Chapter 31

MSMYP3 Functions Of Several Real Variables

(31.1) Limits & Continuity

(31.1.1) Norms

Many ideas from single variable analysis extend readily to more dimensions. The first task when moving from $\mathbb{R}$ to $\mathbb{R}^n$ is to find a suitable way to measure distance.

Definition 1 *For* $\mathbf{a} = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$ *the Euclidean Norm of* $\mathbf{a}$, $\|\mathbf{a}\|$, *by*

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$$

where "$\sqrt{\ }$*" denotes the positive square root.*

This is a good generalisation of the absolute value measure in $\mathbb{R}$, as for $\mathbf{a} = (a) \in \mathbb{R}$ it is clearly the case that $\|\mathbf{a}\| = |a|$. $\mathbb{R}^n$ together with the Euclidean Norm is called n-dimensional Euclidean space.

Lemma 2 (Cauchy's Inequality) *For real numbers* $a_1, a_2, \dots, a_n$ *and* $b_1, b_2, \dots, b_n$

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 = \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)$$

Proof. Let $A > 0$ then

$$\begin{aligned} Ax^2 + Bx + C &= A\left(x^2 + \frac{B}{A}x + \frac{C}{A}\right) \\ &= A\left(x + \frac{B}{2A}\right)^2 + C - \frac{B^2}{4A} \\ &\geqslant 0 \text{ if and only if} \end{aligned}$$

$$\begin{aligned} C - \frac{B^2}{4A} &> 0 \\ B^2 &> 4AC \end{aligned}$$

Using this,

$$\begin{aligned}\sum_{i=1}^{n}(a_i x + b_i)^2 &= \sum_{i=1}^{n} a_i^2 x^2 + b_i^2 + 2a_i b_i x \\ &= \left(\sum_{i=1}^{n} a_i\right)^2 x^2 + \left(2\sum_{i=1}^{n} a_i b_i\right) x + \sum_{i=1}^{n} b_i^2 \\ &\geqslant 0 \text{ if and only if} \\ \left(2\sum_{i=1}^{n} a_i b_i\right)^2 &\leqslant \left(\sum_{i=1}^{n} a_i\right)\left(\sum_{i=1}^{n} b_i\right)\end{aligned}$$

□

Theorem 3 *For $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ the Euclidean Norm has the following properties.*

1. $\|\mathbf{a}\| \geqslant 0$ *with* $\|\mathbf{a}\| = 0 \Leftrightarrow \mathbf{a} = \mathbf{0}$.
2. $\|\lambda\mathbf{a}\| = |\lambda|\|\mathbf{a}\|$.
3. $\|\mathbf{a}+\mathbf{b}\| \leqslant \|\mathbf{a}\| + \|\mathbf{b}\|$. *(Triangle inequality for $\mathbb{R}^n$.)*

Proof. 1. $\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} \geqslant 0$ with equality

$$\begin{aligned}&\Leftrightarrow a_1^2 + a_2^2 + \cdots + a_n^2 = 0 \\ &\Leftrightarrow a_1^2 = a_2^2 = \cdots = a_n^2 = 0 \\ &\Leftrightarrow a_1 = a_2 = \cdots = a_n = 0 \\ &\Leftrightarrow \mathbf{a} = \mathbf{0}\end{aligned}$$

2.

$$\begin{aligned}\|\lambda\mathbf{a}\| &= \|(\lambda a_1, \lambda a_2, \ldots, \lambda a_n)\| \\ &= \sqrt{(\lambda a_1)^2 + (\lambda a_2)^2 + \cdots + (\lambda a_n)^2} \\ &= \sqrt{\lambda}\sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} \\ &= |\lambda|\|\mathbf{a}\|\end{aligned}$$

3. As norms are positive is it sufficient to show that $\|\mathbf{a}+\mathbf{b}\|^2 \leqslant (\|\mathbf{a}\| + \|\mathbf{b}\|)^2$.

$$\begin{aligned}\|\mathbf{a}+\mathbf{b}\|^2 &= \sum_{i=1}^{n}(a_i + b_i)^2 \\ &= \sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 + 2\sum_{i=1}^{n} a_i b_i \\ &\leqslant \sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 + 2\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2} \qquad \text{by Cauchy's Inequality} \\ &= \|\mathbf{a}\| + \|\mathbf{b}\| + 2\|\mathbf{a}\|\|\mathbf{b}\| \\ &= (\|\mathbf{a}\| + \|\mathbf{b}\|)^2\end{aligned}$$

□

(31.1.2) Sequences

Definition 4 *A sequence in* $\mathbb{R}^n$ *is a collection of points* $\{\mathbf{a}^k\}$ *where* $k \in \mathbb{N}$ *indexes the points and where*

$$\mathbf{a}^k = \left(a_1^k, a_2^k, \ldots, a_n^k\right)$$

Definition 5 *A sequence* $\{\mathbf{a}^k\}$ *in* $\mathbb{R}^n$ *converges to a limit* $\mathbf{a} \in \mathbb{R}^n$*, written*

$$\lim_{k\to\infty} \mathbf{a}^k = \mathbf{a}$$

if

$$\forall \varepsilon > 0 \quad \exists N \in \mathbb{N} \quad s.t. \quad k > N \Rightarrow \|\mathbf{a}^k - \mathbf{a}\| < \varepsilon$$

Clearly each co-ordinate of $\mathbf{a}^k$ gives a sequence in $\mathbb{R}$. The following theorem comes as no surprise.

Theorem 6 *Let* $\{\mathbf{a}^k\}$ *be a sequence in* $\mathbb{R}^n$*.*

$$\lim_{k\to\infty} \mathbf{a}^k = \mathbf{a} \Leftrightarrow \lim_{k\to\infty} a_i^k = a_i \quad \forall i\ 1 \leqslant i \leqslant n$$

Proof. ($\Rightarrow$) Suppose $\lim_{k\to\infty} \mathbf{a}^k = \mathbf{a}$ then $\forall \varepsilon > 0\ \exists N$ such that $\|\mathbf{a}^k - \mathbf{a}\| < \varepsilon$ for all $k > N$. Hence for $k > N$

$$\left|a_i^k - a_i\right| \leqslant \sqrt{\left(a_1^k - a_1\right)^2 + \left(a_2^k - a_2\right)^2 + \cdots + \left(a_n^k - a_n\right)^2} = \|\mathbf{a}^k - \mathbf{a}\| < \varepsilon$$

Hence $\lim_{k\to\infty} a_i^k = a_i$.

($\Leftarrow$) Suppose that $\lim_{k\to\infty} a_i^k = a_i$ for $1 \leqslant i \leqslant n$. Let $\varepsilon > 0$ then $\frac{\varepsilon}{\sqrt{N}} > 0$ for $N > 0$. Hence for each i there exists N_i such that

$$k > N_i \quad \Rightarrow \quad \left|a_i^k - a_i\right| < \frac{\varepsilon}{\sqrt{N}}$$

Let

$$N = \max_{1\leqslant i\leqslant n}\{N_i\}$$

then for $k > N$

$$\begin{aligned}\|\mathbf{a}^k - \mathbf{a}\| &= \sqrt{\left(a_1^k - a_1\right)^2 + \left(a_2^k - a_2\right)^2 + \cdots + \left(a_n^k - a_n\right)^2} \\ &\leqslant \sqrt{\frac{\varepsilon^2}{N} + \frac{\varepsilon^2}{N} + \cdots + \frac{\varepsilon^2}{N}} \\ &= \sqrt{\varepsilon^2} = \varepsilon\end{aligned}$$

□

(31.1.3) Cauchy Sequences

Definition 7 *A sequence* $\{\mathbf{a}^k\}$ *in* $\mathbb{R}^n$ *is a Cauchy sequence if*

$$\forall \varepsilon > 0 \quad \exists N \in \mathbb{N} \quad k, l > N \to \|\mathbf{a}^k - \mathbf{a}^l\| < \varepsilon$$

Note that Theorem 3 means that $\|\mathbf{a}^l - \mathbf{a}^k\|$ may be used equally well as $\|\mathbf{a}^k - \mathbf{a}^l\|$ (putting $\lambda = -1$) so that without loss of generality it may be assumed that $k > l$.

Definition 8 *A subset* A *of* $\mathbb{R}^n$ *is bounded if there exists* M *for which* $\|\mathbf{x}\| < M$ *for all* $\mathbf{x} \in A$*.*

Lemma 9 *For* $\mathbf{c}, \mathbf{d} \in \mathbb{R}^n$

$$|\|\mathbf{c}\| - \|\mathbf{d}\|| < \|\mathbf{c} - \mathbf{d}\|$$

Proof. By the triangle inequality

$$\|\mathbf{a} + \mathbf{b}\| - \|\mathbf{b}\| < \|\mathbf{a}\|$$

Putting $\mathbf{a} = \mathbf{c} - \mathbf{d}$ and $\mathbf{b} = \mathbf{d}$ gives $\|\mathbf{c}\| - \|\mathbf{d}\| < \|\mathbf{c} - \mathbf{d}\|$.
Putting $\mathbf{a} = \mathbf{c} - \mathbf{d}$ and $\mathbf{b} = -\mathbf{c}$ gives $\|\mathbf{d}\| - \|\mathbf{c}\| < \|\mathbf{c} - \mathbf{d}\|$. □

Lemma 10 *A Cauchy sequence is bounded*

Proof. Let $\{\mathbf{a}^k\}$ be a Cauchy sequence in $\mathbb{R}^n$. Take $\varepsilon = 1$ then there exists N such that $\|\mathbf{a}^k - \mathbf{a}^l\| < 1$ for all $k, l > N$. Let L be the least natural number bigger than N, then by Lemma 9, for $k > l$,

$$\|\mathbf{a}^k\| - \|\mathbf{a}^l\| < \|\mathbf{a}^k - \mathbf{a}^l\| < 1$$

In particular, this gives $\|\mathbf{a}^k\| < 1 + \|\mathbf{a}^l\|$. Now let

$$N' = \max\left\{\|\mathbf{a}^1\|, \|\mathbf{a}^2\|, \ldots, \|\mathbf{a}^{L-1}\|\right\}$$

Take $M = \max\{N', 1 + \|\mathbf{a}^l\|\}$ then $\|\mathbf{a}^k\| < M$ for all k i.e. the sequence is bounded. □

If a sequence has a limit, then every subsequence of it will also have the same limit. However, there may well be convergent subsequences of sequences which do not themselves converge.

Theorem 11 (Bolzano-Weierstrass) *If* $\{a^k\}$ *is a sequence in* $\mathbb{R}$ *that is contained in the closed interval* $[b, c]$ *then there is a subsequence* $\{a^{k_i}\}$ *which has a limit* $d \in [b, c]$.

Theorem 12 *A sequence* $\{\mathbf{a}^k\}$ *in* $\mathbb{R}^n$ *is a Cauchy sequence if and only if it has a limit.*

Proof. ($\Leftarrow$) Suppose that $\{\mathbf{a}^k\}$ has a limit, $\mathbf{a}$ say. For any $\varepsilon > 0$, $\frac{\varepsilon}{2} > 0$ and so $\exists N$ such that $\|\mathbf{a}^k - \mathbf{a}\| < \frac{\varepsilon}{2}$. Hence for $k, l > N$,

$$\begin{aligned}\|\mathbf{a}^k - \mathbf{a}^l\| &= \|\left(\mathbf{a}^k - \mathbf{a}\right) + \left(\mathbf{a} - \mathbf{a}^l\right)\| \\ &< \|\mathbf{a}^k - \mathbf{a}\| + \|\mathbf{a}^l - \mathbf{a}\| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\end{aligned}$$

($\Rightarrow$) Suppose that $\{\mathbf{a}^k\}$ is Cauchy, then for all $\varepsilon > 0$ there exists N such that $\|\mathbf{a}^k - \mathbf{a}^l\| < \varepsilon$ for all $k, l > N$. So when $\mathbf{a}^k = \left(a_1^k, a_2^k, \ldots, a_n^k\right)$

$$\begin{aligned}\left|a_i^k - a_i^l\right| &= \sqrt{\left(a_i^k - a_i^l\right)^2} \\ &\leqslant \sqrt{\left(a_1^k - a_1^l\right)^2 + \left(a_2^k - a_2^l\right)^2 + \cdots + \left(a_n^k - a_n^l\right)^2} \\ &= \|\mathbf{a}^k - \mathbf{a}^l\| \\ &< \varepsilon\end{aligned}$$

Hence each of the co-ordinate sequences is Cauchy and thus by Theorem 6 it is sufficient to show that all the co-ordinate sequences have a limit.
Let $\{a_i^k\}$ be one of the co-ordinate sequences, so $1 \leqslant i \leqslant n$. By Lemma 10 $\{a_i^k\}$ must be bounded and hence by Bolzano-Weierstrass (Theorem 11) it has a convergent subsequence $\{a_i^{k_j}\}$ with limit a_i say.

For $\varepsilon > 0$, $\frac{\varepsilon}{2} > 0$ and $\exists N_1$ such that $|a_i^{k_j} - a_i| < \frac{\varepsilon}{2}$ for $j > N_1$. But since $\{a_i^{k_j}\}$ is Cauchy, $\exists N_2$ such that $|a_i^{k_j} - a_i^l| < \frac{\varepsilon}{2}$ for $k_j, l < N_2$. Hence for $N = \max\{N_1, N_2\}$

$$\left|a_i^k - a_i\right| \leqslant \left|a_i^k - a_i^{k_j}\right| + \left|a_i^{k_j} - a_i\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$ □

(31.1.4) Subsets of $\mathbb{R}^n$

The ideas of open and closed intervals in $\mathbb{R}$ may be extended to $\mathbb{R}^n$ as follows.

Definition 13 *For $\mathbf{a} \in \mathbb{R}^n$ and $r > 0$ the open ball of radius r around $\mathbf{a}$ is the set*

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

Definition 14 *A subset N of $\mathbb{R}^n$ is a neighbourhood of a point $\mathbf{a} \in \mathbb{R}^n$ if $\exists r > 0$ such that $B_r(\mathbf{a}) \subseteq N$.*

Definition 15 *A subset U of $\mathbb{R}^n$ is an open set if for all $\mathbf{a} \in U$ there exists $r > 0$ such that $B_r(\mathbf{a}) \subseteq U$.*

It is clear that, for example, an open set is a neighbourhood for all of its points, and that an open set is a union of open balls.

Definition 16 *A subset A of $\mathbb{R}^n$ is sequentially compact if every sequence of points of A has a convergent subsequence that has its limit in A.*

Note that the Bolzano-Weierstrass Theorem (Theorem 11) means that every closed interval of $\mathbb{R}$ is sequentially compact. The following "old chestnut" of a theorem can now be presented: note that the condition is weaker than sequential compactness.

Theorem 17 *A subset A of $\mathbb{R}^n$ is closed if and only if each convergent sequence of elements of A has its limit in A.*

Proof. ($\Rightarrow$) Suppose that $A \subseteq \mathbb{R}^n$ is closed and that $\{\mathbf{a}^k\}$ is a convergent sequence in A that has limit $\mathbf{a}$. Suppose $\mathbf{a} \notin A$ then $\mathbf{a} \in \mathbb{R}^n \setminus A$ which is open and thus $\exists r > 0$ such that $B_r\mathbf{a} \subseteq \mathbb{R}^n \setminus A$. But then $\|\mathbf{a}^k - \mathbf{a}\| \geqslant r$ for all k meaning that for $0 < \varepsilon \leqslant r$ the definition of a limit cannot be satisfied, contradicting that $\{\mathbf{a}^k\}$ has limit $\mathbf{a}$. Hence $\mathbf{a} \in A$.

($\Leftarrow$) Let $\{\mathbf{a}^k\}$ be a sequence in A that has limit $\mathbf{a} \in A$. Suppose A is not closed, then $\mathbb{R}^n \setminus A$ is not open and so there exists a point $\mathbf{b} \in \mathbb{R}^n \setminus A$ such that for all $r > 0$, $B_r(\mathbf{b}) \cap A \neq \emptyset$. Hence for each $k \in \mathbb{N}$ construct the sequence $\{b^k\}$ in A with $\mathbf{b}^k \in B_r(\mathbf{b}) \cap A$. Then $\{\mathbf{b}^k\}$ is a sequence in A that does not have its limit in A, which contradicts the hypothesis. Therefore A is closed. □

Theorem 18 (Heine-Borel) *A subset A of $\mathbb{R}^n$ is sequentially compact if and only if it is closed and bounded.*

Proof. ($\Rightarrow$) Suppose that $A \subseteq \mathbb{R}^n$ is sequentially compact, then every sequence in A has a convergent subsequence which has its limit in A. But if a sequence is convergent then any subsequence is convergent to the same limit and so by Theorem 17 A is closed.

Suppose that A is unbounded then for each $k \in \mathbb{N}$ $\exists \mathbf{a}^k \in A$ with $\|\mathbf{a}\| > k$. But then $\{\mathbf{a}^k\}$ is an unbounded sequence and so can have no bounded subsequence. But by Lemma 10 and Theorem 12 convergent sequences are bounded, therefore unbounded sequences cannot converge. As all subsequences of $\{\mathbf{a}^k\}$ are unbounded, none of them can have a limit, contradicting the sequential compactness of A. Thus A must be bounded.

($\Leftarrow$) Suppose that A is closed and bounded. Since A is bounded $\exists M > 0$ such that for all $\mathbf{x} \in A$, $\|\mathbf{x}\| < M$. If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ then also $|x_i| < M$ for $1 \leqslant i \leqslant M$, so $x_i \in [-M, M]$.

Let $\{\mathbf{a}^k\}$ be a sequence in A with $\mathbf{a}^k = (a_1^k, a_2^k, \ldots, a_n^k)$, then each co-ordinate sequence is a sequence in $[-M, M]$. Hence by Bolzano-Weierstrass (Theorem 11) there exists a subsequence, $\{a_1^{k_j^1}\}$, say, of $\{a_1^k\}$ that is convergent with limit a_1 say.

By Bolzano-Weierstrass, the sequence $\{a_2^{k_j^1}\}$ has a convergent subsequence $\{a_2^{k_j^2}\}$, say, with limit a_2.

By the same reasoning, the sequence $\{a_3^{k_j^2}\}$ has a convergent subsequence $\{a_3^{k_j^3}\}$, say, with limit a_3.

Applying this argument n times gives n convergent sequences $\{a_i^{k_j^n}\}$ $(1 \leqslant i \leqslant n)$. Thus by Theorem 6 $\{\mathbf{a}^{k_j^n}\}$ is a convergent sequence with limit $\mathbf{a} = (a_1, a_2, \ldots, a_n)$. Since A is closed Theorem 17 gives that $\mathbf{a} \in A$ and thus A is sequentially compact. □

(31.1.5) Functions

The logical progression from studying sequences is to study functions of $\mathbb{R}^n$ to $\mathbb{R}^m$. As usual, the limit of a function at a point needs to be independent of the value of the function at that point, if indeed the function is defined there.

Definition 19 *Let f be a function $f\colon A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$. f has limit $\mathbf{l}$ (in $\mathbb{R}^m$) as $\mathbf{x}$ tends to $\mathbf{a}$ if*

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad 0 < \|\mathbf{x} - \mathbf{a}\| < \delta \quad \Rightarrow \quad \|f(\mathbf{x}) - \mathbf{l}\| < \varepsilon$$

Theorem 20 *Let f be a function $f\colon A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$.*

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = \mathbf{l} = (l_1, l_2, \ldots, l_n) \quad \Leftrightarrow \quad \lim_{\mathbf{x}\to\mathbf{a}} f_i(\mathbf{x}) = l_i \quad \forall i \; 1 \leqslant i \leqslant m$$

Proof. ($\Rightarrow$) Suppose $\lim_{\mathbf{a}\to\mathbf{a}} f(\mathbf{x}) = \mathbf{l}$ then

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad 0 < \|\mathbf{x} - \mathbf{a}\| < \delta \quad \Rightarrow \quad \|f(\mathbf{x}) - \mathbf{l}\| < \varepsilon$$

So for $0 < \|\mathbf{x} - \mathbf{a}\| < \delta$,

$$\begin{aligned} \|f_i(\mathbf{x}) - l_i\| &\leqslant \sqrt{\sum_{i=1}^{m} (f_i(\mathbf{x}) - l_i)^2} \\ &= \|f(\mathbf{x}) - \mathbf{l}\| \\ &< \varepsilon \end{aligned}$$

and so $\lim_{\mathbf{x}\to\mathbf{a}} f_i(\mathbf{x}) = l_i$.

($\Leftarrow$) Suppose $\lim_{\mathbf{x}\to\mathbf{a}} f_i(\mathbf{x}) = l_i$ for all i, $1 \leqslant i \leqslant m$. Then for any $\varepsilon > 0$, $\frac{\varepsilon}{\sqrt{m}} > 0$ and thus there exists δ_i $(1 \leqslant i \leqslant m)$ such that $\|f_i(\mathbf{x}) - l_i\| < \frac{\varepsilon}{\sqrt{n}}$. Take $\delta < \min_{1\leqslant i\leqslant m}$ then for $0 < \|\mathbf{x} - \mathbf{a}\| < \delta$,

$$\begin{aligned} \|f(\mathbf{x}) - \mathbf{l}\| &= \sqrt{(f_1(\mathbf{x}) - l_1)^2 + (f_2(\mathbf{x}) - l_2)^2 + \cdots + (f_m(\mathbf{x}) - l_m)^2} \\ &< \sqrt{n\frac{\varepsilon^2}{n}} \\ &= \varepsilon \end{aligned}$$

and so $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = \mathbf{l}$. □

It is clear to see that the usual Algebra of Limits results extend naturally from the case $f\colon \mathbb{R} \to \mathbb{R}$.

Continuity

Definition 21 *A function $f\colon A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$ is continuous at the point $\mathbf{a} \in A$ if*

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$$

f is said to be continuous on A if f is continuous at all points of A.

Once again, the Algebra Of Limits results hold for continuity. As is customary, there follows a list of continuous functions.

The Euclidean Norm function is continuous since by the backwards form of the triangle inequality (Lemma 9) $|\|\mathbf{x}\| - \|\mathbf{a}\|| \leqslant \|\mathbf{x}-\mathbf{a}\| < \varepsilon$ simply by choosing $\delta = \varepsilon$.

Noting that the projection function $g_i\colon \mathbb{R}^n \to \mathbb{R}$ given by $g_i\colon \mathbf{x} \mapsto x_i$ is continuous (using $|g_i(\mathbf{x}) - g_i(\mathbf{a})| = |x_i - a_i| < \|\mathbf{a}-\mathbf{a}\|$) a rational function is simply a sum of products of such functions and constants, and so by the Algebra Of Limits is continuous.

Linear transformations are continuous, as choosing bases the transformation may be represented by a matrix (a_{ij}) say so that

$$f(\mathbf{x}) = A\mathbf{x}^T = \left(\sum_{j=1}^n a_{1j}x_j, \sum_{j=1}^n a_{2j}x_j, \ldots, \sum_{j=1}^n a_{mj}x_m\right)$$

which is a polynomial and so is continuous.

Theorem 22 *Let $A \subseteq \mathbb{R}^n$ and $f\colon A \to \mathbb{R}^m$. Let $B \subseteq \mathbb{R}^m$ with $\operatorname{Im} f \subseteq B \subseteq \mathbb{R}^m$ and $g\colon B \to \mathbb{R}^l$. If f is continuous at $\mathbf{a} \in A$ and g is continuous at $f(\mathbf{a})$ then $g \circ f$ is continuous at $\mathbf{a}$.*

Proof. Take $\varepsilon > 0$ then since g is continuous at $f(\mathbf{a})$, $\exists \delta_1 > 0$ such that

$$\|\mathbf{y} - f(\mathbf{a})\| < \delta_1 \quad \Rightarrow \quad \|g(\mathbf{y}) - g\circ f(\mathbf{a})\| < \varepsilon$$

Also, f is continuous at $\mathbf{a}$, so for any $\delta_1 > 0$ there exists $\delta > 0$ such that

$$\|\mathbf{x} - \mathbf{a}\| < \delta \quad \Rightarrow \quad \|f(\mathbf{x}) - f(\mathbf{a})\| < \delta_1$$

Thus for $\|\mathbf{x}-\mathbf{a}\| < \delta$ putting $\mathbf{y} = f(\mathbf{x})$ gives $\|g\circ f(\mathbf{x}) - g\circ f(\mathbf{a})\| < \varepsilon$ and thus $g \circ f$ is continuous at $\mathbf{a}$. □

Theorem 23 *Let $f\colon A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$ and let $\mathbf{a} \in A$. Then f is continuous at $\mathbf{a}$ if and only if $f^{-1}(N)$ is a neighbourhood of $\mathbf{a}$ whenever N is a neighbourhood of $f(\mathbf{a})$.*

Proof. Note that the theorem is of the form $A \Leftrightarrow (B \Rightarrow C)$.

($\Rightarrow$) Suppose that f is continuous at $\mathbf{a}$ and let N be a neighbourhood of $f(\mathbf{a})$. Hence $\exists\varepsilon > 0$ such that $B_\varepsilon(f(\mathbf{a})) \subseteq N$ and thus by continuity $\exists\delta > 0$ such that $\mathbf{x} \in B_\delta(\mathbf{a})$ implies $f(\mathbf{x}) \in B_\varepsilon(f(\mathbf{a}))$. Therefore

$$B_\delta(\mathbf{a}) \subseteq f^{-1}\left(B_\varepsilon(f(\mathbf{a}))\right) \subseteq f^{-1}(N)$$

meaning that $f^{-1}(N)$ is indeed a neighbourhood of $\mathbf{a}$.

($\Leftarrow$) Suppose that if N is a neighbourhood of $f(\mathbf{a})$ then $f^{-1}(N)$ is a neighbourhood of $\mathbf{a}$. For any $\varepsilon > 0$ the open ball $B_\varepsilon(f(\mathbf{a}))$ is a neighbourhood of $f(\mathbf{a})$ and thus $f^{-1}(B_\varepsilon(f(\mathbf{a})))$ is a neighbourhood of $\mathbf{a}$. As this is a neighbourhood, $\exists \delta > 0$ such that

$$B_\delta(\mathbf{a}) \subseteq f^{-1}\left(B_\varepsilon(f(\mathbf{a}))\right)$$

and thus f is continuous at $\mathbf{a}$. □

Generalising from this, it is easy to show that a function is continuous if and only if its inverse preserves open sets. This leads to the subject of general topology: the study of continuous functions with continuous inverses. Finally for this section, the conditions for continuity can be weakened slightly.

Lemma 24 *If $\{\mathbf{a}^k\}$ is a sequence in $A \subseteq \mathbb{R}^n$ with limit $\mathbf{a}$ and $\mathbf{a}^k \neq \mathbf{a}$ for all $k \in \mathbb{N}$, and if $f\colon A \to \mathbb{R}^m$ then $\{f(\mathbf{a}^k)\}$ is a sequence in $\mathbb{R}^m$ that has limit $f(\mathbf{a})$.*

Proof. As f is continuous

$$\forall \varepsilon > 0\ \exists \delta > 0\ 0 < \|\mathbf{a}^k - \mathbf{a}\| < \delta \ \Rightarrow\ \|f(\mathbf{a}^k) - f(\mathbf{a})\| < \varepsilon$$

But as $\{\mathbf{a}^k\}$ is convergent to $\mathbf{a}$,

$$\forall \delta > 0\ \exists N \in \mathbb{N}\ k > N \ \Rightarrow\ \|\mathbf{a}^k - \mathbf{a}\| < \delta$$

Thus for $k > N$, $\|f(\mathbf{a}^k) - f(\mathbf{a})\| < \varepsilon$ and the result is shown. □

Lemma 25 *If C is a closed and bounded subset of $\mathbb{R}$ then it contains its supremum and infimum.*

Proof. As C is bounded, $M = \sup C$ exists. Suppose that $M \notin C$, then $M \in \mathbb{R} \setminus C$ which is open. Hence $\exists r > 0$ such that $B_r(M) \subseteq \mathbb{R} \setminus C$. As this is a subset of $\mathbb{R}$,

$$B_r(M) = (M - r, M + r)$$

and so $\exists y \in B_r(M)$ with $y < M$ which is again an upper bound for C. But this contradicts that M is the least upper bound, and thus by contradiction $M \in C$. □

Note that the proof that $m \in C$ follows a similar form.

Theorem 26 *If A is a sequentially compact subset of $\mathbb{R}^n$ and $f\colon A \to \mathbb{R}$ is continuous, then f is bounded on A and attains its bounds.*

Proof. Let $\{y^k\}$ be a sequence in $f(A)$ then for each k, $y^k = f(\mathbf{x}^k)$ for some $\mathbf{x}^k \in A$. Thus $\{\mathbf{x}^k\}$ is a sequence in A which is sequentially compact. Hence there is a subsequence, $\{\mathbf{x}^{k_i}\}$ say, that has a limit, say

$$\lim_{i\to\infty} \mathbf{x}^{k_i} = \mathbf{x} \in A$$

But then Lemma 24,

$$\lim_{i\to\infty} f\left(\mathbf{x}^{k_i}\right) = f(\mathbf{x})$$

meaning that $\{y^k\}$ has a convergent subsequence. Thus $f(A)$ is sequentially compact. By Theorem 18 $f(A)$ is closed and bounded and so by Lemma 25 $f(A)$ contains its supremum and infimum, M and m say. But as these are in $f(A)$ there must exist $\mathbf{b}$ and $\mathbf{c}$ in A such that $M = f(\mathbf{b})$ and $m = f(\mathbf{c})$. □

(31.1.6) Contraction Mappings

Definition 27 *Let $A \subset \mathbb{R}^n$ and $f\colon A \to \mathbb{R}^m$. f is a contraction mapping if there exists $K \in \mathbb{R}$ with $0 < K < 1$ such that for all $\mathbf{x}$ and $\mathbf{y}$ in A*

$$\|f(\mathbf{x}) - f(\mathbf{y})\| \leqslant K\|\mathbf{x} - \mathbf{y}\|$$

Definition 28 *If $f\colon A \to A$ where $A \subseteq \mathbb{R}^n$ then $\mathbf{a} \in A$ is a fixed point of f if $f(\mathbf{a}) = \mathbf{a}$.*

Theorem 29 (Contraction Mapping, Banach Fixed Point) *If A is a non-empty closed subset of $\mathbb{R}^n$ and $f\colon A \to A$ is a contraction mapping then f has a unique fixed point.*

Proof. First of all, let $\mathbf{x}_0$ and $\mathbf{x}_1$ be fixed points of f. Then

$$\begin{aligned}
\|\mathbf{x}_0 - \mathbf{x}_1\| &= \|f(\mathbf{x}_0) - f(\mathbf{x}_1)\| \\
&\leqslant K\|\mathbf{x}_0 - \mathbf{x}_1\| \quad \text{for } 0 < K < 1 \\
\text{so } (1-K)\|\mathbf{x}_0 - \mathbf{x}_1\| &\leqslant 0 \\
\text{so } (1-K)\|\mathbf{x}_0 - \mathbf{x}_1\| &= 0 \\
\|\mathbf{x}_0 - \mathbf{x}_1\| &= 0 \\
\mathbf{x}_0 &= \mathbf{x}_1
\end{aligned}$$

Thus if f has a fixed point then it is unique. To show that f has a fixed point consider the sequence $\{f^k(\mathbf{x})\}$ where

$$f^k(\mathbf{x}) = f \circ f \circ \cdots \circ f(\mathbf{x}) = f(f^{k-1}(\mathbf{x}))$$

for some particular (fixed) $\mathbf{x} \in A$. Take any $m > 0$ then

$$\begin{aligned}
\|f^{m+1}(\mathbf{x}) - f^m(\mathbf{x})\| &= \|f(f^m(\mathbf{x})) - f^{m-1}(f^{m-2}(\mathbf{x}))\| \\
&\leqslant K\|f^m(\mathbf{x}) - f^{m-1}(\mathbf{x})\| \\
&\vdots \\
&\leqslant K^m\|f(\mathbf{x}) - \mathbf{x}\|
\end{aligned}$$

Hence for $k > l$,

$$\begin{aligned}
\|f^k(\mathbf{x}) - f^l(\mathbf{x})\| &\leqslant \|f^k(\mathbf{x}) - f^{k-1}(\mathbf{x})\| + \|f^{k-1}(\mathbf{x}) - f^{k-2}(\mathbf{x})\| + \cdots + \|f^{l+1}(\mathbf{x}) - f^l(\mathbf{x})\| \\
&\leqslant (K^{k-1} + K^{k-2} + \cdots + K^l)\|f(\mathbf{x}) - \mathbf{x}\| \\
&\leqslant (K^l + k^{l+1} + \cdots + K^k + \dots)\|f(\mathbf{x}) - \mathbf{x}\| \\
&= \frac{K^l}{1-K}\|f(\mathbf{x}) - \mathbf{x}\|
\end{aligned}$$

If $\mathbf{x}$ is a fixed point there is nothing to show, thus assume it is not. As

$$\frac{1-K}{\|f(\mathbf{x}) - \mathbf{x}\|} > 0$$

since $\lim_{l\to\infty} K^l = 0$, there exists N such that for any $\varepsilon > 0$

$$K^l < \frac{(1-K)\varepsilon}{\|f(\mathbf{x}) - \mathbf{x}\|} \quad \text{for } l > N$$

Hence for $k > l > N$, $\|f^k(\mathbf{x}) - f^l(\mathbf{x})\| < \varepsilon$ and so the sequence $\{f^k(\mathbf{x})\}$ is a Cauchy sequence. By Theorem 12 $\{f^k(\mathbf{x})\}$ has a limit $\mathbf{x}_0$ say. But as A is closed, Theorem 17 shows that $\mathbf{x}_0 \in A$. Since f is a contraction mapping

$$\|f^k(\mathbf{x}) - f(\mathbf{x}_0)\| \leqslant K\|f^{k-1}(\mathbf{x}) - x_0\| \tag{30}$$

but

$$\lim_{k\to\infty} f^{k-1}(\mathbf{x}) = \lim_{k\to\infty} f^k(\mathbf{x}) = \mathbf{x}_0$$

and so from equation (30) $f(\mathbf{x}_0) = \mathbf{x}_0$. □

(31.2) Differentiation

(31.2.1) Partial Derivatives

As usual let $\mathbf{e}_i$ denote the vector of zeros except for a 1 in the ith position.

Definition 31 *Let $f\colon A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$. The ith partial derivative at $\mathbf{a} \in A$ is*

$$D_i(f(\mathbf{a})) = \left(\frac{\partial f}{\partial x_i}\right)_{\mathbf{x}=\mathbf{a}} = \lim_{h\to 0} \frac{f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x})}{h}$$

where $\mathbf{e}_i$ is the ith standard ordered basis vector.

Taking the partial derivative considers f as a function of the single variable x_i which is then differentiated in the conventional single variable way at the point a_i. Taking partial derivatives mixed second order partial derivatives can be calculated. However, in general calculating the same mixed partial by differentiating in different orders need not give the same result. There is, however, a sufficient condition for the second mixed partials to be well-defined in this sense. Recall first the mean value theorem.

Theorem 32 (Mean Value) *If $f\colon [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b) then $\exists c \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{b - a} = f'(c)$$

Writing $b - a = h$ then $c = a + \theta h$ for some $0 < \theta < 1$, and so equivalently $\exists\theta$, $0 < \theta < 1$ such that

$$f(a + h) - f(a) = hf'(a + \theta h)$$

Using this a condition for mixed partials to be "well-defined" can be exhibited.

Theorem 33 *Let $f\colon A \to \mathbb{R}$ and $(x_0, y_0) \in A \subseteq \mathbb{R}^2$. If $\exists r > 0$ such that $D_1 f$, $D_2 f$, $D_1D_2 f$, and $D_2D_1 f$ all exist in $B_r(x_0, y_0)$ and are continuous at (x_0, y_0) then $D_1D_2 f(x_0, y_0) = D_2D_1 f(x_0, y_0)$.*

Proof. Define a real function

$$g(h) = \frac{f(x_0 + h, y_0 + h) - f(x_0 + h, y_0) - f(x_0, y_0 + h) + f(x_0, y_0)}{h^2} \tag{34}$$

where h is sufficiently small so that $x_0 + h$ and $y_0 + h$ all lie within $B_r(x_0, y_0)$. Define also the real function

$$F(x) = f(x, y_0 + h) - f(x, y_0) \quad \text{so} \quad g(h) = \frac{F(x_0 + h) - F(x_0)}{h^2}$$

As Df_1 exists in $B_r(x_0, y_0)$, F is differentiable for $x \in (x_0, x_0 + h)$ with derivative

$$F'(x) = D_1 f(x, y_0 + h) - D_1 f(x, y_0)$$

Hence applying the Mean Value Theorem for F on $[x_0, x_0 + h]$,

$$\exists \theta_h \; 0 < \theta_h < 1 \text{ such that } F(x_0 + h) - F(x_0) = hF'(x_0 + h\theta_h)$$

Thus substituting into equation (34),

$$g(h) = \frac{F'(x_0 + h\theta_h)}{h} = \frac{D_1 f(x_0 + h\theta_h, y_0 + h) - D_1 f(x_0 + h\theta_h, y_0)}{h}$$

Define another real function

$$G(y) = D_1 f(x_0 + h\theta_h, y) \quad \text{so} \quad g(h) = \frac{G(y_0 + h) - G(y_0)}{h}$$

then since $D_2 D_1 f$ exists in $B_r(x_0, y_0)$ G is differentiable on $(y_0, y_0 + h)$ and hence by the Mean Value Theorem

$$\exists \phi_h \; 0 < \phi_h < 1 \text{ such that } G(y_0 + h) - G(y_0) = hG'(y_0 + h\phi_h)$$

Thus

$$g(h) = G'(y_0 + h\phi_h) = D_2 D_1 f(x_0 + h\theta_h, y_0 + h\phi_h) \tag{35}$$

The above argument can be repeated treating first y then x, to yield θ'_h and ϕ'_h, both strictly between 0 and 1 such that

$$g(h) = D_1 D_2 f(x_0 + h\theta'_h, y_0 + h\phi'_h) \tag{36}$$

Now, $D_1 D_2 f$ is continuous at (x_0, y_0) and hence by equation (36)

$$\lim_{h \to 0} g(h) = D_1 D_2 f(x_0, y_0)$$

but using equation (35) and the continuity of $D_2 D_1 f$ at (x_0, y_0),

$$\lim_{h \to 0} g(h) = D_2 D_1 f(x_0, y_0)$$

As limits are unique the result is shown. □

Of course, this can be easily generalised for functions of $\mathbb{R}^n \to \mathbb{R}$ simply by replacing 1 by i and 2 by j.

(31.2.2) Differentiability And Derivatives

For a function $f \colon \mathbb{R} \to \mathbb{R}$ to be differentiable at a point a, it is required that there exists some number, denoted "$f'(a)$", such that

$$\lim_{h \to 0} \frac{f(a + h) - f(a)}{h} = f'(a) \quad \text{or equivalently} \lim_{h \to 0} \frac{f(a + h) - f(a) - hf'(a)}{h} = 0$$

Motivated by this, the following definition is made.

Definition 37 *A function $f \colon A \to \mathbb{R}^m$ is differentiable at a point $\mathbf{a} \in A \subseteq \mathbb{R}^n$ if there exists an $m \times n$ matrix M such that*

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

where $M\mathbf{h}$ is used in place of $(M\mathbf{h}^\top)^\top$. M is called the derivative of f at $\mathbf{a}$, written $Df(\mathbf{a}) = M$.

Writing this out using the definition of a limit,

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad 0 < \|\mathbf{h}\| < \delta \quad \Rightarrow \quad \|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - M\mathbf{h}\| < \|\mathbf{h}\|\varepsilon$$

It is usual to use this form when showing functions to be differentiable, though for showing functions to be not differentiable it is useful to leave the main expression divided by $\|\mathbf{h}\|$.

No study of differentiability would be complete without the following old chestnut.

Theorem 38 *If $f\colon A \to \mathbb{R}^m$ is differentiable at $\mathbf{a} \in A \subseteq \mathbb{R}^n$ then f is continuous at $\mathbf{a}$.*

Proof. Let $Df(\mathbf{a}) = M$ and choose $\varepsilon = 1$ in the definition of differentiability so that for some $\delta_1 > 0$

$$0 < \|\mathbf{h}\| < \delta_1 \;\Rightarrow\; \|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - M\mathbf{h}\| = \|\mathbf{h}\|$$

Letting $\|M\|^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} (M)_{ij}$,

$$\begin{aligned}\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})\| &= \|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - M\mathbf{h} + M\mathbf{h}\| \\ &\leqslant \|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - M\mathbf{h}\| + \|M\mathbf{h}\| \\ &< \|\mathbf{h}\|(1 + \|M\|)\end{aligned}$$

Hence for $0 < \|\mathbf{h}\| < \delta = \min\left(\delta_1, \frac{\varepsilon}{1+\|M\|}\right)$, $\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})\| < \varepsilon$ and so the definition of continuity is satisfied. □

(31.2.3) Functions Into $\mathbb{R}$

For functions into $\mathbb{R}$ partial derivatives can be found, and it is not surprising that under certain conditions—fir sufficiently "nice" functions—the derivative of a function is simply the matrix of partial derivatives.

Definition 39 *If $f\colon A \to \mathbb{R}$ and $\mathbf{a} \in A \subseteq \mathbb{R}^n$ and the partial derivatives $D_1 f(\mathbf{a}), D_2 f(\mathbf{a}), \ldots, D_n f(\mathbf{a})$ all exist, then the gradient of f at $\mathbf{a}$ is*

$$\nabla f(\mathbf{a}) = (D_2 f(\mathbf{a}), \ldots, D_n f(\mathbf{a})) \in \mathbb{R}^n$$

Theorem 40 *Let $f\colon A \to \mathbb{R}^m$ and $\mathbf{a} \in A \subseteq \mathbb{R}^n$. Then*

1. *If f is differentiable at $\mathbf{a}$ then $D_1 f(\mathbf{a}), D_2 f(\mathbf{a}), \ldots, D_n f(\mathbf{a})$ all exist and $Df(\mathbf{a}) = \nabla f(\mathbf{a})$.*
2. *If $D_1 f(\mathbf{a}), D_2 f(\mathbf{a}), \ldots, D_n f(\mathbf{a})$ all exist in an open ball $B_r(\mathbf{a})$ for some $r > 0$ and are continuous at $\mathbf{a}$ then f is differentiable at $\mathbf{a}$.*

Proof. 1. Suppose that f is differentiable at $\mathbf{a}$, then $Df(\mathbf{a})$ exists, and suppose that it is a matrix $M = (M_1, M_2, \ldots, M_n)$ where M_i is an $m \times 1$ matrix. By differentiability, $\forall \varepsilon > 0 \; \exists \delta > 0$ such that

$$0 < \|\mathbf{h}\| < \delta \;\Rightarrow\; \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - M\mathbf{h}|}{\|\mathbf{h}\|} < \varepsilon$$

Choosing $\mathbf{h} = h\mathbf{e}_i$ gives $\|\mathbf{h}\| = |h|$ and $M\mathbf{h} = M_i$ and thus

$$0 < |h| < \delta \;\Rightarrow\; \frac{|f(\mathbf{a}+h\mathbf{e}_i) - f(\mathbf{a}) - M_i h}{|h|} < \varepsilon$$

But this is the expression for the ith partial derivative which therefore must exist and have $M_i = D_i f(\mathbf{a})$, as required.

2. Choose $\mathbf{h}$ with $\|\mathbf{h}\| < r$ and such that all the following terms lie in $B_r(\mathbf{a})$: Observe that

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = f(&a_1+h_1, a_2+h_2, \ldots, a_n+h_n) \\ &- f(a_1, a_2+h_2, a_3+h_3, \ldots, a_n+h_n) + f(a_1, a_2+h_2, a_3+h_3, \ldots, a_n+h_n) \\ &- f(a_1, a_2, a_3+h_3, \ldots, a_n+h_n) + f(a_1, a_2, a_3+h_3, \ldots, a_n+h_n) \\ &\vdots \\ &- f(a_1, a_2, \ldots, a_{n-1}, a_n+h_n) + f(a_1, a_2, \ldots, a_{n-1}, a_n+h_n) \\ &- f(a_1, a_2, \ldots, a_n) \end{aligned} \tag{41}$$

where the function $f(a_1, \ldots, a_i x, \ldots, a_n + h_n)$ (of the single variable x) is differentiable for $x \in (0, h_i)$ since $D_i f$ exists there. Hence by the Mean Value Theorem $\exists \theta_i \in (0,1)$ such that

$$f(a_1, \ldots, a_i + h_i, \ldots, a_n + h_n) - f(a_1, a_2, \ldots, a_n) = h_i D f(a_1, \ldots, a_i + \theta_i h_i, \ldots, a_n + h_n)$$

and thus substituting into equation (41)

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \sum_{i=1}^{n} h_i D f(a_1, \ldots, a_i + \theta_i h_i, \ldots, a_n + h_n)$$

But $(\nabla f(\mathbf{a}))\mathbf{h} = D_1 f(\mathbf{a}) h_1 + \cdots + D_n f(\mathbf{a}) h_n$ and so

$$\begin{aligned} |f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - (\nabla f(\mathbf{a}))\mathbf{h}| &= \left| \sum_{i=1}^{n} h_i D_i f(a_1, \ldots, a_i + \theta_i h_i, \ldots, a_n + h_n) - h_i D_i f(a_1, a_2, \ldots, a_n) \right| \\ &\leqslant \sum_{i=1}^{n} |h_i| \, |D_i f(a_1, \ldots, a_i + \theta_i h_i, \ldots, a_n + h_n) - D_i f(a_1, a_2, \ldots, a_n)| \\ &\leqslant \|\mathbf{h}\| \sum_{i=1}^{n} |D_i f(a_1, \ldots, a_i + \theta_i h_i, \ldots, a_n + h_n) - D_i f(a_1, a_2, \ldots, a_n)| \end{aligned}$$

But each $D_i f$ is continuous at $\mathbf{a}$ therefore for any $\varepsilon > 0$, $\frac{\varepsilon}{n} > 0$ and $\exists \delta_i > 0$ such that

$$|D_i f(a_1, \ldots, a_i + \theta_i h_i, \ldots, a_n + h_n) - D_i f(a_1, a_2, \ldots, a_n)| < \frac{\varepsilon}{n}$$

Hence taking $\delta < \min_i \delta_i$

$$|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - (\nabla f(\mathbf{a}))\mathbf{h}| < \mathbf{h} \sum_{i=1}^{n} \frac{\varepsilon}{n} = \|\mathbf{h}\|\varepsilon$$

and thus f is differentiable at $\mathbf{a}$ with derivative $\nabla f(\mathbf{a})$. □

Definition 42 *A function $f\colon A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$ is of class C^1 if all the first order partial derivatives exist and are continuous.*

By Theorem 40 functions of class C^1 are differentiable. However, the converse need not be the case as when a function is differentiable Theorem 40 guarantees only that the first order partial derivatives exist, but not that they are continuous. An example of this is the function

$$f\colon \mathbb{R}^2 \to \mathbb{R} \quad \text{defined by} \quad f\colon (x,y) \mapsto \begin{cases} (x^2+y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

which is differentiable at $(0, 0)$ where also the first order partial derivatives exist but are not continuous.

(31.2.4) Functions Into $\mathbb{R}^m$

Theorem 43 *Let $f\colon A \to \mathbb{R}^m$ and $\mathbf{a} \in A \subseteq \mathbb{R}^n$. f is differentiable at $\mathbf{a}$ if and only if each co-ordinate function $f_i\colon A \to \mathbb{R}$ is differentiable at $\mathbf{a}$. In this case*

$$Df(\mathbf{a}) = \begin{pmatrix} \nabla f_1(\mathbf{a}) \\ \nabla f_2(\mathbf{a}) \\ \vdots \\ \nabla f_m(\mathbf{a}) \end{pmatrix} = \begin{pmatrix} D_1 f_1(\mathbf{a}) & D_2 f_1(\mathbf{a}) & \dots & D_n f_1(\mathbf{a}) \\ D_1 f_2(\mathbf{a}) & D_2 f_2(\mathbf{a}) & \dots & D_n f_2(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f_m(\mathbf{a}) & D_2 f_m(\mathbf{a}) & \dots & D_n f_m(\mathbf{a}) \end{pmatrix}$$

Proof. ($\Rightarrow$) Suppose that f is differentiable with derivative $M = (a_{ij})$, an $m \times n$ matrix. Write

$$M = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_m \end{pmatrix} \quad \text{with} \quad M_i = (a_{i1}, a_{i2}, \dots, a_{in})$$

Since f is differentiable at $\mathbf{a}$, for $\mathbf{h} \in \mathbb{R}^n$, $\forall \varepsilon > 0$, $\exists \delta > 0$ such that for $0 < \|\mathbf{h}\| < \delta$

$$\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - M\mathbf{h}\| < \varepsilon\|\mathbf{h}\|$$

Hence

$$\begin{aligned} |f_i(\mathbf{a}+\mathbf{h}) - f_i(\mathbf{a}) - M_i\mathbf{h}| &\leqslant \sqrt{\sum_{j=1}^{m} \left(f_j(\mathbf{a}+\mathbf{h}) - f_j(\mathbf{a}) - M_j\mathbf{h}\right)^2} \\ &= \|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - M\mathbf{h}\| \\ &< \varepsilon\|\mathbf{h}\| \end{aligned}$$

and thus f_i is differentiable at $\mathbf{a}$ with derivative M_i. By Theorem 40

$$M_i = \nabla f_i(\mathbf{a}) = (D_1 f_i(\mathbf{a}), D_2 f_i(\mathbf{a}), \dots, D_n f_i(\mathbf{a}))$$

($\Leftarrow$) Suppose that each of the co-ordinate functions is differentiable at $\mathbf{a}$, then by Theorem 40 $Df_i(\mathbf{a}) = \nabla f_i(\mathbf{a})$. For any $\varepsilon > 0$, $\frac{\varepsilon}{\sqrt{m}} > 0$ so for each i, $\exists \delta_i > 0$ such that for $0 < \|\mathbf{h}\| < \delta_i$,

$$|f_i(\mathbf{a}+\mathbf{h}) - f_i(\mathbf{a}) - \nabla f_i(\mathbf{a})\mathbf{h}| < \frac{\varepsilon}{\sqrt{n}}$$

Hence taking $\delta < \min_{1 \leqslant i \leqslant n} \delta_i$,

$$\begin{aligned} \|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - M\mathbf{h}\| &= \sqrt{\sum_{i=1}^{m} (f_i(\mathbf{a}+\mathbf{h}) - f_i(\mathbf{a}) - \nabla f_i(\mathbf{a})\mathbf{h})^2} \\ &< \sqrt{\sum_{i=1}^{m} \left(\frac{\varepsilon}{\sqrt{m}}\right)^2 \|\mathbf{h}\|^2} = \varepsilon\|\mathbf{h}\| \end{aligned}$$

and hence f is differentiable at $\mathbf{a}$ with derivative $M = (\nabla f_i(\mathbf{a}))$. □

Definition 44 *For $f\colon A \to \mathbb{R}^n$ where $A \subseteq \mathbb{R}^m$, if the partial derivatives $D_j f_i(\mathbf{a})$ exist then the Jacobian matrix of f*

at **a** *is*

$$Jf(\mathbf{a}) = \begin{pmatrix} D_1 f_1(\mathbf{a}) & D_2 f_1(\mathbf{a}) & \dots & D_n f_1(\mathbf{a}) \\ D_1 f_2(\mathbf{a}) & D_2 f_2(\mathbf{a}) & \dots & D_n f_2(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f_m(\mathbf{a}) & D_2 f_m(\mathbf{a}) & \dots & D_n f_m(\mathbf{a}) \end{pmatrix}$$

Theorem 45 *If $f\colon A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^m$, f is of class C^1 then f is differentiable at all $\mathbf{a} \in A$, and $Df(\mathbf{a}) = Jf(\mathbf{a})$.*

Proof. Immediate from Theorem 40 and Theorem 43. □

Theorem 46 (Chain Rule) *Let $f\colon A \to \mathbb{R}^m$ with $A \subseteq \mathbb{R}^n$. Let $g\colon B \to \mathbb{R}^l$ with $f(A) \subseteq B \subseteq \mathbb{R}^m$. Then if f is differentiable at $\mathbf{a} \in A$ and g is differentiable at $f(\mathbf{a}) = \mathbf{b} \in B$ the the composite function $g \circ f\colon A \to \mathbb{R}^l$ is differentiable at $\mathbf{A}$ with derivative $D(g \circ f)(\mathbf{a}) = Dg(\mathbf{b})Df(\mathbf{a})$.*

Note that $Df(\mathbf{a})$ is an $m \times n$ matrix, $Dg(\mathbf{b})$ is an $l \times m$ matrix, and $D(g \circ f)(\mathbf{a})$ is an $l \times n$ matrix.

Proof. It is required to show that $\forall \varepsilon > 0\ \exists \delta >$) such that

$$0 < \|\mathbf{h}\| < \delta \ \Rightarrow\ \|g \circ f(\mathbf{a}+\mathbf{h}) - g \circ f(\mathbf{a}) - Dg(\mathbf{b})Df(\mathbf{a})\mathbf{h}\| < \varepsilon\|\mathbf{h}\|$$

Now,

$$\begin{aligned} &\|g \circ f(\mathbf{a}+\mathbf{h}) - g \circ f(\mathbf{a}) - Dg(\mathbf{b})Df(\mathbf{a})\mathbf{h}\| \\ &= \|g \circ f(\mathbf{a}+\mathbf{h}) - g \circ f(\mathbf{a}) - Dg(\mathbf{b})(f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})) \\ &\quad + Dg(\mathbf{b})(f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})) - Dg(\mathbf{b})Df(\mathbf{a})\mathbf{h}\| \\ &\leqslant \|g \circ f(\mathbf{a}+\mathbf{h}) - g \circ f(\mathbf{a}) - Dg(\mathbf{b})(f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}))\| + N\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - Df(\mathbf{a})\mathbf{h}\| \end{aligned} \tag{47}$$

where $N = \|Dg(\mathbf{b})\|$.

Now, for $\varepsilon > 0$ and $N > 0$, $\frac{\varepsilon}{2N} > 0$ and so by the differentiability of f, $\exists \delta_1 > 0$ such that

$$0 < \|\mathbf{h}\| < \delta_1 \ \Rightarrow\ N\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - Df(\mathbf{a})\mathbf{h}\| < \frac{\varepsilon}{2}\|\mathbf{h}\| \tag{48}$$

Moreover, if $N = 0$ then $N\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - Df(\mathbf{a})\mathbf{h}\| < \frac{\varepsilon}{2}\|\mathbf{h}\|$ for any δ_1.

Let $\mathbf{k} = f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) \in \mathbb{R}^m$ and $M = \|Df(\mathbf{a})\|$. As f is differentiable at $\mathbf{a}$, choose $\varepsilon = 1$ in the definition of differentiability, then $\exists \delta_2$ such that for $0 < \|\mathbf{h}\| < \delta_2$

$$\begin{aligned} \|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})\| &= \|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) \ - Df(\mathbf{a})\mathbf{h} + Df(\mathbf{a})\mathbf{h}\| \\ &\leqslant \|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - Df(\mathbf{a})\mathbf{h}\| + M\|\mathbf{h}\| \\ &< (1+M)\|\mathbf{h}\| \end{aligned}$$

i.e., $\|\mathbf{k}\| < (1+M)\|\mathbf{h}\|$. Now by the differentiability of g at $\mathbf{b}$, $\frac{\varepsilon}{2(1+M)} > 0$ and $\exists \delta_3 > 0$ such that

$$0 < \|\mathbf{k}\| < \delta_3 \ \Rightarrow\ \|g(\mathbf{b}+\mathbf{k}) - g(\mathbf{b}) - Dg(\mathbf{b})\mathbf{k}\| < \frac{\varepsilon}{2(1+M)}\|\mathbf{k}\| \tag{49}$$

Hence for

$$\delta < \min\left\{\delta_1, \delta_2, \frac{\delta_3}{1+M}\right\}$$

when $0 < \|\mathbf{h}\| < \delta$, in equation 47

$$\begin{aligned}
&\|g \circ f(\mathbf{a}+\mathbf{h}) - g \circ f(\mathbf{a}) - Dg(\mathbf{b})Df(\mathbf{a})\mathbf{h}\| \\
&\leqslant \|g \circ f(\mathbf{a}+\mathbf{h}) - g \circ f(\mathbf{a}) - Dg(\mathbf{b})(f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}))\| + N\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - Df(\mathbf{a})\mathbf{h}\| \\
&< \frac{\varepsilon}{2(M+1)}\|\mathbf{k}\| + \frac{\varepsilon}{2}\|\mathbf{h}\| \quad \text{by euqation (48) and equation (49)} \\
&= \varepsilon\|\mathbf{h}\|
\end{aligned}$$

□

(31.3) Generalising Analytical Results

(31.3.1) The Mean Value Theorem

Recall that if $f: [a,b] \to \mathbb{R}$ is differentiable, then the Mean Value Theorem states that $\exists c \in (a,b)$ such that

$$\frac{f(b)-f(a)}{b-a} = f'(c)$$

If this holds, then $|f(b) - f(a)| = |f'(c)||b-a|$, and so if $|f'(c)| \leqslant M$ for all $x \in [a,b]$ then $|f(b) - f(a)| < M|b-a|$ for all $x \in [a,b]$. This is in the form of Theorem 29 (Contraction Mapping). It a result of this form that can be generalised.

Definition 50 *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *define*

$$[\mathbf{x}, \mathbf{y}] = \{(1-t)\mathbf{x} + t\mathbf{y} \mid 0 \leqslant t \leqslant 1\}$$

Theorem 51 (Generalised Mean Value) *Let* $f: A \to \mathbb{R}^m$ *be a differentiable function with* $A \subseteq \mathbb{R}^n$. *If* $\mathbf{a}, \mathbf{b} \in A$ *with* $[\mathbf{a}, \mathbf{b}] \subseteq A$ *and* $\exists M$ *such that* $\|Df(\mathbf{x})\| \leqslant M \; \forall \mathbf{x} \in [\mathbf{a}, \mathbf{b}]$ *then* $\|f(\mathbf{b}) - f(\mathbf{a})\| \leqslant M\|\mathbf{b} - \mathbf{a}\|$.

Proof. If $f(\mathbf{a}) = f(\mathbf{b})$ then there is nothing to show. Suppose therefore that this is not so, and define

$$\mathbf{u} = \frac{f(\mathbf{b}) - f(\mathbf{a})}{\|\mathbf{b} - \mathbf{a}\|}$$

Define

$$F: [0,1] \to \mathbb{R} \quad \text{by} \quad F: t \mapsto \mathbf{u} \cdot f(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$$

where the dot denotes the inner product on $\mathbb{R}^m$. Using the standard inner product,

$$F(t) = \sum_{i=1}^{m} u_i f_i(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) \quad \text{where} f = (f_1, f_2, \ldots, f_m) \text{ and } u_i = \frac{f_i(\mathbf{b}) - f_i(\mathbf{a})}{\|\mathbf{b} - \mathbf{a}\|}$$

Hence F may be written as a composition $F = h \circ f \circ g$ where

$$\begin{aligned}
&g: [0,1] \to \mathbb{R}^n \qquad && f: \mathbb{R}^n \to \mathbb{R}^m \qquad && h : \mathbb{R}^m \to \mathbb{R} \\
&g: t \mapsto \mathbf{a} + t(\mathbf{b} - \mathbf{a}) && f: \mathbf{x} \mapsto f(\mathbf{x}) && h: \mathbf{x} \mapsto \mathbf{u} \cdot \mathbf{x}
\end{aligned}$$

Now, g is differentiable with derivative $Dg(t) = \mathbf{b} - \mathbf{a}$, and f is differentiable by hypothesis. h is also differentiable, as follows

$$\begin{aligned}
h(\mathbf{x}) &= u_1 x_1 + u_2 x_2 + \cdots + u_m x_m \\
Dh(\mathbf{x}) &= (D_1 h(\mathbf{x}), D_2 h(\mathbf{x}), \ldots, D_m h(\mathbf{x})) \\
&= (u_1, u_2, \ldots, u_m) = \mathbf{u}
\end{aligned}$$

Hence by Theorem 46 (the Chain Rule) F is differentiable with derivative

$$\begin{aligned} DF(t) = F'(t) &= Dh(f \circ g(t))Df(g(t))Dg(t) \\ &= \mathbf{u}Df(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))(\mathbf{b} - \mathbf{a}) \end{aligned}$$

Now applying the Mean Value Theorem for F on $[0,1]$, $\exists c \in [0,1]$ such that

$$\begin{aligned} \frac{F(1) - F(0)}{1 - 0} &= F'(c) \\ F(1) - F(0) &= F'(c) \\ \mathbf{u} \cdot f(\mathbf{b}) - \mathbf{u} \cdot f(\mathbf{a}) &= \mathbf{u}Df(\mathbf{a} + c(\mathbf{b} - \mathbf{a}))(\mathbf{b} - \mathbf{a}) && (52) \\ \text{so } |\mathbf{u} \cdot f(\mathbf{b}) - \mathbf{u} \cdot f(\mathbf{a})| &= |\mathbf{u}Df(\mathbf{a} + c(\mathbf{b} - \mathbf{a}))(\mathbf{b} - \mathbf{a})| \\ &\leqslant \|\mathbf{u}\| \|Df(\mathbf{a} + c(\mathbf{b} - \mathbf{a}))\| \|\mathbf{b} - \mathbf{a}\| \\ &= \|Df(\mathbf{a} + c(\mathbf{b} - \mathbf{a}))\| \|\mathbf{b} - \mathbf{a}\| && (53) \end{aligned}$$

with the last line following because

$$\|\mathbf{u}\| = \| \frac{f(\mathbf{b}) - f(\mathbf{a})}{\|f(\mathbf{b}) - f(\mathbf{a})\|} \| = \frac{\|f(\mathbf{b}) - f(\mathbf{a})\|}{\|f(\mathbf{b}) - f(\mathbf{a})\|} = 1$$

Returning to equation (52), the left hand side is

$$\begin{aligned} \mathbf{u} \cdot f(\mathbf{b}) - \mathbf{u} \cdot f(\mathbf{a}) &= \mathbf{u} \cdot (f(\mathbf{b}) - f(\mathbf{a})) \\ &= \sum_{i=1}^{m} u_i \left(f_i(\mathbf{b}) - f_i(\mathbf{a})\right) \\ &= \sum_{i=1}^{m} \frac{f_i(\mathbf{b}) - f_i(\mathbf{a})}{\|f(\mathbf{b}) - f(\mathbf{a})\|} \left(f_i(\mathbf{b}) - f_i(\mathbf{a})\right) \\ &= \frac{1}{\|f(\mathbf{b}) - f(\mathbf{a})\|} \sum_{i=1}^{m} \left(f_i(\mathbf{b}) - f_i(\mathbf{a})\right)^2 \\ &= \|f(\mathbf{b}) - f(\mathbf{a})\| \end{aligned}$$

By hypothesis $\|Df(\mathbf{x})\| \leqslant M$ for all $\mathbf{x} \in [\mathbf{a}, \mathbf{b}]$ and so equation (53) gives

$$\|f(\mathbf{b}) - f(\mathbf{a})\| \leqslant M\|\mathbf{b} - \mathbf{a}\|$$

as required. □

A more direct generalisation of the Mean Value Theorem is available when $f\colon A \to \mathbb{R}$.

Corollary 54 *If $f\colon A \to \mathbb{R}$ is differentiable with $[\mathbf{a}, \mathbf{b}] \subseteq A \subseteq \mathbb{R}^n$ then $\exists \mathbf{c} \in (\mathbf{a}, \mathbf{b})$ such that*

$$f(\mathbf{b}) - f(\mathbf{a}) = Df(\mathbf{c})(\mathbf{b} - \mathbf{a})$$

Proof. Put $\mathbf{u} = (1)$ (which can be done since F is still a function into $\mathbb{R}$) then equation (52) gives

$$f(\mathbf{b} - f(\mathbf{a}) = Df(\mathbf{a} + c(\mathbf{b} - \mathbf{a}))(\mathbf{b} - \mathbf{a})$$

Hence the result, where $\mathbf{c} = \mathbf{a} + c(\mathbf{b} - \mathbf{a})$. □

The Mean Value Theorem cannot be generalised in this way for functions into $\mathbb{R}^m$ where $m \geqslant 2$. Take for example the function $f(x) = (\cos x, \sin x)$. Then $Df(x) = (-\sin x, \cos x)^\top$ and so $\|Df(x)\| = 1$. However, on the interval $[0, 2\pi]$ say,

$$f(2\pi) - f(0) = (0,0) \neq \|Df(x)\|(2\pi - 0) = 2\pi$$

Definition 55 *A non-empty subset A of $\mathbb{R}^n$ is called convex when if $\mathbf{x}, \mathbf{y} \in A$ then $[\mathbf{x}, \mathbf{y}] \subseteq A$.*

Corollary 56 *If $f\colon A \to \mathbb{R}^m$ is differentiable on A which is a convex subset of $\mathbb{R}^n$, and $\|Df(\mathbf{x})\| \leqslant M$ for all $\mathbf{x} \in A$ then for any $\mathbf{a}, \mathbf{b} \in A$, $\|f(\mathbf{b}) - f(\mathbf{a})\| \leqslant M\|\mathbf{b} - \mathbf{a}\|$.*

Proof. Simply apply Theorem 51 to $\mathbf{f}$ on $[\mathbf{a}, \mathbf{b}]$. □

Note that if $M < 1$ then this corollary gives the conditions to apply Theorem 29 (Contraction Mapping Theorem).

(31.3.2) Taylor's Theorem

Definition 57 *A function $f\colon A \to \mathbb{R}$ where $A \subseteq \mathbb{R}^n$ is said to be of class C^r if all the partial derivatives up to and including those of order r exist and are continuous.*

Definition 58 *Define the operator $\mathbf{x} \cdot \nabla$ to act on a function $g\colon A \to \mathbb{R}$ where $A \subseteq \mathbb{R}^n$ as follows:*

$$(\mathbf{x} \cdot \nabla)g(\mathbf{b}) = \sum_{i=1}^{n} x_i D_i g(\mathbf{b})$$

This is the inner product of $\mathbf{x}$ with the gradient of g at $\mathbf{b}$.

Of course, this operator may be applied more than once. For example

$$\begin{aligned}(\mathbf{x} \cdot \nabla)^2 f(\mathbf{a} + t\mathbf{x}) &= (\mathbf{x} \cdot \nabla) \sum_{i=1}^{n} x_i D_i f(\mathbf{a} + t\mathbf{x}) \\ &= \sum_{i=1}^{n} x_i (\mathbf{x} \cdot \nabla) D_i f(\mathbf{a} + t\mathbf{x}) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j D_j D_i f(\mathbf{a} + t\mathbf{a})\end{aligned}$$

If f is of class C^r then $\mathbf{x} \cdot \nabla$ may be applied up to r times.

Lemma 59 *Let $f\colon A \to \mathbb{R}$ where $A \subseteq \mathbb{R}^n$ is of class C^r and define*

$$F\colon [0,1] \to \mathbb{R}^m \quad by \quad F\colon t \mapsto f(\mathbf{a} + t\mathbf{x})$$

Then F is r times differentiable with $D^s F(t) = (\mathbf{x} \cdot \nabla) f(\mathbf{a} + t\mathbf{x})$.

Note that F is defined on A whenever t is small enough: Since f is differentiable at all $\mathbf{a} \in A$ each $\mathbf{a}$ must have an open ball around it that is contained in A. Hence taking t small enough ensures that F is properly defined on A.

Proof. Since f is differentiable on A,

$$Df(\mathbf{a} + t\mathbf{x}) = \nabla f(\mathbf{a} + t\mathbf{x}) = (D_1 f(\mathbf{a} + t\mathbf{x}), D_2 f(\mathbf{a} + t\mathbf{x}), \ldots, D_n f(\mathbf{a} + t\mathbf{x}))$$

For $k \in \mathbb{R}$,

$$\lim_{k\to 0}\left|\frac{F(t+k)-F(t)}{k} - \mathbf{x}Df(\mathbf{a}+t\mathbf{x})\right| = \lim_{k\to 0}\left|\frac{F(t+k)-F(t)-Df(\mathbf{a}+t\mathbf{x})k\mathbf{x}}{k}\right|$$

Now write $\mathbf{h} = k\mathbf{x}$ so $\mathbf{h} \to \mathbf{0}$ as $k \to 0$ and $\|\mathbf{h}\| = |k|\|\mathbf{x}\|$ to give

$$\begin{aligned} &= \|\mathbf{x}\| \lim_{\mathbf{h}\to\mathbf{0}} \left|\frac{f(\mathbf{a}+t\mathbf{x}+k\mathbf{x}) - f(\mathbf{a}+t\mathbf{x}) - Df(\mathbf{a}+t\mathbf{x})}{\|\mathbf{h}\|}\right| \\ &= 0 \text{ by the differentiability of } f. \end{aligned}$$

Hence

$$\begin{aligned} DF(t) = F'(t) &= Df(\mathbf{a}+t\mathbf{x})\mathbf{x} \\ &= \sum_{i=1}^{n} x_i D_i f(\mathbf{a}+t\mathbf{x}) \\ &= (\mathbf{x}\cdot\nabla) f(\mathbf{a}+t\mathbf{x}) \end{aligned}$$

Now, each partial derivative of f is also a function $A \to \mathbb{R}$, and so replacing f by $D_i f$ in the above argument shows that

$$D^{i+1}F(t) = (\mathbf{x}\cdot\nabla)^{i+1} f(\mathbf{a}+t\mathbf{x})$$

for $2 \leqslant i \leqslant r-1$. □

Before stating and proving a generalised form of Taylor's Theorem, recall the single variable case.

Theorem 60 (Taylor) *Let $a, x \in \mathbb{R}$ be fixed and let $f\colon [a, a+x] \to \mathbb{R}$ be continuously differentiable $r-1$ times, and suppose that the rth derivative exists. Then*

$$f(a+x) = f(a) + xf'(a) + \frac{x^2}{2!}f''(a) + \cdots + \frac{x^{r-1}}{(r-1)!}f^{(r-1)}(a) + \frac{x^r}{r!}f^{(r)}(a+\theta x)$$

for some $0 < \theta < 1$.

Theorem 61 (Generalised Taylor) *Let $F\colon A \to \mathbb{R}$ be of class C^r with $\mathbf{a} \in A \subseteq \mathbb{R}^n$ and let $\mathbf{x}$ be some fixed point of $\mathbb{R}^n$. Then if the line segment $[\mathbf{a}, \mathbf{a}+\mathbf{x}] \subseteq A$ then there exists $\theta \in (0,1)$ such that*

$$f(\mathbf{a}+\mathbf{x}) = f(\mathbf{a}) + (\mathbf{x}\cdot\nabla)f(\mathbf{a}) + \frac{(\mathbf{x}\cdot\nabla)^2}{2!}f(\mathbf{a}) + \cdots + \frac{(\mathbf{x}\cdot\nabla)^{r-1}}{(r-1)!}f(\mathbf{a}) + \frac{(\mathbf{x}\cdot\nabla)^r}{r!}f(\mathbf{a}+theta\mathbf{x})$$

Proof. Define $F\colon [0,1] \to \mathbb{R}$ by $F(t) = f(\mathbf{a}+t\mathbf{x})$ then since f is of class C^r so is F. Hence using Lemma 59 and Theorem 60

$$\begin{aligned} F(1) &= F(0) + DF(0) + \frac{D^2F(0)}{2!} + \cdots + \frac{D^{r-1}F(0)}{(r-1)!} + \frac{D^R F(\theta)}{r!} \quad \text{for some } \theta \in (0,1) \\ \textit{i.e., } f(\mathbf{a}+\mathbf{x}) &= f(\mathbf{a}) + (\mathbf{x}\cdot\nabla)f(\mathbf{a}) + \frac{(\mathbf{x}\cdot\nabla)^2 f(\mathbf{a})}{2!} + \cdots + \frac{(\mathbf{x}\cdot\nabla)^{r-1}f(\mathbf{a})}{(r-1)!} + \frac{(\mathbf{x}\cdot\nabla)^r f(\mathbf{a}+\theta\mathbf{x}}{r!} \end{aligned}$$

□

(31.3.3) Other Generalised Results

Definition 62 *A C^1 function $f\colon A \to \mathbb{R}^n$ where $A \subseteq \mathbb{R}^n$ is locally C^1-invertible at $\mathbf{a} \in A$ if there exist open sets U_1 and U_2 with $\mathbf{a} \in U_1$ and a C^1 function $g\colon U_2 \to U_1$ such that $g \circ f(\mathbf{x}) = \mathbf{x}$ and $f \circ g(\mathbf{y}) = \mathbf{y}$ for all $\mathbf{x} \in U_1$ and*

$\mathbf{y} \in U_2$.

Theorem 63 (Inverse Function) *Let $f\colon A \to \mathbb{R}^n$ be of class C^1 and $f(\mathbf{a}) = \mathbf{b}$ where $\mathbf{a} \in A \subseteq \mathbb{R}^n$. If $Df(\mathbf{a}) = M$ is non-singular then f is locally C^1-invertible at $\mathbf{a}$ with $Df^1(\mathbf{b}) = M^{-1}$.*

Of particular interest is the ability generalise the definition of a norm while retaining most of the results already proven. A norm can be defined on any real vector space, including ones of infinite dimension.

Definition 64 *Let V be a real vector space, then a norm on V is a function $\|\cdot\|\colon V \to \mathbb{R}$ such that for any $\mathbf{v}, \mathbf{w} \in V$ and $\lambda \in \mathbb{R}$,*

N1. $\|\mathbf{v}\| \geqslant 0$ *with equality if and only if* $\mathbf{v} = \mathbf{0}$.

N2. $\|\lambda\mathbf{v}\| = |\lambda|\|\mathbf{v}\|$.

N3. $\|\mathbf{v} + \mathbf{w}\| \leqslant \|\mathbf{v}\| + \|\mathbf{w}\|$.

Example 65 *The uniform norm on $\mathbb{R}^n$ is defined by*

$$\|\mathbf{x}\|_\infty = \max_{i \leqslant i \leqslant n} |x_i|$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Show that this is indeed a norm.

Proof. Solution Verifying the axioms in turn,

N1. Certainly $|x_i| \geqslant 0$ for all $1 \leqslant i \leqslant n$ and hence $\|\mathbf{x}\|_\infty \geqslant 0$. Also, it is clear that $\|\mathbf{x}\|_\infty = 0 \Leftrightarrow x_i = 0$ for $1 \leqslant i \leqslant n$, *i.e.*, $\mathbf{x} = \mathbf{0}$.

N2. Say $\|\mathbf{x}\|_\infty = x_j$ then

$$\|\lambda\mathbf{x}\|_\infty = \max_{i \leqslant i \leqslant n} |\lambda x_i| = |\lambda| \max_{i \leqslant i \leqslant n} |x_i| = |\lambda||x_j| = \lambda\|\mathbf{x}\|_\infty$$

N3. Suppose $\|\mathbf{x}\|_\infty = x_j$ and $\|\mathbf{y}\|_\infty = y_k$. Then

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|_\infty &= \max_{1 \leqslant i \leqslant n} |x_i + y_i| \\ &= |x_l + y_l| \text{ say} \\ &\leqslant |x_j| + |y_k| \\ &= \|\mathbf{x}\|_\infty + \|\mathbf{y}\|_\infty \end{aligned}$$

□

The uniform norm can be generalised slightly to work on the infinite dimensional vector space of continuous functions of $[0,1] \to \mathbb{R}$, $C[0,1]$. Define

$$\|f\|_\infty = \sup_{x \in [0,1]} f(x)$$

An alternative norm for $C[0,1]$ is the L_1-norm,

$$\|f\|_1 = \int_0^1 |f(x)|\, \mathrm{d}x$$

Definition 66 *Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a vector space V are equivalent if $\exists m, M \in \mathbb{R}^+$ such that*

$$m\|\mathbf{v}\|_1 \leqslant \|\mathbf{v}\|_2 \leqslant M\|\mathbf{v}\|_1 \quad \forall \mathbf{v} \in V$$

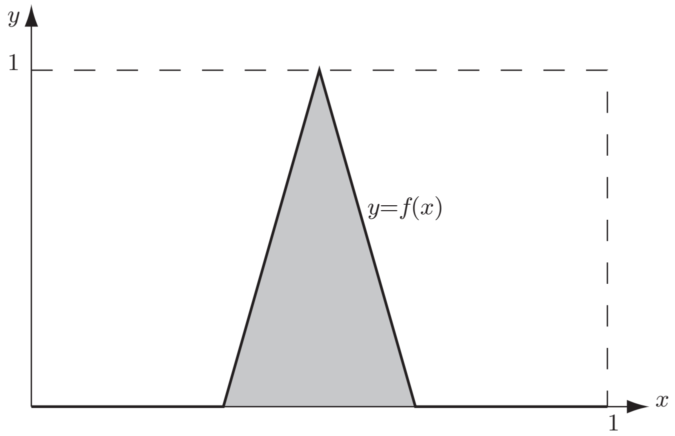

Figure 1: The L_1 norm and the uniform norm are not equivalent on infinite dimensional vector spaces.

On $\mathbb{R}^n$ then Euclidean norm and uniform norms are equivalent. Observe that

$$\|\mathbf{x}\|_\infty = \max_{1\leqslant i\leqslant n} |x_i| \leqslant \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \|\mathbf{x}\|$$

and if $\|\mathbf{x}\|_\infty = |x_j|$ then

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \leqslant \sqrt{nx_j^2} = \sqrt{n}\|\mathbf{x}\|_\infty$$

Hence $\|\mathbf{x}\|_\infty \leqslant \|\mathbf{x}\| \leqslant \sqrt{n}\|\mathbf{x}\|_\infty$ so the Euclidean and uniform norms are indeed equivalent.

Theorem 67 *All norms on $\mathbb{R}^n$ are equivalent.*

However, the same result need not hold for infinite dimensional vector spaces. On $C[0,1]$,

$$\|f\|_1 = \int_0^1 |f(x)|\,\mathrm{d}x \leqslant \int_0^1 \sup_{x\in[0,1]} |f(x)|\,\mathrm{d}x \leqslant \sup_{x\in[0,1]} |f(x)| = \|f\|_\infty$$

However, there is no general inequality in the other direction. Consider the function shown in Figure 31.3.3. Here $\|f\|_\infty = 1$, but $\|f\|_1$ is equal to the shaded area, which may be made arbitrarily small by reducing the width of the base of the peak.

Cauchy sequences may be defined on any normed vector space, and Cauchy sequences are convergent with proof as for Theorem 12. However, Cauchy sequences do not necessarily converge in an arbitrary normed vector space.

Definition 68 *A normed vector space V is said to be complete, or a Banach Space, if every Cauchy sequence in V converges to a limit in V.*

By Theorem 12 and Theorem 67 $\mathbb{R}^n$ with any norm is complete.