MA259

Multivariable Calculus Revision Guide

Written by Sean Middlehurst

Contents

Introduction

This revision guide for MA259 Multivariable Calculus has been designed as an aid to revision, not a substitute for it. This guide is useful for revising through key definitions, theorems and some shorter proofs found in the course. However, a lot of the calculation methods and practical applications of the content of this module are omitted, for which it would be best to refer to the lectures and the online notes for said techniques.

Disclaimer: Use at your own risk. No guarantee is made that this revision guide is accurate or complete, or that it will improve your exam performance.

Authors

Written by Sean Middlehurst.
Based upon the lecture notes for MA259 Multivariable Calculus, written by Dr. Mario Micallef at the University of Warwick.
Any corrections or improvements should be reported by email to comms@warwickmaths.org.

1 Preliminaries

1.1 Notation

- $x \in \mathsf{R}^n$ will denote the n-tuple $(x_1, \dots, x_n), x_i \in \mathsf{R}, 1 \le i \le n$.
- Vectors can either be written as row vectors $(x_1, \dots, x_n)$ or as column vectors $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.
- If $A : \mathsf{R}^n \to \mathsf{R}^k$ is a linear map represented by the matrix:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \dots & a_{kn} \end{pmatrix}$$

then $y := Ax$ is obtained my multiplying A on the ***left*** by column vector x on the ***right***. Thus:

$$y_i = \sum_{j=1}^{n} a_{ij} xj$$

- For vector valued functions $f : U \to \mathsf{R}^k$, where $U \subset \mathsf{R}^n$, then:

$$f(x) \text{ is shorthand for } (f_1(x_1, \dots, x_n), \dots, f_k(x_1, \dots, x_n))$$

1.2 Distances and Convergence

Definition 1.1. The ***Euclidean distance*** between $x, y \in \mathsf{R}^n$ is denoted by $|x - y|$ and is defined as such:

$$|x - y| := (\sum_{i=1}^{n} (x_i - y_i)^2)^{1/2}, x = (x_1, ..., x_n), y = (y_1, ..., y_n).$$

Definition 1.2. A sequence of vectors $x_j \in \mathsf{R}^n$, $n \in \mathsf{N}$, is said to ***converge*** to $x \in \mathsf{R}^n$ if $|x_j - x| \to 0$ as a sequence in R. Equivalently, x_j converges to x if:

$$\forall \varepsilon > 0, \exists N \in \mathsf{N} \text{ such that } j \ge N \implies |x_j - x| < \varepsilon.$$

Definition 1.3. The ***scalar product*** $x \cdot y$, also called the ***dot product*** and ***Euclidean inner product*** of two vectors $x, y \in \mathsf{R}^n$ is defined by:

$$x \cdot y := \sum_{i=1}^{n} x_i y_i$$

Proposition 1.4. The ***Cauchy-Schwarz inequality*** states that:

$$|x \cdot y| \le |x||y|$$

Proof. $0 \le ||y|^2 x - (x \cdot y)y|^2 = |y|^4|x|^2 - (x \cdot y)^2|y|^2$ □

Definition 1.5. For any nonzero pair of vectors x and y, there exists a unique $\theta \in [0, \pi]$, defined as the ***angle*** between x and y, such that:

$$cos\theta = \frac{x \cdot y}{|x||y|}$$

Proposition 1.6. ***(The Triangle Inequality):*** For all $x, y \in \mathsf{R}^n$:

$$|x + y| \le |x| + |y|$$

Proof. $|x + y|^2 = (x + y) \cdot (x + y) = |x|^2 + 2x \cdot y + |y|^2 \le |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2$ □

Corollary 1.7. *(Reverse Triangle Inequality)*: For all $x, y \in \mathbb{R}^n$:

$$||x| - |y|| \leq |x - y|$$

Definition 1.8. Other norms of interest:
$|x|_1 := \sum_{i=1}^{n} |x_i|$
$|x|_\infty := \max\{|x_1|, ..., |x_n|\}$

Proposition 1.9. If $(x_j)_{j \in \mathbb{N}}$ converges to x, then $\exists\, M > 0$ such that $|x_j| \leq M \,\forall\, j \in \mathbb{N}$.

Proof. Given $\varepsilon > 0, \exists\, N \in \mathbb{N}$ such that $j \geq N \implies |x_j - x| < \varepsilon$. Reverse Triangle Inequality then gives:

$$j \geq N \implies ||x_j| - |x|| \leq |x_j - x| < \varepsilon$$

This proves $(|x_j|)_{n \in \mathbb{N}}$ converges to $|x|$, from which the boundedness of $(x_j)_{j \in \mathbb{N}}$ follows. □

1.3 Open and Closed Sets

Definition 1.10. In $\mathbb{R}^n$, given $a \in \mathbb{R}^n$ and $r > 0$, the *open ball with centre a and radius r* is defined by

$$B(a, r) := \{x \in \mathbb{R}^n \mid |x - a| < r\}$$

Definition 1.11. A subset $U \subset \mathbb{R}^n$ is said to be *open* if:

$$\forall\, x \in U, \exists\, r > 0 \,\text{such that}\, B(x, r) \subset U$$

A subset $E \subset \mathbb{R}^n$ is *closed* if $\mathbb{R}^n \setminus E$ is open.

Proposition 1.12.

- An open ball $B(a, r)$ is open.
- Let $U_1, \ldots, U_k$ be open in M. Then $\bigcap_{i=1}^{k} U_i$ is open in M.
- The union of any collection of sets open in M is open in M.

Lemma 1.13. $E \subset \mathbb{R}^n$ is closed if and only if given a sequence $(x_n)_{n=1}^{\infty}$ in E which converges to some point $x \in \mathbb{R}^n$, we have $x \in E$.

1.4 Continuity

Definition 1.14. Given a subset $U \subset \mathbb{R}^n$, a function $f\colon U \to \mathbb{R}^k$ is said to be *continuous at* $p \in U$, if:

$$\forall\, \varepsilon > 0, \exists\, \delta > 0 \text{ such that } (x \in U \text{ and } |x - p| < \delta) \implies |f(x) - f(p)| < \varepsilon$$

If f is continuous at each $x \in U$, we say that f is *continuous*.

Just as in $\mathbb{R}$, we can show that f is continuous at $x \in U$ iff given any sequence $(x_n) \subset U, (x_n) \to x$, we have $f(x_n) \to f(x)$.

Proposition 1.15. Given $f, g : U \to \mathbb{R}^k$ continuous at $p \in U$, $\alpha, \beta \in \mathbb{R}$, then:
$\alpha f + \beta g$ is continuous at p.
For $k = 1, fg$ is continuous at p, where $(fg)(x) := (f(x)) \cdot (g(x))$.

Proposition 1.16. If $U \subset \mathbb{R}^n, V \subset \mathbb{R}^k, f : U \to \mathbb{R}^k$ is continuous at $p \in U, f(U) \subset V, g : V \to \mathbb{R}^m$ is continuous at $f(p) \in V$, then $g \circ f : U \to \mathbb{R}^m$ is continuous at p.

Proposition 1.17. For $U \subset \mathbb{R}^n$, if $f : U \to \mathbb{R}^k$ is written as $f(x) = (f_1(x), f_2(x), ..., f_k(x))$, then f is continuous at $p \in U$ if and only if every component f_i is continuous at p.

Definition 1.18. Given $U \in \mathbb{R}^n, A \subset U$ is *open relative to U* if there exists an open subset $O \in \mathbb{R}^n$ such that $A = O \cap U$.

Theorem 1.19. A function $f : U \to \mathbb{R}^k$ is continuous if and only if for every open set $V \subset \mathbb{R}^k$, the preimage $f^{-1}(V)$ is open relative to U.

Proof. ($\Longrightarrow$) Let V be open in $\mathbb{R}^k$ and take $x \in f^{-1}(V)$. Then $f(x) \in V$ so, as V is open, there is $\varepsilon > 0$ such that $B(f(x), \varepsilon) \subset V$. Since f is continuous at x there exists $\delta > 0$ such that $(y \in U, |x-y| < \delta) \implies |f(x) - f(y)| < \varepsilon$. In other words $y \in B(x, \delta) \implies f(y) \in B(f(x), \varepsilon) \subset V$. Hence $B(x, \delta) \cap U \subset f^{-1}(V)$. Hence $f^{-1}(V)$ is open relative to U.

($\Longleftarrow$) Let $x \in U$ and $\varepsilon > 0$. $B(f(x), \varepsilon)$ is open in $\mathbb{R}^k$ so $f^{-1}(B(f(x), \varepsilon))$ is open in U. Furthermore $x \in f^{-1}(B(f(x), \varepsilon))$ so there is $\delta > 0$ such that $B(x, \delta) \cap U \subset f^{-1}(B(f(x), \varepsilon))$. In other words $(y \in U, |x - y| < \delta) \implies |f(x) - f(y)| < \varepsilon$ so f is continuous at x. Since $x \in U$ was arbitrary, f is continuous. $\square$

1.5 Connectedness

Definition 1.20. $p \in U$ is an ***isolated point of U*** if $\exists\, \varepsilon > 0$ such that $|x - p| > \varepsilon\, \forall\, x \in U \setminus \{p\}$.

Definition 1.21. Given a function $f : U \to \mathbb{R}^k$, a non-isolated point $p \in U$, and $l \in \mathbb{R}^k$, we say that $\lim_{x \to p} f(x) = l$ if:

$$\forall\, \varepsilon > 0, \exists\, \delta > 0 \,\text{such that}\, 0 < |x - p| < \delta \implies |f(x) - l| < \varepsilon$$

Definition 1.22. Let $p, q \in \mathbb{R}^n$. A path from a to b in $\mathbb{R}^n$ is a continuous map $\phi : [a, b] \to \mathbb{R}^n, [a, b] \subset \mathbb{R}$ such that $\phi(a) = p$ and $\phi(b) = q$.

Definition 1.23. $U \subset \mathbb{R}^n$ is called ***path connected*** if any two points in U can be joined by a path in U.

Theorem 1.24. If $U \subset \mathbb{R}^n$ is path connected and $f : U \to \mathbb{R}^k$ is continuous, then $f(U)$ is also path connected.

Proof. Given $v, w \in f(U), \exists\, p, q \in U$ such that $f(p) = v$ and $f(q) = w$. Let $r : [a, b] \to U$ be a path in U that goes from p to q. Then, by continuity of composition of continuous functions, $f \circ r : [a, b] \to f(U) \subset \mathbb{R}^k$ is a path in $f(U)$ that joins v and w. $\square$

Theorem 1.25. $I \subset \mathbb{R}$, is path connected if and only if I is an interval.

Proof. If I is an interval and $a, b \subset I, a < b$ then $r : [0, 1] \to I$ defined by $r(t) := (1 - t)a + tb$ is a path in I joining a and b.

Conversely, if I is path connected and $a, b \in I, a < b$ let $r : [\alpha, \beta] \to I$ be a path in I joining a and b. Then, by the Intermediate Value Theorem, given $t \in (a, b), \exists\, \gamma \in (\alpha, \beta)$ such that $r(\gamma) = t$. In particular, $t \in I$, thereby showing that I is an interval. $\square$

1.6 Sequential Compactness

Definition 1.26. A set $K \subset \mathbb{R}^n$ is ***sequentially compact*** if every sequence in K has a convergent subsequence, which converges to a point of K.

Theorem 1.27. If $K \subset \mathbb{R}^n$ is sequentially compact and $f : K \to \mathbb{R}^k$ is continuous, then $f(K)$ is sequentially compact.

Proof. Let $(y_j)_{j \in \mathbb{N}}$ be a sequence in $f(K)$. Then, for each $j \in \mathbb{N}, \exists\, x_j \in K$ such that $f(x_j) = y_j$. By the sequential compactness of K, there exists a convergent subsequence $(x_{j(l)})_{l \in \mathbb{N}}$ of $(x_j)_{j \in \mathbb{N}}$ such that $\lim_{l \to \infty} x_{j(l)} = x \in K$. By continuity of f at x, $lim_{l \to \infty}\, y_{j(l)} = lim_{l \to \infty}\, f(x_{j(l)}) = f(x) \in f(K)$, i.e., $f(K)$ is sequentially compact. $\square$

Theorem 1.28. A subset $K \subset \mathbb{R}^n$ is sequentially compact if and only if it is closed and bounded.

Proof. ($\Longrightarrow$) If $(x_k)_{k=1}^{\infty} \subset K$ converges, then any subsequence must converge to the same limit. By sequential compactness this limit lies in K, and hence by the fact that sets that contain their limits are closed, K is closed. Furthermore, if K is not bounded, then $\forall\, k \in \mathbb{N}$, $\exists\, x_k \in K$ s.t. $\|x_k\| > k$. Hence any subsequence of $(x_k)_{k=1}^{1}$ is not bounded, and hence cannot converge, contradicting K being sequentially compact. Hence K is bounded.

($\Longleftarrow$) Take any sequence $(x_k)_{k=1}^{1}$ in K and write $x_k = (x_k^1, \ldots, x_k^n)$ for $x_k^j \in \mathsf{R}$. As K is bounded, (x_k) is bounded, so by the Bolzano-Weierstrass Theorem there is a subsequence such that the first component $(x_{k_i}^1)$ converges. Consider (x_{k_i}) and take a subsequence of this such that the second component converges. Repeat n times; the last subsequence converges; since K is closed this limit lies in K. ☐

Theorem 1.29. Let $K \subset \mathsf{R}^n$ be sequentially compact and let $f\colon K \to \mathsf{R}$ be continuous. Then $\exists\, x\ ,x\ \in K$ such that:

$$f(x\ ) \leq f(x) \leq f(x\ )\ \forall\, x \in K$$

Proof. By the previous two theorems, $f(K) \subset \mathsf{R}$ must be closed and bounded. Thus, $M := supf(K)$ and $m := inff(K)$ are both finite because $f(K)$ is bounded. Furthermore, $m, M \in f(K)$ because $f(K)$ is closed. The desired result follows immediately. ☐

Corollary 1.30. Let $K \subset \mathsf{R}^n$ be sequentially compact and let $f : K \to \mathsf{R}^k$ be continuous. Then $\exists\, x\ , x\ \in K$ such that:

$$|f(x\ )| \leq |f(x)| \leq |f(x\ )|\ \forall\, x \in K$$

Proof. Since $x \mapsto |x| : \mathsf{R}^k \to \mathsf{R}$ is continuous (by the triangle inequality $||x| - |y|| \leq |x - y|$), the map $x \mapsto |f(x)| : K \to \mathsf{R}$ is continuous. The result now follows from the theorem on extreme values. ☐

2 Differentiation

2.1 Linear Algebra

- We denote the *space of linear maps*. $L(\mathsf{R}^n, \mathsf{R}^k) := \{A : \mathsf{R}^n \to \mathsf{R}^k \mid A \text{ is linear.}\}$
- We denote the *space of $k \times n$ matrices with real entries,* $M(k \times n, \mathsf{R}) :=$

$$\{\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \ldots & a_{kn} \end{pmatrix} | a_{ij} \in \mathsf{R}, 1 \leq i \leq k, 1 \leq j \leq n\}$$

- $L(\mathsf{R}^n, \mathsf{R}^n)$ shall be abbreviated to $L(\mathsf{R}^n)$.
- There is a linear isomorphism $\mu : L(\mathsf{R}^n, \mathsf{R}^k) \to M(k \times n, \mathsf{R})$, $\mu(A) := (a_{ij})$, where A is such that for $x \in \mathsf{R}^n$:

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto Ax := \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \ldots & a_{kn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathsf{R}^k$$

Definition 2.1. Let $A \in L(\mathsf{R}^n, \mathsf{R}^k)$. The *operator norm* of A is defined by:

$$\|A\| := \sup_{jxj=1} |Ax| = \sup_{x2\mathsf{R}^n nf0g} \frac{|Ax|}{|x|}$$

The following are all properties of the operator norm. Let $A, B \in L(\mathsf{R}^n, \mathsf{R}^k)$ and $\alpha \in \mathsf{R}$:

- $\|A\| = 0 \iff A = 0$
- $\|\alpha A\| = |\alpha| \|A\|$
- **Triangle Inequality:** $\|A + B\| \leq \|A\| + \|B\|$
- **Composition Bound:** If $A \in L(\mathsf{R}^n, \mathsf{R}^k)$ and $B \in L(\mathsf{R}^k, \mathsf{R}^m) \implies BA \in L(\mathsf{R}^n, \mathsf{R}^m)$ and $\|BA\| \leq \|B\| \|A\|$

Definition 2.2. Given an open subset $U \subset \mathsf{R}^n$, $f : U \to \mathsf{R}^k$ is (locally) ***Lipschitz continuous*** at $p \in U$ if there exists $\eta > 0, M > 0$ such that for all $x \in U \cap \mathsf{B}(p, \eta)$:

$$|f(x) - f(p)| \leq M|x - p|$$

Definition 2.3. The ***general linear group over the real numbers***, denoted by $GL(n, \mathsf{R})$, is defined by:

$$GL(n, \mathsf{R}) := \{(a_{ij}) \in M(n \times n, \mathsf{R}) : \det(a_{ij}) \neq 0\}$$

Proposition 2.4. Given $A \in GL(n, \mathsf{R})$, set $\alpha := \frac{1}{kA^{-1}k}$. If $B \in L(\mathsf{R}^n)$ and $\|B - A\| < \alpha$ then B is invertible, i.e., the open ball $\{B \in L(\mathsf{R}^n) : \|B - A\| < \alpha\} \subset GL(n, \mathsf{R})$. Furthermore:

$$\|B - A\| < \alpha \implies \|B^{\ 1}\| \leq \frac{1}{\alpha - \|B - A\|}$$

Proof. $x = A^{\ 1}(Ax) \implies |x| \leq \|A^{\ 1}\||Ax|$, i.e., $|Ax| \geq \alpha|x| \; \forall x \in \mathsf{R}^n$. Therefore, if $x \neq 0$ and $\|B - A\| < \alpha$:

$$|Bx| = |Bx - Ax + Ax| \geq |Ax| - |(B - A)x| \geq (\alpha - \|B - A\|)|x| > 0$$

i.e., $Bx \neq 0$. Therefore, $ker(B) = \{0\}$ and $B \in GL(n, \mathsf{R})$. Finally, on replacing x by $B^{\ 1}x$ as earlier, we see that:

$$|x| = |B(B^{\ 1}x)| \geq (\alpha - \|B - A\|)|B^{\ 1}x|$$

i.e.,

$$|B^{\ 1}| \leq \frac{1}{\alpha - \|B - A\|}|x| \, \forall x \in \mathsf{R}^n$$

□

Proposition 2.5. $A \mapsto A^{\ 1} : GL(n, \mathsf{R}) \to GL(n, \mathsf{R})$ is continuous.

Proof. We need to show that, if B is close to A in GL(n,R) then $B^{\ 1}$ is close to $A^{\ 1}$. So, we consider $A^{\ 1} - B^{\ 1} = A^{\ 1}BB^{\ 1} - A^{\ 1}AB^{\ 1} = A^{\ 1}(B - A)B^{\ 1}$. Therefore:

$$\|A^{\ 1} - B^{\ 1}\| \leq \|A^{\ 1}\|\|B - A\|\|B^{\ 1}\|$$

As above, set $\alpha := \frac{1}{kA^{-1}k}$ and, given $\varepsilon > 0$, set $\delta := \min\{\frac{1}{2}\alpha, \varepsilon\} > 0$. Then, from the previous proposition, we deduce that:

$$\|B - A\| < \delta \implies \|B^{\ 1}\| \leq \frac{2}{\alpha} \implies \|A^{\ 1} - B^{\ 1}\| \leq \frac{2\varepsilon}{\alpha^2}$$

i.e., we have verified the $\varepsilon - \delta$ definition of continuity of $A \mapsto A^{\ 1}$ for $A \in GL(n, \mathsf{R})$. □

Proposition 2.6. $(a_{ij}) \mapsto \det(a_{ij}) : M(n \times n, \mathsf{R}) \to \mathsf{R}$ is continuous with respect to the norm $\| \cdot \|_2$ on $M(n \times n, \mathsf{R})$.

Proof. The determinant is simply a polynomial of degree n in its n^2 variables:

$$a_{11}, \ldots, a_{1n}, a_{21}, \ldots, a_{2n}, \ldots, a_{n1}, \ldots, a_{nn}$$

Therefore, its continuity follows from the identification of $(M(n \times n, \mathsf{R}), \| \cdot \|_2)$ with $(\mathsf{R}^{n^2}, | \cdot |)$ and the usual continuity of polynomials on R^{n^2}. □

2.2 The Derivative

Definition 2.7. Given $p \in (\alpha, \beta) \subset \mathsf{R}$, the ***derivative at p*** of a function $f : (\alpha, \beta) \to \mathsf{R}$ is defined by:

$$f^0(p) = \lim_{h \,!\, 0} \frac{f(p + h) - f(p)}{h}$$

Definition 2.8. Given $p \in U \subset \mathbb{R}^n$, a function $f : \mathbb{R}^n \to \mathbb{R}^k$ is ***(Frechet) differentiable at p*** if $\exists A_p \in L(\mathbb{R}^n, \mathbb{R}^k)$ such that:

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

The linear map, A_p, is called the ***(Frechet) derivative*** of f at p and denoted by $Df(p)$.

Proposition 2.9. Given $f : U \to \mathbb{R}^k, f(x) = (f_1(x), \ldots f_k(x))$, f is differentiable at $p \in U$ if and only if for each $i \in \{1, \ldots, k\}, f_i : U \to \mathbb{R}$ is differentiable at p.

Proof. Exercise. □

Proposition 2.10. If $f : U \to \mathbb{R}^k$ is differentiable at p, i.e., $Df(p)$ exists, then f is locally Lipschitz continuous at p.

Proof. Exercise. □

Theorem 2.11. Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^k$ be open, and let $f : U \to \mathbb{R}^k$ and $g : V \to \mathbb{R}^m$. If f is differentiable at $p \in U$, and g is differentiable at $f(p) \in V$, then $g \circ f$ is differentiable at p and:

$$(D(g \circ f))(p) = (Dg(f(p))) \circ (Df(p))$$

Definition 2.12. Let $f : U \to \mathbb{R}^k$, and let $p, v \in \mathbb{R}^n$. The ***directional derivative*** of f at p in direction v is defined as:

$$(D_v f)(p) := \lim_{h \to 0,\, h \in \mathbb{R}\setminus 0} \frac{f(p+hv) - f(p)}{h} = \frac{d}{dt} f(p+tv)\Big|_{t=0}$$

if the limit exists.

Definition 2.13. Let $\{v_1, \ldots, v_n\}$ be the standard basis of $\mathbb{R}^n$. Then, for $i \in \{1, \ldots, n\}$, the i^{th} ***partial derivative of f at*** $p = (p_1, \ldots, p_n)$ is denoted by $(D_i f)(p)$ and is defined by:

$$\begin{aligned}
(D_i f)(p) &:= \lim_{h \to 0,\, h \in \mathbb{R}\setminus 0} \frac{f(p_1, \ldots, p_{i-1}, p_i + h, p_{i+1}, \ldots, p_n) - f(p_1, \ldots, p_n)}{h} \\
&= \lim_{h \to 0,\, h \in \mathbb{R}\setminus 0} \frac{f(p+hv_i) - f(p)}{h} = \frac{d}{dt} f(p+tv_i)\Big|_{t=0} \\
&= (D_{v_i} f)(p)
\end{aligned}$$

Proposition 2.14. The following algebraic rules apply to partial differentiation:

$$f, g : U \to \mathbb{R}^k \implies D_i(f+g) = D_i f + D_i g$$

$$f : U \to \mathbb{R} \text{ and } g : U \to \mathbb{R}^k \implies D_i(fg) = (D_i f)g + f D_i g$$

Proposition 2.15. If $f : U \to \mathbb{R}^k$ is differentiable at p, then $(D_v f)(p) = (Df(p))(v)$.

Proof. define $r : \mathbb{R} \to \mathbb{R}^n$ by $r(t) := p + tv$. We know from earlier in the course that $Dr(t) = r'(t) = v \,\forall\, t \in \mathbb{R}$, i.e., $(Dr(t))(h) = hv, h \in \mathbb{R}$. Furthermore:

$$\begin{aligned}
\text{by definition of } r(t) \text{ and of the directional derivative}, (D_v f)(p) &= \frac{d}{dt} f(r(t))\Big|_{t=0} \\
&= (f \circ r)'(0) \\
&= (D(f \circ r))(0) \\
&= ((Df)(r(0))) \circ (Dr(0)), \text{by chain rule} \\
&= (Df(p))(v)
\end{aligned}$$

Since Df is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^k$, we see from the above that $D_v f$ depends linearly on v, i.e., $D_{\alpha u + \beta v} = \alpha D_u f + \beta D_v f \,\forall\, \alpha, \beta \in \mathbb{R}, u, v \in \mathbb{R}^n$. □

Corollary 2.16. The Jacobian matrix representation $J_f(p)$ of $Df(p)$ with respect to the standard bases $v_1, \dots, v_n$ and $w_1, \dots, w_k$ of R^n and R^k respectively is given by:

$$J_f(p) := \begin{pmatrix} (D_1 f_1)(p) & \dots & (D_n f_1)(p) \\ \vdots & \ddots & \vdots \\ (D_1 f_k)(p) & \dots & (D_n f_k)(p) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(p) & \dots & \frac{\partial f_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1}(p) & \dots & \frac{\partial f_k}{\partial x_n}(p) \end{pmatrix}$$

where $f(x) = (f_1(x_1, \dots, x_n), \dots, f_k(x_1, \dots, x_n)) = \sum_{i=1}^k f_i(x) w_i$.

Theorem 2.17. Given $f : U \to \mathsf{R}^k$, suppose that the Jacobian matrix J_f exists at all points of U and is continuous at $p \in U$. Then f is differentiable at p.

Note: The existence of the Jacobian matrix at all points does not necessarily guarantee differentiability of a function. The matrix must also be continuous.

Definition 2.18. Suppose that $f : U \to \mathsf{R}^k$ is differentiable on U. Then f is said to be ***continuously di erentiable at*** $p \in U$ if the map $x \mapsto Df(x) : U \to L(\mathsf{R}^n, \mathsf{R}^k)$ is continuous at p.

Theorem 2.19. $f : U \to \mathsf{R}^k$ is continuously differentiable on U if and only if $J_f : U \to M(k \times n, \mathsf{R})$ is continuous on U.

Proposition 2.20. The chain rule for Jacobian matrices takes the form:

$$J_{g \ f}(p) = J_g f(p) \cdot Jf(p)$$

Or in terms of partial derivatives, we see that:

$$\begin{pmatrix} (D_1(g_1 \circ f)(p) & \dots & (D_n(g_1 \circ f)(p) \\ \vdots & \ddots & \vdots \\ (D_1(g_m \circ f)(p) & \dots & (D_n(g_k \circ f)(p) \end{pmatrix}$$

$$= \begin{pmatrix} (D_1 g_1)(f(p)) & \dots & (D_k g_1)(f(p)) \\ \vdots & \ddots & \vdots \\ (D_1 g_m)(f(p)) & \dots & (D_k g_m)(f(p)) \end{pmatrix} \begin{pmatrix} (D_1 f_1)(p) & \dots & (D_n f_1)(p) \\ \vdots & \ddots & \vdots \\ (D_1 f_k)(p) & \dots & (D_n f_k)(p) \end{pmatrix}$$

2.3 The Gradient

Definition 2.21. For a scalar function $f : U \to \mathsf{R}$, the ***gradient of f***, denoted ∇f or grad$f$, is its Jacobian matrix, which is the vector $(D_1 f, \dots, D_n f)$.

Note: For a vector function $f : U \to \mathsf{R}^k$, where $f = (f_1, \dots, f_k)$, the Jacobian matrix is the matrix whose rows are the gradients of each f_i, i.e.:

$$J_f = \begin{pmatrix} \nabla f_1 & \dots \\ \vdots & \\ \nabla f_k & \dots \end{pmatrix}$$

Proposition 2.22. Suppose that $f : U \to \mathsf{R}$ is differentiable on U and that $\exists M \geq 0$ such that $|\nabla f(x)| \leq M \, \forall x \in U$. Given $p, q \in U$, let $r : [\alpha, \beta] \to U$ be a parametrisation of a path C_{pq} joining p to q, i.e., $p = r(\alpha), q = r(\beta)$. Then:

$$|f(q) - f(p)| \leq M \text{length}(C_{pq})$$

Proof.

$$\begin{aligned}
f(q) - f(p) &= f(r(\beta)) - f(r(\alpha)) \\
&= \int_\alpha^\beta \frac{d}{dt} f(r(t))dt && \text{(by FTC)} \\
&= \int_\alpha^\beta ((\nabla f)(r(t))) \cdot r'(t)dt && \text{(by the Chain Rule)} \\
\text{so,} |f(q) - f(p)| &\leq \int_\alpha^\beta |((\nabla f)(r(t))) \cdot r'(t)|dt \\
&\leq \int_\alpha^\beta |(\nabla f)(r(t))||r'(t)|dt && \text{(by Cauchy-Schwarz)} \\
&\leq \int_\alpha^\beta M|r'(t)|dt \\
&= M \text{length}(C_{pq})
\end{aligned}$$

□

Corollary 2.23. Suppose that $f : U \to \mathsf{R}$ is differentiable on U and that $\exists M \geq 0$ such that $|\nabla f(x)| \leq M \, \forall x \in U$. Given $p, q \in U$, suppose that $L_{pq} \subset U$ where L_{pq} is the line joining p to q. Then:

$$|f(q) - f(p)| \leq M|q - p|$$

Proof. Notice that $\text{length}(L_{pq}) = |q - p|$. □

Corollary 2.24. Suppose that $U \subset \mathsf{R}^n$ is path connected and that $f : U \to \mathsf{R}$ satisfies $\nabla f(x) = 0 \, \forall x \in U$. Then f is constant on U.

Proof. Pick $p \in U$. Then, by path connectedness, given $q \in U, \exists$ path C_{pq} joining p to q. So we can apply the earlier proposition with $M = 0$ to conclude that $f(q) = f(p) \, \forall q \in U$, i.e., f is constant on U, as claimed. □

3 The Inverse and Implicit Function Theorems

3.1 Inverse Function Theorem

Definition 3.1. Let $U, V \subset \mathsf{R}^n$ open. A *change of variables* from $(x_1, \dots, x_n) \in U$ to $(y_1, \dots, y_n) \in V$ is achieved by means of a function $\Psi : U \to V$, i.e.:

$$y_1 = \Psi_1(x_1, \dots, x_n), \dots, y_n = \Psi_n(x_1, \dots, x_n)$$

Note: If Ψ is a bijection, we can revert to the original variables using the inverse map $\Psi^{-1} : V \to U$.

Theorem 3.2. Inverse Function Theorem: Let U be an open subset of R^n and suppose that $\Psi : U \to \mathsf{R}^n$ is continuously differentiable. Assume that $D\Psi(p)$ is invertible at a point $p \in U$ and set $\Psi(p) := q$. Then there exists $\mathcal{N}_p, \mathcal{N}_q \subset \mathsf{R}^n$ open such that:

- $p \in \mathcal{N}_p \subset U, q \in \mathcal{N}_q \subset \Psi(U)$,
- $\Psi : \mathcal{N}_p \to \mathcal{N}_q$ is a bijection,
- $\Psi^{-1} : \mathcal{N}_q \to \mathcal{N}_p$ is continuously differentiable,
- $(D\Psi^{-1})(y) = (D\Psi(\Psi^{-1}(y)))^{-1}$.

3.2 Implicit Function Theorem

Theorem 3.3. Implicit Function Theorem: Let U be an open subset of R^{n+l} and suppose that $F : U \to \mathsf{R}^l$ is continuously differentiable on U. Pick $(x_0, y_0) \in U$ and set $c := F(x_0, y_0)$. Suppose that $F_y(x_0, y_0) = DF(x_0, y_0)\Big|_{\{0\}\times\mathsf{R}^l}$ is invertible. Then there exists $\mathcal{N}_{x_0} \subset \mathsf{R}^n$ open and a function $g : \mathcal{N}_{x_0} \to \mathsf{R}^l$ such that:

- $x_0 \in \mathcal{N}_{x_0}, g(x_0) = y_0, \{(x, g(x)) : x \in \mathcal{N}_{x_0}\} \subset U$,
- $F(x, g(x)) = c \,\forall\, x \in \mathcal{N}_{x_0}$,
- g is continuously differentiable on $\mathcal{N}_{x_0}$,
- $F_y(x, g(x))$ is invertible $\forall\, x \in \mathcal{N}_{x_0}$,
- $Dg(x) = -(F_y(x, g(x)))^{-1} \circ F_x(x, g(x)) \,\forall\, x \in \mathcal{N}_{x_0}$.

Definition 3.4. $M \subset \mathsf{R}^{n+l}$ is a ***submanifold*** (without boundary) of dimension n if, $\forall\, p \in M, \exists \mathcal{N}_p \subset \mathsf{R}^{n+l}$ open, $U \in \mathsf{R}^n$ and a continuously differentiable map $r : U \to \mathsf{R}^{n+l}$ such that $p \in \mathcal{N}_p$, $r(x_p) = p$ for some $x_p \in U$, $r : U \to M \cap \mathcal{N}_p$ is a bijection and $\text{rank}(Dr(x)) = n \,\forall\, x \in U$. r is called a (regular) parametrisation of $M \cap \mathcal{N}_p$.

Definition 3.5. The ***tangent space*** $T_{r(x)}M$ of M at $r(x)$ is the image of $Dr(x)$, i.e., the span of $\{D_1 r(x), \dots, D_n r(x)\}$ shifted to $r(x)$. More explicitly:

$$T_{r(x)}M = \{r(x) + (Dr(x))h : h \in \mathsf{R}^n\}$$

Proposition 3.6. Given an open subset U of $\mathsf{R}^{n+l}, F : U \to \mathsf{R}^l$ which is continuously differentiable and a fixed $c \in \mathsf{R}^l$, let $\Gamma_c := \{z \in U : F(z) = c\}$. Suppose that $\text{rank}(DF(z)) = l \,\forall\, z \in \Gamma_c$. Then, the level set Γ_c is a submanifold (without boundary) of dimension n in R^{n+l}. Furthermore:

$$T_z\Gamma_c = z + \ker(DF(z)) = \{z + v : (DF(z))(v) = 0\}$$

4 Vector Analysis

4.1 Vector Fields and Line Integrals

Definition 4.1. A ***vector field*** $\underline{v}$ on $U \subset \mathsf{R}^n$ is simply a function $\underline{v} : U \to \mathsf{R}^n$. Thus, a vector field consists of n functions of n variables:

$$\underline{v}(x) = (v_1(x_1, \dots, x_n), v_2(x_1, \dots, x_n), \dots, v_n(x_1, \dots, x_n)), x \in U$$

Definition 4.2. Given $p, q \in \mathsf{R}^n$, a ***curve*** C_{pq} which goes from p to q is the image of a path $r : [\alpha, \beta] \to \mathsf{R}^n$ such that $r(\alpha) = p$ and $r(\beta) = q$. r is then called a ***parameterization*** of C_{pq}.

Definition 4.3. A path $r : [\alpha, \beta] \to \mathsf{R}^n$ is said to be ***continuously differentiable*** on $[\alpha, \beta]$ if:

- r is continuous on $[a, b]$,
- r is continuously differentiable on (α, β),
- the two limits $\lim_{t\downarrow\alpha} r'(t)$ and $\lim_{t\uparrow\beta} r'(t)$ both exist so that r' can be viewed as a continuous function on $[\alpha, \beta]$.

Definition 4.4. A path $r : [\alpha, \beta] \to \mathsf{R}^n$ is called ***regular*** if $r'(t) \neq 0 \,\forall\, t \in [\alpha, \beta]$.

Definition 4.5. The ***tangential line integral*** $\int_{C_{pq}} \underline{v} \cdot dr$ of a vector field $\underline{v}$ along a curve C_{pq} parameterized by $r : [\alpha, \beta] \to \mathsf{R}^n$ for which $r(\alpha) = p$ and $r(\beta) = q$ is defined by:

$$\int_{C_{pq}} \underline{v} \cdot dr := \int_{\alpha}^{\beta} \underline{v}(r(t)) \cdot \frac{dr}{dt} dt$$

Definition 4.6. If a vector field $\underline{v}$ is the gradient of a function $f : U \to \mathsf{R}$ then $\underline{v}$ is called a ***gradient eld***.

Theorem 4.7. FTC for a Gradient Vector Field: Given a function $f : U \to \mathsf{R}$ and a curve $C_{pq} \subset U$ from p to q parameterized by $r : [\alpha, \beta] \to U$ (in particular, $r(\alpha) = p, r(\beta) = q$) we have:

$$ {}_{C_{pq}} \nabla f \cdot dr = f(q) - f(p) $$

Corollary 4.8. For a given function f, ${}_{c} \nabla f \cdot dr$ depends only on the values of f at the endpoints of p and q but not on the values of f along the portion of C strictly between p and q.

Corollary 4.9.

$$ {}_{C} \nabla f \cdot dr = 0 $$

for all closed curves C, since if C is closed then $p = q$ and so $f(p) = f(q)$.

Definition 4.10. The vector field $\underline{v}$ is ***conservative*** if:

$$ {}_{C} v \cdot dr = 0 $$

for all closed curves C.

Proposition 4.11. A vector field $\underline{v} : U \to \mathsf{R}^n$ is conservative if and only if $\forall p, q \in U$, ${}_{C_{pq}} \underline{v} \cdot dr$ is independent of the choice of path C_{pq} from p to q in U.

Theorem 4.12. A vector field $\underline{v} : U \to \mathsf{R}^n$ is a gradient field if and only if it is conservative.

Definition 4.13. If $\underline{v} = \nabla f$ then f is called a ***scalar potential*** of $\underline{v}$.

Definition 4.14. Given $v = (a, b) \in \mathsf{R}^2$, define $v^{?} := (b, -a)$. $v^{?}$ is v rotated clockwise by 90 . In particular, $v \cdot v^{?} = ab - ba = 0$.

Definition 4.15. The tangent $r^{0}(t)$ of a regular curve C patrameterized regularly by $r(t) := (x(t), y(t))$ is given by $r^{0}(t) := (\frac{dx}{dt}, \frac{dy}{dt})$ and therefore:

$$ \underline{N}(t) := r^{0}(t)^{?} = (\frac{dy}{dt}, -\frac{dx}{dt}) $$

is a ***normal*** to C.

Definition 4.16. The ***ux*** of a vector field $\underline{v}(x, y) = (p(x, y), q(x, y)) \in \mathsf{R}^2$ across a curve C parameterized by $r : [\alpha, \beta] \to \mathsf{R}^2$ is defined to be the integral:

$$ {}_{a}^{b} \underline{v}(r(t)) \cdot \underline{N}(t) dt $$

4.2 Integral Theorems of Vector Calculus

Definition 4.17. A ***region*** in R^n is a subset Ω of R^n for which there exists a function $f : \mathsf{R}^n \to \mathsf{R}$ with the following properties:

- All partial derivatives of f are continuous,
- $\Omega = \{x \in \mathsf{R}^n : f(x) < 0\}$,
- $\nabla f(x) \neq 0 \, \forall x \in \partial\Omega$ where $\partial\Omega := \{x \in \mathsf{R}^n : f(x) = 0\}$ is the boundary of Ω.

f is called a ***de ning function*** of the set Ω.

Note: We define $\overline{\Omega} = \Omega \cup \partial\Omega = \{x \in \mathsf{R}^n : f(x) \leq 0\}$.

Definition 4.18. The ***curl*** of a planar vector field $\underline{v} = (x, y) \mapsto (a(x, y), b(x, y)) : U \to \mathsf{R}^2$ is defined to be the function $\frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}$ and it is denoted by curl$\underline{v}$.

Definition 4.19. Let $n_+(p) = (a(p), b(p))$ be the outward unit normal to the region $\Omega \subset \mathsf{R}^2$ at the point $p \in \partial\Omega$. The ***positively oriented*** unit tangent vector $t_+(p)$ at p is then the vector $(-b(p), a(p))$, which is $n_+(p)$ rotated anticlockwise by 90 . Thus, $t_+ = -(n_+(p))^?$.

Definition 4.20. Let $\Omega \subset \mathsf{R}^2$ be a region. A regular parameterization $r : [a, b] \to \mathsf{R}^2$ of $\partial\Omega$ is ***positively oriented*** if $t_+ := \frac{r'}{jr'j}$ is a positively oriented unit tangent vector to $\partial\Omega$ at $r(t)$.

Theorem 4.21. Green's Theorem: Let $\Omega \subset \mathsf{R}^2$ be a region and let $\underline{v} : U \to \mathsf{R}^2$ be a planar vector field on U, which contains $\overline{\Omega}$. Then:

$$\iint_\Omega \mathrm{curl}\underline{v}(x, y)dA_{x,y} = \oint_{\partial\Omega} \underline{v} \cdot t_+ ds = \oint_{\partial\Omega} \underline{v} \cdot dr$$

where s is the arclength parameter along $\partial\Omega$, r is a positively oriented parametrisation of $\partial\Omega$ and the element of area $dA_{x,y}$ in the plane is often written as just $dxdy$.

Definition 4.22. The ***divergence*** of a vector field $\underline{v}(x) = (v_1(x_1, \ldots, x_n), \ldots, v_n(x_1, \ldots, x_n))$ is denoted by both $div\underline{v}$ and $\nabla \cdot \underline{v}$ and it is defined by:

$$\nabla \cdot \underline{v} := \frac{\partial v_1}{\partial x_1} + \ldots + \frac{\partial v_n}{\partial x_n}$$

Thus, $\nabla \cdot \underline{v}$ is a function.

Theorem 4.23. Gauss' Theorem/Divergence Theorem: Let $\Omega \subset \mathsf{R}^2$ be a region and let $\underline{v} : U \to \mathsf{R}^2$ be a planar vector field on U, which contains $\overline{\Omega}$. Then:

$$\iint_\Omega \nabla \cdot \underline{v}(x, y)dA_{x,y} = \oint_{\partial\Omega} \underline{v} \cdot n_+ ds$$

where n_+ is the unit outward normal to Ω.

Definition 4.24. The *ux* of a vector field $\underline{v}(x, y) \in \mathsf{R}^3$ across a surface C parametrised by $r(u, v) = (x(u, v), y(u, v), z(u, v)) : U \to \mathsf{R}^3$ is defined to be the integral:

$$\iint_S \underline{v} \cdot n_+ dA$$

where n_+ is a unit normal to S and dA is the element of area on S. With respect to parametrisation r, we have:

$$n_+(u, v) := \frac{\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}}{|\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}|}, \qquad dA := |\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}|dudv$$

and therefore:

$$\text{Flux of } \underline{v} \text{ across } S = \iint_U \underline{v}(r(u, v)) \cdot (\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v})dudv$$

Note: Other notation for flux may include:

$$\iint_S \underline{v} \cdot n_+ dA = \iint_S \underline{v} \cdot d\mathbf{A} = \iint_S \underline{v} \cdot n_+ dS = \iint_S \underline{v} \cdot d\mathbf{S}$$

Theorem 4.25. Divergence Theorem in R^3: Let $\Omega \subset \mathsf{R}^3$ be a region and let $\underline{v} : U \to \mathsf{R}^3$ be a vector field on U, which contains $\overline{\Omega}$. Then:

$$\iiint_\Omega \nabla \cdot \underline{v}(x, y, z)dV_{x,y,z} = \iint_{\partial\Omega} \underline{v} \cdot n_+ dA$$

where n_+ is the unit outward normal to $\Omega, dV_{x,y,z}$ is the volume element of Ω (often written as $dxdydz$) and, if $\underline{v}(x, y, z) = (v_1(x, y, z), v_2(x, y, z), v_3(x, y, z))$, then:

$$\nabla \cdot \underline{v}(x, y, z) := \frac{\partial v_1}{\partial x}(x, y, z) + \frac{\partial v_2}{\partial y}(x, y, z) + \frac{\partial v_3}{\partial z}(x, y, z)$$

Definition 4.26. A vector field is said to be *incompressible* if it has zero divergence everywhere.

Definition 4.27.

$$\Delta f := \nabla \cdot (\nabla f) = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}$$

Where if f is incompressible and irrotational (has no circulation), then $\Delta f = 0$.

Proposition 4.28. Suppose $f : U \to \mathsf{R}$ is continuously differentiable. Then, identifying $L(\mathsf{R}^n, \mathsf{R})$ with R^n, we can view Df as ∇f, the map $x \mapsto (D_1 f(x), \ldots, D_n f(x)) : U \to \mathsf{R}^n.Df : U \to \mathsf{R}^n$ is differentiable at $p \in U$ if $\exists\, H_p \in L(\mathsf{R}^n)$ such that:

$$lim_{h\,!\,0} \frac{|Df(p+h) - Df(p) - H_p h|}{|h|} = 0$$

where $H_p = D^2 f(p)$, the *Hessian* of f, i.e.:

$$D^2 f(p) \,\text{is the linear map}\, h \mapsto \begin{pmatrix} (D_{11} f(p) & \ldots & D_{n1} f(p) \\ \vdots & \ddots & \vdots \\ (D_{1n} f(p) & \ldots & D_{nn} f(p) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in \mathsf{R}^n$$

Definition 4.29. C^k spaces are defined as follows:

- $C^0(U, \mathsf{R}^m) = \{f : U \to \mathsf{R}^m : f \text{is continuous}\}$.
- $C^k(U, \mathsf{R}^m) = \{f : U \to \mathsf{R}^m :$ all derivatives of f up to, and including, order k exist and are continuous on $U\}$.
- $C^k(U, \mathsf{R})$ is abbreviated to just $C^k(U)$.

4.3 Laplacian and Harmonic functions

Definition 4.30. The *Laplacian* is the second order partial differential operator defined as:

$$\Delta := \frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_n^2}$$

Definition 4.31. *Laplace's Equation* is $\Delta f = 0$ and its solutions are called *harmonic* functions.

Definition 4.32. A function $f : \mathsf{R}^n \setminus \{0\} \to \mathsf{R}$ is called *radial* if $\exists\, \phi : \mathsf{R}_{>0} \to \mathsf{R}$ such that $f(x) = \phi(|x|)\, \forall\, x \in \mathsf{R}^n \setminus \{0\}$

Proposition 4.33. If f is radial and $x \neq 0$ then, by the chain rule and assuming $\phi \in C^1(\mathsf{R}_{>0})$:

$$\nabla f(x) = \phi^0(|x|)\nabla(|x|) = \phi^0(|x|)\frac{x}{|x|}$$

Proof. Obvious from the fact that $\frac{\partial}{\partial x_i}(x_1^2 + \ldots + x_n^2)^{1/2} = \frac{2x_i}{2(x_1^2+\ldots+x_n^2)^{1/2}}$. □

Proposition 4.34. If $f \in C^2(U)$ is harmonic, and $\overline{\Omega} \subset U$, then $_{\partial\Omega} \nabla f \cdot n_+ dA = 0$.

Proof. If f is harmonic, then $\nabla \cdot (\nabla f) = \Delta f = 0$, so by Divergence Theorem:

$$0 = {}_{\Omega} \nabla \cdot (\nabla f) dV = {}_{\partial\Omega} \nabla f \cdot n_+ dA$$

□

Definition 4.35.

- The *integral average* of an integrable function $f : (a, b) \to \mathsf{R}$ is denoted by ${}_a^b f(x)dx$ and is defined by:

$${}_a^b f(x)dx := \frac{1}{b-a} {}_a^b f(x)dx$$

- Similarly, if we have $E \subset \mathsf{R}^2$ and $f : E \to \mathsf{R}$ integrable, then:

$$\fint_E f(x)dA := \frac{1}{\text{Area}(E)} \int_E f(x)dA$$

- Again, if we have $E \subset \mathsf{R}^3$ and $f : E \to \mathsf{R}$ integrable, then:

$$\fint_E f(x)dV := \frac{1}{\text{Vol}(E)} \int_E f(x)dV$$

Definition 4.36. The ***diameter*** of $E \subset \mathsf{R}^n$ is denoted by $\text{diam}(E)$ and is defined as:

$$\text{diam}(E) := \sup\{\|x - y\| : x, y \in E\}$$

Proposition 4.37. If f is continuous at x then $f(x) = \lim_{E \to (x)} \fint_E f$.

5 Second Order Derivatives

5.1 Second Derivatives

Definition 5.1. Let $f \in C^1(U)$. Suppose further that ∇f is differentiable at $p \in U$, i.e., $\exists\, H_p \in L(\mathsf{R}^n)$ such that:

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

H_p, if it exists, is written as $D^2 f(p)$. If the above limit holds, then the second order partial derivatives $\frac{\mathrm{d}^2 f}{\mathrm{d}x_i \mathrm{d}x_j}(p)$ exist at p and the matrix representation of $D^2 f(p)$ with respect to the standard basis on R^n is called the ***Hessian*** of f at p, denoted by $\text{Hess}f(p)$ and is given by:

$$\text{Hess}f(p) = \begin{pmatrix} \frac{\mathrm{d}^2 f}{\mathrm{d}x_1^2}(p) & \dots & \frac{\mathrm{d}^2 f}{\mathrm{d}x_n \mathrm{d}x_1}(p) \\ \vdots & & \vdots \\ \frac{\mathrm{d}^2 f}{\mathrm{d}x_1 \mathrm{d}x_n}(p) & \dots & \frac{\mathrm{d}^2 f}{\mathrm{d}x_n^2}(p) \end{pmatrix}$$

Proposition 5.2. If $f \in C^1(U)$, $U \subset \mathsf{R}^n$ and $D^2 f(p)$ exists, $p \in U$, then:

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^i \mathrm{d}x^j}(p) = \frac{\mathrm{d}^2 f}{\mathrm{d}x^j \mathrm{d}x^i}(p) \qquad \forall\, i, j \in \{1, \dots, n\}$$

i.e. $\text{Hess}f(p)$ is a symmetric matrix.

Corollary 5.3. If all second order partial derivatives of f at p are continuous, then they also commute at p, because it implies $D^2 f(p)$ exists.

5.2 Second Order Taylor Expansion

Theorem 5.4. The 1-Variable Case: Suppose that $f \in C^2(-\eta, \eta)$ for some $\eta > 0$. Then:

$$\text{for } x \in (-\eta, \eta), \quad f(x) = f(0) + x f'(0) + \frac{1}{2}x^2 f''(0) + R(x) \quad \text{where } \lim_{x \to 0} \frac{|R(x)|}{x^2} = 0$$

Proof. Set $f'(0) := b$, $f''(0) := c$ and $\rho(s) := f''(s) - f''(0) = f''(s) - c$. Then:

$$\begin{aligned} f'(t) &= f'(0) + \int_0^t f''(s)\, ds \\ &= b + \int_0^t \big(c + (f''(s) - c)\big)\, ds \\ &= b + ct + \int_0^t \rho(s)\, ds \end{aligned}$$

Therefore:

$$\begin{aligned} f(x) &= f(0) + \int_0^x f'(t)\,dt \\ &= f(0) + \int_0^x \big(b + ct + \int_0^t \rho(s)\,ds\big)\,dt \\ &= f(0) + bx + \frac{1}{2}cx^2 + \int_0^x\int_0^t \rho(s)\,ds\,dt \\ &= f(0) + xf'(0) + \frac{1}{2}x^2 f''(0) + R(x) \end{aligned}$$

where $R(x) := \int_0^x\int_0^t \rho(s)\,ds\,dt$. By continuity of f'' at 0, $\forall\,\varepsilon > 0\ \exists\ \delta > 0$ such that $|s| < \delta \implies |\rho(s)| < \varepsilon$. Therefore:

$$|x| < \delta \implies |R(x)| \leq \int_0^x\int_0^t |\rho(s)|\,ds\,dt \leq \int_0^x\int_0^t \varepsilon\,ds\,dt = \frac{1}{2}\varepsilon x^2,$$

i.e., $0 < |x| < \delta \implies \dfrac{|R(x)|}{x^2} \leq \frac{1}{2}\varepsilon$, i.e., $\lim\limits_{x\to 0}\dfrac{|R(x)|}{x^2} = 0$, as claimed. □

Theorem 5.5. Second Order Taylor Expansion: Suppose $U \subset \mathbb{R}^n$ is convex and $0 \in U$. If $f \in C^2(U)$ then:

$$f(x) = f(0) + \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}(0) + \frac{1}{2}\sum_{i,j=1}^n x_i x_j \frac{\partial^2 f}{\partial x_i \partial x_j}(0) + R(x) \qquad \text{where } \lim_{x\to 0}\frac{|R(x)|}{|x|^2} = 0.$$

Note: $\sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}(0)$ is often written as $x \cdot \nabla f(0)$ and $\sum_{i,j=1}^n x_i x_j \frac{\partial^2 f}{\partial x_i \partial x_j}(0)$ is often written as:

$$x^T D^2 f(0) x = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(0) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(0) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(0) & \dots & \frac{\partial^2 f}{\partial x_n^2}(0) \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Proof. If $x = 0$, there is nothing to prove. If $x \in U \setminus \{0\}$ then, since U is convex and $0 \in U$, $t\frac{x}{|x|} \in U\ \forall\, t \in [0, |x|]$. Therefore, we may define $g : [0, |x|] \to \mathbb{R}$ by $g(t) := f(t\frac{x}{|x|})$. Then, since $t\frac{x}{|x|} = (t\frac{x_1}{|x|}, \dots, t\frac{x_n}{|x|})$, we have, by the Chain Rule:

$$g'(t) = \sum_{i=1}^n \frac{x_i}{|x|}\frac{\mathrm{d}f}{\mathrm{d}x_i}(t\frac{x}{|x|}) \qquad \text{and} \qquad g''(t) = \sum_{i,j=1}^n \frac{x_i}{|x|}\frac{x_j}{|x|}\frac{\mathrm{d}^2 f}{\mathrm{d}x_j \mathrm{d}x_i}(t\frac{x}{|x|})$$

Since U is open we can, in fact, define g on an open interval that contains $[0, |x|]$ and we can apply the 1-variable case of the second order Taylor expansion to g to get:

$$g(|x|) = g(0) + |x|g'(0) + \tfrac{1}{2}|x|^2 g''(0) + R(|x|)$$

i.e.:

$$f(x) = f(0) + x \cdot \nabla f(0) + \frac{1}{2}x^T \mathrm{Hess} f(0)\, x + R(|x|)$$

where $\lim\limits_{x\to 0}\dfrac{|R(x)|}{|x|^2} = 0$. □

5.3 Critical Points

Definition 5.6. $p \in U$ is called a ***critical point*** of $f \in C^1(U)$ if $\nabla f(p) = 0$.

Proposition 5.7. If f has a local maximum or minimum at p, then $\nabla f(p) = 0$.

Proof. Suppose f has a local maximum/minimum at p. Given $v \in \mathbb{R}^n$, define $g(t) := f(p + tv)$ for $t \in \mathbb{R}$ small enough so that $p + tv \in U$. Then g has a local maximum/minimum at 0 and therefore, $g'(0) = 0$. But $g'(0) = (\nabla f(p)) \cdot v$ and we have shown that $(\nabla f(p)) \cdot v = 0,\ \forall v \in \mathbb{R}^n$. Taking $v = \nabla f(p)$ shows that $\nabla f(p) = 0$, as claimed. □

Definition 5.8. A symmetric matrix P is

- ***positive de nite*** if $x^T P x = x \cdot Px > 0\ \forall x \in \mathbb{R}^n \setminus \{0\}$.
- ***positive semide nite*** if $x^T P x \geq 0\ \forall x \in \mathbb{R}^n$.
- ***negative de nite*** if $x^T P x < 0\ \forall x \in \mathbb{R}^n \setminus \{0\}$.
- ***negative semide nite*** if $x^T P x \leq 0\ \forall x \in \mathbb{R}^n$.
- ***inde nite*** if $x^T P x$ is neither positive semidefinite nor negative semidefinite, i.e., $\exists\ x, y \in \mathbb{R}^n$ such that $x^T P x > 0$ and $y^T P y < 0$.

Theorem 5.9. Every symmetric matrix can be diagonalised by an orthogonal matrix, i.e., if $\lambda_1, \ldots, \lambda_n$ are the real eigenvalues of a symmetric matrix P then there exists an orthogonal matrix $\mathcal{O}$ such that:

$$\mathcal{O}^T P \mathcal{O} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} =: \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

Proposition 5.10. Arrange the eigenvalues of P in increasing order, i.e., $\lambda_1 \leq \cdots \leq \lambda_n$. Then:

$$\lambda_n |x|^2 \geq x \cdot Px \geq \lambda_1 |x|^2\ \forall x \in \mathbb{R}^n$$

Proof. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of P, i.e., $Pe_i = \lambda_i e_i$; e_i is just the i^{th} column of $\mathcal{O}$. Given $x \in \mathbb{R}^n$, let $a_i := x \cdot e_i$. Then $x = \sum_{i=1}^n a_i e_i$ and $|x|^2 = \sum_{i=1}^n (a_i)^2$. It follows that:

$$x \cdot Px = \sum_{i=1}^n (a_i)^2 \lambda_i \geq \lambda_1 \sum_{i=1}^n (a_i)^2 = \lambda_1 |x|^2 \qquad \text{and similarly,} \qquad x \cdot Px = \sum_{i=1}^n (a_i)^2 \lambda_i \leq \lambda_n |x|^2$$

□

Theorem 5.11. Second Order Derivative Test: Suppose that $f \in C^2(U)$ and that $\nabla f(p) = 0$ for some $p \in U$:

- If $\mathrm{Hess} f(p)$ is positive definite then f has a strict local minimum at p.
- If $\mathrm{Hess} f(p)$ is negative definite then f has a strict local maximum at p.
- If $\mathrm{Hess} f(p)$ is indefinite then f has neither a local minimum nor a local maximum at p and p is called a ***saddle point***.
- If $\mathrm{Hess} f(p)$ is positive or negative semidefinite then the test is inconclusive, i.e., f may have a minimum at p, or a maximum or a saddle point.

Test for 2×2 **Symmetric Matrices:** A 2×2 symmetric matrix: $P = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$ is

- positive definite if $\det P = ac - b^2 > 0$ and $a > 0$ or $c > 0$,
- negative definite if $\det P > 0$ and $a < 0$ or $c < 0$,
- indefinite if $\det P < 0$,
- semidefinite if $\det P = 0$.

This guide would not be possible without our wonderful sponsors: