*In matrix form*

$$[Dh(x_0)] = [D_g(f(x_0))][Df(x_0)]$$

**Example 2.19.** $h \colon \mathbb{R}^m \to \mathbb{R}$ with

$$h(u,v) = g(x(u,v), y(u,v))$$

Let $f(u,v) = (x(u,v), y(u,v))$ Therefore

$$[Dh(u,v)] = [Dg(x(u,v), y(u,v))][Df(u,v)]$$
$$\begin{pmatrix} \frac{\partial h}{\partial u} & \frac{\partial h}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$$

Therefore we get that

$$\begin{pmatrix} \frac{\partial h}{\partial u} & \frac{\partial h}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{\partial g}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial u} & \frac{\partial g}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial v} \end{pmatrix}$$

**Example 2.20.** $\gamma \colon [a,b] \to \mathbb{R}^n$ and $f \colon \mathbb{R}^n \to \mathbb{R}$ with

$$h(t) = f(\gamma(t))$$

Then

$$\begin{aligned} [Dh(t)] &= [Df(\gamma(t))][D\gamma(t)] \\ [h'(t)] &= (\nabla f(\gamma(t))) \begin{pmatrix} \gamma_1' \\ \gamma_2' \\ \vdots \\ \gamma_n' \end{pmatrix} \\ &= \nabla f(\gamma(t)) \cdot \gamma'(t) \end{aligned}$$

i.e $h \colon [a,b] \to \mathbb{R}$.

**Consequences.**

1. Mean value theorem for $\mathbb{R}^n \to \mathbb{R}$
2. $Df(x) = 0 \implies f$ is constant (in an open and connected domain).
3. If $\frac{\partial f_i}{\partial x_i}$ exist and are continuous at $x_0$ then $f$ is differentiable at $x_0$.

*Proof.* $f$ is differentiable at $x_0$ then

$$f(x) = f(x_0-) + Df(x_0)(x - x_0)_o(|x - x_0|) \quad (*)$$

and $g$ is diff at $f(x_0)$, therefore

$$g(y) = g(f(x_0)) + Dg(f(x_0))(y - f(x_0)) + o(|y - f(x_0)|) \quad (**)$$

We have

$$\begin{aligned} h(x) &= g(f(x)) \\ &= g(f(x_0)) + Dg(f(x_0))(f(x) - f(x_0)) + o(|f(x) - f(x_0)|) \\ &= h(x_0) + Dg(f(x_0))\left(Df(x_0)(x - x_0) + o(|x - x_0|)\right) + o(|f(x) - f(x_0)|) \\ &= h(x_0) + (Df(f(x_0)) \circ Df(x_0))(x - x_0) + (Dg(f(x_0))(o(|x - x_0|)) + o(|f(x) - f(x_0)|)) \end{aligned}$$

We will be done if we show that

$$R(x) = (Dg(f(x_0))(o(|x - x_0|)) + o(|f(x) - f(x_0)|)) = o(|x - x_0|)$$

but since $Dg(f(x_0))$ is a linear transformation

$$\frac{|Dg(f(x_0))(o(x - x_0))|}{|x - x_0|} = Dg(f(x_0))\left(\frac{o(x - x_0)}{|x - x_0}\right)$$

But using continuity we get as $x \to x_0$

$$Dg(f(x_0))\left(\frac{o(x - x_0)}{|x - x_0}\right) \to Dg(0) = 0$$

On the other hand, $Df(x_0)$ is a continuous linear transformation, then $\exists M$ s.t

$$\|Df(x_0)v\| \leq Mv$$

Let $\varepsilon > 0$, $\exists \delta > 0$ s.t

$$|o(|y - f(x_0)|)| < \frac{\varepsilon}{M+2}|y - f(x_0)|, \quad \forall |u - f(x_0)| < \delta$$

Now $f$ is differentiable at $x_0$, then it is continuous. Hence $\exists \gamma > 0$ s.t for $|x - x_0| < \gamma$

$$|f(x) - f(x_0)| < \delta$$

Therefore for $|x - x_0| < \gamma$ we have

$$o(|f(x) - f(x_0)|) < \frac{\varepsilon}{M+2}|f(x) - f(x_0)|$$

By $(*)$ this is equal to $\frac{\varepsilon}{M+2}|Df(x_0)(x - x_0) + o(|x - x_0|)|$ therefore

$$o(|f(x) - f(x_0)|) < \frac{\varepsilon}{M+2}M\left(|x - x_0| + o(|x - x_0|)\right)$$

by choosing $\gamma$ even smaller

$$o(|x - x_0|) < |x - x_0|$$

Then for $|x - x_0| < \gamma$, we get

$$\begin{aligned} o(|f(x) - f(x_0)|) &< \frac{\varepsilon}{M+2}(M+1)|x - x_0| \\ &< \varepsilon|x - x_0| \end{aligned}$$

□

**Recall.** $\gamma\colon [a,b] \to U \subseteq \mathbb{R}^n$ and $f\colon U \to \mathbb{R}$ differentiable. Then

$$h(t) = f(\gamma(t)) \text{ is differentiable}, \quad h\colon [a,b] \to \mathbb{R}$$

with

$$h'(t) = \nabla(\gamma(t)) \cdot \gamma'(t)$$

**Theorem 2.21** (Mean Value theorem)**.** *$U \subseteq \mathbb{R}^n$ open and convex domain, $f\colon U \to \mathbb{R}$ differentiable and $x, y \in U$ then $\exists c$ on the line segment joining $x$ to $y$ s.t*

$$f(y) - f(x) = \nabla f(c) \cdot (y - x)$$

*Proof.* Take $\gamma\colon [0,1] \to \mathbb{R}^n$ s.t

$$\gamma(t) = (1-t)x + ty$$

notice that $\gamma(1) = y \in U,\ \gamma(0) = x \in U$, therefore by convexity

$$\gamma([0,1]) \subseteq U$$

define

$$h(t) = f(\gamma(t)), \quad h\colon [0,1] \to \mathbb{R}$$

Then by MVT on $\mathbb{R}$, $\exists t_0 \in [0,1]$ s.t

$$f(y) - f(x) = f(\gamma(1)) - f(\gamma(0)) = h(1) - h(0) = h'(t_0)(1-0)$$

But $h'(t_0)$ (by chain rule)

$$\nabla f(\gamma(t)) \cdot \gamma'(t) = f(\gamma(t)) \cdot (y - x)$$

Therefore

$$f(y) - f(x) = f(\gamma(t_0)) \cdot (y - x)$$

letting $c = \gamma(t_0)$ which belongs to the line segment, as desired. □

**Corollary 2.22.** *Let $U \subseteq \mathbb{R}^n$ open, connected domain. Moreover, let $f\colon U \to \mathbb{R}^m$ differentiable s.t*

$$Df(x) = 0 \quad \forall x \in U$$

*Then $f$ is constant.*

*Proof.*

- **Case $m = 1$.** $U$ is open and connected, therefore it must be path connected. Fix $x_0 \in U$ and $y \in U$, let $A = \{x \in U\colon f(x) = f(x_0)\}$, note that $A \neq \varnothing$ and $A$ is closed since

$$A = f^{-1}(\{x_0\})$$

let $x \in A$, $U$ is open then $\exists \delta > 0$ s.t

$$B(x, \delta) \subseteq U$$

but $B(x, \delta)$ is convex, then by the MVT for every $y \in B(x, \delta)$ $\exists c$ in the line joining $x$ to $y$ s.t

$$f(y) - f(x) = \nabla f(c) \cdot (y - x) = 0$$

Therefore

$$f(y) = f(x) = f(x_0)\ \forall y \in B(x, \delta)$$

Therefore

$$B(x, \delta) \subseteq A$$

hence $A$ is open. By connectedness of $U$ we get that $A = U$.

- **General case.** Let $f = (f_1, \cdots, f_m)$ therefore

$$[Df(x)] = \begin{pmatrix} \nabla f_1 \\ \vdots \\ \nabla f_m \end{pmatrix}$$

but $\nabla f_i = 0 \implies f_i = c_i$, therefore $f = c$.

□

**Theorem 2.23.** *$U \subseteq \mathbb{R}^n$ and $f\colon U \to \mathbb{R}$ and fix $a = (a_1, \cdots, a_n) \in U$ If $\frac{\partial f}{\partial x_i}$ exist in a neighborhood of $a$ and $\frac{\partial f}{\partial x_i}$ are continuous at $a$. Then $f$ is differentiable at $a$.*

**Corollary 2.24.** *If $f\colon U \to \mathbb{R}^m$ and $a \in U$ if $\frac{\partial f_i}{\partial x_j}$ exist in a neighborhood of $a$ and are continuous at $a$, then $f$ is differentiable at $a$.*

**Example 2.25.** $f(x, y) = x\sqrt{y}$ is differentiable for $x \in \mathbb{R}$ and $y > 0$ and $f_x = \sqrt{y}$, $f_y = \frac{x}{2\sqrt{y}}$ which both exist and are continuous in a neighborhood of $(x_0, y_0)$ for $x_0 \in \mathbb{R}$ and $y_0 > 0$. So $f$ is differentiable at these points.

*Proof.* Let $B(a,r) \subset U$ and $|h| < R$ with $h = (h_1, \cdots, h_n)$, then

$$\begin{aligned}
f(a+h) - f(a) &= f(a + h_1 e_1) - f(a) \\
&\quad + f(a + h_1 e_1 + h_2 e_2) + f(a + h_1 e_1) \\
&\quad - f(a + h_1 e_1 + h_2 e_2) - f(a) \\
&\vdots \\
&\quad + f(a+h) - f(a + h_1 e_1 + h_2 e_2 + \cdots + h_{n-1} e_{n-1}) \\
&= \frac{\partial f}{\partial x_1}(a + c_1 e_1) \cdot h_1 \\
&\quad \frac{\partial f}{\partial x_2}(a + h_1 e_1 + c_2 e_2) \cdot h_2 \\
&\vdots \\
&\quad + \frac{\partial f}{\partial x_n}(a + h_1 e_1 + \cdots + h_{n-1} e_{n-1} + c_n e_n) h_n
\end{aligned}$$

Where $c_i \in [0, h_i]$, therefore

$$\begin{aligned}
f(a+h) - f(a) &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a + h_1 e_1 + \cdots + h_{i-1} e_{i-1} + c_i e_i) h_i \\
&= \frac{|f(a+h) - f(a) - \nabla f(a) \cdot h|}{h} \\
&= \frac{1}{|h|} \sum_{i=1}^{n} \left( \frac{\partial f}{\partial x_i}(a + h_1 e_1 \cdots + h_{i-1} e_{i-1} + c_i e_i) - \frac{\partial f}{\partial x_i}(a) \right) h_i| \\
&\le \sum_{i=1}^{n} \sum_{i=1}^{n} |\frac{\partial f}{\partial x_i}(a + h_1 e_1 \cdots + h_{i-1} e_{i-1} + c_i e_i) - \frac{\partial f}{\partial x_i}(a)| \frac{|h_i|}{|h|} \\
&\le \sum_{i=1}^{n} \sum_{i=1}^{n} |\frac{\partial f}{\partial x_i}(a + h_1 e_1 \cdots + h_{i-1} e_{i-1} + c_i e_i) - \frac{\partial f}{\partial x_i}(a)|
\end{aligned}$$

Let $h \to 0$ then $a + h_1 e_1 + \cdots + h_{i-1} e_{i-1} + c_i e_i \to a$. and use continuity to get

$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - \nabla f(a) \cdot h}{|h|} = 0$$

so $f$ is differentiable at $a$. □

**Remark 2.26.** The converse is not true, since $f(x,y) = x\sqrt{y}$ is differentiable at $(0,0)$ but $f_y$ doesn't exist at $(x,0)\ \forall x \neq 0$.

**Corollary 2.27.** *$U \subset \mathbb{R}^n$ and $f \colon U \to \mathbb{R}^m$ differentiable,*

$$Df \colon U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$$

*Then $Df$ is continuous on $U$ iff all partial derivatives $\frac{\partial f_i}{\partial x_i}$ of $f$ are continuous on $U$.*

*Proof.* $Df(u) = \begin{pmatrix} \nabla f_1 \\ \nabla f_2 \\ \vdots \\ \nabla f_n \end{pmatrix}$ which is continuous iff every entry is continuous. $\square$

**Definition 2.28.** $V, W$ finite dimensional vector space $U \subset V$ open and $f\colon U \to W$ we say that $f$ is $C^1$ iff $f$ is differentiable and $Df$ is continuous.

Note that this is equivalent to saying that all partial derivatives exist and are continuous.

## 2.3 Higher order derivatives.

**Motivation.** Let $\alpha = \{e_1, \cdots, e_n\}$ basis of $\mathbb{R}^n$ and $\hat{\alpha} = \{\hat{e_1}, \cdots, \hat{e_m}\}$ basis of $\mathbb{R}^m$. Take

$$f\colon U \to \mathbb{R}^m, \quad \text{differentiable}$$

and

$$Df\colon U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$$

Define

$$\gamma = \{T_{11}, \cdots, T_{1m}, T_{21}, \cdots, T_{nm}\}$$

with

$$T_{i_0 j_0}\colon e_{i_0} \to \hat{e_{j_0}},\ e_{i \neq i_0} \to 0$$

Take $x \in U$ then

$$Df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$$

with

$$Df(x) = r_{11}T_{11} + r_{12}T_{12} + \cdots + r_{nm}T_{nm}$$

Then

$$Df(x)e_j = r_{j1}\hat{e_j} + \cdots + r_{jm}\hat{e_m}$$

therefore

$$[Df(x)] = \begin{pmatrix} \nabla f_1 \\ \vdots \\ \nabla f_m \end{pmatrix}$$

Therefore

$$Df(x)e_i = \begin{pmatrix} \frac{\partial f_1}{\partial x_i} \\ \vdots \\ \frac{\partial f_m}{\partial x_i} \end{pmatrix}$$

So we get that

$$r_{ij} = \frac{\partial f_j}{\partial x_i}$$

Therefore

$$[Df]_\gamma = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} \\ \vdots \\ \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Hence we get that

$$D[Df] = \begin{pmatrix} \nabla(f_1)_{x_1} \\ \nabla(f_1)_{x_2} \\ \vdots \\ \partial(f_m)_{x_m} \end{pmatrix}$$

Note that

$$Df\colon U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$$

Hence

$$D(Df)\colon U \to \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$$

Therefore $D[Df]$ is a $mn \times n$ matrix, which we call the Hessian matrix.

**Example 2.29.** $f(x,y) = (x^2y^3, x^2 + y^3, x^3 + y^2)$ we have that

We get that

$$D^2(f) = \begin{pmatrix} (f_1)_{xx} & (f_1)_{xy} \\ (f_1)_{yx} & (f_1)_{yy} \\ (f_2)_{xx} & (f_2)_{xy} \\ (f_2)_{yx} & (f_2)_{yy} \\ (f_3)_{xx} & (f_3)_{xy} \\ (f_3)_{yx} & (f_3)_{yy} \end{pmatrix}$$

**Remark 2.30.** In the particular case where $m = 1$ we have that $f\colon U \to \mathbb{R}$ therefore

$$D^f = \begin{pmatrix} \nabla f_{x_1} \\ \nabla f_{x_2} \\ \vdots \\ \nabla f_{x_n} \end{pmatrix}$$

In the $n = 2$ case we get

$$\begin{pmatrix} \frac{\partial}{\partial x}\frac{\partial f}{\partial x} & \frac{\partial}{\partial y}\frac{\partial f}{\partial x} \\ \frac{\partial}{\partial x}\frac{\partial f}{\partial y} & \frac{\partial}{\partial y}\frac{\partial f}{\partial y} \end{pmatrix}$$

**Theorem 2.31** (Mixed derivatives theorem). *$U \subseteq \mathbb{R}^n$ and $f\colon U \to \mathbb{R}$, take $x_0 \in U$. If $\frac{\partial f}{\partial x_i \partial x_j}$ and $\frac{\partial f}{\partial x_i \partial x_j}$ exist in a neighborhood of $x_0$ and are continuous at $x_0$ then they are equal.*

*Proof.* $n = 2$, assume that $\frac{\partial f}{\partial x_i \partial x_j}$ and $\frac{\partial f}{\partial x_i \partial x_j}$ exist in a neighborhood of $a$ and are continuous at $a = (a_1, a_2)$. Define for $h, k > 0$

$$w(h,k) = f(a_1 + h) - f(a_1 + h, a_2) - f(a_1, a_2 + h) + f(a_1, a_2)$$

Define

$$v(h,k) = f(a_1 + h, a_2 + k) - f(a_1 + h_1 a_2)$$

and

$$u(h,k) = f(a_1 + h, a_2 + k) - f(a_1, a_2 + k)$$

Notice that by MVT

$$\begin{aligned} w(h,k) = v(h,k) - v(0,k) &= v_h(c,k) \quad c \in [0,h] \\ &= (f_{x_1}(a_1 + c, a_2 + k) - f_{x_1}(a_1 + x, a_2))\, h \\ &= (f_{x_1 x_2}(a_1 + c, a_2 + d))\, kh \quad d \in [0,k] \end{aligned}$$

Similarly

$$\begin{aligned} w(h,k) &= u(h,k) - u(h,0) \\ &= u_k(h,d')k \quad d' \in [0,k] \\ &= \left(f_{x_2}(a_1 + h, a_2 + d') - f_{x_2}(a_1, a_2 + d')\right) k \\ &= \left(f_{x_2 x_1}(a_1 + c', a_2 + d')\right) kh \quad c \in [0,h] \end{aligned}$$

Hence

$$f_{x_1 x_2}(a_1 + c, a_2 + d') = f_{x_2 x_1}(a_1 + c', a_2 + d')$$

then let $h, k \to 0$ by continuity we get

$$f_{x_1 x_2}(a_1, a_2) = f_{x_2 x_1}(a_1, a_2)$$

□

**Definition 2.32.** We say that $f \in C^k(U)$ ($k \in \mathbb{N}$) iff all partial derivatives of order $k$ of $f$ (of its components exist) and are continuous.

**Taylor's theorem.** In one variable we have that if $g$ is $C^1$

$$g(t_0 + h) = g(t_0) + g'(t_0)h + o(|h|)$$

if $g$ is $C^2$ then

$$g(t_0 + h) = g(t_0) + g'(t_0)h + \frac{g''(t_0)h^2}{2!} + o(|h|^2)$$

if $g$ is $C^k$ we get that

$$g(t_0 + h) = \sum_{i=0}^{k} \frac{g^{(i)}(t_0)}{i!} h^i + o(|h|^k)$$

Now take $U \subset \mathbb{R}^n$ and take

$$f \colon U \to \mathbb{R}$$

take $x_0 \in U$ and assume that $f$ is $C^1$ in $U$ also take $\hat{u}$ unit vector in $\mathbb{R}^n$. take

$$g(t) = f(x + t\hat{u})$$

which is defined for $t \leq \delta$ small enough. Therefore

$$g \in C^1([-\delta, \delta])$$

Hence

$$g(t) = g(0) + \frac{g'(0)}{1!} \cdot t + o(|t|)$$

But we have that

$$g'(0) = \nabla f(x_0 + t\hat{u}) \cdot \hat{u}$$

Therefore

$$g'(0) = \partial f(x_0) \cdot \hat{u}$$

Hence we get that

$$f(x_0 + t\hat{u}) = f(x_0) + \nabla(f(x_0) \cdot \hat{u})t + o(|t|)$$

Therefore

$$\begin{aligned} f(x_0 + h) &= f(x_0 + |h|\frac{h}{|h|}) \\ &= f(x_0) + \nabla f(x_0) \cdot h + o(|h|) \\ &= f(x_0) + \sum_{i=1}^{n} f_{x_i} h_i + o(|h|) \end{aligned}$$

Assume that $f$ is $C^2$, then $g$ is $C^2$ then

$$g(t) = g(0) + g'(0)t + \frac{g''(0)t^2}{2} + o(|t|^2)$$

but

$$\begin{aligned} g''(x) &= \left( \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0 + t\hat{u}) u_i \right)' \\ &= \left( \sum_{i=1}^{n} \nabla (\frac{\partial f}{\partial x_i})(x_0 + t\hat{u}) \right) \cdot u_i \end{aligned}$$

Therefore

$$g''(0) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial f^2}{\partial^2 x_j x_i} u_j u_i$$

But we have that

$$D^2 f = \begin{pmatrix} \frac{\partial f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial f}{\partial x_n x_1} \\ \frac{\partial f}{\partial x_1 x_n} & \cdots & \frac{\partial f}{\partial x_n x_n} \end{pmatrix}$$

therefore

$$g''(0) = D^2 f \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

This is denoted as $\langle D^2 f(x_0)\hat{u}, \hat{u}\rangle$ therefore we get that

$$g(t) = g(0) + (Df(x_0) \cdot \hat{u})t + \langle Df(x_0)\hat{u}, \hat{u}\rangle \frac{t^2}{2} + o(|t|^2)$$

In general we get that

$$f(x_0 + h) = f(x_0) + \nabla f(x_0) \cdot h + \frac{\langle D^2 f(x_0), h\rangle}{2} + o(|h|^2)$$

Assume that $f = f(x_1, x_2)$, then

$$\begin{aligned} f(x_0 + h) &= f(x_0) + \frac{\nabla f}{\nabla x_1} h_1 + \frac{\nabla f}{\nabla x_2} h_2 + \begin{pmatrix} f_{x_1 x_1} & f_{x_1 x_2} \\ f_{x_2 x_1} & f_{x_2 x_2} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + o(|h|^2) \\ &= f(x_0) + f_{x_1} h_1 + f_{x_2} h_2 + \frac{1}{2}(f_{x_1 x_1} h_1^2 + f_{x_1 x_2} h_2 h_1 + f_{x_2 x_1} h_1 h_2 + f_{x_2 x_2} h_2^2) + o(|h|^2) \\ &= \frac{f(x_0)}{0!} + \left(\frac{f_{x_1} h_1}{1!} + \frac{f_{x_2} h_2}{1!}\right) + \left(\frac{f_{x_1 x_1} h_1^2}{2!} + \frac{f_{x_1 x_2} h_1 h_2}{1!1!} + \frac{f_{x_2 x_2} h_2^2}{2!}\right) + o(|h|^2) \end{aligned}$$

See the pattern here.

**Multi-index notation.** Take $\alpha = (\alpha_1, \cdots, \alpha_n)$ where $\alpha_i \in \mathbb{N} \cup \{0\}$. We denote the order of $\alpha$ by

$$|\alpha| := \alpha_1 + \alpha_2 + \cdots + \alpha_n$$

and the factorial of $\alpha$ by

$$\alpha! = \alpha_1! \alpha_2! \cdots \alpha_n!$$

Then the $\alpha$-partial derivative of a function $f$ with $f \in C^{|\alpha|}$:

$$D^{\alpha = (\alpha_1, \cdots, \alpha_n)} := \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}}$$

Moreover exponent for $x \in \mathbb{R}^n$

$$x^\alpha = x_1^{\alpha_1} \cdots, x_n^{\alpha_n}$$

**Example 2.33.** $\frac{D^{(1,2,1)}(x^2 y^2 z^3)}{(1,2,1)!}$ with $f := f(x, y, z)$ we get

$$\begin{aligned} \frac{D^{(1,2,1)}(x^2 y^2 z^3)}{(1,2,1)!} &= \frac{3 \cdot 2 \cdot 2x \cdot z^2}{1!2!1!} \\ &= 6xz^2 \end{aligned}$$

**Theorem 2.34** (Taylor's theorem). *We have $U$ open subset of $\mathbb{R}^n$ and $x_0 \in U$ with $f\colon U \to \mathbb{R}$ $C^k$ function, then*

$$f(x_0 + h) = \sum_{0 \le |\alpha| \le k} \frac{D^\alpha f(x_0) h^\alpha}{\alpha!} + o(|h|^k)$$

*or*

$$f(x) = \sum_{0 \le |\alpha| \le k} \frac{D^\alpha f(x_0)(x - x_0)^\alpha}{\alpha!} + o(|x - x_0|^k)$$

## 2.4 Extremum

**Definition 2.35.** $f\colon U \to \mathbb{R}$ $U \subseteq \mathbb{R}^n$ and $x_0 \in U$. We say that $f$ has a local minimum at $x_0$ iff $\exists \delta > 0$ s.t $f(x) \ge f(x_0)$ $\forall x \in B(x_0, \delta)$.

We say that $f$ has a local maximum at $x_0$ iff $\exists \delta > 0$ s.t $f(x) \le f(x_0)$ $\forall x \in B(x_0, \delta)$

**Theorem 2.36.** *If $f$ has a local min (or max) at $x_0$ and $f$ is differentiable at $x_0$ then $\nabla f(x_0) = 0$.*

*Proof.* $f$ has a local minimum at $x_0$, then $\exists \delta > 0$ s.t $f(x) > f(x_0)$ $\forall x \in B(x_0, \delta)$. Let $|t| < \delta$ and $\hat{u}$ a unit vector; since $f$ is differentiable at $x_0$, with $x_0 + t\hat{u} \in B(x_0, \delta)$. Hence

$$f(x_0 + t\hat{u}) = f(x_0) + (\nabla f(x_0) \cdot \hat{u})t + o(|t|) \ge f(x_0)$$

Therefore

$$(\nabla f(x_0) \cdot \hat{u})t + o(|t|) \ge 0$$

- if $t \ge 0$ we have that

  $$\nabla f(x_0) \cdot \hat{u} + \frac{o(|t|)}{t} \ge 0$$

  let $t \to 0$

  $$\nabla f(x_0) \cdot \hat{u} \ge 0$$

- if $t \le 0$ we have that

  $$\nabla f(x_0) \cdot \hat{u} + \frac{o(|t|)}{t} \le 0$$

  Let $t \to 0$ we have

  $$\nabla f(x_0) \cdot \hat{u} \le 0$$

Therefore $\forall \hat{u},\ |\hat{u}| = 1$ we have that

$$\nabla f(x_0) \cdot \hat{u} = 0 \implies \nabla f(x_0) = 0$$

□

**Example 2.37.** Find local min/max of the following:

$$f(x, y) = x^2 + y^2$$

We have

$$\nabla f(x, y) = (2x, 2y) = (0, 0)$$

Hence

$$\begin{cases} 2x = 0 \\ 2y = 0 \end{cases} \implies (x, y) = 0$$

Therefore

$$f(0,0)0 \leq x^2 + y^2 = f(x, y) \ \forall (x, y)$$

Therefore at $(0, 0)$ ia have a minimum.

$$f(x, y) = x^2 - y^2$$

Therefore

$$\nabla(x, y) = (2x, -2y) = 0 \implies x = y = 0$$

Notice that $f(0, 0) = 0$ Take $\delta > 0$, $B(0, \delta)$ take

$$f(\delta/2, 0) = \delta^2/4 > 0, \quad f(0, \delta) = -\delta^2/4 < 0$$

Therefore $(0, 0)$ is neither a local min nor a local max (it is a saddle point).

**Second derivative test.** Review the spectral theorem and diagonalization.

Suppose that $A$ is symmetric and let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of $A$ with corresponding orthonormal basis of eigenvectors $v_1, \cdots, v_n$, i.e

$$Av_i = \lambda_i v_i, \quad \langle v_i, v_j \rangle = \delta_{ij}$$

Let $v \in \mathbb{R}^n$, then we can write $v$ as $r_1 v_1 + \cdots + r_n v_n$, therefore

$$Av = (\lambda_1 r_1)v_1 + \cdots + (\lambda_n r_n)v_n$$

Therefore we get that

$$\begin{aligned} \langle Av, v \rangle &= \langle r_1 \lambda_1 v_1 + \cdots + r_n \lambda_n v_n, r_1 v_1 + \cdots + r_n v_n \rangle \\ &= r_1^2 \lambda_1 + r_2^2 \lambda_2 + \cdots + r_n^2 \lambda_n \end{aligned}$$

**Definition 2.38.** Let $A$ be an $n \times n$ symmetric matrix we say that $A$ is positive definite iff

$$\langle Av, v \rangle > 0 \quad \forall v \neq 0$$

iff all eigenvalues of $A$ are positive.

We say that it is positive semi-definite iff

$$\langle Av, v \rangle \geq 0 \quad \forall v \neq 0$$

iff all eigenvalues of $A$ are nonnegative.

We define negative (semi) definite similarly.

Otherwise, we say that $A$ is indefinite (at least two eigenvalues of opposite signs).