Warwick Mathematics Exchange

MA259

Multivariable Calculus

2023, May 1st

Desync, aka The Big Ree

Contents

Table of Contents **i**

1 Notation **1**

2 Convergence and Continuity **5**
2.1 Convergence in $\mathbb{R}^n$ 5
2.2 Continuity 6

3 Topology on $\mathbb{R}^n$ **8**
3.1 Continuity and Topology 8
3.2 Compactness 9

4 The Space $L(\mathbb{R}^n,\mathbb{R}^k)$ of Linear Maps **9**
4.1 Matrix Norms 10
4.2 Convergence and Continuity in $L(\mathbb{R}^n,\mathbb{R}^k)$ 11
4.3 Matrix-Valued Functions 11
4.4 The Space $GL(n,\mathbb{R}) \subset L(\mathbb{R}^n)$ of Invertible Linear Maps 12
4.5 Lipschitz Continuity 12

5 The Derivative **13**
5.1 Partial Derivatives 13
5.2 Directional Derivatives 14
5.3 The Fréchet Derivative 15
5.4 Gradient 16
5.5 Geometric Approximation 17
5.5.1 Graphs 18
5.6 Differentiation of Matrix-Valued Functions 18

6 The Chain Rule **19**
6.1 The Space $C^n(U,\mathbb{R}^k)$ of Continuously Differentiable Functions 20
6.2 Mean Value Inequality 21

7 Vector Fields **23**
7.1 Paths and Curves 23
7.2 Tangential Line Integrals 24
7.3 Flux 25
7.3.1 Flux Across Curves in $\mathbb{R}^2$ 25
7.3.2 Flux Across Surfaces in $\mathbb{R}^3$ 25

8 The Integral Theorems of Vector Calculus **26**
8.1 Green's Theorem for a Rectangle 26
8.1.1 Regions and Unit Normals 27
8.1.2 Boundary Orientation 27
8.2 Green's Theorem for Planar Regions 28
8.3 Flux and Divergence in the Plane 28
8.4 Flux and Divergence in $\mathbb{R}^3$ 29
8.5 Gradient Fields 29
8.5.1 Incompressible and Irrotational Vector Fields 29
8.5.2 Laplacian and Harmonic Functions 30

9 Second Order Derivatives **30**

9.1 Bilinear Forms 30
9.2 The Hessian Matrix 30
9.3 Non-Commutativity of Second Order Partial Derivatives 31

10 Inverse Function Theorem 31
10.1 Change of Variables and Inverse Functions 31
10.2 Local Inverses 32

11 Implicit Function Theorem 34

Introduction

In *Multivariable Calculus*, we extend our work from analysis into higher dimensions. We begin by exploring the notions of convergence and continuity for vector and matrix-valued functions before studying the Fréchet derivative. We then cover vector fields, line and surface integrals, and integral theorems. This document is intended to broadly cover all the topics within the Multivariable Calculus module.

Disclaimer: I make *absolutely no guarantee* that this document is complete nor without error. In particular, any content covered exclusively in lectures (if any) will not be recorded here. This document was written during the 2022 academic year, so any changes in the course since then may not be accurately reflected.

Notes on formatting

New terminology will be introduced in *italics* when used for the first time. Named theorems will also be introduced in *italics*. Important points will be **bold**. Common mistakes will be underlined. The latter two classifications are under my interpretation. YMMV.

Content not taught in the course will be outlined in the margins like this. Anything outlined like this is not examinable, but has been included as it may be helpful to know alternative methods to solve problems.

The table of contents above, and any inline references are all hyperlinked for your convenience.

Scalars are written in lowercase italics, c, or using Greek letters.

Vectors are written in lowercase bold, $\mathbf{v}$, or rarely overlined, $\overleftrightarrow{v}$, where more contrast or clarity is required.

Matrices are written in uppercase bold, $\mathbf{A}$.

Note: transformations represented by matrices may be written in just italics, as functions often are, i.e., $s(\mathbf{v}) = \mathbf{A}\mathbf{v}$.

History

First Edition: 2023-04-20*
Current Edition: 2023-05-01

Authors

This document was written by R.J. Kit L., a maths student. I am not otherwise affiliated with the university, and cannot help you with related matters.

Please send me a PM on Discord @Desync#6290, a message in the WMX server, or an email to Warwick.Mathematics.Exchange@gmail.com for any corrections. (If this document somehow manages to persist for more than a few years, these contact details might be out of date, depending on the maintainers. Please check the most recently updated version you can find.)

If you found this guide helpful and want to support me, you can buy me a coffee!

(Direct link for if hyperlinks are not supported on your device/reader: ko-fi.com/desync.)

*Storing dates in big-endian format is clearly the superior option, as sorting dates lexicographically will also sort dates chronologically, which is a property that little and middle-endian date formats do not share. See ISO-8601 for more details. This footnote was made by the computer science gang.

1 Notation

The notation in this guide has been chosen to be compatible with the lecture notes where possible.

One notable difference is that vectors and vector-valued functions are written in lowercase bold, $\mathbf{v}$, to distinguish them from scalar-valued variables and functions, which will are written in italics, c, or in Greek letters, and matrices will be written in uppercase bold, $\mathbf{A}$. Writing vector-valued functions in bold might be somewhat inconsistent throughout the derivatives section, but will certainly be done for things like parametrisations.

$\|\mathbf{v}\|$	Euclidean norm of the vector $\mathbf{v}$. Also written as $\|\mathbf{v}\|_2$ when discussing other ℓ^p norms. Written as $\lvert\mathbf{v}\rvert$ instead, when matrix-norms are also in use.
$\|\mathbf{v}\|_\infty$	The infinity norm of the vector $\mathbf{v}$.
$\|\mathbf{v}\|_1$	The taxicab or Manhattan norm of the vector $\mathbf{v}$.
$\|\mathbf{A}\|_F$	Frobenius norm of the matrix $\mathbf{A}$. Treats the matrix like a vector, then computes the ordinary Euclidean norm.
$\|T\|$	Operator norm of the linear map $\mathbf{v}$. We will sometimes put a matrix into this norm, as they are isomorphic to linear transformations. Also just written as $\|\cdot\|_{\text{op}}$.
$C(U,\mathbb{R}^k)$	Space of continuous functions $f : U \to \mathbb{R}^k$. Also written as $C^0(U,\mathbb{R}^k)$ (see next entry), or as $C(U)$ when $k = 1$.
$C^n(U,\mathbb{R}^k)$	Space of functions $f : U \to \mathbb{R}^k$ continuously differentiable n times.
$\mathbb{B}_r(\mathbf{a})$	Open ball of radius r centred at a point $\mathbf{a}$. That is, the set $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}-\mathbf{a}\| < r\}$. Also written as $\mathbb{B}(\mathbf{a},r)$.
$\mathbb{B}_r$	Open ball of radius r centred on the origin; $\mathbb{B}_r(\mathbf{0})$.
$\mathbb{B}$	Unit open ball centred on the origin; $\mathbb{B}_1(\mathbf{0})$
$\overline{\mathbb{B}_r(\mathbf{a})}$	Closed ball of radius r centred at a point $\mathbf{a}$. That is, the set $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}-\mathbf{a}\| \leq r\}$. Also written as $\overline{\mathbb{B}(\mathbf{a},r)}$.
$\overline{\mathbb{B}_r}$	Closed ball of radius r centred on the origin; $\overline{\mathbb{B}_r(\mathbf{0})}$
$\overline{\mathbb{B}}$	Unit closed ball centred on the origin; $\overline{\mathbb{B}_1(\mathbf{0})}$.
$S^n(r)$	The n-sphere of radius r; the boundary of $\mathbb{B}_r(\mathbf{0})$; the set $\{\mathbf{x} \in \mathbb{R}^{n+1} : \lvert\mathbf{x}\rvert = r\}$.

S^n	The unit n-sphere; $S^n(1)$.
$L(\mathbb{R}^n,\mathbb{R}^k)$	The space of linear maps $T : \mathbb{R}^n \to \mathbb{R}^k$.
$L(\mathbb{R}^n)$	The space of linear maps $T : \mathbb{R}^n \to \mathbb{R}^n$; isomorphic to and hence interchangable with the space of $k \times n$ matrices with real entries.
$M(k \times n,\mathbb{R})$	The space of $k \times n$ matrices with real entries. Also abbreviated as $\mathbb{R}^{k\times n}$.
$M(n,\mathbb{R})$	The space of $n \times n$ matrices with real entries; $M(n \times n,\mathbb{R})$
$GL(n,\mathbb{R})$	The group of invertible linear maps $T : \mathbb{R}^n \to \mathbb{R}^k$; the group of nonsingular $n \times n$ matrices with real entries.
$\Delta(\mathbf{A})$	The multilinear function that sends a matrix $\mathbf{A}$ to its determinant. Not to be confused with the Laplacian.
M^*	The Lipschitz constant of a function; the upper bound on how quickly a function can vary.
$\dfrac{\partial f}{\partial x_i}$	The partial derivative of a function f with respect to the variable x_i. Also written as $\partial_{x_i} f(x)$; just as $\partial_i f(x)$; or if f has few variables, as f_x, f_y, etc.
$\partial_{\mathbf{v}} f(\mathbf{x})$	The directional derivative of f at $\mathbf{x}$ in the direction of $\mathbf{v}$. Also written as $D_{\mathbf{v}} f(\mathbf{x})$. If $\mathbf{v}$ is one of the basis vectors, then this is the partial derivative.
∇	The del or nabla operator; can be thought of as a vector full of partial differential operators $\left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \dots, \frac{\partial}{\partial x_n}\right]^\top$.
∇f	The gradient of f; the vector $\left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right]^\top$; also written as $\mathrm{grad}(f)$.
$\nabla \cdot \underline{v}$	The divergence of $\underline{v}$; calculated by "dotting" the del operator with the vector field $\underline{v}$; also written as $\mathrm{div}(f)$.
$\nabla \times \underline{v}$	The curl of $\underline{v}$; calculated by "crossing" the del operator with the vector field $\underline{v}$; also written as $\mathrm{curl}(f)$.
$Df(x)$	The Fréchet derivative of f at x.
∂f	The Jacobian matrix of f. Also written as Df, because it's the same thing as the Fréchet derivative in finite dimensions.

$\mathcal{G}_f$	The graph of a function f; if f takes two variables, then $\mathcal{G}_f$ is the surface parametrised by $\mathbf{r}(x,y) = (xy,f(x,y))$.
$\underline{v}$	A vector field; a function $\underline{v} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^n$.
$\mathbf{v}^\perp$	The rotation of the vector $\mathbf{v} \in \mathbb{R}^2$ $90°$ clockwise; if $\mathbf{v} = (x,y)$, then $\mathbf{v}^\perp = (y, -x)$.
$\mathbf{r}(t)$	The parametrisation $r : [a,b] \to \mathbb{R}^2$ of a curve $C \subset \mathbb{R}^2$.
$\rho(s)$	The arclength or unit speed parametrisation $\rho : [0,L] \to \mathbb{R}^2$ of a curve C.
$\dot{\mathbf{r}}(t)$	The tangent to $\mathbf{r}(t)$; given by differentiating $\mathbf{r}$ componentwise.
$\dot{\rho}(s)$	The unit tangent to $\rho(s)$; given by differentiating ρ componentwise.
$\mathbf{N}(t)$	The normal to $\mathbf{r}(t)$; given by $\dot{\mathbf{r}}(t)^\perp$.
$\mathbf{n}(t)$	The unit normal to $\rho(s)$; given by $\dot{\rho}(s)^\perp$.
$\int_0^L \underline{v}\big(\rho(s)\big) \cdot \dot{\rho}(s)ds$	The tangential line integral of $\underline{v}$ along C. Calculated using $\int_a^b \underline{v}\big(\mathbf{r}(t)\big) \cdot \frac{d\mathbf{r}}{dt}\, dt$, where $\mathbf{r}$ is a parametrisation of C.
$\int_C \underline{v} \cdot d\mathbf{r}$	Alternative notation for the tangential line integral of $\underline{v}$ along C.
$\int_0^L \underline{v}\big(\rho(s)\big) \cdot \mathbf{n}(s)ds$	The flux integral of $\underline{v}$ along C; the normal line integral of $\underline{v}$ along C (compare with the tangential line integral above);. Calculated using $\int_a^b \underline{v}\big(\mathbf{r}(t)\big) \cdot \mathbf{N}(t)\, dt$.
$\mathbf{r}(u,v)$	The parametrisation of a surface $S \subset \mathbb{R}^3$.
$\iint_S \underline{v} \cdot \mathbf{n}\, dA$	The flux integral of $\underline{n}$ across a surface S. Calculated using $\iint_U \underline{v}\big(\mathbf{r}(u,v)\big) \cdot \big(\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\big)\, du\, dv$.
$\iint_S \underline{v} \cdot\, d\mathbf{S}$	Alternative notation for the flux integral of $\underline{v}$ across S. Also written as $\iint_S \underline{v} \cdot \mathbf{n}\, dS$, or $\iint_S \underline{v} \cdot\, d\mathbf{A}$.
Δf	The Laplacian of f; calculated as $\nabla \cdot (\nabla f)$, or, the sum of the second derivatives of f; also written as $\nabla \cdot \nabla$ or ∇^2.
$D^2 f(x)$	The Hessian (transformation) of f.
$\partial^2 f(x)$	The Hessian matrix of f; also written as $\operatorname{Hess} f(x)$.

$\mathcal{N}_p$	An (open) neighbourhood of a point p.
Γ_c	The level set of a function set equal to c; the set of inputs to a function such that the output is c.

2 Convergence and Continuity

2.1 Convergence in $\mathbb{R}^n$

A sequence $(\mathbf{x}_i)_{i=1}^{\infty}$ of vectors in $\mathbb{R}^n$ converges to $\mathbf{x} \in \mathbb{R}^n$ if,

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} : i > N \to \|\mathbf{x}_i - \mathbf{x}\| < \varepsilon$$

where $\|\cdot\|$ is the euclidean norm.

Theorem 2.1 (Uniqueness of Limits). *If $(\mathbf{x}_i)_{i=1}^{n}$ converges to both $\mathbf{x}$ and $\mathbf{y}$, then $\mathbf{x} = \mathbf{y}$.*

Theorem 2.2 (Componentwise Convergence). *A sequence $(\mathbf{x}_i)_{i=1}^{n} \subseteq \mathbb{R}^n$ converges to $\mathbf{y}$ if and only if for each $i \in [1,n]$,*

$$\lim_{j\to\infty} \mathbf{x}_{i,j} = \mathbf{y}_i$$

That is, the real number sequences of components of $(\mathbf{x}_i)$ all individually converge to their corresponding component of $\mathbf{y}$.

The *uniform*, *max* or *infinity norm*, denoted by, $\|\cdot\|_\infty$ is defined by,

$$\|\mathbf{x}\|_\infty \coloneqq \max\big(|x_1|, \dots, |x_n|\big), \qquad \mathbf{x} = (x_1, \dots, x_n)$$

The *taxicab* or *Manhattan norm*, denoted by $\|\cdot\|_1$ is defined by,

$$\|\mathbf{x}\|_1 \coloneqq |x_1| + \dots + |x_n|, \qquad \mathbf{x} = (x_1, \dots, x_n)$$

Theorem 2.3. *For all $\mathbf{x} \in \mathbb{R}^n$,*

$$\|\mathbf{x}\|_\infty \le \|\mathbf{x}\| \le \sqrt{n}\|x\|_\infty$$

and

$$\|\mathbf{x}\| \le \|\mathbf{x}\|_1 \le \sqrt{n}\|x\|$$

Theorem 2.4 (Algebra of Limits). *If $(\mathbf{x}_i) \to \mathbf{x}$ and $(\mathbf{y}_i) \to \mathbf{y}$, then,*

- $$\lim_{i\to\infty} (\alpha\mathbf{x}_i + \beta\mathbf{y}_i) = \alpha\mathbf{x} + \beta\mathbf{y}$$

  for all $\alpha, \beta \in \mathbb{R}$;
- $$\lim_{i\to\infty} \langle \mathbf{x}_i, \mathbf{y}_i \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$$

  where $\langle -,- \rangle$ is any inner product, such as the scalar product;
- $$\lim_{i\to\infty} \|\mathbf{x}_i\| = \|\mathbf{x}\|$$

  where $\|\cdot\|$ is any norm.

A sequence $(\mathbf{x}_i)$ is *bounded* if there exists $M > 0$ such that $\|\mathbf{x}_i\| < M$ for all $i \in \mathbb{N}$.

Theorem 2.5 (Boundedness of Convergent Sequences). *If $(\mathbf{x}_i)$ converges to some $\mathbf{x}$, then $(\mathbf{x}_i)$ is bounded.*

Theorem 2.6 (Bolzano-Weierstrass for Vectors). *Any bounded sequence $(\mathbf{x}_i)_{i=1}^{\infty} \subseteq \mathbb{R}^n$ has a convergent subsequence $(\mathbf{x}_{i_j})$.*

2.2 Continuity

A function $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ is *continuous at a point* $\mathbf{p} \in U$ if,

$$\forall \varepsilon > 0, \exists \delta > 0 : \|\mathbf{x} - \mathbf{p}\| < \delta \to \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{p})\| < \varepsilon \qquad (\varepsilon - \delta \text{ Continuity})$$

or if, for all sequences $(\mathbf{x}_i) \to \mathbf{p}$,

$$\big(\mathbf{f}(\mathbf{x}_i)\big) \to \mathbf{p} \qquad \text{(Sequential Continuity)}$$

$\mathbf{f}$ is then said to be *continuous at* U if $\mathbf{f}$ is continuous at all points $\mathbf{p} \in U$.

We write $C(U,\mathbb{R}^k)$ to denote the space of continuous functions $f : U \to \mathbb{R}^k$.

A function $f : U \to \mathbb{R}^k$ has a *(continuous) limit* at $\mathbf{p} \in U$ if there exists a vector $\mathbf{q} \in \mathbb{R}^k$ such that,

$$\forall \varepsilon > 0, \exists \delta > 0 : (x \in \mathbf{U}) \wedge (0 < \|\mathbf{x} - \mathbf{p}\| < \delta) \to \|\mathbf{f}(\mathbf{x} - \mathbf{q})\| < \varepsilon$$

and we write $\lim_{\mathbf{x}\to\mathbf{p}} \mathbf{f}(\mathbf{x}) = \mathbf{q}$.

Just as for limits of sequences, continuous limits are unique. We also have that $\mathbf{f}$ is continuous at $\mathbf{p}$ if and only if $\lim_{\mathbf{x}\to\mathbf{p}} \mathbf{f}(\mathbf{x}) = \mathbf{p}$.

Given a real-valued function $f : \mathbb{R}^2 \to \mathbb{R}$, we define two families of functions g^y and h^x by

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

In computer science terminology, g and h are the partial applications of f in the first and second arguments, respectively.

A function $f : \mathbb{R}^2 \to \mathbb{R}$ is *separately continuous* at (a,b) if g^b is continuous at a and h^a is continuous at b.

Continuity implies separate continuity, but not the converse.

Example. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} 1 & xy \neq 0 \\ 0 & xy = 0 \end{cases}$$

Then, $g^0(x) = 0$ for all x and $h^0(y) = 0$ for all y, so f is separately continuous at $(0,0)$. But,

$$\lim_{(x,y)\to(0,0)} f(x,y) = \lim_{n\to\infty} f\left(0,\frac{1}{n}\right) = 0 \neq 1 = \lim_{n\to\infty} f\left(\frac{1}{n},\frac{1}{n}\right) = \lim_{(x,y)\to(0,0)} f(x,y)$$

so

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

does not have a unique value and hence does not exist.

Theorem 2.7 (Continuity of Sums)**.** *If* $\mathbf{f},\mathbf{g} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ *are both continuous at* $\mathbf{p} \in U$*, then* $\alpha\mathbf{f} + \beta\mathbf{g}$ *is continuous at* $\mathbf{p}$ *for all* $\alpha,\beta \in \mathbb{R}$*.*

Theorem 2.8 (Continuity of Real-Valued Products)**.** *If* $f,g : (U \subseteq \mathbb{R}^n) \to \mathbb{R}$ *are both continuous at* $\mathbf{p} \in U$*, then* $(fg)(x) \coloneqq f(x)g(x)$ *is continuous at* $\mathbf{p}$*.*

Theorem 2.9 (Continuity of Quotients)**.** *If* $f,g : (U \subseteq \mathbb{R}^n) \to \mathbb{R}$ *are both continuous at* $\mathbf{p} \in U$*, and* $g(\mathbf{x}) \neq 0$ *for all* $\mathbf{x} \in \mathbb{U}$*, then* $(f/g)(\mathbf{x}) \coloneqq f(\mathbf{x})/g(\mathbf{x})$ *is continuous at* $\mathbf{p}$*.*

Theorem 2.10 (Continuity of Composition)**.** *If* $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ *is continuous at* $\mathbf{p} \in U$ *and* $\mathbf{g} : (V \subseteq \mathbb{R}^k) \to \mathbb{R}^m$ *is continuous at* $\mathbf{f}(\mathbf{p}) \in V$*, and* $f(U) \subseteq V$*, then* $g \circ f : \mathbf{U} \to \mathbb{R}^m$ *is continuous at* $\mathbf{p}$.

Theorem 2.11 (Componentwise Continuity)**.** *A function* $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ *defined by*

$$(x_1, \dots, x_n) \mapsto \big(f_1(x_1, \dots, x_n), f_2(x_1, \dots, x_n), \dots, f_k(x_1, \dots, x_n)\big)$$

where $(f_j)_{j=1}^k$ *are real-valued functions, is continuous at* $\mathbf{p} \in U$ *if and only if every* f_j *is continuous at* $\mathbf{p}$.

That is, $\mathbf{f}$ is continuous if and only if every component f_j is individually continuous.

Theorem 2.12. *If* $f : \mathbb{R} \to \mathbb{R}$ *is continuous at* $p \in \mathbb{R}$*, then any function* $\mathbb{R}^n \to \mathbb{R}$ *defined by*

$$(x_i)_{i=1}^n \mapsto f(x_j)$$

is continuous on $\{(x_i)_{i=1}^n : x_j = p\}$

That is, any function that is continuous as a function $\mathbb{R} \to \mathbb{R}$ is also continuous as a function $\mathbb{R}^n \to \mathbb{R}$.

A function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^k$ is *continuous along lines* or *linearly continuous* at a point $\mathbf{p} \in \mathbb{R}^k$ if the restriction $\mathbf{f}^L$ of $\mathbf{f}$ to the line L passing through $\mathbf{p}$ is continuous for every such line L.

Continuity implies linear continuity, but not the converse.

Example. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} 1 & 0 < y < x^2 \\ 0 & \text{otherwise} \end{cases}$$

$f = 0$ over any sufficiently short line segment that passes through the point $(0,0)$, so $\lim_{x \to 0} f(x,ax) = 0$ along any straight line path and f is linearly continuous at $(0,0)$. But,

$$\lim_{n \to 0} f\left(n, \frac{1}{2}n^2\right) = 1 \neq 0 = f(0,0)$$

so f is discontinuous at $(0,0)$.

Linear continuity implies separate continuity, but not the converse.

Example. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} 1 & xy = 0 \\ 0 & xy \neq 0 \end{cases}$$

$g^0(x) = 1$ for all x, and $h^0(y) = 1$ for all y, so f is separately continuous at $(0,0)$. But,

$$\lim_{n \to 0} f(n,n) = 0 \neq 1 = f(0,0)$$

so f is not linearly continuous at $(0,0)$.

Continuity → Linear Continuity → Separate Continuity

3 Topology on $\mathbb{R}^n$

The *open ball* of radius $r > 0$ centred at a point $\mathbf{a} \in \mathbb{R}^n$, denoted by $\mathbb{B}_r(\mathbf{a})$ or $\mathbb{B}(\mathbf{a},r)$, to be the the set $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{a}\| < r\}$. We abbreviate $\mathbb{B}_r(\mathbf{0})$ to $\mathbb{B}(r)$, and $\mathbb{B}_1(\mathbf{0})$ (the *unit open ball*) to $\mathbb{B}$.

Similarly, the *closed ball* of radius $r > 0$ centred at a point $\mathbf{a} \in \mathbb{R}^n$, denoted by $\overline{\mathbb{B}_r(\mathbf{a})}$ or $\overline{\mathbb{B}(\mathbf{a},r)}$, to be the the set $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{a}\| \leq r\}$. We abbreviate $\overline{\mathbb{B}_r(\mathbf{0})}$ to $\overline{\mathbb{B}(r)}$, and $\overline{\mathbb{B}_1(\mathbf{0})}$ (the *unit closed ball*) to $\overline{\mathbb{B}}$.

A set $X \subseteq \mathbb{R}^n$ is *closed*, if for every sequence $(\mathbf{x}_i)_{i=1}^{\infty} \subseteq X$ of points in X that converges to a limit point $\mathbf{x} \in \mathbb{R}^n$, we also have $\mathbf{x} \in X$. That is, X is closed (in the algebraic sense) under sequential limits.

A set $U \subseteq \mathbb{R}^n$ is *open* if for all $\mathbf{x} \in U$, there exists $\varepsilon > 0$ such that $\mathbb{B}_\varepsilon(\mathbf{x}) \subset U$.

The empty set and $\mathbb{R}^n$ are both open and closed, or *clopen*.

Theorem 3.1. *A set is open if and only if its complement is closed.*

Theorem 3.2. *Open balls are open sets.*

Theorem 3.3. *Closed balls are closed sets.*

Theorem 3.4 (Arbitrary Union of Open Sets)**.** *If $(U_i)_{i \in I}$ is a (possibly uncountable) collection of open sets, then,*

$$\bigcup_{i \in I} U_i$$

is open.

Theorem 3.5 (Finite Intersection of Open Sets)**.** *If $(U_i)_{i=1}^{n}$ is a finite collection of open sets, then,*

$$\bigcap_{i=1}^{n} U_i$$

is open.

Corollary 3.5.1. *An arbitrary intersection or finite union of closed sets is closed.*

Let $E \subseteq \mathbb{R}^n$. Given $\varepsilon > 0$, the *ε-neighbourhood* $\mathcal{N}(E,\varepsilon)$ of E is defined by,

$$\mathcal{N}(E,\varepsilon) \coloneqq \bigcup_{\mathbf{x} \in E} \mathbb{B}(\mathbf{x},\varepsilon)$$

The ε-neighbourhood of a set is always open.

3.1 Continuity and Topology

Rewriting the $\varepsilon - \delta$ definition of continuity in terms of open sets, a function $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ is continuous at a point $\mathbf{p} \in U$ if,

$$\forall \varepsilon > 0, \exists \delta > 0 : \mathbf{f}\big((\mathbb{B}(\mathbf{p},\delta) \cap U)\big) \subset \mathbb{B}(\mathbf{f}(\mathbf{p}),\varepsilon)$$

Applying the inverse to both sides of the inclusion, we have,

$$\forall \varepsilon > 0, \exists \delta > 0 : \mathbb{B}(\mathbf{p},\delta) \cap U \subset \mathbf{f}^{-1}\big(\mathbb{B}(\mathbf{f}(\mathbf{p}),\varepsilon)\big)$$

This suggests the following alternative characterisations of continuity:

Theorem 3.6. *For any function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^k$, the following statements are equivalent:*

- $\mathbf{f}$ *is continuous at all points of $\mathbb{R}^n$.*

- $\mathbf{f}^{-1}(V)$ *is open whenever* $V \subseteq \mathbb{R}^n$ *is open.*
- $\mathbf{f}^{-1}(\mathcal{F})$ *is closed whenever* $\mathcal{F} \subseteq \mathbb{R}^n$ *is closed.*

Note that this does not imply that the image of an open (closed) set under a continuous function is open (resp. closed): only inverse images preserve the topology of a set.

3.2 Compactness

A set $K \subseteq \mathbb{R}^n$ is *sequentially compact* if every sequence $(\mathbf{x}_i)_{i=1}^{\infty} \subset K$ has a convergent subsequence $(\mathbf{x}_{i_j})$ whose limit is in K.

A set $K \subseteq \mathbb{R}^n$ is *bounded* if there exists some $M > 0$ such that $\|\mathbf{x}\| \leq M$ for all $\mathbf{x} \in K$.

Theorem 3.7. *A set* $K \subseteq \mathbb{R}^n$ *is sequentially compact if and only if it is closed and bounded.*

Theorem 3.8 (Continuity Preserves Sequential Compactness)**.** *If* $\mathbf{f} : K \to \mathbb{R}^k$ *is continuous and* K *is sequentially compact, then* $\mathbf{f}(K)$ *is also sequentially compact.*

Theorem 3.9 (Extreme Value Theorem)**.** *Let* $K \subset \mathbb{R}^n$ *be sequentially compact, and let* $f : K \to \mathbb{R}$ *be continuous. Then, there exists* $\mathbf{x}_*,\mathbf{x}^* \in K$ *in* K *such that*

$$f(\mathbf{x}_*) \leq f(\mathbf{x}) \leq f(\mathbf{x}^*)$$

for all x *in* K*.*

That is, a continuous real-valued function defined over a sequentially compact space attains its extreme values within that space.

Proof. Because f is continuous and K is sequentially compact, $f(K)$ is also sequentially compact, and is hence closed and bounded. Then, the values

$$U := \sup f(K), \quad L := \inf f(K)$$

are both finite and there exists sequences (a_i),$(b_i) \subset K$ such that $(a_i) \to L$ and $(b_i) \to U$. As $f(K)$ is closed, we have L,$U \in f(K)$, so $\mathbf{x}_* := f^{-1}(L)$ and $\mathbf{x}^* := f^{-1}(U)$ exist, and satisfy,

$$f(\mathbf{x}_*) = L \leq f(\mathbf{x}) \leq U = f(\mathbf{x}^*)$$

for all x in K, as required. ■

Corollary 3.9.1. *Let* $K \subset \mathbb{R}^n$ *be sequentially compact and let* $\mathbf{f} : K \to \mathbb{R}^k$ *be continuous. Then, there exists* $\mathbf{x}_*,\mathbf{x}^* \in K$ *in* K *such that*

$$\|f(\mathbf{x}_*)\| \leq \|f(\mathbf{x})\| \leq \|f(\mathbf{x}^*)\|$$

for all x *in* K*.*

Proof. The map $\|\cdot\| : \mathbb{R}^k \to \mathbb{R}$ is continuous, so $\mathbf{x} \mapsto \|f(\mathbf{x})\|$ is a continuous map $K \to \mathbb{R}$. ■

4 The Space $L(\mathbb{R}^n,\mathbb{R}^k)$ of Linear Maps

We denote the space of linear maps $T : \mathbb{R}^n \to \mathbb{R}^k$ by $L(\mathbb{R}^n,\mathbb{R}^k)$. If $n = k$, this is abbreviated to $L(\mathbb{R}^n)$. We denote the space of $k \times n$ matrices with real entries by $M(k \times n,\mathbb{R})$, also abbreviated to $\mathbb{R}^{k \times n}$.

We associate every matrix $\mathbf{A} \in \mathbb{R}^{k \times n}$ with a linear map $T \in L(\mathbb{R}^n,\mathbb{R}^k)$ defined by

$$T(\mathbf{x}) = \mathbf{A}\mathbf{x}$$

and we can write this association as a map $\mu : \mathbb{R}^{k\times n} \to L(\mathbb{R}^n,\mathbb{R}^k)$ that sends a matrix to the linear map it represents under the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^k$. Moreover, μ is a linear isomorphism.

We also have that

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{k,1} & \cdots & a_{k,n} \end{bmatrix} \cong [a_{1,1},\ldots,a_{1,n},a_{2,1},\ldots,a_{2,n},a_{3,1},\ldots,a_{k,1},\ldots,a_{k,n}]^\top$$

is a linear isomorphism, so,

$$\dim(L(\mathbb{R}^n,\mathbb{R}^k)) = \dim(\mathbb{R}^{k\times n}) = \dim(\mathbb{R}^{nk}) = nk$$

4.1 Matrix Norms

To discuss continuity of functions with matrix-valued inputs or outputs, we need to define a norm on $L(\mathbb{R}^n,\mathbb{R}^k)$, or equivalently, on $\mathbb{R}^{k\times n}$. In this section, we will write vector norms as $|\cdot|$, while matrix/linear map norms will be written as $\|\cdot\|$ for contrast.

The first such norm we might think of is to use the matrix-vector isomorphism above, and define the *Frobenius norm* $\|\cdot\|_F : \mathbb{R}^{k\times n} \to \mathbb{R}$ by,

$$\|(a_{i,j})\|_F := \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{n} a_{i,j}^2}$$

That is, treat the matrix as a vector, then calculate the ordinary Euclidean norm.

The *operator norm* $\|\cdot\|_{\text{op}} : L(\mathbb{R}^n,\mathbb{R}^k) \to \mathbb{R}$, also denoted by just $\|\cdot\|$, is defined by,

$$\|T\| := \sup_{\mathbf{x}\in\mathbb{R}^n\setminus\{0\}} \frac{|T(\mathbf{x})|}{|\mathbf{x}|}$$

Informally, the operator norm is the maximum factor by which the transformation lengthens vectors. That is, the operator norm satisfies,

$$|T(\mathbf{x})| \le \|T\||\mathbf{x}|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

Writing T as a matrix multiplication, by linearity, we have,

$$\begin{aligned} \frac{|\mathbf{Ax}|}{|\mathbf{x}|} &= \frac{1}{|\mathbf{x}|}\mathbf{Ax} \\ &= \left|\frac{1}{|\mathbf{x}|}\mathbf{Ax}\right| \\ &= \left|\mathbf{A}\left(\frac{\mathbf{x}}{|\mathbf{x}|}\right)\right| \end{aligned}$$

Because $\left|\frac{\mathbf{x}}{|\mathbf{x}|}\right| = 1$, this gives,

$$\|\mathbf{A}\| = \sup_{|\mathbf{x}|=1} |\mathbf{Ax}|$$

There are some more alternative characterisations of the operator norm for more general normed spaces. For instance, note that the above definitions are not well-defined if the codomain of the linear operator is the trivial space. Let $T : V \to W$ be a linear transformation with matrix $\mathbf{A}$. Then,

$$\|T\| = \inf\{M \ge 0 : |\mathbf{Av}| \le M|\mathbf{v}|,\mathbf{v} \in V\}$$

$$\begin{aligned}
&= \sup\{|\mathbf{Av}| : |\mathbf{v}| \leq 1, \mathbf{v} \in V\} \\
&= \sup\{|\mathbf{Av}| : |\mathbf{v}| < 1, \mathbf{v} \in V\} \\
&= \sup\{|\mathbf{Av}| : |\mathbf{v}| \in \{0,1\}, \mathbf{v} \in V\}
\end{aligned}$$

Theorem 4.1. *The operator norm is a norm. That is, it satisfies,*

- $\|T\| = 0 \leftrightarrow T = 0$ *(Point separating)*
- $\|\alpha T\| = |\alpha|\|T\|$ *(Absolute homogeneity)*
- $\|T + U\| \leq \|T\| + \|U\|$ *(Triangle inequality)*

Theorem 4.2. *For any linear transformation T with matrix $\mathbf{A}$, we have,*

$$\frac{1}{\sqrt{n}}\|\mathbf{A}\|_F \leq \|T\| \leq \|\mathbf{A}\|_F$$

Theorem 4.3 (Composition Bound)**.** *For any $A \in L(\mathbb{R}^n,\mathbb{R}^k)$ and $B \in L(\mathbb{R}^k,\mathbb{R}^m)$, the map $B \circ A \in L(\mathbb{R}^n,\mathbb{R}^m)$ satisfies,*

$$\|B \circ A\| \leq \|A\|\|B\|$$

Proof.

$$\begin{aligned}
|(B \circ A)(\mathbf{x})| &= |B(A(\mathbf{x}))| \\
&\leq \|B\||A(\mathbf{x})| \\
&\leq \|B\|\|A\||\mathbf{x}|
\end{aligned}$$

for all $\mathbf{x} \in \mathbb{R}^n$, so,

$$\begin{aligned}
\frac{|(B \circ A)(\mathbf{x})|}{|\mathbf{x}|} &\leq \sup \frac{|(B \circ A)(\mathbf{x})|}{|\mathbf{x}|} \\
&\leq \|B\|\|A\||\mathbf{x}|
\end{aligned}$$

■

4.2 Convergence and Continuity in $L(\mathbb{R}^n,\mathbb{R}^k)$

These are defined identically to sequences $(\mathbf{x}_i)_{i=1}^{\infty} \subset \mathbb{R}^n$ and functions $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$.

That is, a sequence $(T_i)_{i=1}^{\infty} \subset L(\mathbb{R}^n,\mathbb{R}^k)$ of linear transformations converges to $T \in L(\mathbb{R}^n,\mathbb{R}^k)$ is,

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} : i > N \to \|T_i - T\| < \varepsilon$$

We can also use the Frobenius norm in place of the operator norm here to similarly define convergence of sequences of matrices, and because $\mathbb{R}^{k \times n} \cong \mathbb{R}^{kn}$, this implies that both $\mathbb{R}^{k \times n}$ and $L(\mathbb{R}^n,\mathbb{R}^k)$ are both complete spaces, so every convergent sequence of linear transformations or matrices is also Cauchy.

4.3 Matrix-Valued Functions

A function $f : U \to \mathbb{R}^{k \times n}$ is continuous at $x \in U$ if,

$$\forall \varepsilon > 0 \exists \delta > 0 : |y - x| < \delta \to \|f(y) - f(x)\|_F < \varepsilon$$

Because the Frobenius norm on matrices in $\mathbb{R}^{k\times n}$ is equivalent to the Euclidean norm on vectors in $\mathbb{R}^{nk}$, we also have that a matrix-valued function is continuous if and only if it is componentwise continuous.

This also provides an easy way to check if a linear-transformation-valued function $f : U \to L(\mathbb{R}^n,\mathbb{R}^k)$ is continuous: check if every entry of the matrix representing the linear transformation output is continuous.

Theorem 4.4. *The map $\Delta : \mathbb{R}^{n\times n} \to \mathbb{R}$ that sends a matrix to its determinant is continuous with respect to the Frobenius norm on $\mathbb{R}^{n\times n}$.*

Proof. The determinant is a polynomial of degree n in its n^2 variables, and polynomials are continuous on $(\mathbb{R}^{n^2},|\cdot|) \cong (\mathbb{R}^{n\times n},\|\cdot\|_F)$. ■

4.4 The Space $GL(n,\mathbb{R}) \subset L(\mathbb{R}^n)$ of Invertible Linear Maps

It is clear that if a linear map $T : \mathbb{R}^n \to \mathbb{R}^k$ is a bijection, then $n = k$ and $\ker(T) = \{\mathbf{0}\}$. But by the rank-nullity theorem, the converse also holds. That is, a linear map $T : \mathbb{R}^n \to \mathbb{R}^k$ is a bijection if and only if $k = n$ and $\ker(T) = \{\mathbf{0}\}$.

The *general linear group* over the real numbers, denoted by $GL(n,\mathbb{R})$, is defined by,

$$GL(n,\mathbb{R}) := \{T \in L(\mathbb{R}^n) : T \text{ is invertible}\}$$

with the group operation given by composition. In terms of matrices, this is equivalent to,

$$GL(n,\mathbb{R}) := \{\mathbf{A} \in \mathbb{R}^{n\times n} : \det(\mathbf{A}) \neq 0\}$$

Note that $GL(1,\mathbb{R}) \cong (\mathbb{R}^*,\times)$.

Theorem 4.5. *The space $GL(n,\mathbb{R})$ is an open subset of $\mathbb{R}^{n\times n}$*

Proof. $GL(n,\mathbb{R}) = \Delta^{-1}(\mathbb{R}\setminus\{0\})$, and $\mathbb{R}\setminus\{0\}$ is open, so $GL(n,\mathbb{R})$ is open by the continuity of Δ. ■

$GL(n,\mathbb{R})$ being open means that invertibility of a linear map in $L(\mathbb{R}^n)$ is a stable property: a linear map can be perturbed somewhat, and remain invertible. The next theorem quantifies exactly how much a linear map A can be perturbed, in terms of $\|A^{-1}\|$.

Theorem 4.6. *Given $A \in GL(n,\mathbb{R})$, let $\alpha :== \frac{1}{\|A^{-1}\|}$. If $B \in L(\mathbb{R}^n)$ and $\|B - A\| < \alpha$, then B is invertible. That is, $\mathbb{B}_\alpha(A) \subset GL(n,\mathbb{R})$. Furthermore,*

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

Theorem 4.7. *The map $(\,\cdot\,)^{-1} : GL(n,\mathbb{R}) \to GL(n,\mathbb{R})$ defined by $A \mapsto A^{-1}$ is continuous.*

4.5 Lipschitz Continuity

A map $f : U \to \mathbb{R}^k$ is *Lipschitz continuous* on U if there exists an $M > 0$ such that,

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

for all $x,y \in U$.

The *Lipschitz constant* or *modulus of (uniform) continuity* M^* of f is then defined by,

$$M^* := \sup_{\substack{x\neq y\\ x,y\in U}} \frac{|f(x) - f(y)|}{|x - y|}$$

Intuitively, a Lipschitz continuous function is limited in how fast it can change: for any pair of distinct points, the absolute value of the gradient of the line connecting them is bounded by this Lipschitz constant.

Note that Lipschitz continuity of a function is a very strong form of continuity, and it implies uniform (and hence regular) continuity of the function:

$$\forall \varepsilon > 0 : |x-y| < \frac{\varepsilon}{M} \to |f(x) - f(y)| < \varepsilon$$

Theorem 4.8. *Every linear map T is continuous.*

Proof. By linearity, $T(x) - T(y) = T(x-y)$, and so,

$$|T(x) - T(y)| = |T(x-y)| \leq \|T\||x-y|$$

■

Theorem 4.9. *The map $|\cdot| : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is Lipschitz continuous with Lipschitz constant $M^* = 1$.*

Proof. By the reverse triangle inequality, we have,

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

■

The same holds for any norm, so the operator norm and Frobenius norm are both Lipschitz continuous with Lipschitz constant $M^* = 1$.

5 The Derivative

In this section, $U \subseteq \mathbb{R}^n$ will be an open subset of $\mathbb{R}^n$. This means that if $\mathbf{p} \in U$, then in any limit $\lim_{\mathbf{x}\to\mathbf{p}}$, $\mathbf{x}$ may approach $\mathbf{p}$ from any direction.

5.1 Partial Derivatives

A *partial derivative* of a multivariate function is its derivative with respect to one of its variables, with the other variables held constant.

Let $\{\mathbf{e}_i\}_{i=1}^n$ be the standard basis of $\mathbb{R}^n$. For any function $f : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ with U open, a partial derivative of f at the point $\mathbf{x} \in U$ with respect to the i-th variable x_i is defined as,

$$\begin{aligned}\frac{\partial f}{\partial x_i} &= \lim_{h\to 0} \frac{f(x_1,x_2,\ldots,x_{i-1},x_i+h,x_{i+1},\ldots,x_n) - f(x_1,x_2,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n)}{h} \\ &= \lim_{t\to 0} \frac{f(\mathbf{x}+h\mathbf{e}_i) - f(\mathbf{x})}{h}\end{aligned}$$

Other notations include $\partial_{x_i} f(x)$ or $\partial_i f(x)$. If f is a function of only a few variables, then it is more common to write, say $f(x,y,z)$, rather than $f(x_1,x_2,x_3)$, and we write f_x for the partial derivative of f with respect to x.

Theorem 5.1 (Algebra of Partial Derivatives)**.** *If $f,g : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$, then,*

- $\partial_i(f+g) = \partial_i f + \partial_i g$;
- $\partial_i(fg) = (\partial_i f)g + f\partial_i g$.

5.2 Directional Derivatives

The rate of change of a multivariable function depends on the direction in which the change is measured.

Given a direction vector $\overrightarrow{\mathbf{v}} \in \mathbb{R}^n$ and a point $\mathbf{x} \in \mathbb{R}^n$, the line $L_{\mathbf{x},\overrightarrow{\mathbf{v}}}$ passing through $\mathbf{x}$ in the direction of $\overrightarrow{\mathbf{v}}$ is parametrised by $\mathbf{r}(t) = \mathbf{x} + t\overrightarrow{\mathbf{v}}$. Now, for any function $f : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ with U open, there exists $\tau > 0$ such that the line segment $\mathbf{x} + t\overrightarrow{\mathbf{v}}$ is contained in U for $t \in (-\tau,\tau)$. The restricton $f_{\mathbf{x},\overrightarrow{\mathbf{v}}} : (-\tau,\tau) \to \mathbb{R}^k$ of f to this line segment is defined by,

$$f_{\mathbf{x},\overrightarrow{\mathbf{v}}}(t) \coloneqq f(\mathbf{x} + t\overrightarrow{\mathbf{v}})$$

This is now a function of a single real variable, so we can differentiate it componentwise.

The *directional derivative* of f in the direction of $\mathbf{v}$, denoted by $D_{\mathbf{v}}f(x)$ or $\partial_{\mathbf{v}}f(x)$, is defined by,

$$\begin{aligned}
\partial_{\mathbf{v}}f(\mathbf{x}) &\coloneqq \frac{d}{dt} f_{\mathbf{x},\mathbf{v}}(t)\bigg|_{t=0} \\
&= \frac{d}{dt} f(\mathbf{x} + t\mathbf{v})\bigg|_{t=0} \\
&= \lim_{t\to 0} \frac{f(\mathbf{x} + t\overrightarrow{\mathbf{v}}) - f(\mathbf{x})}{t}
\end{aligned}$$

In practice, you can calculate the directional derivative by multiplying the components of the normalised direction vector by the corresponding partial derivatives, or equivalently, by calculating the scalar product of the gradient and the direction vector: $\partial_{\mathbf{v}} f = \nabla f \cdot \mathbf{v}$ (where $\mathbf{v}$ is a unit vector).

Example. Find the directional derivative of $f(x,y) = x^2 - y^2$ in the direction of $\mathbf{v} = (a,b)$.

Since we are not given values for a and b, we do not modify $\mathbf{v}$, but in general, we would normalise $\mathbf{v}$ first.

We compute the directional derivative from the definition:

$$\begin{aligned}
\frac{d}{dt} f\big((x,y) + t(a,b)\big)\bigg|_{t=0} &= \frac{d}{dt} f(x + ta, y + tb)\bigg|_{t=0} \\
&= \frac{d}{dt}\Big[(x + ta)^2 - (y + tb)^2\Big]\bigg|_{t=0} \\
&= 2a(x + ta) - 2b(y + tb)\Big|_{t=0} \\
&= 2ax - 2by
\end{aligned}$$

Alternatively, we can compute the partial derivatives (the components of ∇f);

$$\begin{aligned}
\frac{\partial}{\partial x} f(x,y) &= 2x \\
\frac{\partial}{\partial y} f(x,y) &= -2y
\end{aligned}$$

then multiply by the components of $\mathbf{v} = (a,b)$,

$$\nabla f \cdot \mathbf{v} = 2ax - 2by$$

For a function $f : \mathbb{R}^n \to \mathbb{R}$, the directional derivative existing for all $\mathbf{v} \in \mathbb{R}^n$ at a point $\mathbf{x}$ does *not* imply that f is continuous at $\mathbf{x}$, similarly to how linear continuity does not imply continuity.

5.3 The Fréchet Derivative

The derivative of a function $f : (a,b) \to \mathbb{R}$ at a point $x \in (a,b)$ is given by

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

This definition cannot be easily extended to functions $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^k$, as there is no notion of division for vectors, unlike for real (or complex) numbers.

Instead, we rearrange the above to,

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

That is, for a fixed x, the nonlinear mapping $h \mapsto f(x+h)$ is locally approximated by the affine linear map $h \mapsto f(x) + f'(x)h$.

Extending this idea to multivariate functions, a function $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ is differentiable at $\mathbf{x} \in U$ if there exists a linear map $T \in L(\mathbb{R}^n,\mathbb{R}^k)$ such that

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

and we say that the linear map T is the *Fréchet derivative* of $\mathbf{f}$, also denoted by $D\mathbf{f}(\mathbf{x})$.

Expanding the $\varepsilon - \delta$ definition of the limit, we equivalently have: a function $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ is differentiable at $\mathbf{x} \in U$ if there exists a linear map $T \in L(\mathbb{R}^n,\mathbb{R}^k)$ such that,

$$\forall \varepsilon > 0, \exists \delta > 0 : |\mathbf{h}| < \delta \to \big|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \big(\mathbf{f}(\mathbf{x}) + T(\mathbf{h})\big)\big| < \varepsilon |\mathbf{h}|$$

Another characterisation is: a function $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ is differentiable at $\mathbf{x} \in U$ if there exists a linear map $T \in L(\mathbb{R}^n,\mathbb{R}^k)$ such that,

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) = \mathbf{f}(\mathbf{x}) + T(\mathbf{h}) + \mathbf{E}(\mathbf{h})$$

where the error $\mathbf{E}(\mathbf{h}) \in o(\mathbf{h})$ grows asymptotically slower than linearly in $\mathbf{h}$.

Theorem 5.2. *If $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ is differentiable at $\mathbf{x} \in U$, then $\mathbf{f}$ is continuous at $\mathbf{x}$.*

Proof. As $\mathbf{f}$ is differentiable at $\mathbf{x}$, for all $\varepsilon > 0$, there exists $\delta > 0$ such that,

$$\begin{aligned} |\mathbf{h}| < \delta &\to |\mathbf{f}(\mathbf{x}+\mathbf{h}) - (\mathbf{f}(\mathbf{x}) + D\mathbf{f}(\mathbf{x})\mathbf{h})| \le \varepsilon |\mathbf{h}| \\ &\to |\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x})| \le |D\mathbf{f}(\mathbf{x})\mathbf{h}| + \varepsilon |\mathbf{h}| \\ &\to |\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x})| \le (\|D\mathbf{f}(\mathbf{x})\| + \varepsilon)|\mathbf{h}| \end{aligned}$$

Set $\delta_* := \min\big(\delta, \epsilon/(\|D\mathbf{f}(\mathbf{x})\| + \varepsilon)\big)$. Then, if $|h| < \delta_*$, we have $|h| < \delta$, so,

$$|h| < \delta_* \to |\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x})| < (\|D\mathbf{f}(\mathbf{x})\| + \varepsilon)\delta_* < \epsilon$$

■

Theorem 5.3 (Componentwise Differentiability)**.** *A function $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ defined by*

$$(x_1, \dots, x_n) \mapsto \big(f_1(x_1, \dots, x_n), f_2(x_1, \dots, x_n), \dots, f_k(x_1, \dots, x_n)\big)$$

where $(f_j)_{j=1}^k$ are real-valued functions, is differentiable at $\mathbf{p} \in U$ if and only if every f_j is differentiable at $\mathbf{p}$.

That is, $\mathbf{f}$ is differentiable if and only if every component f_j is individually differentiable.

Theorem 5.4. *For a function* $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$*, if* $D\mathbf{f}(\mathbf{x})$ *exists, then* $\partial_{\mathbf{v}}\mathbf{f}(\mathbf{x})$ *exists for all* $\mathbf{v} \in \mathbb{R}^n$*, and* $\partial_{\mathbf{v}}\mathbf{f}(\mathbf{x}) = D\mathbf{f}(\mathbf{x})\mathbf{v}$*.*

In particular, if $\mathbf{f}$ *is differentiable at* $\mathbf{x}$*, then* $\partial_{\mathbf{v}}\mathbf{f}(\mathbf{x})$ *is linear in* $\mathbf{v}$*. That is,*

$$\partial_{\alpha\mathbf{v}+\beta\mathbf{w}}\mathbf{f}(\mathbf{x}) = \alpha\partial_{\mathbf{v}}\mathbf{f}(\mathbf{x}) + \beta\partial_{\mathbf{w}}\mathbf{f}(\mathbf{x})$$

for all $\alpha,\beta \in \mathbb{R}$ *and* $\mathbf{v},\mathbf{w} \in \mathbb{R}^n$*.*

Note that the converse of this theorem does not hold – all directional derivatives existing does not guarantee that $\mathbf{f}$ is differentiable.

5.4 Gradient

Let $f : \mathbb{R}^n \to \mathbb{R}$. The *gradient* of f, denoted $\operatorname{grad} f$ or ∇f is the vector,

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

$\nabla : \mathbb{R} \to \mathbb{R}^n$ by itself is the *grad operator*, and is effectively a vector full of partial derivative operators.

The *Jacobian matrix* of a function, $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^k$, denoted $\mathbf{J}$, $D\mathbf{f}$, or $\partial\mathbf{f}$, is the matrix of partial derivatives,

$$\partial\mathbf{f} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3} & \cdots & \frac{\partial f_3}{\partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1} & \frac{\partial f_k}{\partial x_2} & \frac{\partial f_k}{\partial x_3} & \cdots & \frac{\partial f_k}{\partial x_n} \end{bmatrix}$$

This can be more compactly written as,

$$\partial\mathbf{f} = \begin{bmatrix} \frac{\partial\mathbf{f}}{\partial x_1} & \cdots & \frac{\partial\mathbf{f}}{\partial x_n} \end{bmatrix}$$

or,

$$\partial\mathbf{f} = \begin{bmatrix} \nabla f_1 \\ \vdots \\ \nabla f_k \end{bmatrix}$$

Theorem 5.5. *If* $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ *is differentiable at* $\mathbf{x} \in U$*, and* $\mathbf{h} \in \mathbb{R}^n$*, then,*

$$D\mathbf{f}(\mathbf{x})(\mathbf{h}) = \partial\mathbf{f}(\mathbf{x})\mathbf{h}$$

On the left side, we have the linear map $D\mathbf{f}$ given by the Fréchet derivative, and on the right side, we have the Jacobian matrix, so this theorem just says that the Fréchet derivative is represented by the Jacobian matrix if $\mathbf{f}$ is already known to be differentiable at $\mathbf{x}$.

Theorem 5.6. *When f is differentiable at x, $Df(x)(h) = \partial_h f(x) = \partial f(x)h$.*

That is, whenever f is differentiable at x, the Fréchet derivative $Df(x)$ centred at x evaluated at h, the directional derivative $\partial_h f(x)$ of f at x in the direction of h, and the Jacobian matrix evaluated at x multiplied by h are all equal.

If f is not differentiable at x – that is, the Fréchet derivative $Df(x)$ does not exist – then directional derivative $\partial_h f(x)$ and the Jacobian $\partial f(x)$ may both still exist, but may not necessarily be equal.

However, if all partial derivatives are continuous at x (and hence the Jacobian is also continuous at x), then $Df(x)$ is guaranteed to exist:

Theorem 5.7. *Consider $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ and suppose there exists $\mathbb{B}_r(\mathbf{x}) \subset U$ such that the Jacobian matrix $\partial\mathbf{f}(\mathbf{y})$ exists at all points $\mathbf{y} \in \mathbb{B}_r(\mathbf{x})$ and that $\partial\mathbf{f}$ is continuous at $\mathbf{x}$. Then, $\mathbf{f}$ is differentiable at $\mathbf{x}$ and the Fréchet derivative is equal to the Jacobian matrix*

$$D\mathbf{f}(\mathbf{x})(\mathbf{h}) = \partial\mathbf{f}(\mathbf{x})\mathbf{h}$$

for all $\mathbf{h} \in \mathbb{R}^n$:

5.5 Geometric Approximation

Let $\mathbf{r} : [a,b] \to \mathbb{R}^k$ be a continuously differentiable parametrisation of a curve $C = r([a,b]) \subset \mathbb{R}^k$. Furthermore, suppose $\mathbf{r}$ is a *regular* parametrisation – that is, $\mathbf{r}'(t) \neq 0$ for all t. We can then interpret $\mathbf{r}'(t)$ to be the vector tangent to C at $\mathbf{r}(t)$, or alternatively, we can view $\mathbf{r}(t)$ to be the position of a particle tracing out C, and $\mathbf{r}'(t)$ is the velocity of the particle.

The line $L_{\mathbf{r}(t)}$ tangent to C at $\mathbf{r}(t)$ is then parametrised by,

$$\ell(h) = \mathbf{r}(t) + \mathbf{r}'(t)h$$

But, $\mathbf{r}'(t) = \partial\mathbf{r}(t)$, so the affine linear approximation of $h \mapsto \mathbf{r}(t+h)$ by $h \mapsto \mathbf{r}(t) + \big(\partial\mathbf{r}(t)\big)(h) = \ell(h)$ is a parametrisation of the tangent line $L_{\mathbf{r}(t)}$. That is, this approximation using Jacobian, for small h, corresponds to geometrically approximating the curve C by $L_{\mathbf{r}(t)}$ near $\mathbf{r}(t)$. This also holds true for more general parametrisations.

Let $U \subseteq \mathbb{R}^n$ be open, and let $\mathbf{r} : U \to \mathbb{R}^3$ be a continuously differentiable parametrisation of a surface $S = \mathbf{r}(U) \subset \mathbb{R}^3$. Furthermore, suppose $\mathbf{r}$ is a regular parametrisation – that is, $\partial\mathbf{r}(\mathbf{x})$ is of rank 2 for all $\mathbf{x} \in U$. If,

$$\mathbf{r}(u,v) = \begin{bmatrix} x(u,v) \\ y(u,v) \\ z(u,v) \end{bmatrix}$$

then,

$$\mathbf{r}_u = \begin{bmatrix} x_u \\ y_u \\ z_u \end{bmatrix}, \qquad \mathbf{r}_v = \begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix}$$

(recall $\mathbf{r}_u = \frac{\partial \mathbf{r}}{\partial u}$, $x_u = \frac{\partial x}{\partial u}$, etc.) The Jacobian is given by,

$$\partial\mathbf{r} = \begin{bmatrix} x_u & x_v \\ y_u & y_v \\ z_u & z_v \end{bmatrix}$$

So, ∂ is of rank 2 if and only if $\mathbf{r}_u$ and $\mathbf{r}_v$ are linearly independent.

As in the 2-dimensional case, the affine linear approximation of $(h,k) \mapsto \mathbf{r}(u+h,v+k)$ by,

$$\begin{aligned}(h,k) \mapsto \mathbf{r}(u,v) &+ \big(\partial\mathbf{r}(u,v)\big)(h,k) \\ &= \mathbf{r}(u,v) + h\mathbf{r}_u(u,v) + k\mathbf{r}_v(u,v)\end{aligned}$$

is then a parametrisation of the plane $\Pi_{\mathbf{r}(u,v)}$ tangent to S at $\mathbf{r}(u,v)$.

5.5.1 Graphs

Given a function $f : (U \subseteq \mathbb{R}^2) \to \mathbb{R}$, the *graph*, $\mathcal{G}_f$ of f is the surface parametrised by,

$$\mathbf{r}(x,y) = \big(x,y,f(x,y)\big)$$

That is, the height of the surface above the x-y plane is the value of $f(x,y)$, analogous to the 2-dimensional case where we plot the points given by $\big(x,f(x)\big)$.

Note that $\mathbf{r}_x = (1,0,f_x)$ and $\mathbf{r}_y = (0,1,f_y)$ are linearly independent for any function f.

A parametrisation of the plane tangent to the surface $\mathcal{G}_f$ at $\big(x,y,f(x,y)\big)$ is given by,

$$\begin{aligned}(h,k) \mapsto \mathbf{r}(x,y) &+ \big(D\mathbf{r}(x,y)\big)(h,k) \\ &= \begin{bmatrix} x \\ y \\ f(x,y) \end{bmatrix} + h\begin{bmatrix} 1 \\ 0 \\ f_x \end{bmatrix} + k\begin{bmatrix} 0 \\ 1 \\ f_y \end{bmatrix} \\ &= \begin{bmatrix} x+h \\ y+k \\ f(x,y) + hf_x + kf_y \end{bmatrix} \\ &= \begin{bmatrix} x+h \\ y+k \\ f(x,y) + (h,k)\cdot\nabla f(x,y) \end{bmatrix}\end{aligned}$$

so f is not differentiable at $(x_0,y_0) \in U$ if and only if $\mathcal{G}_f$ does not have a tangent plane at $\big(x_0,y_0,f(x_0,y_0)\big)$.

5.6 Differentiation of Matrix-Valued Functions

$L(\mathbb{R}^n,\mathbb{R}^k) \cong \mathbb{R}^{k\times n} \cong \mathbb{R}^{nk}$, so the Fréchet derivative applies similarly to functions with domains and codomains in these spaces, the only difference being that the Euclidean norm $|\cdot|$ in the definition needs to be replaced by the operator norm $\|\cdot\|$ or Frobenius norm $\|\cdot\|_F$, respectively.

Example. Find the derivative of the map $f : L(\mathbb{R}^n) \to L(\mathbb{R}^n)$ defined by $f(T) = T \circ T = T^2$.

We consider $f(A+H) - f(A)$:

$$\begin{aligned}f(A+H) - f(A) &= (A+H)(A+H) - A^2 \\ &= A^2 + AH + HA + H^2 - A^2 \\ &= AH + HA + H^2\end{aligned}$$

The terms linear in H are $AH+HA$, so we should think that $\big(Df(A)\big)(H) = AH+HA$ is the derivative. However, we need to verify that it satisfies the required limit. First, rearrange to obtain,

$$f(A+H) - f(A) - (AH+HA) = H^2$$

Now verify the limit:

$$\begin{aligned}\lim_{H\to 0}\frac{\left\|f(A+H)-f(A)-\big(Df(A)\big)(H)\right\|}{\|H\|} &= \lim_{H\to 0}\frac{\|H^2\|}{\|H\|}\\ &\le \lim_{H\to 0}\frac{\|H\|^2}{\|H\|}\\ &= \lim_{H\to 0}\|H\|\\ &= 0\end{aligned}$$

so $\big(Df(A)\big)(H) = AH + HA$.

If we interpret f to act on matrices, we could also note that the entries of $f(\mathbf{A})$ are quadratic polynomials in the entries of $\mathbf{A}$, and are hence continuous. It then follows that f is differentiable, and $\big(Df(\mathbf{A})\big)(\mathbf{H}) = \partial_{\mathbf{H}} f(\mathbf{A})$, so we could calculate the directional derivative instead:

$$\begin{aligned}\partial_{\mathbf{H}} f(\mathbf{A}) &= \frac{d}{dt} f(\mathbf{A}+t\mathbf{H})\Big|_{t=0}\\ &= \frac{d}{dt}(\mathbf{A}+t\mathbf{H})^2\Big|_{t=0}\\ &= \frac{d}{dt}\mathbf{A}^2 + t\mathbf{AH} + t\mathbf{HA} + t^2\mathbf{H}^2\Big|_{t=0}\\ &= \mathbf{AH} + \mathbf{HA} + 2t\mathbf{H}^2\Big|_{t=0}\\ &= \mathbf{AH} + \mathbf{HA}\end{aligned}$$

6 The Chain Rule

Theorem 6.1 (Chain Rule)**.** *Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^k$ both be open. Suppose $\mathbf{f} : U \to \mathbb{R}^k$ is differentiable at $\mathbf{x} \in U$, and that $\mathbf{f}(\mathbf{x}) \in V$. If $\mathbf{g} : V \to \mathbb{R}^k$ is differentiable at $\mathbf{f}(\mathbf{x})$, then the composition $\mathbf{g}\circ\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^k$ is differentiable at $\mathbf{x}$, and,*

$$D(\mathbf{g}\circ\mathbf{f})(x) = D\mathbf{g}\big(\mathbf{f}(\mathbf{x})\big)\circ D\mathbf{f}(\mathbf{x})$$

Theorem 6.2. *Given $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$, $\mathbf{x} \in U$, and $r > 0$ such that $\mathbb{B}_r(\mathbf{x}) \subset U$ and $T \in L(\mathbb{R}^n,\mathbb{R}^k)$, we define $\Delta_{\mathbf{x},T} f : \mathbb{B}_r(\mathbf{0}) \to \mathbb{R}^k$ by,*

$$\Delta_{\mathbf{x},T} f(\mathbf{h}) = \begin{cases}\frac{\mathbf{f}(\mathbf{x}+\mathbf{h})-\mathbf{f}(\mathbf{x})-T(\mathbf{h})}{|\mathbf{h}|} & \mathbf{h}\neq\mathbf{0}\\ 0 & \mathbf{h}=\mathbf{0}\end{cases}$$

Then, f is differentiable at $\mathbf{x}$ with $D\mathbf{f}(\mathbf{x}) = T$ if and only if $\Delta_{\mathbf{x},T} f$ is continuous at $\mathbf{0}$.

Recall that the linear isomorphism $\mu : L(\mathbb{R}^n,\mathbb{R}^k) \to \mathbb{R}^{k\times n}$ maps linear transformations to their matrices. Applying this to the chain rule above gives a form of the chain rule with Jacobian matrices:

Theorem 6.3 (Chain Rule)**.** *Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^k$ both be open. Suppose $\mathbf{f} : U \to \mathbb{R}^k$ is differentiable at $\mathbf{x} \in U$, and that $\mathbf{f}(\mathbf{x}) \in V$. If $\mathbf{g} : V \to \mathbb{R}^k$ is differentiable at $\mathbf{f}(\mathbf{x})$, then the composition $\mathbf{g}\circ\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^k$ is differentiable at $\mathbf{x}$, and,*

$$\partial(\mathbf{g}\circ\mathbf{f})(\mathbf{x}) = \partial\mathbf{g}\big(\mathbf{f}(\mathbf{x})\big)\partial\mathbf{f}(\mathbf{x})$$

Given functions $f : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, the ith partial derivative of $g\circ f$ can be computed with the above chain rule as,

$$\partial_i(g\circ f)(\mathbf{x}) = g'(f(\mathbf{x}))\partial_i f(\mathbf{x})$$

so,

$$\nabla(g \circ f)(x) = g'(f(\mathbf{x}))\nabla f(\mathbf{x})$$

Example. Calculate $\nabla|\mathbf{x}|$, $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$.

$$|\mathbf{x}| = \sqrt{|\mathbf{x}|^2}$$

so we can apply the chain rule with $f(\mathbf{x}) = |\mathbf{x}|^2 = \sum_{i=1}^n x_i^2$ and $g(t) = \sqrt{t}$. First calculate ∇f and g':

$$\begin{aligned}
\nabla f &= \nabla(x_1^2 + x_2^2 + \cdots + x_n^2) \\
&= \begin{bmatrix} \partial_1(x_1^2 + x_2^2 + \cdots + x_n^2) \\ \partial_2(x_1^2 + x_2^2 + \cdots + x_n^2) \\ \vdots \\ \partial_n(x_1^2 + x_2^2 + \cdots + x_n^2) \end{bmatrix} \\
&= \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} \\
&= 2\mathbf{x} \\
g'(t) &= \frac{1}{2\sqrt{t}}
\end{aligned}$$

$$\begin{aligned}
\nabla|\mathbf{x}| &= \nabla(g \circ f)(\mathbf{x}) \\
&= g'(f(\mathbf{x}))\nabla f(\mathbf{x}) \\
&= \frac{1}{2\sqrt{|\mathbf{x}|^2}} 2\mathbf{x} \\
&= \frac{\mathbf{x}}{|\mathbf{x}|}
\end{aligned}$$

with component form given by,

$$\frac{\partial}{\partial x_i}|\mathbf{x}| = \frac{x_i}{|\mathbf{x}|}$$

6.1 The Space $C^n(U,\mathbb{R}^k)$ of Continuously Differentiable Functions

Suppose $\mathbf{f} : (U \subseteq \mathbb{R}^n) \to \mathbb{R}^k$ is differentiable on U. Then, $\mathbf{f}$ is *continuously differentiable* at $\mathbf{p} \in U$ if the map $D\mathbf{f} : U \to L(\mathbb{R}^n,\mathbb{R}^k)$ defined by $\mathbf{x} \mapsto D\mathbf{f}(\mathbf{x})$ is continuous at $\mathbf{p}$. That is,

$$\forall \varepsilon - 0 \exists \delta > 0 : |\mathbf{x} - \mathbf{p}| < \delta \to \|D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{p})\| < \varepsilon$$

Theorem 6.4. *A function $\mathbf{f} : U \to \mathbb{R}^k$ is continuously differentiable on U if and only if the Jacobian matrix $\partial\mathbf{f} : U \to \mathbb{R}^{k \times n}$ is continuous on U.*

This means that we can check if a function is continuously differentiable by computing all first order partial derivatives $\partial_i f_j$ of $\mathbf{f} = (f_1, \ldots, f_k)$ and verifying that they are all continuous.

6.2 Mean Value Inequality

For any vector-valued function of a single real variable, $\mathbf{f} : [a,b] \to \mathbb{R}^k$, $\mathbf{f}(t) = (f_1(t), f_2(t), \dots, f_k(t))$, we define the integral of $\mathbf{f}$ as,

$$\int_a^b \mathbf{f}(t)\,dt = \begin{bmatrix} \int_a^b f_1(t)\,dt \\ \int_a^b f_2(t)\,dt \\ \vdots \\ \int_a^b f_k(t)\,dt \end{bmatrix}$$

Lemma 6.5. *For any function $\mathbf{f} : [a,b] \to \mathbb{R}^k$,*

$$\left| \int_a^b \mathbf{f}(t)\,dt \right| \leq \int_a^b |\mathbf{f}(t)|\,dt$$

Proof. Let $\mathbf{I} := \int_a^b \mathbf{f}(t)\,dt \in \mathbb{R}^k$. If $\mathbf{I} = \mathbf{0}$, then we have equality. Otherwise,

$$\begin{aligned} |\mathbf{I}| \left| \int_a^b \mathbf{f}(t)\,dt \right| &= |\mathbf{I}|^2 \\ &= \mathbf{I} \cdot \mathbf{I} \\ &= \mathbf{I} \cdot \int_a^b \mathbf{f}(t)\,dt \\ &= \int_a^b \mathbf{I} \cdot \mathbf{f}(t)\,dt \\ &\leq \int_a^b |\mathbf{I}||\mathbf{f}(t)|\,dt \\ &= |\mathbf{I}| \int_a^b |\mathbf{f}(t)|\,dt \end{aligned}$$

Dividing the first and last terms by $|\mathbf{I}|$ provides the result. ■

Theorem 6.6 (Generalised Mean Value Inequality)**.** *Suppose that $\mathbf{x},\mathbf{y} \in U \subseteq \mathbb{R}^n$ can be joined by a continuously differentiable path, $\mathbf{r} : [a,b] \to U$, $\mathbf{r}(a) = \mathbf{x}$, $\mathbf{r}(b) = \mathbf{y}$. Suppose that $f \in C^1(U,\mathbb{R}^k)$, and that there exists $M \geq 0$ such that the Jacobian satisfies $\|\partial \mathbf{f}(\mathbf{x})\| \leq M$ for all $\mathbf{x} \in U$. Then,*

$$|\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})| \leq M \operatorname{length}\big(\mathbf{r}([a,b])\big)$$

Proof.

$$\begin{aligned} \mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x}) &= \mathbf{f}\big(\mathbf{r}(b)\big) - \mathbf{f}\big(\mathbf{r}(a)\big) \\ &= \int_a^b \frac{d}{dt}\mathbf{f}\big(\mathbf{r}(t)\big)\,dt && \text{[Fundamental Theorem of Calculus II]} \\ &= \int_a^b \partial \mathbf{f}\big(\mathbf{r}(t)\big)\mathbf{r}'(t)\,dt && \text{[Chain Rule]} \end{aligned}$$

so by the lemma above,

$$|\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})| = \left| \int_a^b \partial \mathbf{f}\big(\mathbf{r}(t)\big)\mathbf{r}'(t)\,dt \right|$$

$$
\begin{aligned}
&\leq \int_a^b \left|\partial \mathbf{f}\big(\mathbf{r}(t)\big)\mathbf{r}'(t)\right| dt \\
&\leq \int_a^b \left\|\partial \mathbf{f}\big(\mathbf{r}(t)\big)\right\| |\mathbf{r}'(t)| \, dt \\
&\leq \int_a^b M|\mathbf{r}'(t)| \, dt \\
&= M \operatorname{length}\big(\mathbf{r}([a,b])\big)
\end{aligned}
$$

■

Corollary 6.6.1 (Vanishing Derivative)**.** *Suppose that $U \subset \mathbb{R}^n$ is differentiably path-connected and that $\mathbf{f} : U \to \mathbb{R}^k$ satisfies $\partial \mathbf{f}(\mathbf{x}) = 0$ for all $\mathbf{x} \in U$. Then, $\mathbf{f}$ is constant on U.*

Proof. Fix a point $\mathbf{y} \in U$. Then, by differentiable path-connectedness, given $\mathbf{x} \in U$, there exists a continuously differentiable ;ath $\mathbf{r} : [a,b] \to U$ joining $\mathbf{x}$ to $\mathbf{y}$. So, by the generalised mean value inequality with $\partial \mathbf{f} = \mathbf{0}$, $\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{y})$ for all $\mathbf{x} \in U$. ■

This corollary does not hold if U is not path-connected, but the converse is true even if U is not path-connected.

For scalar valued functions $f : U \to \mathbb{R}$, this corollary can be stated as,

Corollary 6.6.2. *If $\nabla f(\mathbf{x}) = \mathbf{0}$ at all points $\mathbf{x}$ of a path-connected open set, then f is constant.*

A set $U \subseteq \mathbb{R}^n$ is *convex* if for all $\mathbf{x},\mathbf{y} \in U$, the line,

$$L_{\mathbf{x},\mathbf{y}} = \{t\mathbf{x} + (1-t)\mathbf{y} : 0 \leq t \leq 1\}$$

is contained within U.

Corollary 6.6.3 (Mean Value Inequality)**.** *Let $U \subseteq \mathbb{R}^n$ be convex, and suppose that $\mathbf{f} \in C^1(U,\mathbb{R}^k)$ satisfies $\|\partial \mathbf{f}(\mathbf{x})\| \leq M$ for all $\mathbf{x} \in U$ for some $M \geq 0$. Then,*

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

That is, $\mathbf{f}$ is Lipschitz continuous.

Proof. The result follows from the generalised mean value inequality with $\operatorname{length}(L_{\mathbf{x},\mathbf{y}}) = |x-y|$. ■

The converse of the mean value inequality does not hold. That is, a function $\mathbf{f}$ being Lipschitz continuous does not imply that $\mathbf{f}$ is differentiable. For example, $(x,y) \mapsto \frac{x^3}{}$ is Lipschitz continuous on all of $\mathbb{R}^n$, but is not differentiable at $\mathbf{0}$ because none of the partial derivatives $\partial_i|\mathbf{x}|$ exist at $\mathbf{0}$. The function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by,

$$(x,y) \mapsto \frac{x^3}{x^2+y^2}$$

is also Lipschitz over all of $\mathbb{R}^2$, and, unlike $\mathbf{x} \mapsto |\mathbf{x}|$, has partial derivatives that exist everywhere, but is still not differentiable at (0,0).

7 Vector Fields

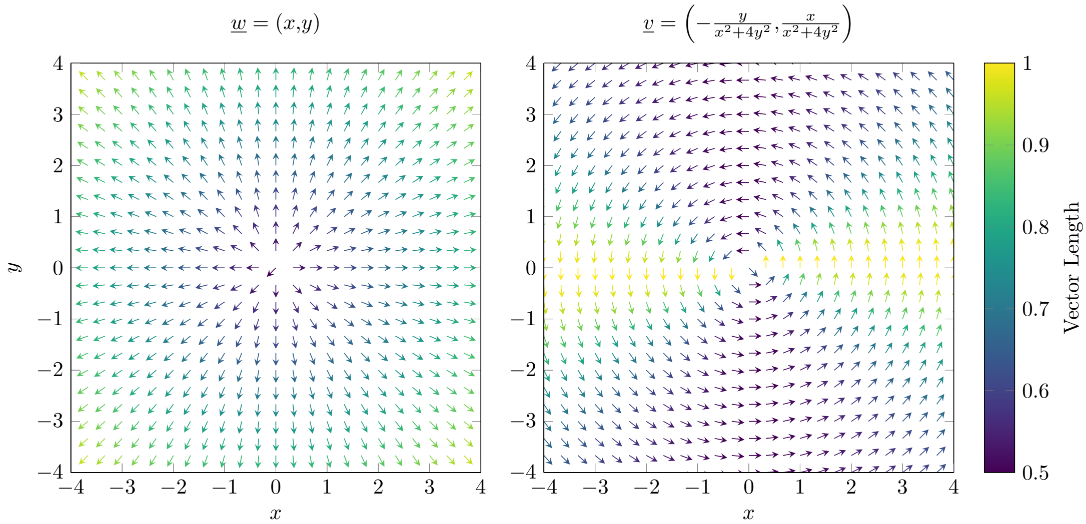
In this section, U will be a path-connected open subset of $\mathbb{R}^n$.

A *vector field* $\underline{v}$ on $U \subseteq \mathbb{R}^n$ is a function $\underline{v} : U \to \mathbb{R}^n$, so a vector field consists of n functions of n variables:

$$\underline{v}(\mathbf{x}) = \begin{bmatrix} v_1(x_1,x_2,\ldots,x_n) \\ v_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ v_n(x_1,x_2,\ldots,x_n) \end{bmatrix}$$

We think of this function as associating a vector to every point in the input space.

Example.



A vector field will always be assumed to be at least continuous, and whenever it is differentiated, it will be assumed to be continuously differentiable.

7.1 Paths and Curves

A path $\mathbf{r} : [a,b] \to \mathbb{R}^n$ is said to be continuously differentiable on $[a,b]$ if,

(i) $\mathbf{r}$ is continuous on $[a,b]$;

(ii) $\mathbf{f}$ is continuously differentiable on (a,b);

(iii) The limits $\lim_{t\to a^+} \mathbf{r}'(t)$ and $\lim_{t\to b^-} \mathbf{r}'(t)$ both exist so $\mathbf{r}'$ is a continuous function $[a,b] \to \mathbb{R}$.

We will always assume that a path $\mathbf{r} : [a,b] \to \mathbb{R}^n$ is continuous and piecewise continuously differentiable in the sense that there are a finite number of points $a_1,\ldots,a_\ell \in (a,b)$ with $a = a_0 < a_1 < a_2 < \cdots < a_\ell < a_{\ell+1} = b$ such that $\mathbf{r}$ is continosly differentiable on $[a_i,a_{i+1}]$ for all $0 \le i \le \ell$. If $\mathbf{r}'(t) \neq 0$ for all t, then $\mathbf{r}$ is *regular*.

Given $\mathbf{p},\mathbf{q} \in \mathbb{R}^n$, a curve $C_{\mathbf{p},\mathbf{q}}$ which goes from $\mathbf{p}$ to $\mathbf{q}$ is the image of some path $\mathbf{r} : [a,b] \to \mathbb{R}^n$ such that $\mathbf{r}(a) = \mathbf{p}$ and $\mathbf{r}(b) = \mathbf{q}$. The path $\mathbf{r}$ is then called a *parametrisation* of $C_{\mathbf{p},\mathbf{q}}$. If a curve C can be parametrised by a regular path, then the curve is also said to be regular.

Note that the parametrisation of a curve is not unique: If $\varphi : [\alpha,\beta] \to [a,b]$ is continuously differentiable, then $r \circ \varphi$ and r parametrise the same curve.

7.2 Tangential Line Integrals

The *component* of a vector $\mathbf{x} \in \mathbb{R}^n$ *in the direction of* a unit vector $\hat{\mathbf{v}}$ is defined as $\mathbf{x} \cdot \hat{\mathbf{v}}$. We also say that $\mathbf{x} \cdot \hat{\mathbf{v}}$ is the *component of* $\mathbf{x}$ *along* $\hat{\mathbf{v}}$.

If $\rho : [0,L] \to \mathbb{R}^n$ is the arclength or unit speed parametrisation of a regular curve $C_{pq} \subseteq \mathbb{R}^n$, then $\dot{\rho}(s) := \frac{d\rho}{ds}(s)$ is a unit vector called the *unit tangent* to C_{pq} at $\rho(s)$.

If $\underline{v}$ is a vector field, then $\underline{v}(\rho(s)) \cdot \dot{\rho}(s)$ is the *tangential component of* $\underline{v}$ *along* C_{pq}.

The *tangential line integral* of $\underline{v}$ along C_{pq} is defined as the integral of the tangential component of $\underline{v}$ along C_{pq}:

$$\int_0^L \underline{v}\big(\rho(s)\big) \cdot \dot{\rho}(s)\, ds$$

Because it is generally almost impossible to parametrise a curve by its arclength, we use a change of variable to actually compute these integrals in practice. Let $\mathbf{r} : [a,b] \to \mathbb{R}^n$ be a parametrisation of C_{pq}. Then, the mapping $\varphi : [0,L] \to [a,b]$ relates ρ and $\mathbf{r}$ by $\rho(s) = \mathbf{r}\big(\varphi(s)\big)$, so,

$$\int_0^L \underline{v}(\rho(s)) \cdot \dot{\rho}(s)\, ds = \int_a^b \underline{v}\big(\mathbf{r}(t)\big) \cdot \frac{d\mathbf{r}}{dt}\, dt$$

This line integral is also denoted by,

$$\int_{C_{pq}} \underline{v} \cdot d\mathbf{r}$$

obtained by cancelling the dt terms in the integral above.

Note that in the formula,

$$\int_0^L \underline{v}\big(\rho(s)\big) \cdot \dot{\rho}(s)\, ds$$

we have $\underline{v}(\rho(s)) \cdot \dot{\rho}(s) > 0$ whenever the angle between $\underline{v}$ and the unit tangent $\dot{\rho}$ is acute. Then,

$$\frac{1}{\text{length}(C_{pq})} \int_0^L \underline{v}\big(\rho(s)\big) \cdot \dot{\rho}(s)\, ds$$

represents the average value of $\underline{v} \cdot \dot{\rho}$ along C_{pq}, so the tangential line integral is a measure of the average rate at which the quantity described by the vector field $\underline{v}$ flows along C_{pq}.

If $\underline{v}$ represents a force, then $\int_C \underline{v} \cdot d\mathbf{r}$ represents the work done by $\underline{v}$ when moving an object along C. If C is a closed curve, then we write $\oint_C \underline{v} \cdot d\mathbf{r}$ instead, and the resulting value is sometimes called the *circulation* of $\underline{v}$ around C, as it measures the rate at which the quantity described by $\underline{v}$ circulates around C.

The value of the integral $\int_{C_{pq}} \underline{v} \cdot dr$ depends on the orientation of the path:

$$\int_{C_{pq}} \underline{v} \cdot d\mathbf{r} = -\int_{C_{qp}} \underline{v} \cdot d\mathbf{r}$$

When C is a closed curve, we write

$$\circlearrowleft\!\!\oint_C \underline{v} \cdot d\mathbf{r} = -\circlearrowright\!\!\oint_C \underline{v} \cdot d\mathbf{r}$$

to indicate the orientation of the path.

7.3 Flux

7.3.1 Flux Across Curves in $\mathbb{R}^2$

Given a vector $\mathbf{v} = (x,y) \in \mathbb{R}^n$, we define $\mathbf{v}^\perp := (y, -x)$. That is, $\mathbf{v}^\perp$ is $\mathbf{v}$ rotated clockwise by 90°. In particular, $\mathbf{v} \cdot \mathbf{v}^\top = 0$, so $\mathbf{v}$ and $\mathbf{v}^\top$ are orthogonal.

The tangent $\dot{\mathbf{r}}(t)$ of a regular curve C with regular parametrisation $\mathbf{r}(t) = (x(t),y(t))$ is given by $\dot{\mathbf{r}}(t) = (\dot{x}(t),\dot{y}(t))$, so the *normal* to C is given by,

$$\mathbf{N}(t) := \dot{\mathbf{r}}(t)^\perp = \big(\dot{y}(t), -\dot{x}(t)\big)$$

If $\rho : [0,L] \to \mathbb{R}^n$ is the arclength parametrisation of C, then $\mathbf{n}(s) := \dot{\rho}(s)^\perp$ is the *unit normal* to C.

The *flux* of a vector field $\underline{v}(x,y) = \big(v_1(x,y),v_2(x,y)\big)$ across a curve C is defined as the integral of the normal component of $\underline{v}$,

$$\int_0^L \underline{v}\big(\rho(s)\big) \cdot \mathbf{n}(s)\, ds$$

Again, due to difficulties with parametrising a curve by its arclength, we compute this integral with another change of variable:

$$\int_0^L \underline{v}\big(\rho(s)\big) \cdot \mathbf{n}(s)\, ds = \int_a^b \underline{v}\big(\mathbf{r}(t)\big) \cdot \mathbf{N}(t)\, dt$$

7.3.2 Flux Across Surfaces in $\mathbb{R}^3$

A surface $S \subset \mathbb{R}^3$ is parametrised by the map $\mathbf{r} : (U \subseteq \mathbb{R}^2) \to \mathbb{R}^3$ with U open, defined by,

$$\mathbf{r}(u,v) = \big(x(u,v),y(u,v),z(u,v)\big)$$

The tangent plane $T_{\mathbf{r}(u,v)}S$ of S at $\mathbf{r}(u,v)$ is spanned by $\frac{\partial \mathbf{r}}{\partial u}(u,v)$ and $\frac{\partial \mathbf{r}}{\partial v}(u,v)$. It follows that the dimension of the tangent plane $T_{\mathbf{r}(u,v)}S$ is 2 if and only if the tangent vectors are linearly independent. If this is the case for all $(u,v) \in U$, then $\mathbf{r}$ is a *regular* parametrisation of S.

The tangent vectors $\frac{\partial \mathbf{r}}{\partial u}$ and $\frac{\partial \mathbf{r}}{\partial v}$ are linearly independent if and only if $\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \neq \mathbf{0}$, in which case,

$$\mathbf{N}(u,v) := \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}$$

is a *normal* to S at $\mathbf{r}(u,v)$.

Similarly to the definition of flux across a curve, the flux of a vector field $\underline{v}$ across a surface S is defined by,

$$\iint_S \underline{v} \cdot \hat{\mathbf{n}}\, dA$$

where $\mathbf{n}$ is a unit normal to S and dA is the area element on S. With respect to a parametrisation $\mathbf{r}$ of S, we have,

$$\mathbf{n}(u,v) := \frac{\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}}{\left|\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right|}, \qquad dA := \left|\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right| du\, dv$$

or,

$$\mathbf{n}(u,v) := \frac{\mathbf{N}}{|\mathbf{N}|}, \qquad dA := |N|\; du\, dv$$

so the flux integral is given by,

$$\iint_S \underline{v} \cdot \hat{\mathbf{n}}\, dA = \iint_U \underline{v}\big(\mathbf{r}(u,v)\big) \cdot \left(\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right) du\, dv$$

The flux of $\underline{v}$ across S is also denoted by,

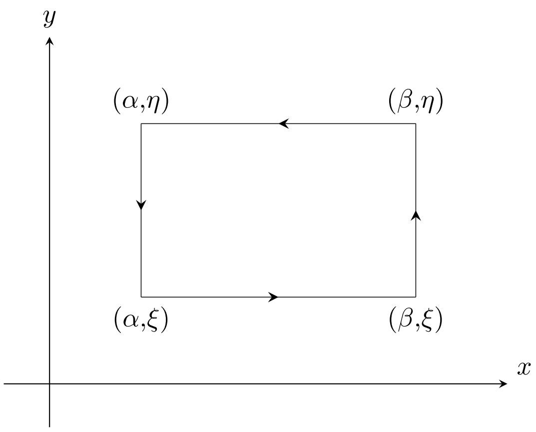
$$\iint_S \underline{v} \cdot d\mathbf{A}, \qquad \iint_S \underline{v} \cdot \mathbf{n}\, dS, \qquad \text{and} \qquad \iint_S \underline{v} \cdot d\mathbf{S}$$

8 The Integral Theorems of Vector Calculus

8.1 Green's Theorem for a Rectangle

Let $U \subseteq \mathbb{R}^2$ be open. A vector field $\underline{v} : U \to \mathbb{R}^2$ is called a *planar* vector field. As usual, we will assume $\underline{v}$ is continuously differentiable over U.

Let $\mathbb{R}$ denote the rectangle $[\alpha,\beta] \times [\xi,\eta]$ that is contained entirely (including the boundary ∂R) within U.



Consider the line integral of $\underline{v}$ around ∂R. If $\underline{v}(x,y) = (a(x,y),b(x,y))$ for some scalar-valued functions $a,b : U \to \mathbb{R}$, then,

$$\oint_{\partial R} \underline{v} \cdot d\underline{r} = \int_\alpha^\beta a(x,\xi)\, dx + \int_\xi^\eta b(\beta,y)\, dy - \int_\alpha^\beta a(x,\eta)\, dx - \int_\xi^\eta b(\alpha,y)\, dy$$

By the fundamental theorem of calculus, we have,

$$\int_\alpha^\beta a(x,\xi)\, dx - \int_\alpha^\beta a(x,\eta)\, dx = \int_\alpha^\beta \int_\xi^\eta -\frac{\partial a(x,y)}{\partial y}\, dy\, dx$$

$$\int_\xi^\eta b(\beta,y)\, dy - \int_\xi^\eta b(\alpha,y)\, dy = \int_\xi^\eta \int_\alpha^\beta \frac{\partial b(x,y)}{\partial x}\, dx\, dy$$

So,

$$\oint_{\partial R} \underline{v} \cdot d\underline{r} = \iint_R \frac{\partial b(x,y)}{\partial x} - \frac{\partial a(x,y)}{\partial y}\, dx\, dy$$

obtaining the statement of Green's theorem for a rectangle.

8.1.1 Regions and Unit Normals

A *region* in $\mathbb{R}^n$ is a bounded open subset Ω of $\mathbb{R}^n$ for which there exists a function $f : \mathbb{R}^n \to \mathbb{R}$ such that,

- All partial derivatives of f are continuous;
- $\Omega = \{x \in \mathbb{R}^n : f(x) < 0\}$;
- $\nabla f(\mathbf{p}) \neq 0 \, \forall \mathbf{p} \in f^{-1}\{0\} = \{\mathbf{p} \in \mathbb{R}^n : f(\mathbf{p}) = 0\}$.

The function f is then the *defining function* of Ω. The set $f^{-1}\{0\}$ is also the *boundary* of Ω, also denoted $\partial\Omega$. We also denote the *closure* $\Omega \cup \partial\Omega$ by $\overline{\Omega}$.

Example. Let $f(x,y,z) = x^2 + y^2 + z^2 - 1$. Then, the region defined by f is,

$$\begin{aligned}\{(x,y,z) : f(x,y,z) < 0\} &= \{(x,y,z) : x^2 + y^2 + z^2 < 1\} \\ &= \mathbb{B}_1(\mathbf{0})\end{aligned}$$

is the unit ball in $\mathbb{R}^3$, and its boundary is the unit 2-sphere $S^2(1) = \{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}| = 1\}$. Note that

$$\begin{aligned}\nabla f(x,y,z) &= 2(x,y,z) \\ &\neq 0\end{aligned}$$

as required.

The last requirement that $\nabla f(\mathbf{p}) \neq 0$ for all $\mathbf{p} \in \partial\Omega$ allows us to define the *outward unit normal* to Ω at $\mathbf{p}$:

$$\mathbf{n}_+(\mathbf{p}) \coloneqq \frac{\nabla f(\mathbf{p})}{|\nabla f(\mathbf{p})|}$$

Unfortunately, this requirement also excludes some well-behaved subsets like polygons and polyhedra which do not have well defined normals at vertices and edges, so we also consider *piecewise regions.*

8.1.2 Boundary Orientation

We now focus on the 2-dimensional case. The boundary of a 2-dimensional, or *planar*, region Ω is a curve, or, if Ω is not simply connected as in the case of an annulus, a system of curves.

Let $\mathbf{n}_+(\mathbf{p}) = \big(h(\mathbf{p}),k(\mathbf{p})\big)$ be the outward unit normal to the region Ω at the point $\mathbf{p} \in \partial\Omega$. The *positively oriented unit tangent vector* $\mathbf{t}_+(p)$ at p is then the vector $\big(-k(\mathbf{p}),h(\mathbf{p})\big)$. That is, the outward unit normal rotated counterclockwise by $90°$, or,

$$\mathbf{t}_+(\mathbf{p}) = -\mathbf{n}_+(\mathbf{p})^\perp$$

Informally, a tangent vector $\mathbf{t}$ to $\partial\Omega$ is positively oriented if, when facing in the direction of the vector, the interior of the region is to our left, and is negatively oriented otherwise. For example, if $\Omega = \mathbb{B}$, then a tangent vector that follows the unit circle in a counterclockwise manner is positively oriented.

However, take an annulus, for example. This region has two boundary curves; an *exterior* and *interior* boundary. A tangent vector on the exterior boundary is positively oriented if it follows the boundary counterclockwise, but a tangent vector on the interior boundary is positively oriented if it follows the boundary clockwise.

8.2 Green's Theorem for Planar Regions

In this section we will assume that all vector fields and functions considered are continuously differentiable on an open set $U \subseteq \mathbb{R}^2$, and that (the closure of) any region Ω lies entirely within U.

The *curl* of a planar vector field $\underline{v} : U \to \mathbb{R}^2$ defined by,

$$\underline{v}(x,y) = \big(a(x,y),b(x,y)\big)$$

is defined to be the function,

$$\operatorname{curl} \underline{v} = \frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}$$

Theorem 8.1 (Green's Theorem for Planar Regions). *Let Ω be a region in $\mathbb{R}^2$ and let $\underline{v} : U \to \mathbb{R}^2$ be a continuously differentiable planar vector field on U that contains $\overline{\Omega}$. Then,*

$$\iint_\Omega \operatorname{curl} \underline{v}(x,y)\, dA_{x,y} = \oint_{\partial\Omega} \underline{v} \cdot t_+ \, ds = \oint_{\partial\Omega} vli\ltimes \cdot d\mathbf{r}$$

where s is the arclength parameter along $\partial\Omega$, $\mathbf{r}$ is a positively oriented parametrisation of $\partial\Omega$, and the area element $dA_{x,y}$ is more often written as $dx\,dy$.

Recall that $\oint_{\partial\Omega} \underline{v} \cdot d\mathbf{r}$ is the circulation of $\underline{v}$ around $\partial\Omega$.

8.3 Flux and Divergence in the Plane

The *divergence* of a vector field $\underline{v}(x_1, \ldots, x_n) = (v_1(x_1, \ldots, x_n), v_2(x_1, \ldots, x_n), \ldots, v_n(x_1, \ldots, x_n))$, denoted by $\operatorname{div} \underline{v}$ and $\nabla \cdot \underline{v}$, is defined by,

$$\nabla \cdot \underline{v} \coloneqq \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i}$$

Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$. Then, $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^\perp \cdot \mathbf{w}^\perp$ and $(\mathbf{v}^\perp)^\perp = -\mathbf{v}$. So, if $\underline{v}$ is a planar vector field and Ω is a region in $\mathbb{R}^2$ that satisfy the hypotheses of Green's theorem, then,

$$\begin{aligned} \underline{v}^\perp \cdot \mathbf{t}_+ &= (\underline{v}^\perp)^\perp \cdot \mathbf{t}_+^\perp \\ &= -\underline{v} \cdot \mathbf{n}_+ \end{aligned}$$

The flux of $\underline{v}$ across $\partial\Omega$ is then given by,

$$\begin{aligned} \oint_{\partial\Omega} \underline{v} \cdot \mathbf{n}_+ \, ds &= -\oint \underline{v}^\perp \cdot \mathbf{t}_+ \, ds \\ &= -\iint_\Omega \operatorname{curl} \underline{v}^\perp \, dx\, dy \\ &= -\iint_\Omega \left(\frac{\partial(-a)}{\partial x} - \frac{\partial b}{\partial x} \right) dx\, dy \\ &= \iint_\Omega \left(\frac{\partial a}{\partial x} + \frac{\partial b}{\partial x} \right) dx\, dy \\ &= \iint_\Omega \nabla \cdot \underline{v}(x,y)\, dx\, dy \end{aligned}$$

Theorem 8.2 (Divergence Theorem for a Planar Region). *Let Ω be a region in $\mathbb{R}^2$ and let $\underline{v} : U \to^2$ be a continuously differentiable planar vector field on U which contains $\overline{\Omega}$. Then,*

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

where $\mathbf{n}_+$ is the outward unit normal to Ω.

Proof. Follows from Green's theorem as shown above. ∎

8.4 Flux and Divergence in $\mathbb{R}^3$

Theorem 8.3 (Divergence Theorem). *Let Ω be a region in $\mathbb{R}^3$ and let $\underline{v} : U \to \mathbb{R}^3$ be a continuously differentiable vector field on U which contains $\overline{\Omega}$. Then,*

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

where $\mathbf{n}_+$ is the outward unit normal to Ω, $dV_{x,y,z}$ is the volume element of Ω, more often written as $dx\,dy\,dz$.

8.5 Gradient Fields

If a vector field $\underline{v}$ is the gradient of a function $f : (U \subseteq \mathbb{R}^n) \to \mathbb{R}$, then $\underline{v}$ is called a *gradient field*, and the function f is called a *scalar potential* of $\underline{v}$.

Theorem 8.4 (Fundamental Theorem of Calculus for Gradient Vector Fields). *Given a continuously differentiable function $f : U \to \mathbb{R}$ and a curve $C_{\mathbf{pq}} \subset U$ from $\mathbf{p}$ to $\mathbf{q}$ parametrised by a continuously differentiable path $\mathbf{r} : [\mathbf{a},\mathbf{b}] \to \mathbf{U}$, we have,*

$$\int_{C_{pq}} \nabla f \cdot d\mathbf{r} = f(\mathbf{q}) - f(\mathbf{p})$$

Corollary 8.4.1. *This means that the value of a tangential line integral of a gradient field depends only on the orientation of C and the endpoints $\mathbf{p}$ and $\mathbf{q}$, and not on the shape of C itself.*

In particular, if the curve is closed, then the endpoints coincide, and we have:

Corollary 8.4.2. *For all closed curves C,*

$$\oint_C \nabla f \cdot d\mathbf{r} = 0$$

These two corollaries are equivalent in that any vector field that satisfies one will satisfy the other. Such a vector field is called a *conservative* vector field.

For example, gravity is a conservative field; it doesn't matter how you climb up a mountain, you gain the same amount of gravitational potential energy regardless of choice of path. Similarly, if you walk around but end up back where you started, you will have zero net gain of gravitational potential energy.

Theorem 8.5. *A vector field is conservative if and only if it is a gradient field.*

8.5.1 Incompressible and Irrotational Vector Fields

A vector field whose divergence is zero everywhere is called an *incompressible* or *divergence-free* vector field.

Theorem 8.6 (Zero Flux Property). *If $\underline{v} \in C^1(U \subseteq \mathbb{R}^3,\mathbb{R}^3)$ is incompressible, and $\overline{\Omega} \subset U$, then,*

$$\iint_{\partial\Omega} \underline{v} \cdot \mathbf{n}_+ \, dA = 0$$

Proof. $\underline{v}$ is incompressible, so $\nabla \cdot \underline{v} = 0$. By the divergence theorem,

$$\iint_{\partial\Omega} \underline{v} \cdot \mathbf{n}_+ \, dA = \iiint_\Omega \nabla \cdot \underline{v}\, dV$$

$$= \iiint_{\Omega} 0\, dV$$
$$= 0$$

■

A vector field whose curl is zero everywhere is called an *irrotational* vector field.

Theorem 8.7. *Every conservative field is irrotational.*

Proof.

$$\begin{aligned}\operatorname{curl}(\nabla f) &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) - \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) \\ &= \frac{\partial f}{\partial x \partial y} - \frac{\partial f}{\partial y \partial x} \\ &= 0\end{aligned}$$

■

8.5.2 Laplacian and Harmonic Functions

Let $\underline{v}$ be a incompressible conservative vector field with scalar potential f. Then, f satisfies the second order partial differential equation $\Delta f = 0$, where,

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

is the *Laplacian* of f.

For $f \in C^2(U)$, the equation $\Delta f = 0$ is called *Laplace's equation*, and its solutions are called *harmonic* functions or harmonic scalar fields.

9 Second Order Derivatives

9.1 Bilinear Forms

A linear map from a vector space to its field of scalars is called a *linear functional* or *covector*. The space $L(\mathbb{R}^n,\mathbb{R})$ of linear functionals on $\mathbb{R}^n$ is denoted by $(\mathbb{R}^n)^*$.

With respect to the standard basis of $\mathbb{R}^n$, $(\mathbb{R}^n)^*$ is identified with $\mathbb{R}^{1\times n}$. That is, every linear functional can be represented by a row vector.

A linear map $T : \mathbb{R}^n \to (\mathbb{R}^n)^*$ can be viewed as a bilinear form $\hat{T} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$:

$$\hat{T}(\mathbf{a},\mathbf{b}) := \mathbf{a}^\top T\mathbf{b}$$

9.2 The Hessian Matrix

Recall that if $f : (U \subseteq \mathbb{R}^n) \to \mathbb{R}$ is differentiable at $\mathbf{x}$, then there exists a linear transformation $Df \in L(\mathbb{R}^n,\mathbb{R}) = (\mathbb{R}^n)^*$ such that,

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\left|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \big(\mathbf{f}(\mathbf{x}) + Df(\mathbf{h})\big)\right|}{|\mathbf{h}|} = 0$$

(and further recall that $Df(\mathbf{x})$ is given by the Jacobian matrix, $\partial f(\mathbf{x})$).

Now, consider the case where $Df(\mathbf{x})$ itself is differentiable. Then, there exists some linear map $T \in L(\mathbb{R}^n,(\mathbb{R}^n)^*)$ such that,

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\left|Df(\mathbf{x}+\mathbf{h}) - \big(Df(\mathbf{x}) + T(\mathbf{h})\big)\right|}{|\mathbf{h}|} = 0$$

This map, if it exists, is called the *Hessian* of f, also denoted by $D^2 f(\mathbf{x})$.

Suppose all second-order partial derivatives of $f : \mathbb{R}^n \to \mathbb{R}$ exist. Then, we define the *Hessian matrix*, denoted by $\mathbf{H}_f$ or $\partial^2 f(\mathbf{x})$, as,

$$\mathbf{H}_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

That is, the (i,j)th entry is given by,

$$(\mathbf{H}_f)_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

This is the matrix that represents the Hessian transformation, if it exists. Note that the converse does *not* hold: even if all second order partial derivatives exist, and hence the Hessian matrix exists, Df may not necessarily be differentiable.

9.3 Non-Commutativity of Second Order Partial Derivatives

Second order partial derivatives do not, in general, commute. That is, for example,

$$\frac{\partial f}{\partial x \partial y} \neq \frac{\partial f}{\partial y \partial x}$$

However, these derivatives can commute under certain restrictions.

Theorem 9.1. *If the Hessian transformation $D^2 f(\mathbf{x})$ exists, then the second order partial derivatives at $\mathbf{x}$ commute. That is,*

$$\frac{\partial f}{\partial x \partial y}(\mathbf{x}) = \frac{\partial f}{\partial y \partial x}(\mathbf{x})$$

for all i,j; or, $\mathbf{H}_f$ is symmetric.

Corollary 9.1.1. *If all second order partial derivatives are continuous at $\mathbf{x}$, then the second order partial derivatives commute at $\mathbf{x}$.*

10 Inverse Function Theorem

10.1 Change of Variables and Inverse Functions

Let U and V be open subsets of $\mathbb{R}^n$. A change from variables $(x_1, \dots, x_n) \in U$ to variables $(y_1, \dots, y_n) \in V$ is achieved using a function $\Psi : U \to V$, with $\Psi = (\psi_1, \dots, \psi_n)$such that,

$$y_1 = (\psi_1(x_1, \dots, x_n))$$

$$y_2 = (\psi_2(x_1, \dots, x_n))$$
$$\vdots$$
$$y_n = (\psi_n(x_1, \dots, x_n))$$

If $\mathbf{\Psi}$ is bijective, then we can change back from y-variables to x-variables with the inverse map $\mathbf{\Psi}^{-1}$.

Theorem 10.1. *Suppose $\Psi : U \to V$ is a bijection differentiable at $\mathbf{x} \in U$, and suppose further that Ψ^{-1} is differentiable at $\mathbf{y} = \Psi(\mathbf{x}) \in V$. Then, $D\Psi(\mathbf{x})$ and $D\Psi^{-1}(\mathbf{y})$ are both invertible and,*

$$D\Psi^{-1}(\mathbf{y}) = \left(D\Psi(\Psi^{-1}(\mathbf{y}))\right)^{-1}$$

Proof. For all $\mathbf{y} \in V$,

$$\Psi\left(\Psi^{-1}(\mathbf{y})\right) = \mathbf{y}$$

Differentiating using the chain rule, we have,

$$D\Psi\left(\Psi^{-1}(\mathbf{y})\right) \circ D\Psi^{-1}(\mathbf{y}) = \mathrm{id}_{\mathbb{R}^n}$$

and the result follows. ■

In the 1-dimensional case, the Fréchet derivative is just the ordinary derivative, so this result is written as,

$$(\Psi^{-1})'(\mathbf{y}) = \frac{1}{\Psi'\left(\Psi^{-1}(\mathbf{y})\right)}$$

or more memorably as,

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$$

10.2 Local Inverses

Does the converse of the previous theorem hold? That is, if $\Psi : U \to V$ is differentiable at $\mathbf{x} \in U$, and $D\Psi(\mathbf{v})$ is invertible – does it then follow that Ψ^{-1} exists, and if it does, is it differentiable?

First, $D\Psi(\mathbf{x})$ depends only on the behaviour of Ψ near $\mathbf{x}$, so if $\left(D\Psi(\mathbf{x})\right)^{-1}$ exists, then Ψ^{-1} can exist at most near $\Psi(\mathbf{x})$, and not on all of $\Psi(U)$.

We will now only consider linear maps in the space $L(\mathbb{R}^n, \mathbb{R}^n)$. From the rank-nullity theorem, any such linear map is injective if and only if it is surjective, which will be useful for showing that an inverse exists.

Let $\mathbf{p} \in U$. If $\mathcal{N}_p \subseteq U$ is an open set containing $\mathbf{p}$, then we say that $\mathcal{N}_p$ is an (open) *neighbourhood* of $\mathbf{p}$.

Then, a function $\Psi : U \to V$ is a *local bijection* at $\mathbf{p} \in U$ if there is an open neighbourhood $\mathcal{N}_\mathbf{p}$ of $\mathbf{p}$ and an open neighbourhood $\mathcal{N}_\mathbf{q}$ of $\mathbf{q} = \Psi(\mathbf{p})$ such that the restriction $\Psi : \mathcal{N}_\mathbf{p} \to \mathcal{N}_\mathbf{q}$ is a bijection. We also say that Ψ is *locally invertible* at $\mathbf{p}$, and that inverse of the restricted function, $\Psi^{-1} : \mathcal{N}_\mathbf{q} \to \mathcal{N}_\mathbf{p}$ is the *local inverse* of Ψ. This local inverse is also called a *branch* of the *global* or *full* inverse Ψ^{-1} (if it exists).

Example. Consider $\Psi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ defined by $x \mapsto x^2$. Ψ is not injective, since $\Psi(x) = \Psi(-x)$.

But, take some $p > 0$, and the open neighbourhood $\mathcal{N}_p = (0,\infty)$. Then, $q = \Psi(p) = p^2 > 0$, and we can take the open neighbourhood $\mathcal{N}_q = (0,\infty)$, and indeed Ψ restricted to these neighbourhoods is bijective, with local inverse given by $\Psi^{-1}(y) = \sqrt{y}$. We can use the previous theorem to calculate the derivative of this inverse:

$$\begin{aligned} \Psi'(x) &= 2x \\ (\Psi^{-1})'(y) &= \frac{1}{\Psi'(\Psi^{-1}(y))} \\ &= \frac{1}{\Psi'(\sqrt{y})} \\ &= \frac{1}{2\sqrt{y}} \end{aligned}$$

If we instead take $p < 0$ with open neighbourhood $\mathcal{N}_p = (-\infty,0)$, and $q = p^2 > 0$ with open neighbourhood $\mathcal{N}_q = (0,\infty)$, Ψ is again bijective, with local inverse given by $\Psi^{-1}(x) = -\sqrt{x}$. This time, the derivative is given by,

$$\begin{aligned} \Psi'(x) &= 2x \\ (\Psi^{-1})'(y) &= \frac{1}{\Psi'(\Psi^{-1}(y))} \\ &= \frac{1}{\Psi'(-\sqrt{y})} \\ &= -\frac{1}{2\sqrt{y}} \end{aligned}$$

However, with our definition of a local inverse, there is no open neighbourhood around $p = 0$ such that Ψ is a bijection, so Ψ is not invertible on any open interval containing 0.

$\sqrt{y}$ and $-\sqrt{y}$ then form the two branches of the multivalued full inverse $\Psi^{-1}(y) = \pm\sqrt{y}$.

Theorem 10.2 (Inverse Function Theorem)**.** *Let $U \subseteq \mathbb{R}^n$ be open, and suppose $\Psi : U \to \mathbb{R}^n$ is continuously differentiable. Suppose that the Fréchet derivative $D\Psi(\mathbf{p})$ is invertible at a point $\mathbf{p} \in U$ (that is, the Jacobian $\partial\Psi(\mathbf{p})$ has non-zero determinant), and define $\mathbf{q} = \Psi(\mathbf{p})$. Then,*

- *There exist neighbourhoods $\mathcal{N}_{\mathbf{p}} \subset U$ and $\mathcal{N}_{\mathbf{q}} \subset \Psi(U)$ of $\mathbf{p}$ and $\mathbf{q}$ respectively, such that the restriction $\Psi : \mathcal{N}_{\mathbf{p}} \to \mathcal{N}_{\mathbf{q}}$ is a bijection;*
- *The inverse of the restriction, $\Psi^{-1} : \mathcal{N}_{\mathbf{q}} \to \mathcal{N}_{\mathbf{p}}$, is continuously differentiable, and furthermore,*

  $$D\Psi^{-1}(\mathbf{y}) = \big(D\Psi(\Psi^{-1}(\mathbf{y}))\big)^{-1}$$

  for all $\mathbf{y} \in \mathcal{N}_{\mathbf{q}}$.

A map $\Psi : U \to V$ between two open subsets of $\mathbb{R}^n$ is called a *diffeomorphism* if it is bijective, continously differentiable on U, and its inverse is continuously differentiable on V.

Ψ is called a *local diffeomorphism near* $\mathbf{p} \in U$ if there exists a neighbourhood $\mathcal{N}_{\mathbf{p}} \subset U$ of $\mathbf{p}$ such that the restriction $\Psi : \mathcal{N}_{\mathbf{p}} \to \mathcal{N}_{\mathbf{q}}$ is a diffeomorphism, where $\mathbf{q} := \Psi(\mathbf{p})$ and $\mathcal{N}_{\mathbf{q}} := \Psi(\mathcal{N}_{\mathbf{p}}) \subset V$ is a neighbourhood of q.

Theorem 10.3. *Let $U \subseteq \mathbb{R}^n$ be open. Given a continuously differentiable function $\mathbf{f} : U \to \mathbb{R}^k$, suppose $D\mathbf{f}(\mathbf{p})$ is injective at some point $\mathbf{p} \in U$. Then there exists $\delta > 0$ such that $\mathbb{B}_\delta(\mathbf{p}) \subset U$ and such that $\mathbf{f}$ is injective on $\mathbb{B}_\delta(\mathbf{p})$.*

11 Implicit Function Theorem

All functions in this section will be assumed to be continuously differentiable so we will write the Jacobian for the derivative instead of the Fréchet derivative.

Suppose we have a function $F : \mathbb{R}^2 \to \mathbb{R}$. If we set $F(x,y) = c$ for some $c \in \mathbb{R}$, then this equation defines y *implicitly* in terms of x, or x implicitly in terms of y.

For instance, if $F(x,y) = x^2 + y^2$ and $c > 0$, then $x^2 + y^2 = c$ describes a relation between x and y implicitly. If $c \neq \pm\sqrt{c}$, then the equation has two solutions for y in terms of x; namely, $y(x) = \sqrt{c - x^2}$ and $y(x) = -\sqrt{c - x^2}$, $-\sqrt{c} < x < \sqrt{c}$. Each of these solutions is called an *explict determination* of y in terms of x by the means of the functions $\sqrt{c - x^2}$ and $-\sqrt{c - x^2}$.

Let U be an open subset of $\mathbb{R}^{n+\ell} = \mathbb{R}^n \oplus \mathbb{R}^\ell$. We will write (x,y) for points in $\mathbb{R}^{n+\ell}$, where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^\ell$.

Given a continuously differentiable function $\mathbf{F} : U \to \mathbb{R}^\ell$, we will write the Jacobian $\partial\mathbf{F}(x,y) \in \mathbb{R}^{\ell\times(n+\ell)}$ as $(\partial_x\mathbf{F}(x,y) \quad \partial_y\mathbf{F}(x,y))$ where $\partial_x\mathbf{F} \in \mathbb{R}^{\ell\times n}$ and $\partial_y\mathbf{F} \in \mathbb{R}^{\ell\times\ell}$

So, a matrix $\Lambda \in \mathbb{R}^{\ell\times(n+\ell)}$ can be written as $\Lambda = [\mathbf{A} \quad \mathbf{B}]$, where $\mathbf{A} \in \mathbb{R}^{\ell\times n}$ and $\mathbf{B} \in \mathbb{R}^{\ell\times\ell}$. If we then write a vector $\mathbf{z} \in \mathbb{R}^{n\times\ell}$ as $\mathbf{z} = (\mathbf{x},\mathbf{y})$ where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^\ell$, then we can write a linear map $\mathbf{F} : \mathbb{R}^{n+\ell} \to \mathbb{R}^\ell$ defined by $\mathbf{F}(\mathbf{z}) = \Lambda\mathbf{z}$ as,

$$\mathbf{F}(\mathbf{x},\mathbf{y}) = [\mathbf{A} \quad \mathbf{B}] \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}$$

Given some $\mathbf{c} \in \mathbb{R}^\ell$, we can then rewrite the equation $\mathbf{F}(\mathbf{x},\mathbf{y}) = \mathbf{c}$ as,

$$\mathbf{B}\mathbf{y} = \mathbf{c} - \mathbf{A}\mathbf{x}$$

This linear system of equations can be solved for the ℓ variables in $\mathbf{y}$ explicitly in terms of the n variables in $\mathbf{x}$ if $\mathbf{B}$ is invertible:

$$\mathbf{y} = \mathbf{B}^{-1}(\mathbf{c} - \mathbf{A}\mathbf{x})$$

If $\mathbf{B}$ is not invertible, then the system either has infinitely many solutions $\mathbf{y}$ if it is consistent, or no solutions if it is inconsistent. In either case, $\mathbf{y}$ cannot be written uniquely as a linear function of $\mathbf{x}$ if $\mathbf{B}$ is not invertible.

The implicit function theorem for a general continuously differentiable function $\mathbf{F} : U \to \mathbb{R}^\ell$ asserts that, if we have one solution $(\mathbf{x}_0,\mathbf{y}_0)$ of the equation $\mathbf{F}(\mathbf{x},\mathbf{y}) = \mathbf{c}$, and if $\partial_y\mathbf{F}(\mathbf{x}_0,\mathbf{y}_0) \in \mathbb{R}^{\ell\times\ell}$ is invertible, then we can solve for $\mathbf{y}$ in terms of $\mathbf{x}$ for $\mathbf{x}$ sufficiently near $\mathbf{x}_0$. The implicit function theorem is therefore concerned with converting an implicit relation $\mathbf{F}(\mathbf{x},\mathbf{y}) = \mathbf{c}$ to an explicit relation $\mathbf{y} = \mathbf{g}(\mathbf{x})$ such that the relation $\mathbf{F}(\mathbf{x},\mathbf{g}(\mathbf{x})) = \mathbf{c}$ holds for all $\mathbf{x}$ in some open neighbourhood $\mathcal{N}_{\mathbf{x}_0}$ containing $\mathbf{x}$.

Theorem 11.1 (Implicit Function Theorem)**.** *Let $U \subseteq \mathbb{R}n + \ell$ be open and let $\mathbf{c} \in \mathbb{R}^\ell$. Suppose that $\mathbf{F} : U \to \mathbb{R}^\ell$ is continuously differentiable, and that theequation $\mathbf{F}(\mathbf{x},\mathbf{y}) = \mathbf{c}$ has a solution $(\mathbf{x}_0,\mathbf{y}_0) \in U$ such that $\det\big(\partial_y\mathbf{F}(\mathbf{x}_0,\mathbf{y}_0)\big) \neq 0$. Then, there exists an open neighbourhood $\mathcal{N}_{\mathbf{x}_0} \subseteq \mathbb{R}^n$ of $\mathbf{x}_0$ and a continuously differentiable function $\mathbf{g} : \mathcal{N}_{\mathbf{x}_0} \to \mathbb{R}^\ell$ such that,*

- $\mathbf{g}(\mathbf{x}_0) = \mathbf{y}_0$*,* $\{(\mathbf{x},\mathbf{g}(\mathbf{x})) : \mathbf{x} \in \mathcal{N}_{\mathbf{x}_0}\} \subset U$*, and* $\mathbf{F}(\mathbf{x},\mathbf{g}(\mathbf{x})) = \mathbf{c}$ *for all* $\mathbf{x} \in \mathcal{N}_{\mathbf{x}_0}$*;*
- *Furthermore,* $\partial_y\mathbf{F}(\mathbf{x},\mathbf{g}(\mathbf{x}))$ *is locally invertible over all* $\mathbf{x} \in \mathcal{N}_{\mathbf{x}_0}$*, and the derivative of* $\mathbf{g}$ *is given by,*

$$\partial\mathbf{g}(\mathbf{x}) = -\big(\partial_y\mathbf{F}(\mathbf{x},\mathbf{g}(\mathbf{x}))\big)^{-1} \cdot \partial_x\mathbf{F}(\mathbf{x},\mathbf{g}(\mathbf{x}))$$

for all $\mathbf{x} \in \mathcal{N}_{\mathbf{x}_0}$*.*

Example. In the previous example, we had $F(x,y) = x^2 + y^2$. Given a point (x_0,y_0) on the circle $F(x,y) = c > 0$ such that $y_0 > 0$, we have seen that $g(x) = \sqrt{c - x^2}$. The implicit function then gives the derivative of g to be,

$$\begin{aligned}\partial_x F(x,y) &= [2x]\\ \partial_y F(x,y) &= [2y]\\ g'(x) &= -\big(\partial_y F(x,g(x))\big)^{-1} \cdot \partial_x F(x,g(x))\\ &= -\frac{2x}{2g(x)}\\ &= -\frac{x}{\sqrt{c-x^2}}\end{aligned}$$

Geometrically, the implicit function theorem asserts that if $\det\big(\partial_y \mathbf{F}(\mathbf{x}_0,\mathbf{y}_0)\big) \neq 0$, then near a point $(\mathbf{x}_0,\mathbf{y}_0)$, the level set $\Gamma_\mathbf{c} := \{(\mathbf{x},\mathbf{y}) : \mathbf{F}(\mathbf{x},\mathbf{y}) = \mathbf{c}\}$ is the graph $\mathcal{G}_\mathbf{g}$ of a function $\mathbf{y} = \mathbf{g}(\mathbf{x})$. That is, near $(\mathbf{x}_0,\mathbf{y}_0)$,

$$\Gamma_\mathbf{c} = \mathcal{G}_\mathbf{g} := \{(\mathbf{x},\mathbf{g}(\mathbf{x})) : \mathbf{x} \in \mathcal{N}_{\mathbf{x}_0}\} = \{(\mathbf{x},\mathbf{y}) : \mathbf{y} = \mathbf{g}(\mathbf{x}), \mathbf{x} \in \mathbf{x}_0\}$$

For instance, for $F(x,y) = x^2 + y^2 = c > 0$, $\partial_y F(x_0,y_0) \neq 0$ whenever $y_0 \neq 0$. If $y_0 > 0$, then (x_0,y_0) is contained in the upper semicircle which is the graph of $y = \sqrt{c - x^2}$, and if $y_0 < 0$, then (x_0,y_0) is contained in the lower semicircle which is the graph of $y = \sqrt{c - x^2}$.

The graph of a continuously differentiable function of one variable is a special case of a regular parametrisation of a curve, and the graph of a continuously differentiable function of two variables is a special case of a regular parametrisation of a surface. Regular parametrisations can be pieced together to form spaces that are known as *submanifolds* of Euclidean spaces. The circle in this case is a 1-dimensional submanifold of $\mathbb{R}^2$ that can be viewd as being pieced together along overlaps from the four semicircles which are the graphs of,

$$\begin{aligned}y(x) &= \sqrt{c - x^2}, x \in (-\sqrt{c},\sqrt{c}) & x(y) &= \sqrt{c - y^2}, y \in (-\sqrt{c},\sqrt{c})\\ y(x) &= -\sqrt{c - x^2}, x \in (-\sqrt{c},\sqrt{c}) & x(y) &= -\sqrt{c - y^2}, y \in (-\sqrt{c},\sqrt{c})\end{aligned}$$

A set $M \subset \mathbb{R}^{n+\ell}$ is a *submanifold* (without boundary) of dimension n if, for each $\mathbf{p} \in M$, there exists an open neighbourhood $\mathcal{N}_\mathbf{p} \subset \mathbb{R}^{n+\ell}$ of $\mathbf{p}$, an open set $U \subset \mathbb{R}^n$ and a continuously differentiable function $\mathbf{r} : U \to \mathbb{R}^{n+\ell}$, such that $\mathbf{r}(\mathbf{x}_\mathbf{p}) = \mathbf{p}$ for some $\mathbf{x}_\mathbf{p} \in U$, $\mathbf{r} : U \to M \cap \mathcal{N}_\mathbf{p}$ is a bijection, and $\operatorname{rank}(\partial \mathbf{r}(\mathbf{x})) = n$ for all $\mathbf{x} \in U$.

The function $\mathbf{r}$ is then called a (regular) parametrisation of $M \cap \mathcal{N}_\mathbf{p}$. The *tangent space* $T_{\mathbf{r}(\mathbf{x})}M$ of M at $\mathbf{r}(\mathbf{x})$ is the image of $\partial \mathbf{r}(\mathbf{x})$ shifted by $\mathbf{r}(\mathbf{x})$; that is, $\mathbf{r}(\mathbf{x}) + \operatorname{span}\big(\partial_1 \mathbf{r}(\mathbf{x}), \dots \partial_n \mathbf{r}(\mathbf{x})\big)$, or,

$$T_{\mathbf{r}(\mathbf{x})}M = \{\mathbf{r}(\mathbf{x}) + (\partial \mathbf{r}(\mathbf{x}))\mathbf{h} : \mathbf{h} \in \mathbb{R}^n\}$$

Thus, the tangent space is identified with the image of the affine linear approximation of $\mathbf{r}$.

Theorem 11.2. *Given a continuously differentiable function $\mathbf{F} : (U \subseteq \mathbb{R}^{n+\ell}) \to \mathbb{R}^\ell$ with U open, and some fixed $\mathbf{c} \in \mathbb{R}^\ell$, define the level set $\Gamma_\mathbf{c} := \{\mathbf{z} \in U : \mathbf{F}(\mathbf{z}) = \mathbf{c}\}$.*

Suppose that $\operatorname{rank}\big(\partial \mathbf{F}(\mathbf{z})\big) = \ell$ for all $\mathbf{z} \in \Gamma_\mathbf{c}$. Then, $\Gamma_\mathbf{c}$ is a submanifold (without boundary) of dimension n in $\mathbb{R}^{n+\ell}$. Furthermore, $T_z\Gamma_\mathbf{c} = \mathbf{z} + \ker(\partial \mathbf{F}(\mathbf{z})) = \{\mathbf{z} + \mathbf{v} : \partial \mathbf{F}(\mathbf{z})\mathbf{v} = \mathbf{0}\}$.

In the special case that $\ell = 1$, then $\Gamma_\mathbf{c}$ is called a *hypersurface* and,

$$\partial \mathbf{F}(\mathbf{z}) = (\partial_1 \mathbf{F}(\mathbf{z}), \dots \partial_{n+1} \mathbf{F}(\mathbf{z})), \qquad \nabla \mathbf{F}(\mathbf{z}) = \begin{bmatrix}\partial_1 \mathbf{F}(\mathbf{z})\\ \vdots\\ \partial_{n+1} \mathbf{F}(\mathbf{z})\end{bmatrix}$$

so,

$$\begin{aligned} \mathbf{v} \in \ker\big(\partial \mathbf{F}(\mathbf{z})\big) &\longleftrightarrow (\nabla \mathbf{F}(\mathbf{z})) \cdot \mathbf{v} = 0 \\ &\longleftrightarrow \nabla \mathbf{F}(\mathbf{z}) \perp T_{\mathbf{z}}\Gamma_{\mathbf{c}} \end{aligned}$$

so $\nabla \mathbf{F}$ is orthogonal to the level set $\Gamma_{\mathbf{c}}$, so the gradient of a function is the normal to the hypersurface it describes.

If we also have $n = 1$, then Γ_c is called a *level curve* in $\mathbb{R}^2$, and if $n = 1$, then $\Gamma_{\mathbf{c}}$ is called a *level surface* in $\mathbb{R}^3$.